UNIVERSITAT DE BARCELONA

Task-based pronunciation teaching and learning of L2 vowels in EFL learners: task complexity effects

Ingrid Mora Plaza

TASK-BASED PRONUNCIATION TEACHING AND LEARNING OF L2 VOWELS IN EFL LEARNERS: TASK COMPLEXITY EFFECTS

Tesi doctoral presentada per

Ingrid Mora Plaza

per a l'obtenció del grau de doctor

Programa de doctorat: *Ciència cognitiva i llenguatge*
Línia de recerca: *Lingüística teòrica i aplicada*
Departament de Llengües i Literatures Modernes i d'Estudis Anglesos

Directors: **Joan Carles Mora Bonilla**
Roger Gilabert Guerrero
Tutor: **Joan Carles Mora Bonilla**

2023

UNIVERSITAT DE BARCELONA

Als meus pares,

i a la meva germana.

Acknowledgements

As I sit down to write these pages, I am filled with a sense of gratitude that words cannot fully express. The journey towards a PhD has been a challenging yet incredibly rewarding experience, and I could not have done it without the generous support and encouragement of so many wonderful people. To each and every one of you, I offer my deepest appreciation.

First and foremost, I would like to express my heartfelt thanks to my PhD directors, Dr. Joan C. Mora and Dr. Roger Gilabert. Their unwavering support throughout my PhD has been a constant source of inspiration. I am deeply grateful for their contribution to my academic and personal growth, and the knowledge and skills I gained from them will always be treasured. Joan Carles taught me how to approach phonetics with a critical mindset, instilled in me a passion for teaching pronunciation, and showed me the importance of conducting research with the utmost rigor. Roger taught me how to approach linguistics with a practical, applied mindset, emphasizing the importance of establishing connections between theoretical knowledge and real-world classroom situations. Through his guidance, I developed a love for designing and implementing effective language learning tasks. It is difficult to express in just a few lines the depth of my gratitude for the incredible time and effort you invested in this thesis. Your commitment has made this a truly unforgettable and remarkable experience for me.

My second acknowledgement goes to my good friends and colleagues, Dr. Mireia Ortega and Dr. Cristina Aliaga. I am extremely thankful to have had the opportunity to share countless hours in the L2 speech lab, present at numerous conferences, and collaborate on research projects with them. Throughout this journey, they have been my constant support, offering invaluable advice, words of encouragement, and innumerable moments of laughter.

A special thanks to Dr. Carme Muñoz who, from the very beginning, kindly integrated me in the *Grup de Recerca en Adquisició de Llengües* (GRAL) from the University of Barcelona. I would like to thank all the members for your advice and feedback throughout these years. In particular, I am thankful to Dr. Lena Vasylets and Dr. Julia Barón, with whom I have spent invaluable moments.

Furthermore, my sincere gratitude goes again to Dr. Joan C. Mora who provided me with the opportunity to participate in the Barcelona L2 Speech meetings and share my research with other like-minded scholars such as Dr. Juli Cebrián, Dr. Angélica Carlet, Dr. Lucrecia Rallo, Dr. Natalia Fullana and Dr. Celia Gorba, amongst others.

I would also like to express my gratitude to the colleagues from the Department of Modern Languages and Literatures and English Studies, which has been my home for the last five years. In particular, to the junior researchers with whom I have had the pleasure to share fantastic moments in McDermott's room. To Montse, Mercedes and Olga, for their help when I needed it, and their practical advice.

I am deeply grateful to the head teacher, coordinators of the linguistic team and English teachers for warmly welcoming the project on L2 oral abilities and pronunciation in the secondary school *I.N.S. Vilatzara* in *Vilassar de Mar*. I would like to express my immense gratitude to Maite and Montse for their willingness to open their classrooms and allow me to collect data from their Year 11 students for a period of 6 months. They also generously assisted me in the development of the project without any hesitation.

A very special thank you to Daniela, Valeria and Miren, as their invaluable contribution was crucial in enabling me to gather data from nearly one hundred students. Without their assistance, it would not have been possible to collect such a large dataset. They have been key to this investigation. I would also like to thank Gonzalo and Dr. Mireia Ortega, for their help in data analyses and editing of the manuscript, along with the native speakers, the teachers and inter-raters who contributed significantly to the value of this dissertation.

I am indebted to my PhD follow-up committee for their constructive feedback at the onset of this project, and the members of the examining committee for having accepted to be part of the board and taking the time to read the present dissertation.

I would like to extend my sincere gratitude to Dr. Kazuya Saito for his exceptional hospitality and support during my research stay at the Institute of Education (University College London). I am truly grateful for the opportunity to have collaborated in an L2 speech project with Dr. Kazuya Saito's team (Dr. Adam Tierney, Dr. Yui Suzukida, Dr. Magdalena Kachlicka, and Ms. Yaoyao Ruan) and for the invaluable insights and knowledge they shared with me.

The completion of this PhD thesis was made possible thanks to the generous support of the Spanish *Ministerio de Educación y Formación Profesional*, through a four-year predoctoral research grant (FPU17/03739) and a 3-month stay-abroad grant (EST21/00061). Their assistance was crucial in the successful completion of this work.

The research experience I have acquired throughout these years is ascribed in great deal to my participation in the research projects directed by Dr. Joan C. Mora (PID2019-107814GB-I00) and Dr. Pilar Prieto (*RecerCaixa* 2017ACUP 00249). Thank you for the financial support for making this investigation possible. A special thanks to Dr. Pilar Prieto and Ïo Valls, who counted on me for co-leading a project on L1/L2 oral abilities in secondary schools in Catalonia.

I am extremely grateful for the funding sources that have enabled me to attend international and national conferences, where I have had the privilege of meeting and learning from leading experts in the field of L2 speech acquisition and pronunciation teaching. In particular, I have been deeply inspired by the groundbreaking work of Dr. Alice Henderson, Dr. Anastazija Kirkova-Naskova, Dr. Anna Jarosz, Dr. Danielle Daidone, Dr. Isabelle Darcy, Dr. Miquel Llompart, Dr. Natalia Kartushina, Dr. Pavel Trofimovich, and many other esteemed scholars, whose insights and guidance have been invaluable to me.

Last but not least, I would like to express my heartfelt appreciation to my friends and family for their unwavering love, support, and encouragement throughout my PhD journey. *En especial, vull donar les gràcies als meus pares, Frederic i Rosa, i a la meva germana, Diana, per recolzar-me en totes les meves decisions i per estar al meu costat en els moments bons i els no tan bons. Sense vosaltres, arribar fins aquí no hagués estat possible. Us agraeixo molt la paciència i els sacrificis que heu hagut de fer per mi durant aquests anys per tal de que jo pogués dedicar tot el temps i energia a aquest projecte. Aquesta tesi també és vostra*. Ross, you have always been there for me, whether it was to celebrate a small victory or to provide a listening ear during a moment of frustration. Thank you for being my rock, my breath of fresh air, my source of inspiration, and for always standing by my side.

Barcelona, April 26th 2023

Abstract

Second language (L2) pronunciation learning in foreign language instructional contexts is particularly challenging due to insufficient quantity and quality of L2 input. Decades of L2 speech acquisition research have demonstrated that L1 phonology exerts a strong influence in L2 phonological learning (Best & Tyler, 2007; Flege & Bohn, 2021). One way to mitigate the effects of L1 phonological interference may be to raise learners' awareness of the relevant L2 pronunciation targets through tasks that encourage attention to pronunciation within meaning-based interaction. Although form-focused communicative approaches have been gaining interest in pronunciation teaching research (Darcy & Rocca, 2023), few investigations have applied task-based language teaching (TBLT) principles, which involve real-world processes of language use, to L2 pronunciation learning (Gurzynski-Weiss et al., 2017a). So far, none have assessed gains in L2 vowel perception, lexical encoding and production for the same participants after a pedagogical intervention. Furthermore, studies investigating whether the predictions of the Cognition Hypothesis (Robinson, 2011b) can extend to L2 phonology are relatively scarce. This doctoral dissertation seeks to contribute to task-based pronunciation teaching (TBPT) research by (1) examining the effectiveness of task design manipulation and (2) exploring the effects of increased task complexity on the pronunciation of two difficult English vowel contrasts (/iː-ɪ/, /æ-ʌ/); (3) assessing to what extent individual differences may mediate L2 vowel performance and gains; and (4) obtaining learners' perceptions of TBPT and L2 vowel learning.

A total of 92 bilingual Catalan/Spanish learners of English were divided into two experimental groups and a control group. Sixty-three experimental learners carried out 20 dyadic problem-solving tasks over 7 weeks. Participants were assigned to either simple (*N*=31) or complex (*N*=32) cognitive task complexity groups depending on the number of reasoning demands along resource-directing dimensions the tasks involved. Task completion required the distinction of the target minimal pairs (e.g., *bean-bin, cat-cut*), which learners were exposed to during the pre-task and consolidated in the post-task. Improvement in L2 vowel perception and lexical encoding was gauged through ABX discrimination and forced lexical choice and lexical decision tasks (accuracy and reaction time) respectively, and in production though delayed word and sentence repetition tasks (Mahalanobis distances between contrastive vowels and between non-native and native speakers' productions). Additionally, we calculated the occurrence of pronunciation-focused language related episodes (P-LRE) and assessed learners' individual differences in L2 experience, L2 proficiency, working memory and auditory selective attention.

Results showed that the TBPT intervention enhanced the discrimination and lexical encoding of L2 vowel contrasts, and resulted in significantly more distinct and accurate vowel productions when they were elicited in words in isolation and sentences. Gains in vowel discrimination and vowel quality generalized to untaught lexical contexts and were retained 11 weeks after the intervention. Although simple and complex task groups improved L2 pronunciation after the TBPT intervention, performing cognitively demanding tasks led to greater long-term gains in the discrimination, lexical encoding and production of L2 vowels than performing simple tasks. However, increased task complexity did not have a significant impact on the frequency and duration of P-LRE. As for individual differences, working memory and selective attention explained larger inter-individual variation in L2 vowel performance than English experience, but learner factors were only weakly associated to L2 vowel gains. Last, learners expressed a general sense of enjoyment and learning after the TBPT intervention, but self-perceived L2 pronunciation improvement was especially evident in the complex group. Globally, the present study suggests that orienting learners' attention to phonological form during oral communication is beneficial for L2 pronunciation development, and paves the way for further research in task design and manipulation so as to promote pedagogical practices for pronunciation learning in foreign language classrooms.

Keywords

Task-based language teaching; L2 pronunciation learning; vowel perception; vowel production; individual differences; English as a foreign language; classroom research

Resum

L’aprenentatge de la pronunciació d’una segona llengua (L2) en contextos d'ensenyament de llengües estrangeres és particularment difícil a causa de la quantitat i la qualitat insuficients d’input en L2. Dècades de recerca en adquisició de la parla en L2 han demostrat que la fonologia de la L1 dels aprenents exerceix una forta influència en l'aprenentatge fonològic de la L2 (Best & Tyler, 2007; Flege & Bohn, 2021). Una manera d’atenuar els efectes de la interferència fonològica i de sensibilitzar els aprenents en quant a la importància d'una bona pronunciació en anglès és a través de tasques interactives que fomentin l'atenció a la pronunciació. Malgrat que els estudis en l'ensenyament de la pronunciació a través d'una metodologia centrada en la comunicació han proliferat en els darrers anys (Darcy & Rocca, 2023), pocs han estat els estudis que han aplicat els principis de l'ensenyament de llengües basat en tasques (TBLT), els quals impliquen un ús real de la llengua a l'aprenentatge de la pronunciació de la L2 (Gurzynski-Weiss et al., 2017a). Fins al moment, cap estudi ha avaluat guanys en la percepció, codificació lèxica i producció de les vocals de la L2 per als mateixos participants després d'una intervenció pedagògica. Així mateix, són relativament escassos els estudis que investiguen si les prediccions de la *Cognition Hypothesis* (Robinson, 2011b) es poden aplicar a la fonologia de la L2. Aquesta tesi doctoral pretén contribuir a la recerca sobre l'ensenyament de la pronunciació basat en tasques (TBPT) (1) examinant l'efectivitat de manipular el disseny de tasques i (2) explorant els efectes de l'increment de la complexitat de les tasques en la pronunciació de dos contrastos vocàlics difícils de l'anglès (/iː-ɪ/, /æ-ʌ/); (3) avaluant en quina mesura les diferències individuals poden interferir en la percepció/producció i en els guanys de les vocals de la L2; i (4) obtenint les percepcions dels aprenents sobre TBPT i l'aprenentatge de les vocals de la L2.

Un total de 92 aprenents d’anglès bilingües català/castellà es van dividir en dos grups experimentals i un grup de control. Seixanta-tres aprenents experimentals van dur a terme 20 tasques basades en resolució de problemes en parelles durant 7 setmanes. Es van assignar a grups de complexitat cognitiva simple (*N*=31) o complexa (*N*=32) pel que fa al nombre de raonaments que exigien les tasques. La realització de les tasques forçava la distinció de parells mínims (p. ex., *bean-bin, cat-cut*), els quals es van presentar als aprenents durant la pre-tasca i es van consolidar a la post-tasca. Millores en la percepció de les vocals de la L2 i la codificació lèxica es van mesurar mitjançant tasques de

discriminació ABX i de decisió lèxica (precisió i temps de reacció), respectivament, i millores en producció a través de tasques de repetició de paraules i frases amb demora (distàncies Mahalanobis entre vocals que contrasten i entre les produccions dels parlants no-nadius i nadius). A més, es va calcular l'ocurrència d'episodis lingüístics relacionats amb la pronunciació (P-LRE) i es van avaluar les diferències individuals dels aprenents en relació a l'experiència en la L2, el nivell de competència en la L2, la memòria operativa i l'atenció selectiva auditiva.

Els resultats van mostrar que la intervenció TBPT va millorar la discriminació i la codificació lèxica dels contrastos vocàlics en L2, i va generar produccions vocàliques significativament més dissemblants i precises tant en paraules aïllades com en frases. Els guanys en discriminació i qualitat vocàlica es van generalitzar a nous contextos lèxics i es van mantenir 11 setmanes després de la intervenció. Tot i que els grups simple i complex van millorar la pronunciació de la L2 després de la intervenció TBPT, realitzar tasques cognitivament exigents va produir més guanys a llarg termini en la discriminació, codificació lèxica i producció de les vocals de la L2 que realitzar tasques simples. No obstant això, la complexitat de les tasques no va tenir un impacte significatiu en la freqüència i durada dels P-LRE. Pel que fa a les diferències individuals, la memòria operativa i l'atenció selectiva van explicar una major variabilitat interindividual en la percepció/producció de les vocals de la L2 que l'experiència en anglès; en canvi, els factors individuals dels aprenents només van estar feblement relacionats amb els guanys de les vocals de la L2. Per últim, els estudiants van manifestar la seva percepció d'haver gaudit i haver après amb la intervenció TBPT, però l'autopercepció de la millora de la pronunciació de la L2 va ser especialment evident en el grup complex. En general, aquest estudi suggereix que orientar l'atenció dels aprenents cap a la forma fonològica durant la comunicació oral és beneficiós per al desenvolupament de la pronunciació de la L2, i obre camí per a més investigacions sobre el disseny i la manipulació de tasques per tal de promoure pràctiques pedagògiques per a l'aprenentatge de la pronunciació a les aules de llengua estrangera.

Paraules clau

Ensenyament de llengües basat en tasques; Aprenentatge de la pronunciació de la L2; Percepció de vocals; Producció de vocals; Diferències individuals; Anglès com a llengua estrangera; Investigació a l'aula

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACCESS	Automatization in Communicative Contexts of Essential Speech Segments
AOL	Age of Onset of Learning
ASA	Auditory Selective Attention
CAF	Complexity, Accuracy, Fluency
CAPT	Computer-Assisted Pronunciation Teaching
CEFR	Common European Framework of Reference for Languages
CG	Complex Group
CLT	Communicative Language Teaching
CSd'A	*Consell Superior d'Avaluació del Sistema Educatiu*
CTG	Control Group
DR	Duration Ratio
DSR	Delayed Sentence Repetition
DWR	Delayed Word Repetition
EFL	English as a Foreign Language
EIT	Elicited Imitation Task
FL	Foreign Language
FLeC	Forced Lexical Choice
FonF	Focus on Form
FonFS	Focus on FormS
FonM	Focus on Meaning
GLMM	Generalized Linear Mixed-effects Models
GRAL	*Grup de Recerca en Adquisició de Llengües*
HVPT	High Variability Phonetic Training
ID	Individual Differences
LBQ	Language Background Questionnaire
LD	Lexical Decision
LMM	Linear Mixed Models
LRE	Language-Related Episodes
L1	First Language
L2	Second Language
NRV	Natural Referent Vowel

PAM	Perceptual Assimilation Model
PAM-L2	Perceptual Assimilation Model for the Second Language
PSTM	Phonological Short-Term Memory
P-LRE	Pronunciation Language-Related Episodes
RQ	Research Question
RT	Reaction Times
SBE	Southern British English
SG	Simple Group
SLA	Second Language Acquisition
SLM	Speech Learning Model
SLM-r	Revised Speech Learning Model
SSARC	Stabilize, Simplify, Automatize, Restructure, Complexify
TBLT	Task-Based Language Teaching
TBPT	Task-Based Pronunciation Teaching
T1	Pre-test
T2	Post-test
T3	Delayed Post-test
VKS	Vocabulary Knowledge Scale
WM	Working Memory

INTRODUCTION

Learning the pronunciation of a second language (L2) is fundamental for communicating meaning, which, in fact, is one of the main goals of second language acquisition (SLA). In Foote and Trofimovich's (2018) words, "pronunciation [is believed to] permeate all sphere of human life, lying at the core of oral language expression and embodying the way in which the speaker and the hearer work together to produce and understand each other's utterances" (p.85). Therefore, learning L2 pronunciation is not simply about mastering a series of decontextualized phonetic forms and imitating a model, but it is about learning to convey the intended message in an intelligible and comprehensible way (Levis, 2018, 2022). Beyond its obvious connection with speaking, L2 pronunciation is intrinsically related to listening comprehension, reading and written production because pronunciation helps identify aspects of connected speech in careful and spontaneous speech (Cauldwell, 2018), as well as to be able to connect the written form of words and sentences to their spoken forms. "Spoken language is sound—and sound gives life to grammar and vocabulary. Without the sound (that is, the phonology, or the pronunciation), one cannot bring the rest of language to life" (Darcy, 2018, p.13). Despite the considerable importance of mastering pronunciation for L2 communication (Moyer, 2014), oral skills and pronunciation are not systematically taught in Foreign Language (FL) classrooms (Fouz-González, 2020; Henderson et al., 2015; Rallo, 2022; Tragant et al., 2010), where traditional approaches to FL teaching tend to relegate pronunciation to the background of L2 students' learning.

For many years, the English proficiency level has not been satisfactory in the instructional context of Catalan schools (Tragant, 2009), with a total of 65% of learners obtaining a B1 level in the Common European Framework of Reference for Languages

(CEFR). Teachers have reported placing excessive emphasis on grammar and vocabulary and too little to oral production, and have claimed the need for more functional and communicative approaches (Tragant et al., 2010). More recently, state evaluation reports periodically conducted by *Consell Superior d'Avaluació del Sistema Educatiu* (CSd'A, 2020) on learners' oral abilities in English as a Foreign Language (EFL) at the end of secondary school (Year 11) have shown an increase in learners' pragmatic, sociolinguistic and linguistic competence in English. In general, CSd'A (2020) reports have shown significant differences between high and low complexity schools in terms of learners' English oral abilities: whereas 50% of learners do not reach sufficient level of intelligibility in oral production in high complexity schools, 22.5% and 5% do not reach it in mid- and low-complexity schools, respectively.[1] Interestingly, 60% of evaluated learners had high scores on interactions but 41.5% of learners were found to have a low-to-mid level in segmental (individual sounds, lexical stress) and suprasegmental features (sentence stress, intonation) of speech. Despite the considerable increase in exposure to native English input through original television series and participation in extra-curricular activities (CSd'A, 2020), which has undoubtedly aided the development of listening and speaking, state-run schools do not yet provide much practice in pronunciation and oral skills (Rallo Fabra, 2022; Tragant et al., 2010).

The small quantity and low quality of input in FL instructional contexts (Muñoz, 2014; Tragant et al., 2010) is one of the main factors affecting L2 pronunciation (Carlet & Rato, 2015). On the one hand, the quantity of language input is usually restricted to the teacher's input, which is sometimes delivered in the learners' L1 (Muñoz, 2014), and limited outside the FL classroom; on the other hand, the quality of the input received

[1] According to CSd'A (2020), the level of complexity in schools is defined in terms of student diversity (e.g., students with special needs or low socioeconomic status), teacher and student mobility, student absenteeism and demand for schooling.

tends to be of a foreign-accented nature due to the massive exposure to input from non-native peers (García Lecumberri & Gallardo del Puerto, 2003).

Another often-cited barrier for L2 pronunciation development is the lack of pronunciation focus in EFL instruction because teachers prioritize the teaching of other skills (grammar, reading, writing) over communicative competences. Most published materials do not integrate pronunciation into speaking activities or are presented as an extra skill which can be skipped (Levis & Sonsaat, 2016); many teachers often lack confidence in their ability to teach pronunciation (Henderson et al., 2015; Kirkova-Naskova et al., 2021; Uchida & Sugimoto, 2018); curricular time pressure makes it hard to incorporate pronunciation in the EFL lessons (Burri, 2021; Darcy, 2018); and there is a lack of external pressure from standardized pre-university tests (*Proves d'Accés a la Universitat* - PAU) to assess L2 speaking skills or pronunciation. In addition, methodological drawbacks such as too many students per class and great heterogeneity in terms of L2 proficiency (Tragant, 2009; Tragant et al., 2010) make the teaching of L2 oral abilities especially challenging. Beyond the influence of the L1 phonological system as well as non-linguistic factors (e.g., motivation, anxiety, aptitude), FL instruction in pronunciation is crucial to prevent pronunciation errors from fossilization and compromise intelligibility (Rallo Fabra, 2022) in the long term.

Research conducted in the Catalan/Spanish FL context has shown that accurate perception and production of L2 vowels is difficult to achieve by only receiving input from the FL classroom (Fullana, 2006; Mora & Fullana, 2007). Difficulty in non-native categorization lies in the degree of cross-language perceptual similarity between L2 sounds and L1 categories. In other words, L1-based perception limits L2 phonetic learning, especially when L2 sounds are perceptually mapped onto single categories (Best & Tyler, 2007), as it is the case of English /æ/ and /ʌ/ assimilated to the Catalan low vowel category /a/ (Cebrian, 2006; 2021). Reliance on duration cues to distinguish L2 English vowel contrasts, which are primarily distinguished qualitatively in terms of spectral cues, often result in inaccurate L2 vowel pronunciation (Aliaga-Garcia & Mora, 2009; Cebrian, 2006; Mora & Fullana, 2007).

Considering the lack of pronunciation improvement in EFL classrooms after years of instruction (Fullana, 2006), there is an urgent need to investigate effective methods for L2 pronunciation learning. Unless a lot of L2 experience is received (Saito, 2015), accompanied by corrective feedback (Saito & Lyster, 2012), or learners undergo phonetic training (Aliaga-Garcia & Mora, 2009; Thomson, 2012), explicit (Gordon & Darcy, 2016; Kissling, 2013) or incidental (Mora-Plaza et al., 2018; Solon et al., 2017) pronunciation instruction, learners' attention is likely not to be oriented to phonetic form.

Empirical studies on L2 pronunciation instruction have found that including a communicative component in the teaching of segmental and suprasegmental aspects of L2 pronunciation is more beneficial for the development of intelligibility, comprehensibility and L2 pronunciation accuracy in spontaneous speech than a decontextualized focus on L2 phonological forms (Darcy & Rocca, 2023; Darcy et al., 2019; Saito 2012, 2015). When L2 pronunciation is practiced repetitively, under meaningful interaction, L2 phonological and phonetic processing becomes more easily automatized and learners are better equipped to generalize what they have learned in the classroom to real-world conversations (Segalowitz & Hulstijn, 2005; Trofimovich & Gatbonton, 2006). In form-focused communicative instruction, L2 phonological features are usually contextualized in meaningful tasks after receiving explicit pronunciation instruction (e.g., Gordon, 2021), instead of being practiced incidentally during communicative interaction following a task-based approach, as the current thesis proposes.[2]

Task-based language teaching (TBLT) is now firmly established as an approach rooted in psychological approaches to SLA and educational principles (Long, 1985, 2015). TBLT has been subject to extensive research exploring the effect of task design, manipulation and their implementation on L2 acquisition (Ellis et al., 2019). Even though a wealth of research has focused on the effectiveness of tasks for L2 oral performance (CAF: complexity, accuracy and fluency) and development (lexico-grammatical and pragmatic), much less research has investigated the potential of tasks to promote attention to phonetic form and lead to L2 pronunciation development. Studies conducted under a task-based pronunciation teaching (TBPT) approach (e.g., Gurzynski et al., 2017a) have shown positive results about the integration of pronunciation in meaningful tasks. Nevertheless, further investigations are needed to understand the effectiveness of task design and task manipulation for L2 phonological development, as well as the most appropriate methods to assess L2 pronunciation proficiency (Mora & Levkina, 2017).

This thesis seeks to contribute to the growing TBPT line of research by examining the effects of tasks and task complexity on the development of two challenging L2 vowel contrasts (/iː-ɪ/-/æ-ʌ/) by adolescent Catalan/Spanish learners of English. First, the findings from this study intend to provide a better understanding of the potential of tasks for L2 pronunciation learning, beyond lexico-grammatical and pragmatic development (Gurzynski et al., 2017b). Second, this study follows up on Solon et al. (2017) and Gordon (2021) and expands them by investigating whether the Cognition Hypothesis' predictions (Robinson, 2011b) can apply to the perception, lexical encoding and production of L2

[2]Form-focused instruction is an umbrella term which refers to any pedagogical technique that draws learners' attention to language (Long, 1997), through any pedagogical technique (e.g., reactive, proactive, explicit, implicit).

vowels after a considerably long intervention. Third, this study contemplates the role of individual differences (ID) in mediating L2 pronunciation gains, as previous L2 speech studies have assessed (Mora & Mora-Plaza, 2019; Mora-Plaza et al., 2022; Suzukida & Saito, 2023). Last but not least, this dissertation seeks to complement quantitative results by offering insights into learners’ beliefs and impressions about learning pronunciation through tasks and qualitative comments about self-perceived improvement.

Organization of the dissertation

The current dissertation is organized into six main chapters. Chapter 1 presents the theoretical background to this study, including an overview of the speech acquisition processes and second language development, a cross-linguistic comparison between L1 and L2 vowel inventories, a review of pronunciation instruction, TBLT and TBPT empirical investigations and an exploration of learner factors mediating L2 speech acquisition. Chapter 2 states the main contributions, describes the main objectives and research questions, presents the target sounds of the study and, subsequently, formulates hypotheses grounded on previous literature. Chapter 3 explains the methodology of the current study, describing the participants, the experimental design and timeline, the materials, the general procedure and the analyses conducted to analyse the data. Chapter 4 reports the results of the investigation and is divided into four sub-chapters corresponding to the four main research questions (i.e., the effectiveness of TBPT, task complexity effects, the role of ID and learners’ perceptions of the intervention). Chapter 5 provides the reader with the discussion and interpretation of the study’s results, comparing them with previously published empirical evidence. Finally, Chapter 6 summarizes the main findings of the study, acknowledges some study limitations, makes suggestions for further research and proposes several pedagogical implications.

01

LITERATURE REVIEW

CHAPTER 1. LITERATURE REVIEW

The aim of Chapter 1 is to explore the processes involved in speech perception, production and lexical encoding and frame them in speech acquisition models as well as to examine the role of input and attention (Section 1.1.); present the Catalan and Standard British English vowel inventories and show cross-linguistic similarities and differences to understand inaccuracies in L2 pronunciation learning (Section 1.2.); summarize main findings in pronunciation training and instruction in the context of second/foreign language learning and its potential challenges (Section 1.3.); review the literature on TBLT and task complexity together with its implementation challenges (Section 1.4.); introduce TBPT as a methodology to focus on phonetic form during interaction (Section 1.5.); and assess ID (i.e., experiential, socio-psychological and, especially, cognitive) that may contribute to explain gains in L2 vowel perception and production (Section 1.6.).

1.1. Speech acquisition and second language development

The first section of Chapter 1 focuses on the processes involved in L2 speech acquisition. In particular, Section 1.1. presents how speech perception and speech production processes work in native and non-native language systems; how theoretical models explain perception and production difficulties in L2 speech development in second and FL contexts; to what extent perceiving L2 sounds accurately entails their lexical encoding and updating of phonolexical representations; in what way speech production models explain production difficulties; and how input noticing, selection, attention and memory capacities are relevant in SLA processes.

1.1.1. Speech perception and production

The present dissertation assesses L2 pronunciation improvement mainly in terms of perception and production. Therefore, we detail L1/L2 perception and production processes and which tasks and scoring methods have been used to measure speech perception and production in L2 research. In addition, we describe three different scenarios concerning the existing link between perception and production.

Concerning perception, research has shown that up to 6 months of life, infants can discriminate nearly any phonetic contrast in any of the world's languages, but after that point, infants begin losing this ability, namely, ignoring phonetic information which is not required to categorize L1 speech sounds (Kuhl & Iverson, 1995). This coincides with the emergence of robust L1 specific phonological categories. By the age of 4, children's ability to perceive foreign language contrasts has diminished to a point they do not perform better than adults (Werker, 2018). In terms of production, at 10 months the infant's babble closely reflects the surrounding input so learners become attuned to their L1 and are less capable of learning new categories. In fact, infants modify their productions in real time to approximate their interlocutors' pronunciation (Thomson, 2022).

At the articulatory level, L1 speech involves automatized articulatory gestures such as the production of accurate vowels and consonants concerning their voicing, place and manner of articulation or duration, as well as correct lexical stress placement, rhythm and intonation patterns (Mora & Levkina, 2017). Nevertheless, L1 phonological knowledge is hard to be inhibited when processing L2 input. As a result, L2 articulation gestures may not be accurate or fast enough to produce nativelike L2 vowels and consonants (segmentals) and lexical stress, rhythm and intonation (suprasegmentals), often leading to accented speech.

The reality is that adult L2 learners are not blank slates, they already possess strongly established L1 speech categories (Flege et al., 2003) so they tend to perceive non-native sounds in terms of their native categories. This phenomenon, called *L1-based processing*, is characterized by transfer from L1 phonological knowledge during L2 processing. Although the effects of early language experiences negatively affect our perceptual flexibility, humans remain capable of some degree of phonetic learning. Munro & Derwing (2008) present longitudinal data for L2 vowels, confirming that during early periods of intensive L2 exposure, L2 speech production can improve without explicit instruction. Within a few months, however, average improvement seems to plateau far short of native-like ability.

On the one hand, L2 segmental perception has been typically measured in research through a Categorization task (Llompart, 2021a); Forced Choice Identification task (Carlet & De Souza, 2018), a discrimination (ABX) task (Melnik-Leroy et al., 2022) or, the more challenging version, oddity task (Darcy & Daidone, 2021). In most cases, a proportion of accurate responses and averaged reaction times has been calculated. The use of tasks containing multiple talkers (Thomson, 2011) allows us to obtain mean perceptions scores that average out listener responses to speech produced by individual talkers.

On the other hand, measuring production requires eliciting and recording speech, which can be obtained through spontaneous (monologic or dialogic) or more controlled tasks. One way is to employ immediate word/sentence repetition tasks (Kabakoff et al.,

2020) which measure phonological short-term memory (PSTM[3]) and mirror properties of the stimulus rather than the speakers' typical ability to produce the target sounds. Alternatively, delayed word/sentence repetition tasks (Munro & Derwing, 2008; Mora et al., 2022) are believed to gauge the state of the speakers' long-term phonological representations, and avoid reproducing the input's phonetic properties from sensory memory (Thomson, 2022). Concerning L2 production assessment, perceptual dimensions of L2 speech can be evaluated subjectively through listeners' ratings, such as comprehensibility[4], intelligibility[5], fluency[6] and accentedness[7] (Derwing & Munro, 2015; Munro & Derwing, 1995). Alternatively, specific constructs like segmental accuracy, lexical stress, intonation, or breakdown fluency can be operationalized through subjective (i.e., listener-based identification tasks; Rallo Fabra & Romero, 2012) or objective (i.e., acoustic analysis of phonetic and phonological properties of speech; Kartushina et al., 2015; Melnik-Leroy et al., 2022) measures. Objectively, for instance, accuracy in distinguishing vowels can be measured through acoustic distance in the F1 (height) × F2 (frontness) vowel space (in Frequencies or Bark), where greater acoustic distinctiveness of a non-native contrast implies a better command of L2 (Kartushina & Frauenfelder, 2014). See Mora (2021) and Saito and Plonsky (2019) for comparisons between global and specific subjective and objective measures, and Section 3.3.3.3. for a discussion on the acoustic distance measures employed in the present dissertation.

[3] PSTM is a short-term phonological store for verbal information which allows individuals to encode phonological units and their sequential order in the form of auditory traces that can be sustained in memory for further processing through silent articulation mechanism (Baddeley, 2007).

[4] Comprehensibility is defined as the degree of effort required of a listener to understand a speaker's message (Derwing & Munro, 1995; 2022).

[5] Intelligibility means the degree to which a listener understands the speaker's intention (Levis, 2005; Derwing & Munro, 1995; 2022).

[6] Fluency refers to how quickly and smoothly the speech flows (Derwing & Munro, 2022).

[7] Accentedness is defined as the degree to which one's own speech deviates from that of a local/native community (Derwing & Munro, 1995; 2022).

Provided that the present dissertation explores the relationship between L2 perception and production, three diverging views are presented: perceptual learning precedes production learning but accurate perception does not entail accurate production; the two skills develop simultaneously; production changes might take place prior to perceptual changes.

Regarding the *perception leads production* hypothesis, extensive literature assumes that accuracy in the perception of L2 speech sounds precedes their accurate production (Bradlow et al., 1997; Flege, 1995; Melnik-Leroy et al., 2022) or consider a time-lagged approach to the perception–production link, wherein production would lag behind perception but eventually catch up with it (Casillas, 2020; Nagle, 2021). However, numerous factors such as age of L2 learning (AOL), L1 background or frequency of L2 use may affect L2 perception and, consequently, L2 production (Bohn & Flege, 1990; Kuhl & Iverson, 1995). Furthermore, perception-based training with no production practice component have found gains in segmental production (Huensch, 2016; Sakai & Moorman, 2018 for meta-analysis).

Flege and Bohn's (2021) Revised Speech Learning Model proposed that segmental *production and perception coevolve* without precedence. Such co-evolution hypothesis is supported by Turner (2022) who found that, English learners of L2 French with higher processing scores for French vowels /y/ and /u/ were more likely to mark a greater acoustic distinction in their production of this contrast.

Regarding the *production leads perception* hypothesis, Sheldon and Strange (1982) found that Japanese learners of English showed better production than perception of /r/-/l/ so perceptual mastery of L2 sounds did not necessarily precede mastery in production. Few studies have shown evidence of positive carry-over effects of production-based training/instruction to perception gains (Carlet & De Souza, 2018; Kartushina et al., 2015).

1.1.2. Speech perception models

Having reviewed the processes and assessment of L2 speech perception, this section aims at describing recent theories which link the ability to categorize L2 sounds accurately to the ability to discern differences between native and non-native sounds: the Speech Learning Model (Flege, 1995; Flege & Bohn, 2021), the Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007) and the Natural Referent Vowel framework (Polka & Bohn, 2003, 2011). Even though the present study is not directly testing the ability to discern L1 from L2 sounds, it is concerned with assessing L2 sounds that are highly influenced by L1 sounds after pronunciation instruction, in light of the following L2 speech models[8].

Flege's (1995) Speech Learning Model (**SLM**) and Flege and Bohn's (2021) expanded Revised Speech Learning Model (**SLM-r**) argue that L2 learners retain the ability to modify existing phonetic categories and develop new ones for L2 sounds if they are provided with favourable input conditions. In other words, their phonological system remains malleable over the life span and adapts to the phonetic input that is received. Within this perception-based account of L2 speech learning, learners' accuracy in L2 speech production reflects the nature of the developing L2 phonetic representations stored in long-term memory. Flege (1995) predicts that, when an L2 sound is distinct from perceptually similar L1 sounds and the L2 learner is sensitive to such differences, it is

[8] Other L2 speech models such as the Native Language Magnet model (Kuhl & Iverson, 1995) and the Second Language Linguistic Perception model (Escudero, 2009) are not included in this section as they are not relevant to the present investigation, but should be definitely considered in other L2 speech perception and production research.

likely to be perceived as a "new" sound and a separate L2 sound category may be established (e.g. L2 English /ɜː/ having no clear match in L1 Catalan/Spanish). If sounds are almost indistinguishable from the L1 counterparts, they are predicted not to pose any problems to L2 learners, as no category formation is required (e.g., L2 English /iː/ mapped onto L1 Catalan/Spanish /i/). Conversely, when an L2 sound is highly "similar" perceptually to an L1 sound, it is less likely that a separate L1 category is formed (e.g., L2 English /ʌ/ mapped onto L1 Catalan /a/). According to SLM-r (Flege & Bohn, 2021), when a new category is not formed for L2 sounds that differ phonetically from the closest L1 sound, a composite L1–L2 phonetic category will develop and this will lead to L1-based foreign-accented productions.

As far as production is concerned, the SLM attributes L2 phonological errors mostly to incorrect perception; conversely, the SLM-r (Flege & Bohn, 2021) posits that perception and production coevolve in time, being mutually influential. However, the quantity and quality of input play a role in the perception, and subsequent production, of cross-language differences. In FL contexts, where input is often scarce (Muñoz, 2014), formal instruction should begin at an early age, learners should be exposed to high quality input, and provide some kind of perception and production training to orient learners' attention on L1 and L2 phonetic differences (Tyler, 2019).

Whereas the SLM focuses mainly on experienced listeners (Flege, 1995), the Perceptual Assimilation Model (**PAM**) by Best (1995) focuses primarily on naive listeners and assumes a learner who is actively acquiring an L2 in an immersion setting. PAM's core principle is grounded on the fact that perception is directly created from the environment. Therefore, during the first months of life, sound perception is attuned to the L1 inventory and, consequently, identifying familiar sounds becomes natural. After that, perceiving speech sounds that differ from those in terms of gestures becomes more difficult. Best & Tyler (2007) extend the original model to L2 perception (**PAM-L2)** which relies on the notion of cross-language similarity to predict the likelihood of acquiring new L2 sound contrasts when a learner is learning a non-native language. PAM-L2 postulates that the L1 does not necessarily hinder L2 perception. More specifically, PAM-L2 posits that “categorisation” takes place when a non-native sound is perceived as existing in the native language, whereas it is “uncategorised” if it fails to be ascribed to it. PAM/PAM-L2 describes several patterns of perceptual relationships between L1 & L2 sounds: (1) Single-Category Assimilation: Contrasting L2 sounds are perceptually identified as a single L1 sound (i.e. there is no L1 phonological contrast to support discrimination), and neither is really a good fit (*Discrimination is poor*), (2) Category-Goodness Assimilation: two L2 sounds map onto the same category in the native sound system, with one judged better exemplar of this category (*Discrimination is moderate to good*); (3) Two-Category Assimilation: each contrasting sound is perceptually linked to a unique L1 sound (*Discrimination is excellent*); (4) Uncategorized-Categorized assimilation: one L2 sound is assimilated to a native category and the other one is uncategorized (*Discrimination is good to very good*); (5) Uncategorized-Uncategorized: two non-native sounds are uncategorized so difficulty may depend on how close the two non-native sounds are from each other and from the L1 phones (*Discrimination ranges from poor to excellent*).

Best and Tyler (2007) suggested that perceptual attunement should happen prior to the establishment of a large L2 vocabulary so that an increasing vocabulary drives perceptual reattunement to the L2 phonology. Whereas L2 learners in an immersion environment with rich native-speaker input may have ample opportunity for perceptual learning, classroom FL learners may not have the same opportunities and may change the predicted outcomes.

Tyler (2019) mentions several factors affecting phonological category acquisition in the FL classroom. First and foremost, interactions in instructed SLA are likely to be foreign-accented because the teacher and peers are likely to speak the target language with a non-native accent. Hence, this may imply that the average learner is not often exposed to speech they can use as perceptual referents for identifying L2 phonological contrasts. Unless speech maintains a phonological distinction between all L2 phonemes, learners' likelihood of acquiring them may be rather low. Secondly, FL instruction is based on the use of written language to teach vocabulary and grammar, with very little time dedicated to teach pronunciation. Tyler (2019) claims that this may reduce the window of time available for discriminating L2 contrasting sounds. If words have been learned in the absence of spoken input, the learner may have already developed a large vocabulary of L2 words via orthography and, consequently, fossilization of L2 phonological forms may occur. Taking these factors into account, it is predicted that category-goodness and uncategorized-categorized assimilations are less likely to be acquired in the classroom than in immersion contexts, especially if learners are exposed to foreign-accented spoken input, or a vast amount of lexis is learned on the basis of their written form (Tyler, 2019).

Last but not least, the Natural Referent Vowel (**NRV**) framework (Polka & Bohn, 2003, 2011) is based on the directional asymmetries noticed in how infants and adults perceive a shift from one vowel sound to another. These vowel asymmetries can be predicted taking into account the position of each vowel in the vowel space in terms of vowel height (first formant [F1] frequency) and frontness (second formant [F2] frequency). The main tenet is that identifying a change from repeating a vowel that is acoustically and articulatory more peripheral in the vowel space to a more centralized one (e.g., English /iː/→/ɪ/) is more difficult than detecting the same change in the opposite direction (e.g., English /ɪ/→/iː/). The authors claim that more peripheral vowels act as perceptual anchors to guide the development of native vowel categories.

Polka and Bohn (2003) state that infants are born with a reference point for each vowel category that is based on the natural resonant frequencies of their vocal tract. Infants use these reference points to form categories and differentiate between vowel sounds, and their discrimination ability is influenced by the distance between the reference points for the different vowel categories. Asymmetries in infant vowel discrimination are robust, predictable, and do not dependent on specific language experience. It is hypothesized that, in adults, asymmetries will fade for native contrasts and be maintained or enhanced for non-native contrasts. Polka and Bohn (2011), who tested Danish-speaking adults' perception of L2 English vowel contrasts, showed that proportion of correct responses from a discrimination task differed significantly as a function of direction of presentation: adults made significantly more errors when the vowel changed from the more peripheral to the less peripheral vowel compared to the opposite direction. The authors also revealed that bias favouring peripheral vowels could change if attention to phonetic detail was provided to L2 learners.

All in all, the NRV framework posits that (1) vowels with extreme articulatory-acoustic properties (peripheral) act as natural referent vowels which guide perceptual learning; (2) with experience, the perceptual bias favouring these natural referent vowels may be adapted to language-specific vowel categories but will be enhanced when perceivers are mapping out a new vowel system to learn a second language; and (3) vowels that approach formant convergence limits (focal vowels) are more acoustically salient and stable than less-focal vowels, whose formant frequencies do not converge. Therefore, these natural reference vowels may act as production targets that infants can aim for when they are expanding their emerging vowel production skills.

All in all, the three models (i.e. SLM, PAM-L2, NRV) agree that L2 language experience is fundamental in determining L2 phonological development as it might enhance the capacity to establish new categories of a target language (SLM, NRV), or re-organize the L1 categories (PAM-L2). The findings of the current dissertation are mainly discussed in view of the assimilation patterns described in PAM-L2 (Best & Tyler, 2007), with occasional reference to the SLM-r (Flege & Bohn, 2021) and NRV (Polka & Bohn, 2003) models.

Finally, developing sensitivity in the perception or production of L2 phonological contrasts at pre-lexical phonetic and phonological level does not necessarily mean that such sensitivity is present when sounds and phonological contrasts appear in the lexicon, as will be discussed in Section 1.1.3.

1.1.3. Lexical encoding and phonolexical representations

One of the current dissertation's objective is to assess learners' lexical encoding of L2 sounds after task-based pronunciation instruction and explore potential links with L2 perception gains. Therefore, Section 1.1.3. explores the symmetries and asymmetries between perception and lexical encoding through various theoretical perspectives; presents factors that may affect the update of imprecise phonological representations; and offers instructional techniques on how they could be possibly updated.

It is well known that the phonetic properties of phonological representations are maintained in the phonological representation of lexical items. As vocabulary grows, the phonological form of words (i.e. phonolexical representations) develops and it is activated and retrieved during spoken word recognition processes. However, in the L2, these representations are not always precise because the L2 phonology may be greatly influenced by the L1, hence, L2 word forms are represented in terms of our native language phonological representations. Imprecision at the phonolexical level may not only lead to problems at word recognition levels (i.e., listening and reading comprehension), but also negatively impact production (i.e., speaking) in the L2.

Two theoretical views, “category first” and “lexicon first” (Darcy et al., 2013) have developed to explain “fuzzy” phonolexical representations. On the one hand, the “categories first” view (Ota et al., 2009; Pallier et al., 2001) assumes that perception determines phonolexical representations so what cannot be perceived, cannot be learned, thus, improving perception might be the way to accurate representations. In fact, most models of L2 speech acquisition (i.e., SLM, PAM-L2) propose a direct link between perception abilities and accuracy of phonological representations in the lexicon. The first to show a one-to-one correspondence between phonetic distinction and lexical contrasts was Pallier et al. (2001) who investigated the lexical encoding of the /e/- /ɛ/ Catalan contrast by Spanish-dominant bilinguals. Results showed that Spanish-dominant bilinguals treated /e/- /ɛ/ minimal pairs (“té” /te/– “tè” /tɛ/) as homophones (perceived only /e/) and related the lacking lexical distinction to the fact of not having established distinct phonetic categories for the Catalan phones. Similarly, Ota et al. (2009, p. 267) found that “the lexicon of late bilinguals indeed fail[ed] in completely separating L2 lexical entries that involve non-native phonological contrasts”, giving support to the assumption that the ability to establish phonetic categories is a pre-requisite for encoding the contrast lexically.

On the other hand, the “lexicon first” view (Cutler et al., 2006; Darcy et al., 2013; Weber & Cutler, 2004) advocates for an asymmetric mapping from phonetic to lexical representations, hence, learners can establish a lexical contrast even if the perceptual categorization of the contrast is still not robust. For instance, Weber and Cutler (2004), following a visual-word paradigm, challenged the “homophones-creation view” from Pallier et al. (2001) as they found that Dutch learners of English had stored separate lexical entries containing the L2 vowels /æ/, /e/ (e.g., “panda” /pæ̲ndə/- “pencil” /pɛ̲nsl/) without reliable support from perception. Darcy et al. (2013) corroborated the findings from Weber and Cutler (2004) and Cutler et al. (2006) in that L2 learners could distinguish relevant L2 contrasts with high accuracy, and yet, showed an asymmetrical lexical encoding. In the same line, Darcy et al. (2012b) reported that advanced learners of French had established distinct lexical entries for /y/, /u/, /oe/, /ɔ/ based on all tested French minimal pairs in the lexical decision task with repetition priming; however, exhibited persistent perceptual errors when categorizing the contrasts in the ABX categorization task, similar to the ones produced by the intermediate proficiency group. Darcy et al. (2013) extended this line of research by providing further evidence that L2 learners’ lexical representations are quite detailed even if the lexical encoding of contrasts still receives the influence of L1 dominant categories.

This lack of symmetry between sensitivity in perceiving a phonological contrast in a sound discrimination task and a discrimination task based on a lexical contrast (i.e., phonetic perception and lexicon being independent) has also been shown in correlational analyses (Darcy & Holliday, 2019) where the strength of the correlations between categorization, identification, discrimination tasks and lexical-decision, forced-lexical choice, word-picture matching, word-recognition tasks has been weak to moderate, or non-existent (Daidone & Darcy, 2021). Llompart (2021a) argues that perceptual categorization abilities do not account for individual differences in lexical encoding in very advanced learners, so perception of L2 sound contrasts may be necessary initially but not sufficient to develop accurate phonolexical representations. In sum, accurate discrimination may not guarantee accurate lexical encoding, hence, maintaining a robust phonetic difference and storing separate lexical representations for words involving that phonetic difference may involve different processes (Darcy & Holliday, 2019; Darcy et al., 2013; Hayes & Masuda, 2008).

As far as production is concerned, Hayes-Harb and Masuda (2008) investigated English–Japanese learners' lexical representations of Japanese words containing geminate consonants and concluded that they could lexically encode a geminate /tt/ consonant as /t*/ even if they had not yet determined specifically how /t/ and /t*/ differed (see also Cutler et al., 2006). In addition, Simonchyk and Darcy (2021) found that English learners of Russian could improve their production of palatalization even in the absence of precise phonolexical representations. They concluded that the relationship between lexical encoding and production may not be very transparent as it may be affected by a myriad of linguistic and extralinguistic factors.

Initial problems with difficult L2 phonological contrasts may lead to fuzzy L2 phonolexical encoding, hence, fuzzy phonolexical representations that can persist over time (i.e., *Fuzzy Lexicon Hypothesis*: Cook & Gor, 2015). Given that some imprecise representations seem to be very resistant to change, L2 speech researchers have investigated several factors which may explain this lack of malleability:

a) Orthography: L2 learners may deploy orthographic knowledge to learn sound contrasts. Hayes-Harb et al. (2018), using a word-learning paradigm, demonstrated that learners who had been exposed to the letters of test words were more likely to produce final voiced stops when naming pictures, suggesting that orthographic input can interfere with the L2 acquisition of allophonic variation. Similarly, Charoy and Samuel's (2020) findings suggested that listeners built phonological representations from the words' printed forms, and that these representations guided their spoken word recognition.

b) Vocabulary: Vocabulary size is hypothesized to be a potential predictor of the encoding of difficult L2 phonological categories in the lexicon. Llompart (2021a) found that, for advanced German learners of L2 English, /ɛ/-/æ/ nonword rejection results (i.e., identify "thank" */θɛŋk/ as a nonword) could be predicted by their scores in the vocabulary test. Similarly, Daidone and Darcy (2021) revealed that vocabulary size (vs. perception, phonological short-term memory, attention and inhibitory control) was the factor with the largest impact on L2 lexical encoding for most of the tested contrasts.

c) Familiarity, age of words, lexical frequency, neighbourhood density: Cook and Gor (2015) predicted that low level of familiarity resulting from low exposure could result in weak and fuzzy phonolexical representations. In addition, Darcy and Thomas (2019) and Llompart and Reinisch (2021) suggested that older lexical representations could be more resistant to updates than recently acquired lexical representations because learners' phonological knowledge was presumably less accurate at the time they learned the first words but they could rapidly be updated. Finally, Llompart (2021b) showed that nonwords with low-frequency word counterparts were more easily rejected (i.e., identified as nonwords) than those with high-frequency word counterparts and nonwords made of words with more lexical neighbours were more easily rejected than those with fewer neighbours. For example, L2 English high-frequency word *thank* was more often heard as *th[ɛ]nk by L1 speakers than not non-frequent words such as *habit* hear as *h[ɛ]bit.

Individual-learner factors such as L2 proficiency (Llompart 2021a, 2021b) and attention control (Darcy et al., 2013; Daidone & Darcy, 2021) may well also be the cause of "fuzziness" in learners' L2 phonolexical representations.

Finally, whereas pronunciation instruction (Lee et al., 2015; Saito and Plonsky, 2019) and phonetic training (Aliaga-Garcia & Mora, 2009) have demonstrated to be effective at improving the perception and production of L2 sounds, evidence for the effectiveness of instructional techniques at improving the encoding of L2 sound contrasts and updating phonological representations is scarce.

On the one hand, Llompart and Reinisch (2021) assessed the effect of phonological specificity training during word learning on the lexical encoding of the English /e/-/æ/ contrast into novel L2 minimal pairs by German learners of English. It appeared that recently established L2 lexical representations were flexible enough to allow for their updating (in line with Darcy & Holliday, 2019) once learners' phonological knowledge improved thanks to the phonologically-focused training, that is, once learners were prompted to make use of contrastive information regarding the challenging L2 vowel distinctions.

On the other hand, the effects of high variability phonetic training (HVPT) on the lexical encoding of L2 sounds have been mixed and lower than on perceptual discrimination/identification. Whereas Mora and Mora-Plaza (2019) and Mora-Plaza et al. (2022a) reported little or no improvement in the sensitivity of L2 English /æ-ʌ/ (i.e., rejecting */sæn/ as a real word) embedded in the target words after four 40-min phonetic training sessions, Adrian and Mora (2022) found that HVPT helped learners improve how precisely they encoded the lexical contrast, especially those who initially had a poor lexical encoding. Finally, in Melnik and Peperkamp (2021), French learners of English completed eight online sessions of HVPT and were tested on their prelexical perception and their lexical processing of stimuli of English word-initial /h/. Findings indicate that HVPT could improve not only L2 learners' prelexical processing, but also their lexical processing, and such positive training effects were retained 4 months after the post-test.

Nevertheless, learners' initial good performance (i.e., over 70% in the lexical decision task) and talker familiarity (i.e., training and testing tasks shared the same talkers) could have played a role in the improvement of phonolexical representations. To our knowledge, the present study represents an initial effort to investigate whether form-focused communicative tasks may lead to the lexical encoding of L2 sounds in FL classroom settings.

1.1.4. Speech production models

Although L2 speech acquisition models have attempted to explain how perception may limit production accuracy and possibly lead to foreign accent (Flege et al., 1995; Flege, et al., 1997), less attention has been paid to how articulation may account for accurate L2 production in speech learning. However, the control of oral-motor movements and self-monitoring of phonological errors are of utmost importance for the accurate production of sounds. For this reason, this section outlines the main tenets of Levelt's (1989), De Bot's (1992) and Kormos' (1999) models of speech production, with special attention to the role of articulators and phonological monitoring.

Firstly, the most widely accepted and influential model in SLA and psycholinguistic research is Levelt's (1989) model of L1 production which consists of a number of autonomous modules which are responsible for different aspects of speech production, namely, the conceptualizer, the formulator, the articulator, the audition/ acoustic-phonetic processor and the speech comprehension system/parser (Figure 1.1.).

1. The conceptualiser is responsible for generating a communicative intention and conceptualizing the message, as well as monitoring the whole production process (pre-verbal, pre-articulatory, overt speech). Levelt distinguishes two stages in

message planning: macroplanning (i.e., generating the speech act intentions such as expressing an opinion) and microplanning (i.e., dividing smaller conceptual ‘chunks’ and give them the correct propositional shape and informational perspective).

2. The formulator is in charge of giving the grammatical and phonological shape to messages by activating the items (lemmas and lexemes) that best correspond to the intended message.
3. The articulator is a simple motor output consisting of a computational neural system that controls the motor system (i.e., speech organs such as the lungs, larynx and vocal tract). It is in the articulator that internal speech (i.e., the phonetic plan) becomes overt speech by articulating the message.
4. The audition/acoustic-phonetic processor transforms the acoustic signal into phonological representations.
5. The speech comprehension system/parser anatomizes the message produced by one-self or others. In terms of internal speech, the preverbal plan is checked in working memory (WM) before being articulated. In terms of overt speech, the acoustic-phonetic processor detects the articulated words and speech comprehension retrieves meaning.

These components do not interact and each contains its own procedural knowledge and the only possible feedback is that provided by internal or overt speech by means of monitoring.

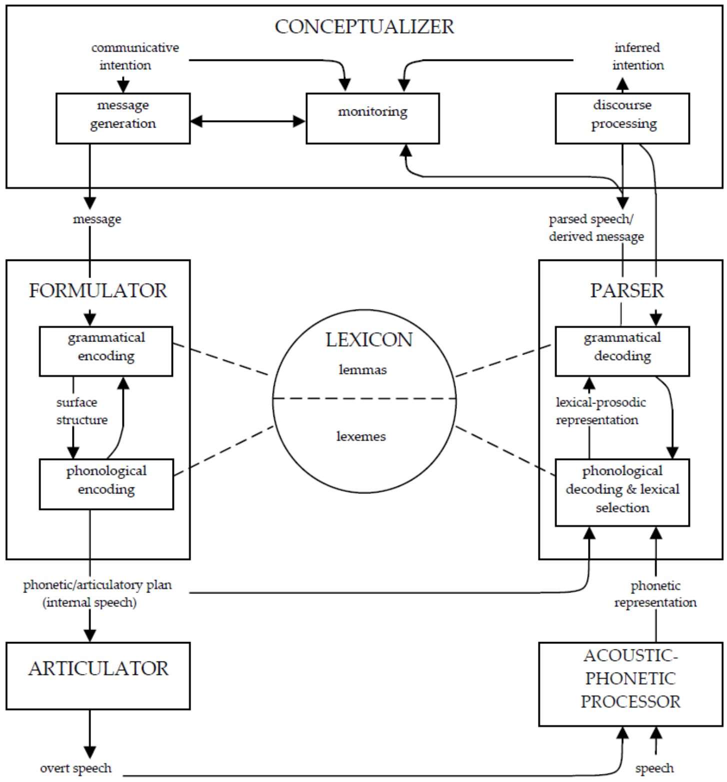


Figure 1.1. Levelt's (1989) model of language production.

Secondly, De Bot (1992) proposed the bilingual version of Levelt's model for monolingual production. Focusing on the articulator, De Bot claimed that we have one single articulatory system for the L1 and L2, hence, the interference from L1 causes accented productions. In addition, L1/L2 production differences were hypothesized to be related to the size and specification of the lexicon, the degree of automaticity in processing -especially at morphophonological levels- and the presence of traces of L1 in L2 production, De Bot explained that phonological interference was precisely because both languages could be accessed in parallel and L2 phonology was less automatic and narrow than L1's. The less the proficiency in the L2, the more competition between the two language phonologies.

Last, Kormos (1999) postulated that L1 and L2 phonemes coexist in one single module, and inaccurate productions may be ascribed to those articulatory gestures in the

L2 that have not been yet automatized, although this may vary according to the learners' proficiency. In addition, Dörnyei and Kormos (1998) and Kormos (1999) argued that the main cause of problems in L2 speech production derived from resource deficits (i.e., incomplete lexicon or insufficient morphological, or phonological specification), processing time pressure, and perceived trouble with their own and the interlocutor's output.

When oral-motor movements are relatively new in the L2 and they do not match the phonological representations, motor-sensory feedback systems are necessary to intercept erroneous output. The perceptual loop theory of monitoring (Levelt, 1989) adopted the idea that pre-articulatory output could be inspected and the same mechanism could be applied both for revising one's own message and for perceiving and revising other speakers' messages. Within the three monitor loops, the first was in charge of the preverbal plan against intentions; the second checked the articulatory plan against the overall plan; and a third one monitored overt speech through the acoustic-phonetic processor. Attention to the third loop is given considering the articulatory nature of the present dissertation's intervention.

During self-monitoring at the articulation level, L2 learners perceive an error in the articulation, halt the speech flow, this alarms the system, where motor adjustments are made to repair the erroneous utterance. Therefore, their production is compared to what would be correct according to their phonological knowledge. Kormos (1999) complemented the perceptual loop theory of monitoring with theories of consciousness and awareness. In other words, Kormos was concerned with the fact that many errors may be unnoticed because, contrary to L1 monitoring which involves controlled processing (Levelt, 1989), L2 processing is less efficient and most errors are automatized. Kormos (1999, 2000) claimed that L2 erroneous utterances are due to attentional deficits and limited resources available during self-monitoring, and learners' WM capacity and the demands of tasks are tightly linked to learners' allocation of attention to monitoring. Thus, repetition of a task may reduce its cognitive demands and allow for more attention for monitoring learners' speech (Bygate et al., 2001).

When comparing the different types of self-repairs in L2 production studies, it was generally found that phonological self-repairs were always less present than lexical, morphological, syntactic or semantic self-repairs (Lennon, 1984). When task demands of communicative form-focused tasks were high, Kormos (2000) found that grammatical and lexical self-repairs occurred more frequently (47%) than phonological self-repairs (7.6%), although the length of *reparatum* was less for phonological self-repairs because it usually takes place first in bottom-up models (compared to rephrasing the original utterance, for instance). The fact that most repairs are lexical in meaning-oriented tasks is because they carry most of the relevant information in the message and making errors may result in serious misunderstandings. All in all, L2 phonological errors may be originated at the stage of articulation, especially when task demands are high, and attention limitations may affect the efficiency of the monitoring systems. This dissertation assesses whether a pronunciation-focused task-based intervention may help allocate attentional resources to the phonetic encoding of L2 words and monitoring L2 phonetic form after articulation by assessing negotiation of phonetic form during interaction.

1.1.5. Noticing, selection, attention and memory capacities

It is especially relevant for this dissertation to review the notions of noticing and selection, —underlying Schmidt's (1990, 2001) theory of noticing— the limitations of our attentional capacity, as well as the important role of memory for second language development.

Whereas attentional mechanisms underlying production in the L1 are completely assumed in Levelt's model (1989), noticing, selection of input and attention capacity are especially relevant to understand second language development. Attention models (e.g., Wickens, 1989) agree that a sensory system detects the information that comes into the system for further processing, supervises response selection and execution and short-term memory semantically processes the selected information, that eventually will be stored in long-term memory.

Schmidt's noticing hypothesis (1990, 2001) claimed that, for learning to take place, certain features need to be noticed. In other words, learners need to demonstrate a conscious apprehension and awareness of some particular form in the input before any processing (and further intake) of that form can take place. Schmidt (1990, p.26) differentiated noticing from understanding: "Noticing is crucially related to the question of what linguistic material is stored in memory...understanding relates to questions concerning how that material is organized into a linguistic system." Leow's (2000) study showed that learners who demonstrated awareness of the targeted morphological forms during the experimental exposure contributed to more recognition and accurate written production of noticed forms compared with learners who demonstrated a lack of such awareness. In other words, allocating attention to form allowed learners to take in and retrieve the grammatical information in a more efficient manner compared to no awareness to form.

Conversely, Tomlin and Villa (1994) questioned Schmidt's theory in that conscious awareness (noticing) is not necessary for second language learning. They understood attention in terms of alertness (i.e., overall readiness to deal with incoming stimuli), orientation (i.e., direction of attentional resources to a certain type of stimuli) and detection (i.e., cognitive registration of the stimuli), but claimed that only detection was essential for acquisition. Robinson (1995, p.296) attempted to reconcile these two positions by proposing to define the concept of noticing to mean "detection plus rehearsal in short-term memory, prior to encoding in long-term memory".

Robinson (2003b) added that it is noticing and focusing processes complemented by memory mechanisms (i.e., preservation and rehearsal) that can account for learning. In short, despite some researchers defending the idea that learning may take place with or without awareness, all agree that attention is needed for second language development and, even if noticing is not necessary, it certainly contributes to learning and retention (Robinson, 2003b). In Sections 1.4.2. and 1.5.2., we will explore how focusing learners' attention on form has implications for second language learning, specifically, second language pronunciation learning (Mora & Levkina, 2017).

Secondly, there is the idea of the capacity limitations of attentional resources, which has been useful to explain interference in second language learning. Wickens' (1989) model of multiple resource pools proposes a series of dichotomical dimensions where the spatial is opposed to the verbal, auditory perception is opposed to visual perception, and responses can be manual or vocal. Provided that different dimensions draw on different resource pools, competition for attention may not necessarily happen, unless two tasks or two dimensions within the same task feed on the same pool. For instance, speaking to two people simultaneously would lead to poor performance, but speaking to one person while cooking would not pose a problem for attentional demands as the actions would draw on two dimensions (i.e., vocal vs. manual). However, it is still not clear to what extent different dimensions of oral production (fluency, accuracy and complexity) draw on different pools or on the same verbal pool. Whereas some SLA researchers advocate limited-capacity models of attention (Skehan, 1998; Skehan & Foster, 2001), others like Robinson (1995, 2011) advocate for a combined multiple-resource and non-limited capacity interference model of attention. See Section 1.4.4. on how attentional capacity models interact with increased cognitive demands.

Finally, it is relevant to mention how memory processes have been linked to L2 learning. Robinson (1995) states that without awareness, thus, peripheral attention, information can only enter short-term memory and activate stored information in the long-term memory; however, if information is noticed, it receives focal attention and enters WM. Therefore, WM is the part of short-term memory that receives attention, and both are activated in the long-term memory. It is of upmost importance to consider the role of memory in L2 learning: learners need to have enough resources in their working and long-term memory to be able to compare their own (accurate or inaccurate) utterances to the utterances of more competent interlocutors, and be able to restructure their interlanguage.

Section 1.1. has reviewed the processes involved in the perception and production of L2 speech as well as their relationship, and listed the main type of objective and subjective assessments of L2 perception and production, to inform the L2 perception-production link investigated in this study as well as justify our assessment method. In addition, the main tenets of speech perception and production models have been outlined to understand the potential sources of perceptual and production difficulties stemming from cross-language similarities and attentional deficits. The present dissertation is also interested about assessing learners' lexical encoding of L2 sounds and has reviewed previous instructional techniques that have enhanced phonolexical encoding. The last subsection has been dedicated to highlight the importance of noticing and attention for L2 learning, which are key elements in the current pedagogical intervention. The next section deals with L2 vowels that have been particularly challenging to perceive and produce for this study's population, connecting its difficulty to perception, production and lexical encoding processes already reviewed in Section 1.1.

1.2. Central Catalan and Southern British English (SBE) vowel inventories: a cross-language comparison

The second section of Chapter 1 is dedicated to describe the vowel inventories of Central Catalan and Southern British English (SBE); identify similarities and differences between the two vocalic systems; and explore potential causes of difficulty in the perception, production and lexical encoding of L2 vowels. Since participants of the present study are bilingual Spanish and Catalan speakers, and the Spanish inventory constitutes a subset of the Catalan vowels (i.e. five Spanish out of eight Catalan vowels), the object of study will be Central Catalan and SBE vowels.

1.2.1. Central Catalan vowel inventory

Catalan is a Western Romance language spoken in Alghero (Sardinia), Andorra, the Balearic Islands, Catalonia and the Valencian community, but the Catalan variety spoken by our participants is central Catalan, specifically from Maresme (Barcelona province). The Catalan vowel inventory contains eight vowels, seven stressed phonemes /a ɛ e i ɔ o u/ (Prieto, 2004; Recasens, 1993). In unstressed position, the vowel sounds /ɛ e a / are reduced to /ə/[9] (e.g. “acta/acte” /aktə̱/) and the vowel sounds /ɔ o/ are raised to /u/ (e.g. “pontet/puntet” /puntet/). Finally, Catalan has a variety of rising (/i̯a i̯ə i̯ɛ i̯e i̯ɔ i̯o i̯u u̯a u̯ə u̯ɛ u̯e u̯i u̯ɔ u̯o u̯u/) and falling (/ai̯ əi̯ ɛi̯ ei̯ ɔi̯ oi̯ ui̯ au̯ əu̯ ɛu̯ eu̯ iu̯ ɔu̯ ou̯/) diphthongs (Recasens, 1993).

1.2.2. SBE vowel inventory

[9] The Catalan variety spoken in Valencia, unlike the variety spoken in Barcelona, has no neutral vowel /ə/

The present investigation targets SBE vowels because learners have been mostly exposed to it through formal instruction but the target contrasts object to study are common to most varieties of English and do not differ in spectral and temporal cues across varieties (e.g., General American and Received Pronunciation). The English vowel inventory is constituted of 11 stressed phonemes /iː ɪ e ɜː æ ʌ ɑː ɒ ɔː uː ʊ/ and one unstressed phoneme /ə/ (Cruttenden, 2014; Mott, 2011). In unstressed position, the same vowels tend to be reduced to /ə ɪ ʊ/. SBE is often described as having five falling diphthongs /eɪ əʊ aɪ aʊ ɔɪ/ and three centring diphthongs /ɪə eə ʊə/ (Cruttenden, 2014; Mott, 2011). SBE vowels can be classified into tense /iː ɜː ɑː ɔː uː/ and lax /ɪ e æ ʌ ɒ ʊ/.

1.2.3. Commonalities and differences between Catalan & SBE vowel inventories

The vowel system of Southern British English (SBE) differs considerably from the Catalan system in size (i.e., 11 versus 7), as well as in the type of cues used for vowel discrimination and identification (i.e., spectral [*height and frontness-backness*] vs. temporal [*duration*] acoustic cues). Although none of the native vowel categories of Catalan (Spanish) is acoustically or articulatory identical to any SBE vowel category (Figure 1.2.), certain commonalities and differences can be outlined.

- Unlike Catalan's system formed by tense vowels, SBE vowels are classified into tense /iː ɜː ɑː ɔː uː/ and lax /ɪ e æ ʌ ɒ ʊ/, but lax vowels can never be found in word-final position (Mott, 2011).
- Whereas English has 4 high vowels /iː ɪ uː ʊ/, Catalan only has 2 /i u/. Catalan front /i/ is closer to English /iː/ than /ɪ/ in terms of spectral distance (F1 and F2), but is similar in duration to /ɪ/. English back vowels /uː ʊ/ have been identified as being acoustically closer to Catalan /u/ in height (F1) although

English /ʊ/ is more central and slightly lower than Catalan /u/ (McDougall & Nolan, 2007).

- In terms of mid vowels, English /e/ is positioned between Catalan /e/ and /ɛ/, although it is closer to Catalan /ɛ/ in terms of height. English central /ɜː/ has no correspondence with any Catalan vowel and it may only be related to the Catalan sounds /ɛ e o ɔ ə/ with low degrees of perceptual assimilation (Cebrian et al., 2011). Catalan and English unstressed vowel /ə/ share almost the same vowel space (i.e., F1 & F2).
- English mid vowel /ɔː/ is almost identical to Catalan /o/ (Cebrian et al., 2011) and English low vowel /ɒ/ is approximates Catalan /ɔ/ in terms of backness (F2) but it is in between English /ɔː - ɒ/ for height (F1).
- English open vowels /æ ʌ ɑː/ are slightly different in height (F1), being /ʌ/ less open and more central than their counterparts, but greatly different in terms of F2, being /æ/ the most front and /ɑː/ the most back.

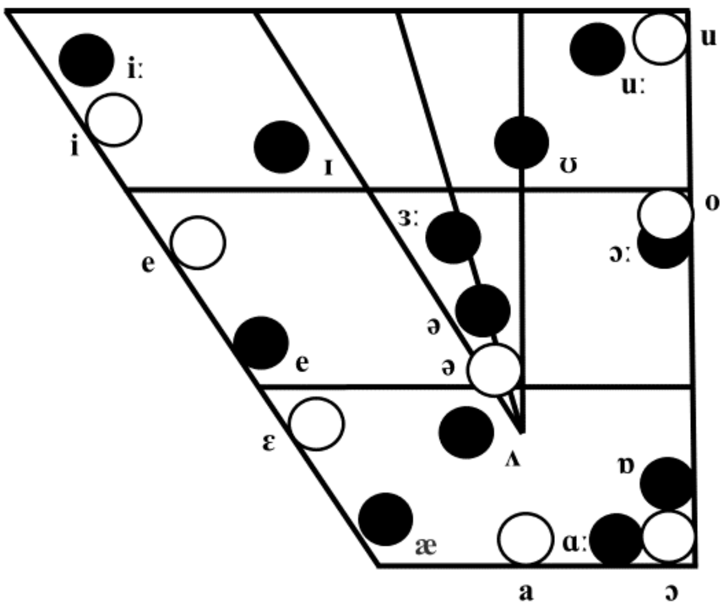


Figure 1.2. Central Catalan (white) and SBE (black) vowel systems (adapted from Mott, 2011).

Whereas English vowels that are very different (e.g., /ɜː/) or identical (e.g., /ə/) to Catalan vowels do not usually pose any difficulties to Catalan learners of English, L2 vowels which do not have one-to-one correspondence (e.g., /iː, ɪ, æ, ʌ/) in the L1 are typically confused in perception and are difficult to produce given the strong influence from the L1 (Kuhl & Iverson, 1995), as we will discuss in the next section.

1.2.4. Difficulties in vowel perception, production and lexical encoding: comparing Catalan and SBE phonological systems

The following section is dedicated to explain the expected segmental difficulties regarding the four target vowel sounds of the present study (i.e., /iː, ɪ, æ, ʌ/) in light of the acoustic (dis)similarity between Catalan and English vowels (Cebrian 2019, 2021; Cebrian et al., 2011), L1-L2 cue weighting differences (Cebrian, 2006; Iverson & Evans, 2007) and L2 speech perception (e.g., SLM, PAM-L2, NRV), production (e.g., SLM) and lexical encoding theories.

In order to explore perceptual vowel assimilation across languages, Cebrian et al. (2011) used a perceptual assimilation task and a rated discrimination task to compare L1 and L2 vowels. In the case of the tense /iː/ - lax /ɪ/ contrast, they found that the English tense vowel /iː/ was strongly assimilated to Catalan /i/ more than 90% of the time and obtained high goodness of fit ratings (4.6 out of 7 or higher), hence, was being perceived as the corresponding native category /i/. The Catalan /i/ - English /iː/ pair did not differ significantly from the dissimilarity scores for the English same-vowel pair /iː/-/iː/ (2.3), suggesting that the L1 /i/ and L2 /iː/ may be perceived as instances of the same vowel category. Similar results were reported in other studies testing the perceptual similarity between Catalan and SBE (Cebrian, 2006, 2015, 2021; Rallo Fabra & Romero, 2012).

In Cebrian et al. (2011), lower goodness ratings obtained for English vowels in the perceptual assimilation task consistently corresponded to higher dissimilarity scores in the rated discrimination task. Nevertheless, whereas the English lax vowel /ɪ/ was predominantly assimilated to Catalan /i/ (82% of the time vs. Catalan /e/, 15% of the time), it was perceived as more dissimilar from Catalan /i/ (3.2 of goodness rating) than from Catalan /e/ (3.4). In a similar study with naïve listeners (Cebrian, 2021), the modal response was Catalan /e/ (80% vs. 12% as /i/), with higher goodness ratings (5.6 as /e/, 3.8 as /i/). Therefore, English /ɪ/ seems to have a poor match in Catalan (Figure 1.3.), as it assimilates to both Catalan /e/ and Catalan /i/ to varying degrees but consistent lower goodness of fit ratings (Cebrian, 2006, 2021; Cebrian et al. 2011; Rallo Fabra, 2005). The lack of consistency across studies regarding this assimilation may be related to cross-study methodological differences in population, variety of English, phonetic context of the stimuli, and response options used (Cebrian, 2021).

The English vowel contrast /æ/-/ʌ/ is also complicated to perceive and produce for Catalan speakers (Aliaga-Garcia & Mora, 2009) due to the presence of one single vowel in the Catalan acoustic space, the low central vowel /a/ (Cebrian, 2021). Rallo Fabra and Romero (2012) stated that Catalan learners experienced great difficulty when discriminating between the Catalan and American English vowel pairs /a/-/æ/ and /a/-/ʌ/. In terms of /æ/, Cebrian (2021) found a high degree of perceptual similarity between the Catalan central vowel /a/ and the English /æ/ (99% of the time); however, whilst English /ʌ/ was also frequently assimilated to Catalan /a/, /ʌ/ had a lower degree of perceptual similarity than its counterpart /æ/ (Figure 1.3.).

	English vowel stimuli												
Cat. resp.	/iː/	/ɪ/	/eɪ/	/ɛ/	/æ/	/ʌ/	/aɪ/	/aʊ/	/ɔʊ/	/ɑː/	/ɒ/	/ɔː/	/uː/
/i/	**87 (5.7)**	12 (3.8)					3 (1.9)						
/e/		**80 (5.6)**		8 (5.1)									
/ej/	12 (4.2)		**76 (4.2)**										
/ɛ/		7 (4.8)		**91 (5.6)**									
/a/					**99 (5.6)**	**98 (4.7)**				31 (3.9)	3 (3.7)		
/ai/			21 (4.3)				**95 (3.6)**						
/au/								**99 (4.9)**	19 (2.6)	11 (2.3)	3 (1.9)		8 (1.2)
/ou/									**79 (3.7)**			8 (3.4)	10 (2.4)
/ɔ/										**54 (4.8)**	**82 (5)**	7 (3.5)	
/o/										4 (4.5)	11 (4.6)	**77 (4.8)**	
/u/												6 (3.9)	**82 (3.7)**

Figure 1.3. Assimilation patterns of English vowels to Catalan vowels (in percentages) and goodness of fit (in parenthesis) (adapted from Cebrian, 2021).

Inaccuracies in the acquisition of L2 vowels can also be well understood through theories of acoustic cue weighting (e.g. how a listener weights the spectral and durational information when perceiving non-native vowel stimuli). Native English speakers have been shown to rely mostly on spectral cues (F1 and F2) when identifying tense and lax vowels, with duration only playing a secondary role. In contrast, Catalan has no tense-lax temporal contrast (Recasens, 1993), and supporting evidence has shown that learners of English whose L1 does not make use of duration exploit temporal cues to a greater extent than spectral cues in differentiating between tense and lax vowels (e.g., English /iː/ and /ɪ/) (Cebrian, 2006; Escudero & Boersma, 2004; Flege et al., 1997).

For instance, Catalan EFL learners have been often shown to fail to distinguish English /iː/ from /ɪ/ in terms of spectral cues, hence, they rely mostly on temporal cues (i.e., duration) to distinguish these two sounds in perception and production (Cerviño & Mora, 2009). Cebrian (2006) tested the perception of English /iː/ and /ɪ/ by means of a synthetic /iː/-/ɪ/-/ɛ/ vowel continuum varying in temporal and spectral steps. Unlike English speakers, Catalan learners of English tended to rely heavily on duration to distinguish between /iː/ and /ɪ/. Mora and Fullana (2007) replicated earlier findings of learners' overuse of temporal cues (Flege, 1995), and proved that, irrespective of AOL and L2 experience, non-native speakers of English did not produce significant spectral differences between vowels (/iː/-/ɪ/, /æ/-/ʌ/); instead, they used duration to differentiate them.

Provided that vowel duration is no distinctive feature in Catalan, leaners' use of duration as the main cue to distinguish L2 vowels (e.g., /iː/-/ɪ/) may not be attributed to negative transfer from the L1, but to lack of experience with spectral differences, resulting in the need to rely on duration as a compensatory strategy. Bohn's (1995) Desensitization Hypothesis claims that the formation of a new category for English /ɪ/ may be based on different duration values of /iː/ because Catalan learners of English have become linguistically "desensitized" to the spectral differences between /iː/ and /ɪ/ and vowel duration is probably more acoustically salient. Furthermore, factors such as instructional biases or the effect of L2 orthography may have an impact on EFL learners' reliance of duration to distinguish English /iː/ and /ɪ/ (Cebrian, 2006).

Finally, since the present study assesses gains in terms of perceptual discrimination and production, it is important to look into the predictions made by the main speech models in relation to the target sounds of the current thesis. Following SLM hypotheses (Flege, 1995), a new category may be established for the least similar category (e.g., English /ɪ/), but discriminating the English /iː/-/ ɪ/ contrast accurately may take time for Catalan (Spanish) learners of English.

Recall that PAM-L2 model (Best & Tyler, 2007) claims that accurate discrimination of L2 phones is determined by the degree to which pairs of target phones are assimilated to one or more L1 categories. Therefore, two target phones that are assimilated to the two different L1 phones (two category assimilation) will be more accurately discriminated than two target phones that are assimilated to a single L1 category. In the latter case, the two target phones may be perceived as equally good or bad versions of the L1 category (single category assimilation – poor discrimination) or one target phone may be perceived as a better match for the L1 category than the other target phone (category-goodness assimilation – good discrimination). In PAM-L2 terms (Best, 1995; Best & Tyler, 2007), Catalan learners assimilate English /iː/ and /ɪ/ to Catalan/Spanish /i/ via an uncategorized-categorized assimilation pattern[10]. Thus, whilst the English vowel /iː/ has been found to be highly comparable to the Catalan /i/, the English vowel /ɪ/ has been found to be perceived as a poorer fit of the same L1 sound, sometimes being identified as Catalan /i/ or /e/. Discrimination is expected to be good to very good for uncategorized-categorized assimilation (Cebrian, 2019) according to PAM-L2 (Best & Tyler, 2007); however, Cebrian (2019) found that the results for /iː-ɪ/ did not conform to the predictions as this pair was not better discriminated than the category-goodness pair /æ/-/ʌ/. One possible explanation is the different L2-L1 assimilation patterns have been evidenced for English /iː-ɪ/ in previous studies with L1-Catalan participants. As for the low-central vowel pair /æ/ and /ʌ/, PAM-L2 would classify it as a category-goodness assimilation as the two SBE vowels are consistently assimilated to Catalan /a/ (Figure 1.4.), but /æ/ is being perceived as closer to Catalan /a/ than /ʌ/ to Catalan /a/ (Cebrian, 2021; Tyler, 2021). This scenario predicts difficulty of discrimination leading to the perception of words such as "cat" and "cut" as homophonous (Rallo Fabra & Romero, 2012). Finally, the NRV framework (Polka & Bohn, 2003, 2011) would predict that English /iː/ and /æ/ would act as the natural reference vowels, guiding

[10] The English high-vowel pair /iː-ɪ/ has been patterned as an uncategorized-categorized type of assimilation in some studies (Cebrian 2019; Cebrian et al., 2021) because, while English /iː/ was assimilated to Spanish /i/, English /ɪ/ was assimilated above chance (below 70%, which has been proposed as the threshold for categorization; Antoniou et al., 2012; Tyler et al., 2014) to more than one L1 phone (Spanish /e/ and /i/) and thus illustrated an uncategorized clustered type of assimilation. Other studies (e.g., Carlet, 2017) have described English /iː-ɪ/ in the frame of the category-goodness assimilation since vowel /iː/ has been found to be strongly assimilated to Spanish /i/ and /ɪ/ has been found to be perceived as a poorer fit of Spanish /i/. The present dissertation treats English /iː-ɪ/ as an uncategorized-categorized type of assimilation according to PAM-L2 terms; however, a perceptual assimilation task would have provided valuable information about the type of assimilation pattern the participants of this study had.

the perception of English /ɪ/ and /ʌ/, which are more centralized in the vowel space, as observed in Figure 1.4.

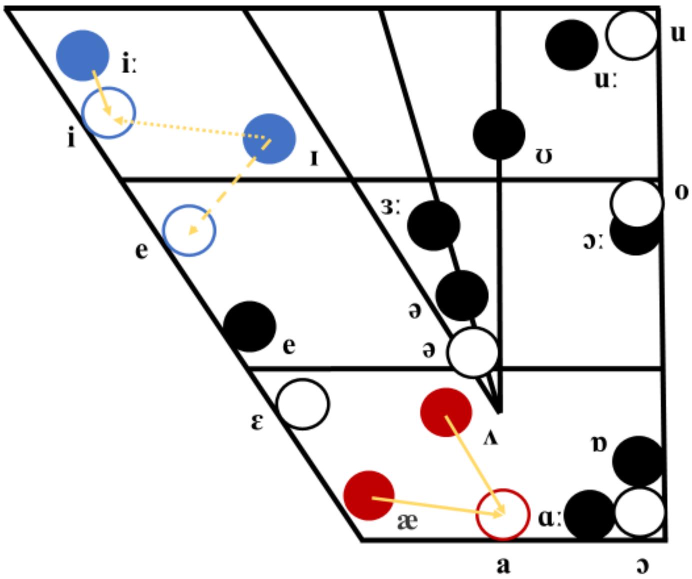


Figure 1.4. A visual example of an uncategorized-categorized pair (blue circles) and a category-goodness assimilation pair (red circles) (Best & Tyler, 2007). Filled circles represent SBE and unfilled circles Catalan vowel categories.

Concerning the production of English vowels /iː, ɪ, æ, ʌ/, L2 research has suggested that once learners become aware of the spectral differences between English /iː/ and /ɪ/ (Bohn, 1995), it may be possible that an increase in accuracy for lax /ɪ/ results in a decrease of accuracy in tense /iː/ until the tense-lax contrast is perceived as a duration and spectral contrast (Escudero & Boersma, 2004). Flege et al. (1997) showed that accurate productions of /ɪ/ were significantly related to poor productions of /iː/ so their learners were reanalysing the tense-lax contrast by redirecting their attention away from duration and produce larger spectral differences. Similarly, Cebrian (2007), who investigated the use of spectral and temporal cues in the production of the L2 English tense-lax vowel contrasts, found that /iː/ tokens with longer duration, lower F1 and greater F2-F1 difference were better identified and rated, whereas in the case of vowel /ɪ/, better results corresponded to the opposite characteristics, that is, shorter vowel duration, higher F1, lower F2 and smaller F2-F1 difference. In terms of the low-vowel contrast, L2 learners struggle to produce a front-central distinction for English /æ-ʌ/ because it is missing in Catalan and Spanish, whose only open vowel is a relatively front /a/. Although English /æ/ often reaches native-like accuracy levels (Rallo Fabra & Romero, 2012) in production when delivered in /æ/-/e/ contrasts, L2 learners struggle to produce distinct vowels in the production of the English /æ-ʌ/ contrast (Mora & Fullana, 2007). In addition, focal English vowels /iː/ and /æ/ may be easier to pronounce than /ɪ/ and /ʌ/ due to their saliency coming from formant convergence in the vowel space (Polka & Bohn, 2003, 2011).

Last but not least, difficulties in the perceptual discrimination of English vowel contrasts /iː-ɪ/ and /æ-ʌ/ may lead learners to inaccurately encode the contrasts, thus, develop homophony in the interlanguage lexicon (Pallier et al., 2001) such as "sheep" /ʃ**i**p/ - "ship" /ʃ**i**p/ or "cap" /k**a**p/ - "cup" /k**a**p/ that are contrastive in English (/ʃiːp-ʃɪp/ /kæp-kʌp/). In addition, learners may also develop phonolexical representations for other words containing these vowels that may be "fuzzy" (Darcy, Daidone & Kojima, 2013) in that they are misrepresented in the mental lexicon (e.g., sun /san/) despite being able to distinguish them phonetically in perception.

Section 1.2. has presented the main commonalities and differences between Central Catalan and SBE vowel inventories to be able to understand learners' perceptual assimilation patterns as well as cross-linguistic differences in cue weighting (e.g., Catalan speakers' overreliance of temporal cues when producing the English /iː-ɪ/ contrast). Having identified learners' difficulties in perceiving and producing differences between L1 and L2 vowels in relation to main L2 speech models, the following section sets out to provide instructional techniques that help to notice and orient attention to challenging L2 phonological features which, due to L1-based processing, are difficult to acquire for second language learners.

1.3. Pronunciation training and instruction in L2 contexts

Section 1.3. provides an insight into previous work on phonetic training and pronunciation instruction in second and FL contexts. The objective is to explore various aspects related to L2 pronunciation development, including the role of attention, the effectiveness of phonetic training and computer-assisted pronunciation teaching, the historical beginnings of pronunciation instruction, the effects of explicit instruction and form-focused communicative instruction, and the challenges of conducting pronunciation instruction in the FL classroom along with potential solutions.

1.3.1. Pronunciation learning and attention in instructional settings

In this subsection, we highlight the important role of attention and focus on phonetic form (vs. solely focusing on meaning) to overcome L2 phonological difficulties, and we show how several instructional methodologies (i.e., phonetic training, explicit instruction) as well as the use of corrective feedback may promote attention to L2 phonological features, and lead to L2 pronunciation development.

One major constraint in L2 pronunciation learning is lack of quantity and quality of input (García Mayo & García Lecumberri, 2003; Muñoz, 2014) which, as discussed in Section 1.1., may lead to learners' inability to perceptually identify phonetic and phonological cross-language differences (Flege, 1995; Best & Tyler, 2007) and, consequently, result in inefficient phonological processing in the recognition and production of L2 words (Bradlow, 2008), and more accented productions in the L2.

Research has shown that, for pronunciation learning to take place, the relevant properties of the speech input need to be attended to and noticed, hence, awareness of the key phonological/acoustic feature at some level of processing is necessary (Francis & Nusbaum, 2002). In other words, an exclusive focus on meaning (FonM)[11] does not suffice for the acquisition of L2 phonological features (DeKeyser, 1998). For instance, Guion and Pederson (2007) and Pederson and Guion-Anderson (2010) assessed the effect of different orienting instructions during a training period. Rather than drawing conscious attention to particular cues, Guion and Pederson (2007) manipulated endogenous[12] orienting via varying instructions during the training of unfamiliar phonetic categories from Hindi. Their experiment oriented one group of participants' attention to the phonetic form of the stimuli and another group of participants to the meaning of the same stimuli. Results showed that phonetic learning occurred when learners were instructed to attend to the phonetic properties of the speech stimuli in the training but not when instructed to attend to their meaning. In a similar vein, Pederson & Guion-Anderson (2010) explored the effects of differential attention within the phonetic domain. English monolinguals were randomly assigned to either of two groups: a consonant-attending group was given explicit instructions to attend to word-initial consonants and a vowel-attending group to attend to word-medial vowels of the same Hindi stimuli. Their results confirmed that orienting attention during phonetic training facilitated learning of the specific class of stimuli to which the participants were instructed to attend. In a way, the type of attentional orienting regulated the selective uptake and storage of phonetic information (Pederson &

[11] Focus on Meaning (FonM) refers to the incidental learning of a second language when learners are presented with comprehensible samples of L2 use. In FonM pedagogical approaches, L2 rules are deduced by the learner from mere exposure to the L2 input (Long & Robinson, 1998).

[12] According to Guion and Pederson (2007), participants' orienting of attention can be manipulated endogenously or exogenously. Exogenous orienting is the attraction of attention by external stimuli (e.g. a loud beep) and endogenous orienting is the result of the individuals' directing their attention according to their goals or expectations.

Guion-Anderson, 2010). These findings are in line with Carlet and Cebrian (2022), who found that performing an identification task during phonetic training succeeded in orienting L2 learners' focal attention solely to the target sounds, while ignoring other cues present in the input. Finally, Park (2000) showed that form-focused + meaning-focused (but not a solely meaning-focused) instructional treatment proved to be effective in improving L2 lexical stress assignment. Similarly, in an auditory priming experiment with learners differing in L2 pronunciation accuracy, Trofimovich and Gatbonton (2006), following the ACCESS[13] framework, found a focus on meaning to be detrimental to the priming of L2 words for low-accuracy learners but not for high-accuracy learners, who obtained priming effects in both form-focused and meaning-focused conditions. In sum, it seems that, in the absence of a focus on phonetic form, development of L2 pronunciation is unlikely to take place in meaningful contexts (Derwing et al., 1998), where FonM may not be sufficient to notice the relevant phonetic properties in the input.

Even though Moyer (2013) posits that "phonological instruction is predicated on the learner's ability to *detect* the differences between their own output and the native(like) model provided to them" (p. 154, my emphasis), and Derwing and Munro (2005) state that "students learning L2 pronunciation benefit from being explicitly taught phonological form to help them *notice* the difference between their own productions and those of proficient speakers in the L2 community" (p. 388, my emphasis), the role of directing learners' attention to specific phonetic features of the input in L2 speech learning is still under-researched beyond the domain of phonetic training. This is partly because of the difficulty in developing learners' L2 phonological awareness, even at high-proficiency levels, in comparison to grammar and lexis.

[13] Gatbonton and Segalowitz (2005) proposed a theoretical teaching framework called Automatization in Communicative Contexts of Essential Speech Segments (ACCESS). It elicited the necessary repetition to promote automatic fluency in a communicative framework that integrates attention to form.

Orienting attention to critical acoustic features in the acquisition of a given L2 sounds has been done through manipulating characteristics of the stimuli or conditions during phonetic training (e.g., Aliaga-García & Mora, 2009; Bradlow et al., 1997; Mora et al., 2022), explicit pronunciation instruction and oral practice on the articulation of specific sound segments (e.g., Kissling, 2013; Saito, 2013), and the use of explicit or implicit feedback targeting pronunciation-related errors (e.g., Saito, 2015; Saito & Lyster, 2012; Ruan & Saito, 2023). The use of phonetic training and pronunciation instruction as a medium for drawing learner attention to phonetic and phonological form and how it may influence L2 pronunciation development will be explored in Sections 1.3.2, 1.3.3. and 1.3.4. The last part of this section is dedicated to assess the role of corrective feedback to raise learners' awareness of different phonetic forms.

Corrective feedback[14] has been mainly used in instructional contexts as something integrated into teaching through planned diagnostics and tasks that bring learners' attention to particular linguistic difficulties. Although the impact of oral corrective feedback on pronunciation errors shows a pattern of low rates of occurrence (compared to morphosyntax or lexis), the high rates of uptake (Brown, 2016) may be connected to how these errors are more likely to seriously cause a breakdown of communication (Scheuer & Horgues, 2021). A key study investigating recasts and pronunciation development is Saito and Lyster (2012), who found that, learners who received recasts of instances of mispronunciation of the target form, made demonstrable gains while those in an experimental group in which the target form was highlighted and practiced but

[14] Corrective feedback is defined as "a response to an error that occurs during interaction in meaning primary tasks" (Li & Vuono, 2019, p. 99). In the FL classroom, it can be taken as something integrated into teaching through planned diagnostics and activities to bring a learner's attention to particular difficulties. Some examples are recasts, metalinguistic feedback, elicitation, repetition, explicit correction, and clarification requests.

without recasting made no such gains. In fact, students tend to benefit from the opportunities afforded by recasts, first to notice the negative evidence directed at the intelligibility of their output and second, to practice the correct form in response to their teachers' model pronunciation (Lyster, et al., 2013), only if there is a specific focus on pronunciation. Indeed, drawing learners' attention to phonological problems through corrective feedback has been found most effective when it relates back to prior formal teaching and when learners are prompted to provide peer feedback under teacher guidance (Martin & Sippel, 2021). Finally, drawing learners' attention to perceptual or production errors during phonetic training has also found to contribute to L2 learning (Thomson, 2011, 2012).

To conclude, a growing body of research on L2 pronunciation learning is demonstrating that helping learners notice relevant L2 phonological targets through instructional techniques and corrective feedback may increase the likelihood of detection of the relevant phonetic information, processing in short term memory, and transfer to long-term memory. The following section shows how phonetic training has been found effective at enhancing attention to phonetic form and lead to robust L2 speech learning.

1.3.2. Phonetic training in a second language

This section presents the effectiveness of phonetic training to develop stable, generalizable (novel contexts, talkers and tokens) and robust L2 phonetic categories that lead to L2 speech development, and how it can be manipulated to lead to greater pronunciation gains. Apart from HVPT, other explicit and implicit phonetic training methods are described. Lastly, not only are the benefits of phonetic training exposed, but also its challenges, especially in relation to instructed L2 speech learning.

1.3.2.1. Effectiveness, generalization and retention of learning

A wealth of evidence from a considerable amount of studies has shown that phonetic training can be beneficial for different L1-L2 target structures, particularly L2 consonants and vowels (Aliaga-García & Mora, 2009; Carlet & Cebrian, 2019; Iverson & Evans, 2007, 2009; Rato, 2014; Thomson, 2012, 2018). In an attempt to mirror the variability that characterizes L2 input in a natural environment, Logan et al. (1991) were the first to adopt an HPVT approach to train L2 speech in highly-controlled laboratory settings.

HVPT is a perceptual (and/or production) training paradigm in which learners are exposed to new and/or partially acquired L2 sounds that are embedded in diverse lexical or pre-lexical contexts though multiple talkers. For each token, learners identify and discriminate target sounds in minimal pairs, with feedback presented after each trial. In general, HVPT has been found effective to develop stable, generalizable and robust speech categories in the target language (for comprehensive reviews, see Sakai & Moorman, 2018; Thomson, 2018), leading to gains in L2 speech perception (Bradlow, 2008; Carlet & Cebrian, 2019) and also L2 production (Kartushina et al., 2015; Thomson, 2018; Wang et al., 2003). The effect of HVPT has been investigated for different areas of pronunciation training, such as tone (Wang et al., 2003), syllable structure (Huensch, 2016); consonants (Iverson et al., 2005) and vowels (Carlet & Cebrian, 2019; Rato & Rauber, 2015).

One of the main strengths of HVPT as a teaching technique has been the generalizability of perceptual learning to novel talkers (Lee & Baese-Berk, 2021; Thomson, 2018), untrained tokens (Iverson et al., 2005; Mora & Mora-Plaza, 2019; Ortega et al., 2021; Thomson & Derwing, 2016), new contexts (Thomson, 2011), untrained sounds (Carlet & Cebrian, 2019), and new modalities (i.e., perception→production[15]: Bradlow et al., 1997; Rato & Rauber, 2015 / production→perception: Hirata (2004) / perception → phonolexical representations: Darcy & Thomas, 2019; Llompart & Reinisch, 2021). In addition, robust learning from HVPT has been demonstrated further when the observed improvement is retained sometime after training has ended (Carlet & Cebrian, 2019; Rato, 2014; Rato & Rauber, 2015; Wang & Munro, 2004). A number of key studies that have found robust learning of HVPT through generalization and/or retention are briefly summarized in Table 1.1.

So as to direct learners' attention to phonetic properties of speech and generate more L2 speech training gains, stimuli type (Ortega et al., 2021; Thomson & Derwing, 2016) and training conditions (Cooke & García Lecumberri, 2018; Kartushina et al., 2015) have been manipulated in the HVPT paradigm.

In addition, other types of phonetic training such as "phonological specificity training" —a training paradigm used to teach minimal pairs that is designed to enhance the distinctiveness of words' phonological representations— have been shown to have positive effects on phonological awareness and vocabulary in adolescents/adults learning an L2 (Van de Ven et al., 2019; Llompart & Reinisch, 2021). In the same vein, Saito et al. (2022b) showed that auditory-only training helped 98 Japanese speakers of English to improve both auditory sensitivity (F2 discrimination of 1200-1600 Hz) and L2 speech proficiency (the identification of English [æ] and [ʌ]) whether it was combined with phonetic training or not. In contrast, the gains of phonetic training (Phonetic-Only) were limited to speech perception (English [æ] and [ʌ]).

[15] However, "improvement in perception as a result of training is greater than accompanying changes in production, consistent with a lag between the development of perception and its impact in production". (Thomson, 2022, p.380)

Table 1.1. Summary of five key high variability phonetic training (HVPT) studies on L2 English vowel acquisition.

Study	**Participants**	**HVPT design**	**L2 Target vowels**	**Findings**	**Generalization**	**Retention**
Wang & Munro (2004)	17 Mandarin and 4 Cantonese speakers in Canada (length of residence: 24 months)	~15-20 hours 3200 trials (160 trials x ~20 sessions)	/iː ɪ æ ɛ uː ʊ/	10–15% gains	To novel tokens To novel talkers	3 months after training
Iverson & Evans (2009)	17 Spanish[a] and 16 German[b] 9 Spanish and 11 German	4 hours (225 x 5 sessions)	/iː ɪ ɜː e ɑː æ ʌ ɒ ɔː ʊ/ /eɪ aɪ aʊ əʊ/	10–15% gains	To novel tokens To novel talkers	2 to 6 months after training 12 months after training
Thomson (2012)	26 Mandarin Chinese speakers in Canada (length of residence: 11.6 months)	2.5 hours 1600 trials (200 trials × 8 sessions)	/iː ɪ e ɛ æ ʌ ɒ ɔː uː ʊ/	10–20% gains	To novel phonetic contexts To novel talkers	1 month after training
Rato (2014)	34 Portuguese speakers in Portugal	5 hours 420 trials (84 trials x 5 sessions)	/iː ɪ æ ʌ ɛ uː ʊ/	15–30% gains	To novel tokens To novel talkers	2 months after training
Carlet & Cebrian (2019)	54 Catalan/Spanish bilingual speakers in Spain	2.5 hours 1400 trials (288 trials x 5 sessions)	/iː ɪ æ ʌ ɜː/	10–15% gains	To novel non-words and words To novel talkers	2 months after training

[a] Spanish speakers in English speaking countries (length of residence:18 months)
[b] German speakers in Germany. None had lived in English-speaking countries

Although explicit HVPT has been proved to help learners' reattunement of L2 sounds, it may not necessarily lead to automatization in the processing of these sounds. Lim and Holt's (2011) study examined the role of *incidental* and multimodal learning in the context of L2 segmental learning through an alien-based video game. The amount of gains obtained mirrored what Japanese speakers normally exhibited after receiving intensive explicit HVPT. Similarly, Saito et al. (2022c), using 58 Japanese English-as-a-foreign-language speakers, wanted to test whether three hours of game-based incidental training could impact their acquisition of the English [ɾ] and [l] contrast and the English [æ] and [ʌ] contrast. Findings revealed relatively low improvement relative to the effectiveness of explicit phonetic training—5% gains for English [ɾ] and [l] and 5–10% gains for English [æ] and [ʌ]—. Therefore, explicit HVPT may be useful for learners to notice and attend to the perceptual characteristics of target sounds. Once auditory representations have been partially established, learners might then be more likely to benefit from incidental HVPT and develop more automatized knowledge of L2 sounds.

1.3.2.2. Challenges

In the previous section, the multiple benefits of explicit and implicit phonetic training for L2 speech development have been outlined. From a research perspective, using HVPT in controlled lab-based experiments allows researchers to use methods that would not be feasible in a classroom setting; investigate particular research questions while removing confounding factors; and control that improvement is only due to the experimental intervention (Martin & Inceoglu, 2022). Nevertheless, one major constraint of lab-based studies is their lack of ecological validity[16] compared to classroom-based studies.

[16] Ecological validity refers to whether the experiment is similar to the context to which it aims to generalize, and whether it has educational relevance (Rogers & Cheung, 2021).

Derwing & Munro (2015) state that the technical nature of such research, and the fact that it is often conducted under controlled laboratory conditions, makes the immediate relevance and applicability of such research to the classroom less clear. In fact, Fraser (2011; p. 12) refers to HVPT as “the deprecated ‘drill and kill’ training”. Decontextualized L2 training may not only be dissociated from the use and learning of language in real-life contexts but it could negatively affect students’ motivation, enjoyment, and attitude toward L2 speech learning. On the one hand, language instructors are likely to be more willing to use methods that have been proved in real classroom settings (Martin & Inceoglu, 2022; Thomson & Derwing, 2015) than lab environments. On the other hand, it may be very challenging to methodologically implement a personalized HVPT in class given the recording process alone is a massive obstacle to overcome (Thomson, 2022).

Nevertheless, the present dissertation shows that findings from this line of research can directly inform classroom-based pronunciation instruction by highlighting the importance of including various L2 talkers and phonetic contexts in listening tasks that are used in the classroom to develop learners’ L2 speech perception and production, and testing generalization and retention of learning after any classroom intervention (Lee et al., 2015 for meta-analysis). Empirically-validated HVPT computer-assisted systems such as the *Golden Speaker* (Ding et al., 2019), or the *English Accent Coach* (Thomson, 2012, 2018), amongst others, can be used as a complement to pronunciation instruction.

1.3.3. Computer-assisted pronunciation teaching

With the increase of technology use in the L2 classroom, another methodology which has been employed for instructed L2 pronunciation learning in the last few years is computer-assisted pronunciation teaching (CAPT). In this section, some advantages and challenges of CAPT are presented in relation to opportunities and accessibility of learning, feedback and assessment methods.

1.3.3.1. Advantages for L2 pronunciation learning

In the last decades, the field of CAPT has grown exponentially, with an expansion of web-based and mobile apps and resources. The benefits of technology for pronunciation teaching and learning are evident and have been empirically demonstrated (Fouz-González, 2015; Pennington & Rogerson-Revell, 2019; Rogerson-Revell, 2021). For instance, Foote and McDonough (2017) found that the use of shadowing with mobile technology improved participants' comprehensibility and fluency of speech and Fouz-González (2020) showed that the use of the *English File Pronunciation* app improved both perception and production of segments by Spanish learners of English. The same author, in 2017, reported results in favour of Twitter's potential for educational purposes. By receiving tweets with sound-awareness activities involving explanations and links to videos and audio recordings, participants significantly improved in their oral production of the target words.

To start with, CAPT resources have the potential to provide individualized, self-paced learning environments where learners can practice L2 pronunciation comfortably, at their own pace, and in places where they feel at ease. This helps them overcome FL anxiety and enhances learners' motivation to learn pronunciation (Fouz-González, 2015). In addition, technology allows for rapid accessibility: learners have free access to a large variety of authentic materials, such as podcasts on Spotify, films in their original language, online dictionaries or YouTube audio/video files. Therefore, it maximizes opportunities for exposure to a broader variety of spoken language, including different

L2 accents and speech genres and styles, which develops learners' awareness of phonological variation (e.g. *English Accent Coach* by Thomson, 2012). Apart from providing opportunities for controlled pronunciation practice, such as the repetition of sounds or structured responses (Rogerson-Revell, 2021), rapid advances are permitting more meaningful, authentic interactions with computers through the use of speech recognition and natural language processing technologies (e.g., chatbots, talking heads, virtual reality headsets).

CAPT resources also have the potential to provide immediate, customised feedback that is targeted to the individual through automatic speech recognition software, or the use of visual displays (e.g., spectrograms to help students monitor their productions and enhance feedback: Olson, 2014). Finally, the use of CAPT has raised the interest of many scholars as a tool for eliminating human bias and error in pronunciation assessment (Rogerson-Revell, 2021). Whereas CAPT had long been used to assess perception and production of sounds, the incorporation of ASR has only recently enabled automated assessments of comprehensibility (Saito et al., 2022a).

1.3.3.2. Challenges

Despite the potential of CAPT tools in terms of accessibility, feedback provision and assessment many CAPT tools appear to be technology-driven rather than pedagogy-led (Rogerson-Revell, 2021). In fact, several scholars (Levis, 2018; Pennington & Rogerson-Revell, 2019) have called for collaboration between pronunciation experts and CAPT-resources developers so as to minimize the tension between pedagogy and technology. Whereas many teachers and researchers are now accepting intelligibility as a much more achievable goal for many learners than "native-like pronunciation", the "nativism" approach is still much present in CAPT resources which promote native-like evaluations and corrections (Levis, 2018). As Rogerson-Revell (2021, p.191) points out, "as technology progresses, pedagogy appears to regress, returning to audiolingual approaches of repetition, mimicry, and drilling", which do not enhance the development of L2 phonology for communicative purposes.

Another criticism of CAPT is that content and feedback is not tailored for each individual's needs (Derwing & Munro, 2015; Levis, 2018), hence, little guidance is given concerning which pronunciation features may be affecting learners' intelligibility (but see ELSA (https://elsaspeak.com/en/). Comparisons of waveforms or spectrograms with native speakers is not ideal either as two speakers can produce the same utterance with different acoustic features (Rogerson-Revell, 2021). Finally, an important challenge for CAPT is defining the appropriate criteria with which to evaluate pronunciation proficiency. Many CAPT tools tend to assess "nativelike" over "comprehensible" pronunciation and have not found empirical support from research (but see *Cool Speech*, and *Sounds of Speech* apps).

Although CAPT resources have the potential to provide multimodal input and ample practice opportunities, systems need to ensure that learners can improve the comprehensibility and intelligibility of their L2 speech. Lastly, teaching approaches that incorporate the advantages of technology (such as ASR-based feedback) with form-focused teaching (e.g., pronunciation awareness activities) have been found to lead to better L2 pronunciation development than just the use of CAPT (Sardegna & Jarosz, 2022). Using the *English Accent Coach* (Thomson, 2012) web application as complementary practice to classroom instruction (Section 1.3.4.) or watching captioned videos (Wisniewska & Mora, 2020) outside the EFL classroom may accelerate L2 speech development.

1.3.4. Pronunciation instruction

Although L2 pronunciation learning can occur in the absence of teaching, given the right timing, environment, and positive attitudes, this kind of improvement is limited and is most obvious within the first year being surrounded by the L2 (i.e., window of maximal opportunity by Derwing & Munro, 2015). In FL contexts, pronunciation instruction metanalyses have shown that pronunciation instruction is effective (Lee et al., 2015; Saito & Plonsky, 2019); however, pronunciation gains in these studies really depend on the scope and duration of instruction, type of instruction, outcome measures and feedback. Lee et al. (2015) found that pronunciation instruction effects were strongest for longer periods of instruction, interventions that provided feedback on learner language, and for more controlled language use. Also, pronunciation instruction effectiveness should consider transferability of learning beyond the classroom (Darcy, 2018).

This section starts with a summary of the main historical changes in the teaching of pronunciation to understand current pronunciation teaching practices. Then, the definition and functions of explicit pronunciation instruction are mentioned to understand its effectiveness and potential challenges for L2 pronunciation development. The third subsection focuses on the importance of form-focused instruction for the development of pronunciation accuracy and comprehensibility in genuinely communicative situations, and current L2 pronunciation assessment methods. The forth subsection identifies potential obstacles in the teaching of pronunciation (i.e., time, method and focus) and the final subsection is dedicated to provide solutions to the aforementioned challenges though integration, dual-approach methods and prioritization of certain L2 pronunciation aspects.

1.3.4.1. The history of pronunciation teaching

In order to understand current practices in L2 pronunciation teaching, it is useful to go back to pre- and early- communicative language teaching (CLT) era (for a full historical perspective, see Murphy & Baker, 2015). In light of the four waves of pronunciation instruction during the classical period (pre-1850), intuitive-imitative practices under teacher-centredness (i.e., oral repetition, minimal pairs, imitated pronunciation, reading aloud) were maintained during the first wave (1850-1880). It was not until the second wave (1880-1980) that analytic-linguistic approaches to pronunciation instruction were applied, which consisted of listening, audiolingualism[17] and situational language teaching, and development of pronunciation teaching philosophies, among others. As a reaction to the monotony of audiolingualism, other methods that also targeted native-like pronunciation and were fully teacher-centred were born (e.g., The Silent Way, Total Physical Response). All in all, pronunciation had been a central element in the 60s with the audiolingual and oral-situational methods; however, its popularity was lost from the 70s, at the time that the CLT method was welcomed by many teachers (Derwing & Munro, 2022).

The emergence of CLT implied a drastic shift in the language teaching paradigm, where authentic and meaningful communicative competence gained relevance and learners' fluency was prioritized over pronunciation accuracy (Levis & Sonsaat, 2017). As a result, teacher preparation programmes rarely included pronunciation in their curricula, hence, teachers were not trained and did not feel adequately prepared to teach

[17] The Audiolingual Method's goal was to promote native-like pronunciation and accurate grammar. It was based on the imitation of exemplars until they were memorized, and it was considered extremely monotonous for learners (Flanders & Nuthall, 1972).

pronunciation (Derwing & Munro, 2005). Indeed, Kelly (1969) referred to pronunciation as the Cinderella of language teaching[18].

The first attempts to develop methods for teaching pronunciation under CLT were proposed by Celce-Murcia (1983) and Pica (1984), who marked the initial steps to a future communicative approach to pronunciation teaching. Celce-Murcia (1983) method was based on the following steps: (1) identification of learners' problematic areas, (2) location of grammatical/lexical contexts where sound problems persist; (3) incorporation of these words in communicative tasks, and (4) creation of exercises to practice the challenging sounds in new contexts. Instead, Pica (1984) realized comprehensibility could be affected by inaccurate production of words during communication, therefore, her method consisted of: (1) introduction of a pronunciation point and representative examples for class repetition, (2) incorporation of the pronunciation rule in a communicative activity, (3) monitoring of the communicative activity, and (4) oral presentations of the activities and final report.

From there, Celce-Murcia et al. (2010) proposed a Communicative Framework to teach pronunciation relying on CLT principles where there was an emphasis on pronunciation accuracy while contextualizing it in extemporaneous conversations. The rationale for the framework was to help learners attain sufficient control of the pronunciation features taught (i.e., low cognitive load) before learning to use them for more genuinely communicative purposes (high cognitive load).

---

[18] This view has been challenged by Levis & Sonsaat (2017) who claim that, although pronunciation teaching fell into a period of decline during the CLT era, there was never a time when pronunciation was completely neglected in language teaching. Because of the emergence of "intelligibility" (Levis, 2005), many suprasegmental features were prioritized. Moreover, an important number of professional publications continued to include pronunciation articles throughout the CLT era (see Levis & Sonsaat, 2017, p. 275-276).

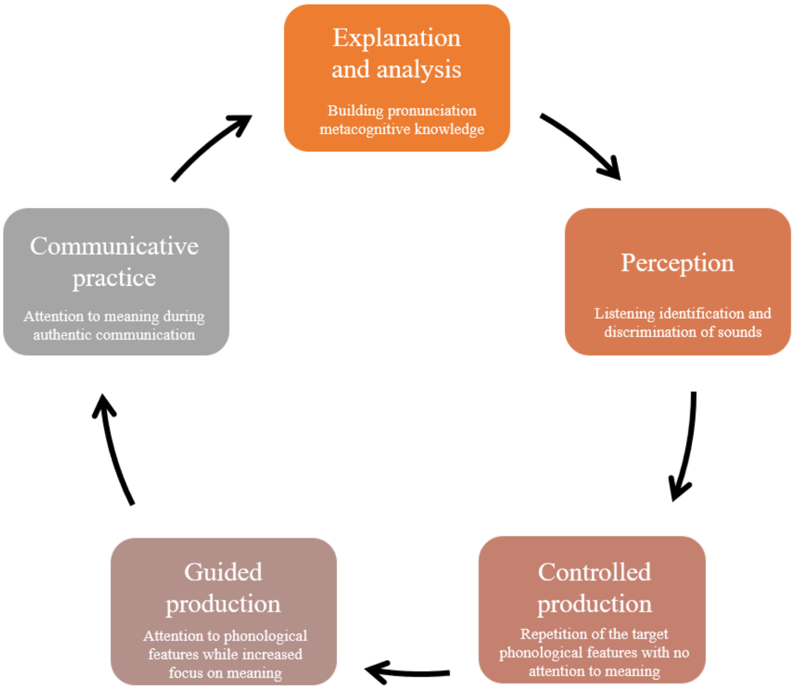


Figure 1.5. Cyclical representation of the 5 stages of the Communicative Framework for teaching pronunciation (Celce-Murcia et al., 2010).

As seen in Figure 1.5., the first three stages of Celce-Murcia et al.'s (2010) framework were based on accuracy-oriented pronunciation-teaching techniques. The first stage made learners aware of certain phonological forms (e.g., explanation with handouts, flipped video); the second stage aimed at providing listening exercises for learners to discriminate key differences between sounds (e.g., listening exercises, odd one out, dictations); and the third stage helped learners attend to form in structured activities (e.g., elicitation, tongue twisters, controlled map tasks). The last two stages constituted a transition from accuracy-focused practice to fluency and meaning-focused use, hence, learners focused on meaning while paying attention to form (e.g., information gaps, minimal pair map tasks) and did fluency-building communicative activities (e.g., discussions, storytelling).

Since then, comprehensibility-based studies have become increasingly common as well as others focusing on intelligibility and choosing pronunciation models for instruction and assessment (Levis, 2018). Integrating pronunciation instruction within communication is crucial to ensure that L2 learners acquire the appropriate pronunciation of new words, along with their meanings (Tyler, 2019). In order to have better control of the phonological forms, many teachers still keep using explicit (rather than form-focused communicative) approaches to pronunciation teaching.

1.3.4.2. Explicit instruction

Explicit instruction (or Focus on FormS [FonFS][19]) means equipping learners with metalinguistic information such as teaching the voicing, place and manner of sound articulations or explaining the acoustic characteristics of segmental, suprasegmental and connected speech features (Sardegna & McGregor, 2022). Meta-analyses (Lee et al., 2015; Saito, 2012; Saito & Plonsky, 2019; Thomson & Derwing, 2015) and individual studies (Derwing et al., 1998; Elliot, 1997; Gordon & Darcy, 2016, 2022; Saito & Lyster, 2012; Sturm, 2013) have provided empirical evidence that explicit instruction is effective in making learners aware of L2 phonological features, resulting in the development of pronunciation accuracy, intelligible (actually understandable) and comprehensible (easy to understand) L2 speech in classroom contexts. In addition, positive results have been shown for both short (Gordon & Darcy, 2016) and long interventions (Darcy et al., 2019; Derwing et al., 1998).

In an attempt to integrate the findings from laboratory studies into L2 classroom pronunciation instruction, Gordon et al. (2013) conducted a study with 30 ESL learners who were separated into three intact classes where learners received 4-hour explicit instruction in either segmentals (vowels) or suprasegmentals (rhythm, stress, linking, and reductions), or no explicit instruction. Findings from a comprehensibility rating task

[19] In L2 learning, Focus on FormS [FonFS] refers to a synthetic approach to language teaching where items of grammar; lexis, phonology, among others, are presented one at a time to the learners via teacher-centred lessons (Long, 1998; Long & Robinson, 1998).

provided compelling evidence for the benefits of explicit instruction for pronunciation, especially for the suprasegmental group whose learners could focus their attention on global aspects that may directly impact their L2 comprehensibility. Similarly, Gordon and Darcy (2016) investigated the effects of short-term (i.e., three 25-minute sessions per 3 weeks) explicit pronunciation intervention in suprasegmental (i.e., stress, rhythm, reductions and linking) and segmental features (i.e., /i/, /ɪ/, /æ/, /ɛ/) on the comprehensibility of ESL learners' production, and compared it to a non-explicit pronunciation instruction group. Explicit phonetic instruction resulted in comprehensibility gains over the course of 3 weeks, but again, only for the group trained on suprasegmentals. Finally, Gordon and Darcy (2022) replicated and expanded their 2016 study with a longer treatment (i.e., 30 minutes per 10 weeks) and more target segmentals and suprasegmentals. Following an explicit instruction plus communicative method (Celce-Murcia et al., 2010), 3 intact classes (i.e., only segmentals, only suprasegmentals, mixed segmentals and suprasegmentals) were compared in terms of comprehensibility, fluency and accentedness. In line with Darcy et al. (2019), Derwing et al. (1998) and Gordon and Darcy (2016), only the suprasegmental group demonstrated comprehensibility and fluency gains in spontaneous speech after the intervention. Accented ratings remained the same before and after the treatment. Lack of comprehensibility gains for the segmental group may be explained by the fact that learners focused their attention on accuracy of segmental production, to the detriment of other aspects that are also necessary to develop comprehensibility in spontaneous speech (e.g., fluency, lexical and sentence stress, rhythm, appropriate pauses, intonation; Derwing et al., 1998). However, other studies (e.g. Kissling, 2013) have found small effect sizes of explicit instruction (around 5%) and no improvement for accentedness or comprehensibility. Table 1.2. shows five key studies on explicit instruction of English segmentals and suprasegmentals.

Research has also suggested that the effects of explicit instruction are maximized when it is combined with corrective feedback, as it raises awareness of different linguistic forms and has very positive effects on L2 speech development in the classroom (e.g., Saito, 2013, 2015; Saito & Lyster, 2012). The best way to correct students' productions is if the problem has been addressed previously in formal teaching rather than an *ad hoc* response (see Couper's 2022 proposal for a framework for pronunciation feedback).

To conclude, although explicit instruction helps learners notice L1-L2 phonological differences (Derwing & Munro, 2005), explicit instruction usually entails a decontextualized focus on the accuracy of specific forms, using mainly controlled practice and the generalizability of such instructional gains to real-life contexts remains contentious. Given the emphasis on grammar-based lessons and ignorance about what aspects of pronunciation should be taught (Section 1.3.4.4.), researchers must inform teachers and trainers about what aspects of L2 pronunciation should be prioritized, and how they can be integrated in content-based lessons. Conducting interactive tasks in dyads has the enormous potential of increasing learners' awareness of pronunciation's communicative impact because learners are able to incidentally focus on form while communicating meaning.

Table 1.2. Summary of five key explicit instruction studies on L2 English pronunciation.

Study	Participants	Target features	Method	Duration	Findings
Derwing, Munro & Wiebe (1998)	48 speakers in Canada (1) 16 speakers (9 European; 1 Asian; 3 Spanish; 3 non-specified) -length of residence: 3.9 years- (2) 16 speakers (10 European; 2 Asian; 3 Spanish; 1 non-specified) -length of residence: 2.7 years- (3) 16 speakers (8 European; 4 Asian; 2 Spanish; 2 non-specified) -length of residence: 3.2 years-	Segmentals (vowels and consonants) Suprasegmentals (speaking rate, intonation, rhythm, projection, word stress, and sentence stress)	3 groups: (1) explicit segmental instruction (identification, discrimination and repetition tasks); (2) global instruction (general speaking habits and prosodic factors: discourse-level tasks); or (3) no pronunciation specific instruction.	20 hours x 12 weeks	Significant improvement for comprehensibility, and accentedness in the two instructional groups. Larger and more generalized improvement in comprehensibility obtained by the global instruction group on extemporaneous productions.
Couper (2006)	71 speakers in New Zealand Experimental: 21 speakers (14 Chinese; 1 Korean; 6 other Asians) -length of residence: 2.5 years- Control: 50 speakers (26 Chinese; 13 Korean; 2 other Asians) -length of residence: 3 years-	Epenthesis (the addition of an extra sound, usually a schwa, after a consonant) and absence (the inappropriate dropping of a consonant sound)	Experimental: Diagnostic test; Listening of the learners' recordings; Listen and repeat; Explicit explanations on the rule and nature of the English syllable. Baseline: No explicit focus on pronunciation.	30 minutes x 12 sessions	No significant effect of instruction on perception. Instruction was effective for production (compared to the baseline group) and learning was retained 12 weeks after.
Saito (2011)	20 Japanese speakers in the US (length of residence: 2.3 months)	Segmentals: /æ,f,v,θ,ð,w,l,ɹ̂/	Experimental: Identification, discrimination and production of sounds in a controlled manner (similar to phonetic training). Control: No pronunciation instruction.	60 minutes x 4 weeks	Explicit instruction benefited comprehensibility in the experimental group especially at the controlled speech level. No significant improvement in accentedness.
Gordon & Darcy (2019)	22 Spanish speakers in Costa Rica (1) 7 speakers, (2) 8 speakers, (3) 7 speakers	Segmentals: vowels /i, ɪ, ɛ, æ, ʌ, ɑ, u, ʊ/ and consonants /p, t, k, b, d, g/ Suprasegmentals: word	3 groups: (1) Segmental, (2) Suprasegmental, (3) Segmental + Suprasegmental. Sequence: Description & Analysis / Listening Discrimination;	30 minutes x 10 weeks	Explicit instruction on suprasegmentals during a short period of time can enhance comprehensibility.

		and sentence stress, rhythm, intonation, linking, contractions, and vowel reduction	Controlled & Guided Practice; Communicative Practice (Celce-Murcia et al., 2010).		Mixed and Suprasegmental groups were rated as significantly more comprehensible than the Segmental group after the intervention.
Yeldham & Choy (2022)	38 Cantonese speakers in Hong Kong Experimental: 18 speakers Control: 20 speakers	Segmentals: /iː/-/eɪ/; /z/-/v/; /t/-/d/	Experimental: 1- direct instruction: exposure to the teacher's model pronunciation; illustrations of relevant articulatory configurations (explanations, diagrams); demonstrations of the various articulatory placements; introduction to the relevant International Phonetic Alphabet symbols/sounds; articulatory–abdominal instruction; minimal pairs exercises. 2- controlled practice: recitation and reading aloud; pronunciation squares activities . 3- pedagogical tasks: role-plays, short speeches, dialogue-based interactions. Control: no direct attention to pronunciation.	4 hours	The experimental group greatly outperformed the comparison group separately in all three segmental categories examined.

1.3.4.3. Form-focused instruction and communication

Following Long (1998), and Long and Robinson's (1998) definitions, Focus on Form (FonF) refers to drawing learners' attention to linguistic elements in context, as they incidentally arise as a consequence of comprehension or production problems in meaning-driven interactions. These definitions were later adopted by Saito (2012), who refers to FonF instruction when learners' attention is drawn to form in communicative contexts, hence, when practicing pronunciation form while being involved in contextualized meaning-oriented communicative activities. In contrast, FonFS refers to the practice of pronunciation form in a decontextualized manner, via controlled practice (i.e., mechanical drills and choral repetition) and no elaboration.

For instance, Saito (2013) tested the pedagogical efficacy of providing explicit instruction prior to FonF (relative to only FonF) to boost the effects of FonF on familiar/unfamiliar lexical items across various task and phonetic conditions. After 1-hour lessons distributed over 2 weeks, Japanese EFL learners who received FonF without explicit instruction demonstrated medium improvement of English /ɹ/, especially in familiar lexical contexts. However, learners in the FonF + explicit instruction group showed across controlled and spontaneous production tasks and vowel conditions, and generalization of gains to novel lexical contexts. Following the same methodology and materials, Ruan and Saito (2023) instructed intermediate Chinese learners of English to pay attention to the phonological differences between /iː/-/ɪ/ through explicit instruction (i.e., awareness raising task) and a FonF communicative task (i.e., debate with enhanced target items containing the vowel contrast) for 1.5 hours. Results from a forced-choice identification task showed an overall 4% instructional gain in the identification of /iː/-/ɪ/, but the benefits of FonF were unclear for learners with low levels of auditory processing.

On the other hand, instruction-method comparison studies have found a superiority of FonF over FonFS instruction to develop intelligibility, comprehensibility and L2 pronunciation accuracy (Abe, 2011; Darcy et al., 2019, 2021; Park, 2000; Saito, 2015). Recently, Darcy and Rocca (2023) compared FonFS "explicit-only" pronunciation instruction, which consisted of listening and guided practice of suprasegmental features of L2 English, to FonFS + "communicative" (FonF) pronunciation instruction, which combined explicit repetitive instruction with communicative activities that enhanced proceduralization —a mixture of the CLT and ACCESS frameworks—. Their findings revealed that whereas the "explicit-only" learners only improved L2 comprehensibility in the controlled reading task, the "explicit + communicative" group improved consistently in both controlled and spontaneous tasks, and noticeably in the group-discussion task. In fact, Saito's (2012) synthesis of 15 quasi-experimental studies showed that, whereas FonFS studies resulted in improvement only at a controlled level, FonF studies enabled learners to improve at controlled and spontaneous speech levels (See Table 1.3. for a summary of FonF studies).

First, according to Darcy et al. (2019), two key components form effective learning: automaticity of L2 phonological and phonetic processing and generalization from classroom practice to actual behaviour. FFI can merge both, by using activities that are inherently repetitive yet have a high communicative value (Gatbonton & Segalowitz, 1988). Second, integrating language focus into meaning-oriented classrooms (FonF) is hypothesized to help students establish form-meaning mappings as well as to promote proceduralization of their declarative knowledge (Lyster, 2007). In L2 phonology, Trofimovich and Gatbonton (2006) suggest that pre-planned form-focused activities that occur during genuinely communicative L2 interaction could be considered as contextualized repetitive practice, resulting in impacts not only on accuracy but also on fluency. In contrast, instruction with focus exclusively on forms (FonFS) does not allow students to transfer what they learn in classroom to outside of the classroom.

In order to facilitate a carryover from targeted practice into spontaneous, real-word communication, Grant (2014) and Darcy et al., (2019) suggested shifting the communicative load from more to less controlled in an instructional sequence, conveying this progression by plotting activity types on a form-meaning continuum. The work by Sardegna (2022) advocates for a transition from FonFS to FonF in a pronunciation strategy-based methodology. Sardegna and McGregor (2022) recommend that, first, students can practice their oral skills through decontextualized oral drills (FonFS) such as reading aloud isolated words, sentences or paragraphs repeatedly with the help of pronunciation learning strategies, repetition and speech models. Once they are aware of their pronunciation difficulties, they can move on to practice their oral skills through contextualized tasks (FonF) such as role-plays, interviews, information-gap activities, picture-based storytelling, etc. Nevertheless, the danger is that FonF tasks may inadvertently lapse into FonFS due to previous explicit instruction. In such scenario, instruction would lose its communicative orientation and become predominantly linguistically focused.

The current dissertation adopts a task-based FonF approach to L2 pronunciation learning (See Section 1.5.) and assesses L2 pronunciation instruction gains through acoustic measures, namely, Mahalanobis distances (Mahalanobis, 1936) between learners' productions of contrastive L2 vowels, and between learners' vs native-speakers' vowel productions (Kartushina et al., 2016). We advocated for the use of acoustic analyses, rather than impressionistic judgements of comprehensibility and accentedness, because the main aim was to examine the direct impact of task-based pronunciation instruction on specific aspects of L2 pronunciation (i.e., L2 vowels), which would have been difficult to observe if we had assessed learners' global pronunciation proficiency (Saito & Plonsky, 2019). Previous work has demonstrated that listeners not only rely on phonological and temporal information but also on lexico-grammatical information when making comprehensibility judgements (Isaacs & Trofimovich, 2012). In addition, provided that rater assessment of segmental accuracy has been found to be subject to variation (Saito & Plonsky, 2019), acoustic rather than subjective analyses of L2 vowels were conducted. A potential disadvantage of listener-based analyses (relative to acoustic analyses) is that listeners may differ in their accuracy ratings depending on their previous listening experience, familiarity with a particular accent, or the use of different acoustic information to analyse the same speech targets. The last sections deal with potential challenges and solutions in the implementation of pronunciation instruction.

Table 1.3. Summary of five key FonF instruction studies on L2 English pronunciation.

Study	**Participants**	**Target features**	**Method**	**Duration**	**Findings**
Park (2000)	44 speakers in the US Experimental: 32 speakers. 3 intact classes: (1), (2), (3) Control: 12 speakers	Suprasegmentals: Lexical stress	(1) FonF: forms-focused practice into meaning-based communicative activities and received feedback. (2) FonM: purely meaningful teacher-fronted activities. (3) FonFS: teacher-led drills and controlled pair practice such as reading dialogues.	65 minutes x 10 sessions	Both FonFS and FonF groups showed significant improvement after treatment. FonF Group obtained significantly more improvement than the FonFS group in perception and production.
Abe (2011)	60 Japanese speakers in Japan (1) 30 speakers (2) 30 speakers	Suprasegmentals: Weak forms	(1) FonF: negotiation-of-form treatment, which was comprised of noticing and form-negotiation task in dyads. (2) FonFS: explanation of English weak forms and listen-and-repeat exercises.	4 weeks	FFI treatment significantly improved learners' perception and production of English connected speech and gains were retained 1 month after. The FonFS group demonstrated small improvement only at the immediate post-tests.
Saito (2015)	49 Japanese speakers in Japan (1) 18 speakers (2) 17 speakers (3) 14 speakers	Segmentals: /ɹ/	(1) FonF (embedded segmental feature in argument critique, English debating, argument creation, and public speaking tasks) with corrective feedback. (2) FonF (embedded segmental feature in argument critique, English debating, argument creation, and public speaking tasks) without corrective feedback. (3) Control: no focus on phonetic form.	4 hours x 2 weeks	Communicative FonF is facilitative of the development of L2 speech perception and production of /ɹ/. Corrective feedback was not effective for acquisition.

Darcy, Hancock, Lee & Rocca (2019)	25 speakers in the US with mixed L1s: Experimental: 15 speakers (1), (2) Control: 10 speakers	Suprasegmentals: Word stress and vowel reduction	(1) FonFS + FonF: explicit instruction and feedback + genuinely communicative activities -> merger of the Communicative Framework (Celce-Murcia et al., 2010) and ACCESS (Gatbonton & Segalowitz, 2005). (2) FonFS: explicit instruction and feedback, controlled practice that is purely metalinguistic, not communicative. (3) Control: no specific pronunciation instruction.	6 weeks	Both FonFS + FonF and FonF improved comprehensibility and vowel reduction, whereas the control group remained unchanged. *See task effects* Only FonFS + FonF group showed more consistent comprehensibility and vowel reduction improvements and transferred gains to spontaneous speech.
Darcy, Rocca & Hancock (2021)	3 EFL instructors from a content-based Intensive English Program in the US. 6 intact classes, 25 students with mixed L1s: (1) 9 speakers (2) 6 speakers (3) 10 speakers	Suprasegmentals: word stress, sentence stress, intonation, connected speech, thought groups, and reduced function words Segmentals: through English Accent Coach (Thomson, 2012)	(1) FonF: focus on phonetic form embedded in meaning-based activities -> merger of the Communicative Framework (Celce-Murcia et al., 2010) and ACCESS (Gatbonton & Segalowitz, 2005). (2) FonFS: decontextualized focus on the accuracy of specific forms; controlled practice and little elaboration. (3) Control: no specific pronunciation instruction.	250 minutes x 7 weeks	FonFS & FonF groups became more comprehensible compared to the control. FonF had the greatest positive change in comprehensibility. Pronunciation integration in communicative classes shows improvement in oral production

1.3.4.4. Challenges: when, how, what?

Several studies have demonstrated that both teachers and students recognize the importance of pronunciation (Zielinski & Yates, 2014). However, when asked how much they actually teach pronunciation, teachers report seldom teaching pronunciation (Darcy, et al, 2012a) mostly due to their lack of time (Burri, 2021), confidence (Uchida & Sugimoto, 2018), probably attributed to the absence of pronunciation-specific training opportunities in second language teacher education programs (Burri, 2021; Couper, 2016). Following Darcy's (2018) taxonomy of the factors contributing to the challenges of pronunciation instruction in classroom context, I will address the *when*, *how* and *what* obstacles.

Time

Popular teacher concerns include having little time for pronunciation instruction, when grammar and vocabulary go first, or wondering why they should teach pronunciation if oral skills are not assessed. First, speaking and pronunciation are usually taught separately from grammar and vocabulary —unless the school follows a TBLT or CLT and coursebooks integrate these two components— thus, adding the pronunciation component unfeasible in intensive language curricula. Second, L2 oral production and pronunciation are neglected in the FL classroom (Tragant, 2009; Tragant et al., 2010) because many teachers do not see a compelling reason to do so (Levis, 2022). Teachers always find time to teach what they believe is important, either because a teacher finds a topic essential to what they think learners should master, a curriculum prioritizes it or school assessment requires it. In fact, in the last few decades, most publishing houses have been pushing grammar and vocabulary-syllabi, which tend to minimize the importance of pronunciation, and relegate it to a few exercises at the end of coursebooks.

Method

Another key obstacle in pronunciation instruction has to do with the “how”: Teachers wonder about how to go about teaching pronunciation, whether they are entitled to do it as non-native speakers or consider teaching pronunciation as boring and ineffective. First, lack of teacher training on phonetics and phonology (Baker, 2014; Murphy, 2014) results in an increased widespread insecurity (Kirkova-Naskova et al., 2021) about which methods are effective for improving pronunciation instruction. Second, lack of empirically validated classroom materials, hence, inadequate models of practice (Levis & Sonsaat, 2016), have casted some doubt on how effective pronunciation can be addressed in the classroom. Being unfamiliar with recent trends in pronunciation instruction yields a lack of transfer from traditional decontextualized pronunciation practices (i.e. mechanical drills and choral repetition) to communicative opportunities using authentic language (Grant, 2014). In addition, given that most of teachers’ phonetic and phonological knowledge comes from outdated methods, pronunciation teaching may result in demotivating lessons for their students (Levis, 2022) and teachers may end up abandoning it.

Focus

The last obstacle has to do with pedagogical priorities that teachers need to make for pronunciation. In other words, teachers may be concerned about which phonological aspects of speech they should focus on, or which should be the target of pronunciation instruction if learners are heterogenous in L1 and L2 proficiency (Darcy, 2018). Although the goal of pronunciation instruction has shifted from achieving "nativelikeness" toward achieving comfortable intelligibility (Levis, 2005, 2018), teachers struggle with selecting an appropriate focus (i.e., setting priorities) for the pronunciation curriculum. All learners within a given level might differ in terms of how intelligible their spontaneous speech is, depending on their L1 background, L2 proficiency or the quality and intensity of their interactions (Darcy, 2018) inside and outside the classroom. This makes the task of shortlisting and prioritizing target pronunciation features especially challenging.

1.3.4.5. Solutions: integration and prioritization

In this section, I will address potential solutions to the time (*when*), method (*how*) and focus (*what*) obstacles for pronunciation instruction through integration and prioritization principles.

Integration

A potential solution to the constraints of time to dedicate to pronunciation in instructional contexts is to make pronunciation instruction a curricular component that is integrated with the other areas of language by giving it regular attention into every lesson (see Darcy et al. 2012a; 2021; Darcy & Rocca, 2023; Jones, 2016; Levis, 2022). Benefits of integrating pronunciation include (1) ensuring pronunciation is actually addressed in a contextualized manner during spontaneous conversations (Mora-Plaza et al., 2018); (2) a positive impact on other language skills (Levis & Echelberger, 2022); in other words, it may serve as a "value-added" factor in language instruction (Pennington & Rogerson-Revell, 2019); (3) overall better fluency, comprehensibility and intelligibility. However, guidance for teachers on how to effectively integrate pronunciation instruction into listening and speaking or all-skills courses and how to provide feedback on pronunciation features is lacking (but see Jones, 2016), and descriptions of what integration looks like in practice are missing.

Levis and Echelberger (2022) proposed two approaches of integrating pronunciation: pronunciation-focused integration (i.e., making pronunciation preeminent and connect it to other language skills) or other skills-focused integration (i.e., emphasizing other language skills and identifying where pronunciation is essential to teaching them), but there are multiple ways of integrating the teaching of pronunciation segmentals and suprasegmentals into the other skills (Gordon, 2022; Pennington & Rogerson-Revell, 2019). Empirical research has provided evidence of the multiple benefits of integrated pronunciation instruction, even for a short period of time (Darcy et al., 2021). Darcy and Rocca (2023), who defined integrated pronunciation instruction in their study as “features that were previously taught in isolation are recycled by the teacher and become connected to the curriculum when the instructor draws attention to them in context, either helping students notice the use of the form in context and/or by providing explicit feedback” (p.11-12), also found beneficial effects of integrating pronunciation instruction for overall L2 comprehensibility.

Finally, pronunciation instruction should be introduced as early as possible (i.e., in the window of maximal opportunity (Derwing & Munro, 2015) in SL contexts), so students can learn the phonological form of words as their vocabulary grows (Tyler, 2019) and teachers develop confidence at teaching pronunciation as it becomes part of a systematic routine. Pronunciation integration would be even more efficient if oral skills were assessed by school teachers and university entrance exams.

Dual-focus approach and research-informed materials

It is well-known and empirically attested that when teachers have students pay explicit attention to pronunciation features and dedicate class time to meaningful and communicative pronunciation practice, L2 pronunciation improves over time. However, teacher training programs rarely include a pedagogical component addressing how to teach pronunciation (Couper, 2016; Henderson et al., 2015).

One way is to implement a dual focus on form[20] and meaning (Darcy, 2018), which contains explicit and communicative activities in which repetition and elaboration are key techniques to learn L2 pronunciation robustly. Using the Communicative Framework (Celce-Murcia et al., 2010), ACCESS (Gatbonton & Segalowitz, 2005; Trofimovich & Gatbonton, 2006) or Strategy-based (Sardegna, 2022) frameworks for pronunciation teaching has the potential to progressively guide learners to pay attention first to form (i.e., pronunciation awareness activities), and then to both form and meaning (i.e. automatization activities). Another possible strategy to ensure that attention to form is indeed maintained as learners focus more on meaning is to include tasks whose successful completion depends on sufficient control of the pronunciation targets (Gordon, 2021; Mora-Plaza et al., 2018; Mora & Levkina, 2018; Solon et al., 2017).

In addition, a dual-focus from form to meaning can be provided with many innovative (e.g., Hancock, 2017) and authentic materials (e.g., *Sounds of Speech* <soundsofspeech.uiowa.edu>) that include recordings from both native and non-native speakers, or from speakers of different English varieties, and put it into practice with learners with the same or other L1 (Sonsaat-Hegelheimer & McCrocklin, 2022).

[20] This would be referred as "Focus on FormS" according to Long (1997) and Long and Robinson (1998).

Setting priorities: Diagnosis and functional load

It is now generally accepted that the goal of achieving "nativeness"[21] (promoted by the audiolingual method in the mid-20th century) is unrealistic and, in most cases, unachievable (Levis, 2005). As a result, contemporary research and teaching has been motivated to promote comfortable intelligibility and comprehensibility of learners' spontaneous speech (Derwing & Munro, 2022) through instructional approaches that involve communicative and contextualized tasks. Given the time constraints that characterize pronunciation instruction, there is a need to determine which segmental (Jenkins, 2000) and suprasegmental (Derwing et al., 1998; Gordon & Darcy, 2016) speech features matter the most for comprehensibility and intelligibility.

One potential solution is to diagnose learner needs based on learners' spontaneous productions in class through dictations or listening comprehension exercises (Darcy, 2018). For instance, Isbell (2020) elaborated the Korean Pronunciation Diagnostic to include a detailed coverage of Korean's phonemic inventory assessed via perceptive (i.e., pronunciation judgement; nonword identification) and productive (i.e., picture naming; nonword reading) skills to improve learners' intelligibility in L2 Korean. Levis and Echelberger (2022)'s diagnosis approach consisted of recording students' speech for around 30 seconds, listening to it, and evaluating following a simple rubric that contained a holistic scale (few errors → many errors) for several pronunciation features.

Apart from conducting a diagnosis, a fruitful strategy to help teachers prioritize specific segmental contrasts (instead of focusing on them all) that may affect intelligibility, comprehensibility and overall proficiency (Suzukida & Saito, 2022) could

[21] The "nativeness" principle argues that the primary goal of L2 pronunciation teaching and learning is the attainment of native-like pronunciation, hence, learners can and should achieve native-like accents (Levis, 2005; 2018).

be to consider functional load[22]. Some sounds have been found to cause more difficulties than others (Munro & Derwing, 2006), based on the number of pairs of words differing in one sound (i.e., minimal pairs); the number of minimal pairs that a particular phonemic distinction differentiates at the beginning or end of a word; the frequency of occurrence of each word in a minimal pair; and the part of speech of the two words in a minimal pair (Alnafisah et al., 2022). For instance, English consonant contrasts such as /t/ and /d/ (e.g., *tip/dip*) and the vowels /e/ and /æ/ (e.g., *men/man*) have a high functional load because of the large number of minimal pairs with these sounds, hence, would be more likely to warrant emphasis in the curriculum. In contrast, the dental fricatives /θ/ and /ð/ or /ʤ/ and /j/, have a relatively low functional load because few English words could be misunderstood due to their mispronunciation (e.g., *thigh-thy; juice-use*). Munro and Derwing (2006)'s study revealed that Cantonese-accented read-aloud English sentences that contained a single high functional load error were rated as significantly harder to understand than those with a low functional load error. For example, substituting /d/ for /z/ were rated as significantly less comprehensible than sentences containing low functional load divergence (substitution of dental /d/ by alveolar /d/). They also found evidence of a cumulative effect in that sentences with more than 1 high functional load error (but not multiple low functional load errors) were rated as significantly more accented and less comprehensible than those with only one. Together with the studies by Suzukida and Saito (2022) and Alnafisah et al. (2022), these findings shed light on the fact that high functional load pairs should be prioritized in language pedagogy in order to improve learners' comprehensibility of their spoken language.

[22] Functional load is defined as the hierarchy of segmental contrasts that determines which vowels or consonants are more prone to cause problems in communication if they are mispronounced (Brown, 1991).

Finally, language teachers should pay attention to key differences between the L1 and L2 that may explain difficulties to achieve intelligible segmental production (see Section 1.1.2. for theoretical speech models that provide explanations), target them in form-focused instruction and order appropriately in an instructional syllabus. Given the lack of time and resources, published basic pronunciation diagnoses (McAndrews & Thomson, 2017) can be used for the aforementioned purpose.

Section 1.3. has reviewed several instructional approaches that have been shown to draw learners' attention to L2 phonological form and lead to L2 pronunciation development. Considering the advantages and challenges of each one of the pedagogical approaches (i.e., phonetic training, CAPT, explicit and form-focused communicative instruction), the present dissertation assesses the effectiveness of pronunciation-focused integration and a dual focus approach where relevant phonetic forms are part of L2 communication, with the aim of promoting intelligible and comprehensible L2 speech. The following section introduces task-based language teaching as a pedagogical approach that, through task design and manipulation, promotes the development of L2 linguistic forms under fully communicative circumstances (Long, 1985, 2015, 2016) that reflect real-world interactions.

1.4. Task-based language teaching (TBLT) and task complexity

This section revises TBLT theoretical underpinnings in relation to SLA, thus, it attempts to present the foundations of TBLT; relate the concept of FonF to TBLT; revisit definitions of "task" from the TBLT literature; assess possible ways in which tasks can be manipulated and sequenced; investigate the effects of task complexity on L2 oral production; and identify potential challenges associated with the implementation of TBLT along with potential solutions.

1.4.1. TBLT: origins and definitions

This section's aim is to briefly revise the origins of TBLT since the beginnings of the structural approach to understand its main tenets. [23]

The dominant approach to teaching foreign languages in the 1980s was the structural one, where a language was broken down into grammatical, lexical and phonological bits that were presented and practised individually and sequentially (Ellis, 2021). CLT arose as a response to the belief that synthesis of linguistic features was not enough to be able to communicate in a second language. Although this approach claimed to be "analytic" rather than "synthetic", the notional syllabus was based on a fragmentation process and the methodology used to implement it remained essentially the same -drills and situational exercises (Ellis, 2021). Within CLT, the Presentation-Practice-Production methodological approach emerged with the final *P* consisting of a task intended to provide opportunity for learners to communicate freely using the target feature. Byrne (1986) suggested that the order of the three stages could be flexible, with the free production stage preceding the presentation and practice stages. Nevertheless, in all the aforementioned proposals the focus of the lesson remained on pre-selected target features that learners had to use accurately and freely in communication.

It was not until the mid to late 1980s that the first proposals for a task-based approach appeared. These early proposals (Breen, 1987; Candlin, 1987; Long, 1985) focused on the rationale for a task-based syllabus and outlined how to design and evaluate a task-based curriculum. Prabhu (1987) was the first to provide a complete account of a task-based course, as a result of his dissatisfaction with grammar-based methods (i.e., the Structural–Oral–Situational Method) that were built on the assumption that language was

[23] See Section 1.4.3. for numerous definitions of "tasks" according to the TBLT literature.

a process of successive input-assimilation so what the teacher taught was what students would learn. Prabhu (1987) argued that "the development of competence in a second language requires not systematization of language input or maximization of planned practice, but rather the creation of conditions in which learners engage in an effort to cope with communication" (p. 1) through tasks. Nunan (1989) offered teachers a practical introduction to the design and use of tasks, as a means through which learners could learn to communicate through interaction in the target language, by drawing on their linguistic resources in a creative way. Finally, Candlin (1987) highlighted the educational standpoint of tasks as learners would become more aware of their own personalities and social roles and this would generate an increase in self-fulfilment and self-confidence.

Contrary to previous approaches to language teaching, TBLT was based on a true "analytic approach", where learners were presented with holistic samples of language and learners' role was to analyse the samples and induce rules for themselves. The emphasis was on "natural and authentic representations of target language communication as possible", while learners were engaged in meaningful target language production (Long, 2015, p. 20). In contrast with grammar-based syllabus that were product-oriented, the focus of TBLT was on learners and their learning process. FonF arose incidentally on lessons when learners faced a communication problem; thus, it depended on the learner's developing language system instead of pre-selected language forms (Long, 1998; Long & Robinson, 1998). Long (1985), Skehan (1988) and Ellis (2003)'s approaches to TBLT shared the need of focus on form in natural language contexts, and agreed that tasks should be the organizing principle of a syllabus, being the driving force in lessons and instruction should be learner-centred. However, they were they were different in terms of how attention to form was achieved (i.e., Long underlying corrective feedback, Skehan task design and pre-task planning, and Ellis several ways in all 3 stages of a task-based lesson); focused vs. unfocused tasks (i.e., while Skehan and Long reject focused tasks – which typically promote a covert structural syllabus – Ellis is more flexible about using tasks to teach grammatical structures or other linguistic units) and rejection of traditional approaches (i.e., Long and Skehan regarded structural teaching as theoretically unsupportable).

In sum, TBLT advocates for a more natural approach where attention to language is not artificially imposed but rather driven by the context the task is placed in. Tasks are real-world communicative activities to which learners will put the L2 beyond the classroom (Long, 2016) and they constitute the organizing units of the syllabus[24]. Task detection, description, and selection during needs analyses determine every other aspect of syllabus design, such as task sequencing, pedagogic task design, methodological implementation, assessment and even program evaluation (Gilabert & Malicka, 2021).

1.4.2. Focus on form, interaction and L2 development in TBLT

This section further explains why TBLT is an analytic FonF approach and presents empirical evidence of the benefits of negotiation for meaning during interaction and FonF techniques for L2 oral production and development.

Following up on Long (1996, 1998) and Long and Robinson's (1998) definitions, Long (2015, p. 317) defined FonF as a "reactive use of a wide variety of pedagogic procedures to draw learners' attention to linguistic problems in context, as they arise during communication (in TBLT, typically as students work on problem-solving tasks),

[24] However, there is no single way of doing "TBLT" (Ellis, 2009, p.224). The approach needs to be adapted to the teachers' own beliefs, classrooms and situations, as long as they enhance student motivation and communicative proficiency, differential understandings are welcomed (Willis & Willis, 2007).

thereby increasing the likelihood that attention to code features will be synchronized with the learner's internal syllabus, developmental stage and processing ability." TBLT researchers (e.g., Ellis, 2003; Long, 1985, 2015; Long & Robinson, 1998; Robinson, 2001b, 2011; Skehan, 1998) believed that FonF within communicative tasks was a central component of task-based frameworks. Contrary to the FonFS used in structural methods; FonF in TBLT had to be incidental, implicit, and purely reactive brought about in interactional settings (Sudharshana & Mukhopadhyay, 2021).

According to Gilabert et al. (2016), performing FonF tasks involved SLA learning processes of input processing, intake processing and L2 knowledge processing as well as output production (see Leow's, 2015 model). First, SLA research, studies in immersion contexts and psycholinguistic studies on the role of consciousness in language acquisition highlighted the need for conscious attention to forms particularly those which are infrequent, or perceptually non-salient. As a result, there was a need to direct learners' conscious attention towards language form "without losing the values of tasks as realistic communicative motivators, and as opportunities to trigger acquisitional processes" (Skehan, 1996, p. 42). Through the manipulation of task design to FonF, tasks allowed for the noticing of relevant linguistic information, as well as the establishment of form-meaning connections, hypothesis formation and testing, and analyses of L2 internal representations (Gilabert et al., 2016). In fact, maintaining learners' attention on both meaning and form at the same time was thought to aid proceduralization of learners' declarative knowledge in long-term memory (DeKeyser, 1998).

Interactionist researchers considered that conversational interaction provided crucial opportunities for learners to refine and restructure their interlanguage by drawing their attention to linguistic code features during negotiation for meaning (Gass, 1997; Pica, 1994). According to the interaction hypothesis (Long, 1985, 1996), when negotiating for meaning, linguistic resources mediated by selective attention and L2 processing capacity are brought together, which provokes adjustments to linguistic form and message content. When learners are communicating with peers or the teacher, learners can engage in language-related episodes (LRE)[25] (i.e., clarification requests, confirmation checks and comprehension checks) that result in reformulations, repetitions, prompts or paraphrases to resolve communication breakdowns (Swain & Lapkin, 2001). In fact, interaction enables input to become "uniquely tailored to individual's strengths, weaknesses and communicative needs, providing language that suits their distinct developmental levels" (Mackey, 2012, p.12). Therefore, negotiated interaction offers valuable opportunities for learners to produce modified output (Output Hypothesis: Swain, 1985, 1995) because learners can pay attention to their own productions in order to communicate successfully. Three functions of learning are involved in output production: noticing (i.e., learners notice gaps and analyse their linguistic knowledge therefore gaining awareness about the knowledge they lack about the target language), hypothesis testing (i.e., learners can test whether they are understood during communication and whether their utterances are well-formed, usually through LRE) and metalinguistic function (i.e., learners externalise their hypotheses about form and meaning and may detect problematic issues in their interlanguage) (Swain, 1995). In addition, increasing task demands fosters automatization of form-function mappings and helps learners progressively engage in tasks that approximate real task performance (Robinson, 2011b). According to Gilabert et al. (2016), output practices through tasks involve moving from semantic processing in comprehension to syntactic processing,

[25] Language-related episodes (LRE) are defined as instances in which learners "talk about language they are producing, question their language use, or other- or self-correct their language production" (Swain & Lapkin, 2001, p. 104).

which entails higher attention to form and deeper language analyses, hence, a beneficial effect on L2 language development.

Over the years, different ways of FonF have been proposed (see Sudharshana, 2021 for a detailed review). Proactive FonF techniques include input enhancement, input flooding, input elaboration, task essential language, priming, etc. during the pre-task and reflection, metalinguistic explanations, awareness raising activities during the post-task (Doughty & Williams, 1998, Ranta & Lyster, 2018), and reactive FonF techniques refer to recasting, negotiation of form/meaning, output enhancement during the task cycle[26] (Saito & Lyster, 2012). Both proactive and reactive approaches have been found to enhance the probability that language learners pay attention to particular aspects of the language code while primarily focused on producing and understanding messages in a purely meaningful situation (Doughty, 2001; East, 2012).

L2 development in TBLT has been conceptualized as gains in pre- to post-tests (e.g., Révész & Han, 2006) and has been often discussed in terms of measuring L2 performance in complexity, accuracy, and/or fluency (for a thorough review of CAF measures see Housen et al., 2012). In addition, empirical research has found task manipulations (type, complexity, mode, etc.) to promote the development of L2 linguistic accuracy. Whereas TBLT has been considered successful for promoting L2 development, this statement is applicable primarily for lexical (e.g., locative prepositions, Gurzynski-Weiss & Baralt, 2014), grammatical (e.g., imperfect subjunctive, Baralt, 2014; past progressive, Révész & Han, 2006), and more recently, for pragmatic linguistic targets (e.g., speech acts of giving opinion, agreeing/disagreeing, interrupting, and acknowledging the interlocutor, Barón et al., 2020; e.g., requests, Kim & Taguchi, 2015).

[26] Proactive FonF refers to the creation and manipulation of tasks wherein learners are required to pay attention to form for successful task completion. Reactive FonF refers to the provision of feedback (e.g., recasts) in response to the occurrence of linguistic inaccuracies.

To the best of our knowledge, only a special issue in the *Studies in Second Language Acquisition* journal (Gurzynski-Weiss et al., 2017a) set out to examine whether task design and manipulation could direct learners' attention to targets beyond L2 grammar, lexis, and pragmatics (See Section 1.5.).

1.4.3. Tasks

In this section, definitions of "task" will be listed and contrasted in terms of "language learning goals" and "educational activities"; task types will be presented; criteria for task selection will be proposed; and well-known frameworks for task-based syllabus design will be presented, with special attention on Willis' (1996) Task Based Learning Framework.

1.4.3.1. Definitions

In the literature, various "task" definitions have been offered that differ quite widely in scope and formulation (for overviews see Bygate et al., 2001; Ellis, 2003; Van der Branden, 2006). See Table 1.4. for definitions of "task" over the years.

Definitions from Long (1985), Crookes (1986), Carroll (1993), Bachman & Palmer (1996), Bygate et al. (2001), among others, define tasks with a language learning goal, that is "an activity in which a person engages in order to attain an objective, and which necessitates the use of language" (Van den Branden, 2006, p. 4). These definitions emphasize that tasks are activities which are goal-directed (linguistic or non-linguistic) and language is a means to completion.

Table 1.4. Definitions of "task" ordered by year of publication.

Author (year)	Definition
Long (1985)	The hundred and one things people do in everyday life, at work, at play and in between. Tasks are the things people will tell you they do if you ask them and they are not applied linguists.
Crookes (1986)	A piece of work or an activity, usually with a specified objective, undertaken as part of an educational course, or at work.
Candlin (1987)	One of a set of differentiated, sequenceable, problem posing activities involving learners and teachers in some joint selection from a range of varied cognitive and communicative procedures applied to existing and new knowledge in the collective exploration and pursuance of foreseen or emergent goals within a social milieu.
Krahnke (1987)	An activity that learners has to do for non-instructional purposes outside the classroom as opportunities for language learning.
Prabhu (1987)	An activity which required learners to arrive at an outcome from given information through some process of thought, and which allowed teachers to control and regulate that process.
Breen (1987)	A range of workplans which have the overall purpose of facilitating language learning from the simple and brief exercise type to more complex and lengthy activities such as group problem-solving or simulations and decision-making.
Nunan (1989)	A piece of classroom work which involves learners in comprehending, producing or interacting in the target language while their attention is principally focused on meaning rather than form.
Carroll (1993)	Any activity in which a person engages, given an appropriate setting, in order to achieve a specifiable class of objectives.
Bachman & Palmer (1996)	An activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation.
Willis (1996)	A goal-oriented activity in which learners use language to achieve a real outcome.
Skehan (1998)	An activity in which meaning is primary; there is some communication problem to solve; there is some sort of relationship to comparable real-world activities; task completion has some priority; the assessment of tasks is in terms of outcome.
Lee (2000)	A classroom activity or exercise that has an objective obtainable only by the interaction among participants; a mechanism for structuring and sequencing interaction; and a focus on meaning exchange. It is also a language learning endeavor that requires learners to comprehend, manipulate, and/or to produce the target language as they perform some set of workplans.
Bygate, Skehan & Swain (2001)	An activity which requires learners to use language, with emphasis on meaning, to attain an objective.
Ellis (2003)	A workplan which requires learners to give primary attention to meaning and to make use of their own linguistic resources, although the design of the task may predispose them to choose particular forms. A task is intended to result in language use that bears a resemblance, direct or indirect, to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills, and also various cognitive processes.
Van den Branden (2006)	An activity in which a person engages in order to attain an objective, and which necessitates the use of language.
Willis & Willis (2007)	Successful tasks are characterized by the following listing criteria framed as questions: 1. Does the activity engage learners' interest?

	2. Is there a primary focus on meaning? 3. Is there an outcome? 4. Is success judged in terms of outcome? 5. Is completion a priority? 6. Does the activity relate to real world activities?
Samuda & Bygate (2008)	A task is a holistic activity which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning through process or product or both.
Ellis (2009)	A workplan whose primary goal of a task is on "meaning"; there is some kind of gap; learners largely have to rely on their own resources (linguistic and non-linguistic) in order to complete the activity; there is a clearly defined outcome other than the use of language (i.e. the language serves as the means for achieving the outcome, not as an end in its own right).

The rest of definitions from Table 1.4. define "task" as an educational activity. Firstly, these definitions agree that the activities learners do with the target language in the classroom (i.e., pedagogic tasks) should be related to what learners are supposed to be able to do with the target language in the real world (i.e., target tasks). Moreover, classroom tasks should provide many opportunities to process meaningful input and produce meaningful output through interaction. In addition, some definitions point to the fact that negotiation for meaning may demand from the learners that they draw on their linguistic and cognitive resources. Van der Branden (2006) states "task-based language teaching naturally evokes a wide diversity of cognitive operations that people need to perform in order to function in real life" (p. 9). Although there is a primary focus on meaning, according to some authors (e.g., Skehan, 1998), the marriage between meaning and form (i.e., FonF) constitutes one of the key features of TBLT. The meaningful use of language will necessary imply the establishment of relevant form-meaning mappings, the learner will need to manipulate and pay some attention to form. Another key feature of classroom tasks is learner-centeredness, where learners are asked to reach certain goals by making functional use of their own linguistic resources (contrary to structural language teaching methodologies which were purely teacher-centred).

The current thesis adopts the definition by Ellis (2003, 2009) and Skehan (1998) by designing tasks whose objective is to direct learners' attention to meaning while using their own linguistic resources but the tasks predispose them to focus on challenging L2 phonological forms. In addition, the performance of these tasks is thought to result in language that bears some kind of resemblance to the way is used in the real world and prepares learners for task completion outside of the L2 classroom. Most importantly, these tasks involve several cognitive processes that are believed to promote second language (and especially pronunciation) development.

1.4.3.2. Types

From a cognitive perspective, Prabhu (1987) classified tasks based on how the information in the task was handled by learners. He distinguished information-gap (i.e., transferring information from one person, form or place to another), reasoning-gap (i.e., deriving new information from given information through inference or deduction), and opinion-gap tasks (i.e., expressing a personal feeling, preference or attitude).

From a pedagogic perspective, Willis (1996) distinguished six types of tasks based on the cognitive operations they involve, namely, listing, ordering and sequencing, comparing, problem-solving, sharing personal experiences and creative. Other task-based researchers (e.g., Pica et al., 1993) have sought to distinguish task types in terms of the interactive and cognitive processes involved in performing them. This has given rise to a widely accepted set of task types that can be psycholinguistically classified according to:

- Interactant relationship: One way vs. two-way (e.g., Gass & Varonis, 1985). One learner possesses all the information that the other needs to complete the task (e.g., information-gap map-task); vs. each learner has only part of the information (e.g., building a detective story).

- Goal orientation: Divergent vs. convergent (e.g., Duff, 1986). Learners have independent or even opposite goals to accomplish (e.g., a controversial-issue debate) vs. learners share the goal of jointly finding an acceptable solution (e.g., the desert island task).
- Outcome options: Open vs. closed (e.g., Loschky & Bley-Vroman, 1993). A wide range of solutions can be accepted (e.g., a controversial-issue debate); vs. only a solution or limited set of solutions is possible (e.g., a spot-the-difference task).
- Information distribution: Shared vs. split (e.g., Newton & Kennedy, 1996). Information is shared by all learners (e.g., organizing the layout of a zoo); vs. information is equally split among team members (e.g., organizing different people in a table for a meeting).
- Information exchange: Optional vs. required[27] (e.g., Doughty & Pica, 1986). Completion of the task does not require sharing information (e.g., open-ended discussion); vs. information needs to be necessarily exchanged to reach a solution (e.g., organizing a group trip).

Two-way, convergent, closed, split and required task features have been found to generate more episodes of negotiation of meaning, thus increase the quantity and quality of input and output, and contribute to the development of overall proficiency. Therefore, the tasks in the current study were designed on the basis of the aforementioned criteria. Apart from this task typology, tasks can be monologic or dialogic, can involve different rhetorical modes (e.g., instructions, description, narrative, argument (Swales, 1990)) and can be input-based (i.e. involve listening or reading) or output-based (i.e. involve speaking or writing) (Ellis et al., 2019).

[27] The required/optional distinction is sometimes used synonymously with Long's (1985) one-way/two-way distinction.

One question to consider when designing tasks is whether classroom tasks should be true copies of the target tasks or rather increasingly complex approximations to the target tasks (Long, 1985). TBLT researchers have distinguished between “real-world” (i.e., a piece of work done by oneself or others in everyday life such as going out to a restaurant or taking a hotel reservation) or “pedagogic” (i.e., tasks that mainly occur in a classroom context such as ‘spot-the-difference’) tasks. Willis and Willis (2007) defined “real-world” tasks as those which make learners produce meanings and realize discourse acts that mirror real-world conversations. In contrast, “artificial” tasks as those that may not offer a precise reflection of the real world, but they oblige learners to engage in real-world meanings and real-world discourse acts. For Ellis (2003) and Long (2015), replicating real-life scenarios is less important than promoting interactional authenticity, meaning that learners utilize the type of skills that they might use in any real-life interactional situation beyond the task. What is crucial is to prepare learners to deal with a wide range of language encounters they may experience in the real world.

Another distinction that requires special attention, especially because there is controversy surrounding which type is compatible with TBLT is unfocused versus focused tasks (Ellis, 2003, 2009). An unfocused task is “designed to provide learners with opportunities for using language in general communicatively” (Ellis, 2009, p. 223). A focused task is “designed to provide opportunities for communicating using some specific linguistic feature” (Ellis, 2009, p. 223). Advocates of TBLT differ in whether they think tasks should be entirely unfocused or can be both unfocused and focused (Ellis, 2021). For instance, Ellis (2017) argues that focused tasks are necessary because they (1) target some L2 features that may be difficult for learners even at advanced levels, (2) direct learners’ attention towards meaning and use of linguistic features, (3) allow for negative feedback, and (4) are useful in language assessment since the outcomes of the task are clearly specified and learner performance can be judged against them. In contrast, Skehan (2007) believes that focused tasks "meet the definition of a task... [but] enable specific structures to be "forced" into use" (p.295); Willis and Willis (2007) reject the notion of 'focused' (or, as they call them, 'metacommunicative') tasks and Long (2015) advocates the use of reactive FonF rather than pre-emptive FonF. However, Loschky and Bley-Vroman (1993) stated that the use of certain forms could greatly facilitate the completion of a task (i.e., task-essential language). This means it is essential to attend to the relevant L2 feature in order to perform the task successfully; and it is impossible to succeed unless the (grammatical, lexical or phonological) knowledge is attended to. Task essentialness was one of the task conditions that Keck et al. (2006) found to positively influence L2 development.

The tasks of this dissertation involved the following operations: information-gap, reasoning-gap, comparing, problem-solving, and were defined as two-way, convergent, closed, split, information-required, real-world and focused. Tasks were performed dialogically, enhanced arguments and discussions and were output-based. Last, task-essential language was used to draw learners' attention to phonological form while communicating.

1.4.3.3. Selection

Long (2015) argued that task selection is a necessary step in designing a TBLT syllabus or curriculum. Task selection should be based on a needs analysis that establishes course content in terms of the real-world target tasks a specific group of learners need to be able to perform outside the classroom. Gilabert and Malicka (2021) define it as "a thorough empirical investigation of learner needs in occupational, social, academic or professional context" (p. 95). In the study of this dissertation, tasks were designed around topics related to travelling and spending days in a FL country which is what most secondary school students do during their school period. As a matter of fact, the learners of the present experiment were getting ready for their end-of-school trip to London.

1.4.3.4. Frameworks for pedagogical implementation

Long (2015) proposed a six-stage TBLT program development/syllabus design that consisted of needs analysis, task selection and sequencing, pedagogic design, pedagogic implementation, task assessment, program evaluation. In terms of frameworks for task-based units, Nunan (2004) proposed a six-stage framework that brought together both meaning and form through: (1) schema building activities, (2) controlled practice of the target language elements, (3) intensive listening practice where the target language elements could be incorporated, (4) controlled practice in a communicative context, (5) free production tasks, and (6) communicative tasks such as information gap activities.

Inspired by Prabhu’s (1987) three-stage framework (i.e., pre-task, task and marking) and research findings in the field of SLA (e.g., Skehan, 1996; Skehan & Foster, 2007), in 1996, Willis proposed a Framework for Task-Based Learning, which was predominantly meaning-focused with FonF appeared in the post-task stage. The present dissertation adopts Willis’ (1996) pedagogical framework to implement pronunciation-focused tasks.

In Willis’ (1996) framework, learners begin with a holistic experience of language in use, where learners deploy any language they have learned to express themselves, and end with an analytical look at specific features of language form. Teachers adopt the role of leader/organizer of the discourse, manager of pair work, facilitator of tasks, motivator, language adviser and eventually, language teacher. The Task-Based Learning framework consisted of a pre-task phase, a task cycle, and a language focus phase (Willis, 1996, 2021).[28]

Pre-task: The teacher (T) helps students understand the theme and objectives of the task (e.g., brainstorming ideas using pictures, personal experiences, etc.). T highlights useful words and phrases (i.e., priming stage with language focus). Students (S) listen to a recording of a parallel task being done and may do some comprehension activity with it. T clarifies goal of the task, groups S and sets a time limit to perform it. S are given preparation time to plan the performance of the task and can look up words in the dictionary or ask T.

Task cycle: Learners do the task and then they are asked to tell the rest of the class about their resolution, but they are given time to plan their speech before the "public" report. It is divided into three main stages:

1) **Task:** T's role is to encourage S to use their L2 to express themselves in order to complete the task. The emphasis is on promoting spontaneous, exploratory talk and confidence-building in small groups. Therefore, T monitors, making sure all learners are doing the task, and encouraging everyone's attempts at communication in the target language, in a supportive and positive way. T helps them to formulate what they want to say, but will not intervene to correct errors of form unless there is a communication breakdown, and acts as time-keeper.
2) **Planning:** T's overall role is to encourage learners to work out appropriate ways to express the task outcome. S draft and rehearse what they want to say during the

[28] Later on, Ellis (2003), simplified Willis' (1996) framework into three stages -a pre-task stage, a main-task stage and a post-task stage- with various options available in each stage. However, the only stage that was essential was the main-task stage.

report stage. T goes around to advise S on language, and the emphasis is on the clarity, organization and accuracy of the oral discourse.

3) **Report:** T asks several groups to report the findings to the class. S listen and take notes. T acts as time-keeper and comments on the overall content of each report without focusing specifically on problematic language items. T notes them down. As extra practice, T can play S the recording of fluent speakers doing the same task and engage in a comparison on the similarities and differences in their speech.

Language focus: More concentrated focus on pre-selected language forms coming from the pre-task and task cycle. During the consciousness-raising activities, individuals are free to discover target features at their own pace. Its purpose is to help learners systematize what they know, and to expand their knowledge of words and patterns. This phase divides into two stages:

1) **Focus on Form Analysis:** T sets some language-focused tasks or consciousness-raising activities (e.g., underline, classify, choose some of the target words, etc.). T encourages S to explore language for themselves, and to develop an awareness of aspects of grammar, lexis (and pronunciation), to clarify concepts and to notice new things. For instance, S could record themselves performing a task, transcribe a short section and compare their interactions with that of more fluent speakers doing the same task.
2) **Practice:** T conducts practice activities as needed, based on the language analysis work or using examples from the transcript. These include choral repetition, focus on pronunciation, memory challenge games, etc. T may also draw attention to typical learners' errors noticed during the task cycle. S take note of salient features. Optionally, at the end of the framework, S can do a similar task with different partners, take notes of relevant phrases they have

noticed during task performance, or discuss with T how they felt during the task cycle and what they might like to work on next.

To sum up, Willis (1996) framework (1) helps students recall relevant words and phrases and to recognize new ones that will help them get their meaning across, and plan their speech during the pre-task phase. During the task cycle, (2) learners carry out the communication task, using language they have been exposed to in the previous phase. TBLT teachers ask learners to plan and report back how they completed the task. In this phase, learners move from spontaneous language and a focus on fluency to planned language that has a focus on fluency, accuracy and clarity in the organization of the "public" discourse. Only after that, (3) is their attention directed towards useful grammatical, lexical or pronunciation features that occur naturally in the conversations they have had during the task cycle (Willis, 2021). Willis and Willis (2007) opted for treating the task-based syllabus as cyclical so that items were not simply taught once (i.e., traditional approaches) but they were revisited several times. Having defined the concept of "task" and presented the task design features, selection criterion and framework used in the present dissertation, the following section is dedicated to explore how tasks can be further manipulated and sequenced in a task-based syllabus.

1.4.4. Task manipulation and sequencing

Manipulating task features has been empirically shown to enhance learners' attention to target forms during communication; promoting L2 fluency, accuracy and complexity. Task-based manipulations that aim at promoting L2 oral development have been investigated in relation to task type (e.g., Révész & Han, 2006), interlocutor proficiency (e.g., Kim & McDonough, 2008), task mode (face-to-face vs. synchronous computer-mediated communication; e.g., Baralt, 2014; Gurzynski-Weiss & Baralt, 2014) and, especially, task complexity (e.g., Gilabert et al., 2009; Kim, 2012; Nuevo et al., 2011; Révész, 2011; Révész et al., 2011). As a result of the preoccupation with grading and sequencing tasks in a principled way (e.g., from simple to complex) in a task-based syllabus, researchers proposed a set of criteria for evaluating the complexity of a task supported by theoretical frameworks and conducted studies to investigate whether task complexity on L2 production were predicted by those theories.

Skehan's (1996, 1998) limited capacity model, founded on theories of WM (Gathercole & Baddeley, 1993), conceptualizes attention as a single volume which runs out of resources (Kahneman, 1973). It is based on attentional resource allocation to different aspects of task performance, i.e. fluency, which is content-focused, and accuracy and complexity, which are form-focused. Given that human attentional resources are limited, Skehan believes that attention can only be allocated to certain aspects of performance to the detriment of others. That is, when task demand increase, learners first allocate attentional resources to the content of the task, and what remains is assigned to linguistic form (i.e., complexity and accuracy). However, if the content demands are extremely high, complexity and accuracy may compete for attention. According to studies in Skehan's paradigm (e.g., Foster & Skehan, 1996), for cognitively simple tasks (with planning time), there is a marked increase in fluency, relatively strong effect on complexity, and very little effect on accuracy. Skehan (1998) suggested three factors contributing to the difficulty[29] of the task, namely, *code complexity, cognitive complexity* and *communicative stress*, and other learner factors. Nevertheless, Skehan's (1996, 1998) model was unable to explain the phenomenon of dual-task performance and divided

[29] "Task difficulty has to do with the amount of attention the task demands from the participants. Difficult tasks require more attention than easy tasks" (Skehan & Foster, 2001, p. 196).

attention nor was concerned with how tasks should be sequenced to prepare leaners for language use outside the L2 classroom (Robinson, 2011). Sections 1.4.4.1. and 1.4.4.2. deal with a multiple processing model, which is the basis of Robinson's framework, and the underlying model for task complexity in the current dissertation.

1.4.4.1. Robinson's model of task complexity

One of the most notable strands of TBLT research attempting to manipulate learners' attention is the work within the Cognition Hypothesis (Robinson, 2001a, 2005, 2007a, 2011a, 2011b). This hypothesis, together with the Triadic Componential Framework (Robinson, 2001b; 2005; 2007a; 2011b; Robinson & Gilabert, 2007) and SSARC (*Stabilize*, *Simplify*, *Automatize*, *Restructure*, *Complexify*) model (Robinson, 2010), offer a theoretical foundation for maximizing L2 learning through the manipulation of cognitive task demands and subsequent sequencing of tasks according to cognitive complexity.

Integrating information-processing theories (Schmidt, 2001), interactionist explanations of L2 task effects (Long, 1985) and psychological models such as Wickens' model of dual-task performance (Wickens, 1989), Robinson claims that learners can simultaneously access multiple and non-competitional resource pools (see Robinson, 2003b for an illustration). The fundamental pedagogic claim of Robinson's **Cognition Hypothesis** (2001a, 2001b, 2005, 2007a, 2011a, 2011b) is that tasks should be designed and sequenced according to task complexity, defined as "the result of attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (Robinson, 2001a, p. 29) because it mirrors the sequences in which children meet the cognitive demands of task during L1 acquisition[30] (Long, 1985; Robinson, 2001a, 2001b, 2011a, 2011b). The predictions of the Cognition Hypothesis —based on cognitive linguistics (Talmy, 2000), developmental psychology and SLA theories— indicate that increasing the cognitive demands of a task can push learners to greater complexity and accuracy of L2 production in order to meet the greater functional/communicative demands tasks place on the learner (Robinson, 2001a, 2003a, 2007a), as long as learners draw from different pools of attentional resources. Therefore, differences in terms of fluency, accuracy and complexity may be better explained by concepts such as interference rather than by limited resources (Gilabert, 2005). In addition, provided that "task demands are a powerful determinant of what is noticed" (Schmidt, 1990, p.143), increasing task complexity would yield changes both in learner production (moving from fluency to a focus on accuracy and complexity) and would promote interaction and negotiation of form and meaning (i.e., LRE), ultimately leading to incorporation of forms made salient in the input (i.e., uptake in L2 production) and long-term retention effects (Robinson, 2011b). Also, performing simple-to-complex sequences will also lead to automaticity and efficient scheduling of the components of complex L2 task performance. However, ID in cognitive abilities and affective factors may affect task-based performance (e.g., Kormos & Trebits, 2011) and learning when complexity is increased.

Robinson (2001b; 2005; 2007a; 2011b; Robinson & Gilabert, 2007) proposed a three-dimensional model (i.e. the **Triadic Componential Framework**) that elucidated specific task factors to be modified during design so that versions of tasks could be more or less cognitively demanding. Three different types of factors were distinguished:

[30] Robinson's Cognition Hypothesis agrees with Long (1985) that tasks should be developed and sequenced to increasingly approximate the demands of real-world target tasks, rather than on the basis of linguistic grading (as in traditional structural syllabuses, Nunan, 1989).

cognitive (i.e., task complexity: resource-directing vs. resource dispersing variables), *interactive* (i.e., task condition: participation vs. participant variables) and *learner* (i.e., task difficulty: ability vs. affective variables) (Figure 1.6).

Task complexity (Cognitive factors)	Task condition (Interactive factors)	Task difficulty (Learner factors)
(Classification criteria: cognitive demands)	(Classification criteria: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: information-theoretic analyses)	(Classification procedure: behavior descriptive analyses)	(Classification procedure: ability assessment analyses)
Sub categories: a. resource-directing variables making cognitive/conceptual demands	Sub categories: a. participation variables making interactional demands	Sub categories: a. ability variables and task relevant resource differentials
± here and now (Gilabert, 2007) ± few elements (Kuiken et al., 2005) ± spatial reasoning (Becker & Carroll, 1997) ± causal reasoning (Robinson, 2005a) ± intentional reasoning (Ishikawa, 2008) ± perspective-taking (MacWhinney, 1999)	± open solution (Lambert & Engler, 2007) ± one way flow (Pica et al., 1993) ± convergent solution (Duff, 1986) ± few participants (Crookes, 1986) ± few contributions needed (McGrath, 1984) ± negotiation not needed (Gass & Varonis, 1985)	h/l working memory (Mackey et al., 2002) h/l reasoning (Stanovitch, 1999) h/l task-switching (Monsell, 2003) h/l aptitude (Robinson, 2005b) h/l field independence (Skehan, 1998) h/l mind-reading (Langston et al., 2002)
b. resource-dispersing variables making performative/procedural demands	b. participant variables making interactant demands	b. affective variables and task relevant state-trait differentials
± planning time (Skehan, 1998) ± prior knowledge (Urwin, 1999) ± single task (Robinsonet al., 1995) ± task structure (Skehan & Foster, 1999) ± few steps (Fleishman & Quaintance, 1984) ± independency of steps (Romiszowski, 2004)	± same proficiency (Yule & MacDonald, 1990) ± same gender (Pica et al., 1991) ± familiar (Plough & Gass, 1993) ± shared content knowledge (Pica et al., 1993) ± equal status and role (Yule & MacDonald, 1990) ± shared cultural knowledge (Brindley, 1987)	h/l openess (Costa & Macrae, 1985) h/l control of emotion (Mayer et al., 2000) h/l task motivation (Dörnyei, 2002) l/h anxiety (MacIntyre & Gardner, 1994) h/l willingness to communicate (MacIntyre, 2002) h/l self-efficacy (Bandura, 1997)

Figure 1.6. Robinson's (2011b) Triadic Componential Framework.

First, the triadic framework presented two distinct dimensions of *task complexity*: resource-directing and resource-dispersing. On the one hand, (a) resource-directing dimensions are those in which the demands on language use -triggered by task complexity- can be met by manipulating the manner (directing their resources) in which the information is presented. For example, a task which requires to justify your beliefs, provide reasons with solid arguments, infer from given set of facts and provide evidence for interpretations (e.g., Gilabert, 2007b), is more cognitively complex than a simple descriptive or narrative task. When a task becomes cognitively more challenging, more attentional resources are directed towards the linguistic form. Since formal aspects of language receive attention, consciously or unconsciously, complexity and accuracy would also be pushed up, though fluency might be affected negatively. Task complexity with respect to resource-directing dimension can be manipulated on the following factors:

±[31] here and now, ± few elements and ± no reasoning demands. In the current dissertation, we looked at the effects of ± no reasoning demands on perception and production. On the other hand, (b) resource-dispersing (or resource-depleting in Robinson, 2001a) dimensions (e.g., ± planning time, ± prior knowledge) are those where increasing complexity replicates the processing conditions under which real-time language is often used (e.g., unexpected questions during a job interview). For instance, having no planning time, the linguistic output of L2 performers suffers (resulting in the lowering of linguistic complexity, accuracy and fluency) due to the extra cognitive load of a complex task that focuses the attention on other task features than language.

Robinson (2003b) claims that the aforementioned two categories play an important role in resource allocation during L2 task performance and consequently in the quality of L2 production and comprehension. Increasing task complexity along resource-directing dimensions (± elements, ± here-and-now, ± reasoning demands) directs learners' attention to a wide range of functional and linguistic requirements, thus, "has the potential to connect cognitive resources, such as attention and memory, with effort at conceptualization and the L2 means to express it" (Robinson, 2011, p.14), leading to a greater accuracy and complexity in production (Robinson, 2003a, 2005). In contrast, complex tasks along resource-dispersing dimensions (± planning time, ± prior knowledge, ± single task) make greater demands on attention and WM (i.e., depleting attention to features of language code) with negative consequences for production, in agreement with Skehan (1998). These two categories of task complexity are likely to interact and affect L2 production. For instance, designing a task simple along resource-dispersing variables (e.g., allowing planning time) and complex along resource-directing

[31] "±" refers to whether a component is present or absent. According to Robinson (2001a), they may be also thought as continua, along which there is relatively more, vs. relatively less of a component such as planning time, elements, etc.

variables (e.g., requiring reasoning) frees up attentional and memory resources and allows optimum resource allocation to satisfy the linguistic demands, compared to making the task complex on both dimensions at the same time (Robinson, 2001a, 2005) [32]. In contrast, as found in Robinson (2007b), if tasks are complex along resource-dispersing dimensions (- planning time, and +dual task), it may mitigate attempts to respond to the increase in resource-directing reasoning demands using more accurate and complex speech. It is particularly worth highlighting that task complexity manipulations should be validated by subjective ratings and/or expert judgements that can reliably estimate task-generated cognitive load (Révész et al., 2016).

In monologic tasks, Robinson (1995, 2001a, 2001b, 2003a) hypothesized that increasing complexity along resource-directing variables (and initially keeping resource-dispersing variables simple) could have as a consequence lower fluency but higher complexity and accuracy. This prediction is based on a number of acquisitional arguments: the functional communicative demands of the task may (1) cause greater structural complexity; (2) force learners to move beyond "the basic learner variety" and more explicit lexico-grammatical forms will be used (Perdue, 1993); (3) gear learners' attention to overlap o divergence of L1 and L2 form-meaning mappings (Talmy, 2000); and (4) push production, stretch interlanguage, and destabilize fossilized forms (Robinson, 2001a; Gilabert, 2005 for considerations). In contrast, increasing complexity along resource-dispersing variables would lead to reduced fluency, accuracy and complexity. In interactive tasks with increased resource-directing variables (and decreased resource-dispersing), Robinson (2003a, 2007b) predicted less fluent, less

[32] However, progressively increasing resource-dispersing variables (e.g., -planning time and -prior knowledge) would be beneficial for L2 development as it would approximate the performance constraints of real-world task activity (Robinson, 2010). See Robinson (2005) for progression from low performative and low developmental complexity tasks to high performative and high developmental complexity.

structurally complex language but more accurate and more negotiation of meaning (i.e., LRE) (Table 1.5.).

Table 1.5. Predicted effects of task complexity on accuracy, fluency and complexity along resource-directing dimensions.

Task		measure		
		accuracy	complexity	fluency
monologue	simple	–	–	+
	complex	+	+	–
dialogue	simple	–	–	+
	complex	+	–	–

Second, *task conditions*, meaning the interactive demands of tasks, are concerned with (a) the participation dimension, for instance, the information flow in classroom interactions (one-way or two-easy tasks), open/closed, convergent /divergent, etc., and (b) participant variables such as gender of participants in a group/pair, familiarity/unfamiliarity with each other, etc.

Last, *task difficulty* concerns what learners bring to the task (Robinson, 2001b) and refers to individual differences in the resources they draw on to address task demands. Resource pools can be affected in a fixed way by (a) abilities and skills learners bring to the task, i.e. intelligence, WM, aptitude or (b) affective learner factors such as motivation, anxiety or self-confidence. For instance, greater motivation for a learner can result in a temporary expansion of resource pools to meet the demands of a task (e.g., heightened attention), in comparison to a learner with superior aptitude but less motivation, thus, fewer temporarily available resources.

Interactions between complexity, condition and difficulty are bound to occur during task performance. In Robinson's (2001a) view, syllabus designers should use task complexity as the criterion to create and organize pedagogic tasks, since task conditions (participation and participant variables) may vary and task difficulty (affective and ability variables) only informs on-line decisions (e.g., grouping students with differences in anxiety or motivation), and can be determined after diagnosis (e.g., organizing students in terms of cognitive styles/proficiency).

Finally, Robinson's **SSARC** (Simplify, Stabilize, Automatize, Restructure, Complexity) model (2010) specifies the sequence in which task versions (more or less complex) can be completed to maximize L2 learning opportunities (see Baralt et al., 2014; Robinson & Gilabert, 2007; for empirical research on task sequence). The rationale for SSARC model can be described in the following steps: (1) tasks simple on all dimensions are performed (e.g. + planning, − intentional reasoning); (2) complexity on resource-dispersing dimensions is increased (e.g., − planning, − intentional reasoning); (3) complexity on both resource-dispersing and resource-directing dimensions is increased (e.g., − planning, + intentional reasoning). Therefore, the SSARC model postulates that applying its principles theoretically help learners to efficiently schedule their cognitive and communicative resources in order to progressively meet task demands on the way to highly complex real task performance. According to Robinson (2010), in the first step, task performance draws on simple and stable state of the interlanguage; in the second step, complexity promotes faster access to the current interlanguage; in the third step, the interlanguage system is restructured, form-function mappings develop and language complexifies to meet the task demands.

In sum, the Cognition Hypothesis, the Triadic Componential Framework, and the SSARC model specify how tasks may be designed and sequenced to encourage learners' awareness of certain linguistic targets within meaning-based task completion and, ultimately, to promote the acquisition of new L2 knowledge and restructuring of existing interlanguage (i.e., L2 development).

1.4.4.2. Task complexity, language-related episodes and effects on language production

Given that increasing the cognitive demands of the task leads to more attention to, incorporation and rehearsal of input in WM (see Robinson 2003-Study 3 for empirical evidence) and greater attention to and modification of output -restructuring and reanalysis of current linguistic resources- (Swain, 1995), most empirical work (e.g., Michel, 2011) has focused on the effects of task complexity manipulation on learner production using measures of fluency, complexity, and accuracy (see Table 1.6. for a summary of five key studies on the effect of task complexity on oral CAF in monologic tasks, and Table 1.7. for five key studies on the effect of task complexity on oral CAF and interaction moves in interactive tasks). In terms of CAF performance, findings from the systematic review by Jackson and Suethanapornkul (2013) on the effects of task complexity on CAF measures reveal (1) a statistically non-significant difference for *syntactic complexity* between simple and complex conditions in 8 independent studies ($d = -0.02$); (2) a small positive effect size for *accuracy* ($d = 0.28$) in complex tasks, in line with the Cognition Hypothesis, from 9 studies; (3) a negligible but positive effect size for lexical complexity across 7 studies ($d = 0.03$) suggesting that increases in task complexity result in more lexical variety/diversity/density; and (4) a detrimental effect of task complexity on fluency based on 6 studies ($d = -0.16$), as predicted by the Cognition Hypothesis.

As for L2 development, some studies have investigated the potential links between task complexity on L2 accuracy of specific grammatical or lexical targets. In terms of grammar, Révész (2009) found a relationship between task complexity (i.e., here-and-now) and learners' development of English past progressive during interaction with a researcher. Likewise, Baralt (2013) similarly found a link between task complexity (i.e.,

+reasoning) and learners' development of Spanish past subjunctive when learners interacted with a researcher in a face-to-face mode. Concerning learner-learner task-based interaction, Kim and Tracy-Ventura (2011) and Kim (2012) found that more complex tasks (Kim & Tracy-Ventura: +reasoning; Kim: +reasoning and -few elements) led to greater development of the English past tense and English question formation, respectively. In addition, Baralt (2014) found that English-speaking learners of Spanish produced more LRE and demonstrated more development of the past subjunctive in task sequences containing more complex tasks (+reasoning demands) than in those containing simpler tasks (-reasoning demands). Similarly, Révész (2011) and Kim (2009, 2012) found higher LRE production in the more complex task (i.e., -few elements, + reasoning demands) conditions for English conjoined clauses (Révész) and advanced question development (Kim). In terms of lexis, Nuevo et al. (2011) revealed that the development of L2 English locatives could be promoted by task complexity, when operationalized as increased reasoning demands, and together with Gilabert (2007b), found that increased task complexity triggered self-repair behaviour in L2 learners. Following this line of research, Kim et al. (2018) explored the effect of task complexity and repetition on L2 lexicon use (i.e., word familiarity, word age of acquisition, and word frequency). Regarding task complexity, their results revealed that complex tasks elicited fewer familiar words and less frequent words than simple tasks, especially in procedural repetition tasks. Róg (2021) also showed a significant positive impact of task complexity on L2 lexical complexity only when the complex task was repeated.

Gilabert and Barón (2013) extended this line of research to L2 pragmatics by analysing suggestions and requests produced by EFL learners through problem-solving tasks, and their results showed that learners produced more suggestion and request strategies in complex tasks (i.e. -few elements) but not with greater variety. Similarly, Kim and Taguchi’s (2015) study suggests that a complex task (e.g. +reasoning demands while creating a drama script) promoted negotiation of pragmatic elements (i.e., pragmatic-related episodes), namely, L2 English learners’ development of request expressions. More recently, Márquez and Barón (2021) suggested that increased task complexity (i.e. -few elements, +social distance, + imposition) may promote accuracy and complexity of pragmatic moves.

Table 1.6. Summary of five key studies on the effects of task complexity on CAF in monologic tasks.

Study	**Participants**	**Task type**	**Task complexity factor**	**Findings (complex vs. simple task)**
Robinson (1995)	5 Japanese, 5 Korean, 1 Indonesian and 1 Filipino speakers in Hawaii, US	Narrative	+There-and-Then	+Accuracy* +Lexical complexity and density* +Syntactic complexity -Fluency*
Gilabert (2007a)	48 Spanish speakers in Spain	Narrative	+There-and-Then -Planning time	+Accuracy* +Lexical complexity and density* -Fluency*
Ishikawa (2008)	24 Japanese speakers in Japan	Problem-solving	+Intentional reasoning	+Accuracy* +Lexical complexity and density* -Fluency*
Kuiken & Vedder (2011)	44 Dutch speakers in the Netherlands (oral task group only)	Decision-making	+Reasoning demands	+Accuracy* -Syntactic complexity
Levkina & Gilabert (2012)	21 Russian and 21 Spanish speakers in Spain	Decision-making	-Few elements -Planning time	-Accuracy +Lexical complexity (only* in unplanned conditions) -Syntactic complexity -Fluency*

Note: (*) indicates significance

Table 1.7. Summary of five key studies on the effects of task complexity on CAF and interaction moves in dialogic tasks.

Study	**Participants**	**Task type**	**Task complexity factor**	**Findings (complex vs. simple task)**
Robinson (2001a)	44 Japanese speakers in Japan	Map task	-Few elements -Prior knowledge -Planning time	+Accuracy +Lexical complexity and density* +Syntactic complexity -Fluency* +Confirmation checks* +Clarification requests
Michel, Kuiken & Vedder (2007)	29 Moroccan and 15 Turkish speakers in The Netherlands	Decision-making	-Few elements	-Accuracy +Lexical complexity +Syntactic complexity -Fluency
Robinson (2007b)	42 Japanese speakers in Japan	Narrative	+Intentional reasoning -Planning time +Dual performance	+Accuracy -Lexical complexity and density* +Syntactic complexity -Fluency +Confirmation checks* +Clarification requests*
Gilabert, Barón & Llanes (2009)	60 Spanish speakers in Spain	Narrative, instruction-giving and decision-making	+There-and-Then +Reasoning demands	+Confirmation checks* +Clarification requests* (except for decision-making task)
Révész (2011)	43 Spanish, Korean and Japanese speakers in the US (length of residence 11 months)	Decision-making	-Few Elements +Reasoning demands	+Accuracy* +Lexical diversity* -Syntactic complexity* +LRE*

Note: (*) indicates significance

Finally, several studies (Robinson, 2001a; 2007b) have shown that increase in task complexity manipulated along resource-dispersing and resource-directing variables have been accompanied by significantly higher learner ratings of task difficulty, and stress, poorer performance, and non-significant differences in task interest, or task motivation (but see Révész, 2011). Testing the relationship between self-perception of task difficulty and performance is outside the scope of this study but would be interesting to explore in future analyses. Also, investigations on task sequencing, which is at the base of models for task (and syllabus) design, are still very scarce. The lack of robust findings regarding the impact of task complexity on L2 performance and development has stalled the sequencing research agenda.

So far, SLA studies based on the Cognition Hypothesis and SSARC models have made predictions regarding the effect of modifying task complexity on learner production in terms of CAF performance and L2 grammatical, lexical, and (more recently) pragmatics. Despite advances in understanding how manipulating task complexity may impact L2 interlanguage development (e.g., as observed through LRE), the current dissertation seeks to explore if and how the predictions of the Cognition Hypothesis apply for learners' development of L2 phonetics and phonology, and takes into account potential challenges in the implementation of TBLT in a FL context.

1.4.5. Implementation challenges and potential solutions

This section deals with challenges that have been identified regarding the implementation of TBLT in school contexts and evidence-supported solutions.

Whereas research on the benefits of TBLT for L2 oral and written performance and development has been abundant, studies on teacher beliefs and practices of TBLT approaches in real classrooms are scarcer, and the role of the teacher has usually received little attention. Studies such as East (2012), Ellis (2018; 2021), Erlam and Tolosa (2022), Willis and Willis (2007), among others, have identified real challenges in the implementation of TBLT in the classroom. According to Ellis (2018), these problems can be grouped according to whether they concern the teacher, the students, or structural issues within the education system. Concerning teachers, they usually have little time to prepare tasks and sometimes have a clear misunderstanding of what constitutes a task; they struggle to abandon the traditional role of knowledge-transmitter and adopt the role of “co-learner”; and some fear losing control in large TBLT classes. As far as students are concerned, problems are related to the limited communicative abilities of low-level students, lack of confidence in more “incidental” language learning —being accustomed to grammar-oriented experiences—, use or minimal language or overuse of L1 when planning or doing tasks and lack of perceived progress, amongst others. Finally, structural problems in educational systems concern a lack of task-based teaching resources and limited time for teachers to develop their own task-based resources (but see recent Anderson & McCutcheon, 2019 textbook); teaching is preferred when based on items listed in a syllabus rather than on tasks; and assessing learning by means of system-referenced tests that prepare for exams which are not task-based rather than testing implicit knowledge.

This lack of TBLT implementation in the classroom can only be addressed through carefully-designed initial and in-service teacher training/education programmes that take account of the latest innovations (Van den Branden, 2006; Erlam & Tolosa, 2022). Training courses with task-based *input sessions* would help teachers experience TBLT for themselves, while the TBLT framework and rationale were made explicit (Ellis, 2018). Other potential solutions that may help to ameliorate the problems that arise in the implementation of TBLT may be to (1) provide teachers with task examples as concretization of TBLT (e.g. The Task Bank: https://tblt.indiana.edu/); (2) help teachers identify, adapt and ‘taskify’ activities from their regular coursebooks; (3) facilitate video-based examples of TBLT; (4) engage teachers in Action Research so they can familiarize with the learners’ needs and gain awareness about the potential benefits of tasks; (5) promote teacher exchange to be able to connect with other teachers and researchers implementing TBLT approaches; and (6) to reward teachers with financial support for trying innovative ideas in the classroom to improve teaching effectiveness. However, collaboration between researchers, teachers and teacher trainers is maximally needed to understand the realities of TBLT in actual classrooms to be able to cope better with the aforementioned challenges (Erlam & Tolosa, 2022).

Section 1.4. has focused on TBLT as a learner-based analytic approach that enhances conscious attention to form incidentally while focusing on meaning and, supported by interactionist research, we have presented the benefits of negotiation of form and meaning for input noticing and producing modified input. In addition, we have reviewed the definitions of "task", task types, task selection and the framework in order to justify the criterion we followed for the design of tasks in the present dissertation (Section 3.3.2). In order to increase the noticing of the linguistic features, the current study manipulated task complexity following Robinson's Cognition Hypothesis and Triadic Componential Framework, which have been shown to promote L2 lexico-grammatical and pragmatic accuracy. Finally, the aforementioned challenges for TBLT implementation were considered: in this study, we familiarized teachers with the task procedure and provided them with real examples, and created tasks that were suitable for different L2 proficiency learners and emphasized the importance of L2 oral abilities and pronunciation. Provided that TBLT enhances the learning of L2 forms during meaningful interaction, the next section aims at reviewing investigations on the learning of L2 pronunciation under a task-based approach.

1.5. Task-based pronunciation teaching (TBPT)

This section's aim is to review previous work on L2 speaking and pronunciation within the field of TBLT. Specifically, Section 1.5. aims to investigate whether L2 pronunciation is adequately represented in the area of L2 speaking and whether pronunciation-focused language related episodes are present in interaction-based research. Additionally, this section seeks to explore the origins of TBPT and identify task features that may influence L2 pronunciation accuracy. Furthermore, this section aims to investigate how task complexity affects L2 pronunciation performance and development, and finally, review current research agenda in TBPT.

1.5.1. Oral production and pronunciation

In this section, the role of pronunciation is explored in relation to speaking research. Since the 80s-90s, tension has existed between pronunciation and speaking. Whereas it seemed that the emphasis on communicative speaking reenergized during CLT era, some practitioners overlooked the importance of systematic pronunciation teaching. Despite the resurgence of teaching pronunciation in recent decades, the neglect of pronunciation in the teaching of speaking still continues to happen (Newton, 2018). For instance, Ellis (2003) book, which provided an extensive review of TBLT, did not include any references to pronunciation or phonetics/phonology. Therefore, it seems that despite there being a large body of TBLT research dedicated to speaking, pronunciation has received very little attention and is still under-represented in the TBLT literature.

Given that oral production has been mainly assessed in terms of CAF, one wonders whether pronunciation has been represented in measures of spoken accuracy and fluency. In general, accuracy has been mostly defined in relation to morphosyntactic and lexical features, but pronunciation and prosody have not been included in measures of fluency beyond measures of speech rate and pausing. It could well be that this research is highly cognitivist in orientation and so focuses on how task factors influence the conceptualization and formulation stages of speech production rather than the articulatory stage (Newton, 2018).

De Ridder et al. (2007) did include an independent measure of pronunciation (i.e., distinctness, intelligibility, naturalness) and intonation, (i.e., clearness, intelligibility, naturalness, melodic phrasing) apart from other CAF measures but were only assessed subjectively by two raters. Finally, Derwing and Rossiter (2003) investigated the CAF framework and Skehan's (1998) trade-off hypothesis to inform research on pronunciation instruction. Learners on the segmental instruction condition reduced the number of phonological errors but got significantly worse in comprehensibility and fluency. In contrast, learners in the suprasegmental condition improved significantly on both. Whereas an exclusive focus on vowel and consonant production prevented attentional resources to be directed to fluency and complexity, the close relationship between the communicative meaning of speech and prosodic features brought benefits to both complexity and fluency in mutually supportive ways. The only other representation of pronunciation in L2 speaking research has been through analyses of interactive features targeting pronunciation.

1.5.2. Interaction and pronunciation-focused language-related episodes (P-LRE)

The minor presence that pronunciation has had in interaction studies has been mainly through references to pronunciation-focused language-related episodes (P-LRE) arising from negotiations for meaning during spontaneous conversations.

In comparison to grammar and vocabulary, pronunciation seems to play a limited role in classroom interaction. For example, Ellis et al. (2001) reported that out of 429 FonF episodes that they identified in some 12 hours of TBLT, 163 were related to grammar, 159 addressed lexical problems and 76 pronunciation problems. In the same vein, Gurzynski-Weiss and Baralt (2014) observed a focus on pronunciation in only 3% of LRE in face-to-face task-based interaction, and Bowles et al. (2014) only 1% (L2-L2 pairs) and 4% (L2-heritage speaker pairs), probably because learners have never been taught about L2 segmental and suprasegmental features of speech. Notably, despite P-LRE being infrequent, they are more likely to result in successful uptake and higher rates of self-reported learning than grammatical or lexically-based LRE (Bitchener, 2004).

Indeed, the comprehension skills of the listener and the pronunciation skills of the speaker are both critical in the achievement of communicative success (Moyer, 2014) and, hence, the construct of intelligibility (Levis, 2005).

In addition, Horgues and Scheuer (2014) and Loewen and Isbell (2017), and Matsumoto (2011) found that over 90% of intelligibility breakdowns were related to segmental features (vs. 8% on suprasegmentals, in Loewen & Isbell, 2017). Kennedy et al. (2015) also found that most intelligibility issues were related to segmentals (N = 21), with suprasegmentals (N = 3) rarely leading to breakdowns in understanding. Given the considerable importance of pronunciation for intelligibility and comprehensibility, task design and manipulation should take into account challenging L2 phonological forms that are crucial during negotiation for meaning

1.5.3. TBLT and pronunciation instruction

This section deals with the first approaches to TBPT and how it has been assessed; as well as summarizes main findings from studies dealing with task manipulation (especially task complexity) and suggests future studies that could be conducted in the area of task-based instruction and L2 pronunciation.

1.5.3.1. Origins and assessment of TBPT

Techniques for encouraging noticing, attention, and conscious analysis and understanding of pronunciation forms and patterns are especially relevant in L2 speech learning (Pennington & Rogerson-Revell, 2019). Therefore, task-based methodologies that promote attention and noticing to differences between L1 and L2 phonetic and phonological systems may create opportunities for learners to acquire new sounds, adjust vowel space, acquire intonational patterns, etc. (e.g. Mora & Levkina, 2017).[33] Attention to phonetic form and L2 pronunciation development in the context of TBLT can be achieved through reactive (e.g., providing negative feedback) or proactive form-focused instructional techniques (e.g., repetition or complexity) that have proved effective in the development of grammar and lexis (Baralt et al., 2014; Lyster, et al., 2013). Before the TBLT 2015 Leuven colloquium (*TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis?*, Gurzynski-Weiss et al., 2017b) on TBLT and pronunciation, only Sicola (2008) had investigated a focus on phonological forms during interactive tasks within the TBLT literature. Sicola's intermediate-level ESL students were exposed to an interactive map task that necessitated accurate production of the interdental fricative /θ/ for successful task completion. Sicola examined whether learners used corrective feedback to draw their attention to phonological form; whether productions of /θ/ varied after feedback; and whether these modifications yielded more target-like realizations. Results showed that learners used pausing or rising intonation, to make their production of /θ/ more salient to their partners, provided a variety of corrective feedback and, in over 65% of the cases, learners modified their production of /θ/ in a target-like direction.

After this, the findings of the studies included in a special issue in the *Studies in Second Language Acquisition* journal (Gurzynski-Weiss et al., 2017a) were the first to exemplify research on the role of tasks in raising awareness of pronunciation features during communicative task performance and assess improvement in L2 pronunciation accuracy. Authors explored the crucial role of task features in the enhancement of

[33] Ellis (2017) argues that non-salient and complex features, especially when these are 'blocked' by the learner's L1, are unlikely to be acquired even with the help of focus on form. Therefore, he suggests a hybrid syllabus consisting primarily of a task-based component but supported by a task-supported component to address challenging phonological issues.

attention to phonetic form such as task repetition, L1 pairing in collaborative interactive tasks, task complexity, and task modality (Gurzynski-Weiss, et al., 2017b; Mora & Levkina, 2017). See Sections 1.5.3.2. and 1.5.3.3. below for their research summaries).

In TBPT studies, phonology-specific definitions of pronunciation accuracy and fluency will necessarily differ from the way these dimensions of oral production have been operationalized for grammar and lexical development. Phonetic measures of accuracy may include assessing phonetic distances between L2 contrasting sounds, or speaking fluency measures based on connected speech processes of assimilation, epenthesis or weakening (Mora & Levkina, 2017). Measuring the perception and production of specific L2 phonological contrasts entails controlled-production tasks which have been lab-validated in L2 speech research. Segmental perception has been usually assessed through identification and discrimination tasks (Carlet & Cebrian, 2019) presenting high-variability stimuli. Segmental production has been measured through reading and auditory repetition tasks and assessed through various spectral or duration distance measures such as Euclidean distance scores (e.g., Flege, et al., 1997), Pillai scores (e.g., Hall-Lew, 2010) or Mahalanobis distances (e.g., Kartushina et al., 2015; Melnik-Leroy et al., 2022). In terms of global production measures (Saito & Plonsky, 2019), L2 speech has also been evaluated in terms of ratings of comprehensibility, intelligibility, accentedness and fluency (e.g., Gordon, 2021).

The great challenge of pronunciation-focused task-based research is to adapt the design features (e.g., the use of high-variability stimuli) and assessment methods of phonetic training (e.g., pronunciation proficiency measures) that are effective to evaluate gains in L2 phonetic training, for their use in interactive communicative tasks and assessment of L2 phonetic form. Either through open-ended informal conversations or more controlled production tasks that elicit particular sounds in specific phonetic contexts, empirical studies measuring L2 pronunciation learning need to consider high-quality audio-recorders and quiet (better isolated) places that allow for subsequent accurate acoustic analyses and interpretation. Quantifying phonetic and phonological development in a longitudinal study happening in a real classroom context is undoubtedly ambitious but this dissertation proves that it is achievable and provides conclusive evidence of L2 speech development.

1.5.3.2. TBLT techniques for L2 pronunciation

Several factors may have an impact on learners' attention to phonological form during interaction. In this section, we will explore widely accepted and extensively researched TBLT techniques concerning participant variables (i.e., the role of the L1), context (i.e., task modality) and task manipulation effects (i.e., task repetition, task complexity) that Gurzynski-Weiss et al. (2017a) took into consideration in the design of focus-on-phonological form activities to draw learners' attention to phonetic form and to assess the potential for developing L2 pronunciation.

Concerning participant variables, Loewen and Isbell's (2017) study investigated the role of the L1 as a source of interference in developing L2 pronunciation accuracy. In pronunciation-unfocused studies, interacting with L1-different interlocutors has been found to result in more L2 production and to improve learners' oral L2 communication skills (Bueno-Alastuey, 2011), but this was yet to be explored in pronunciation-focused studies. Although the average of overall P-LRE was twice as high among the different-L1 pairs (*M*=16, *SD*=1.6) than same L1-pairs (*M*=8.5, *SD*=3.2), L1-pair differences in percentages (19.4% vs. 12.3%) did not reach significance (see Kennedy et al. 2015 for similar results). Despite the small sample size (15 dyads), the current results lent some support to the notion that familiarity with one's own L2 accent may decrease the need for negotiation for meaning that is triggered by pronunciation issues. As evidenced in Bueno-Alastuey's (2013) study, 49% of P-LRE took place in learner-native speaker dyads, followed by 40% for same-L1 (Spanish) learner-learner dyads, and 35% for different-L1 (Spanish-Turkish) learner-learner dyads. Thus, overall it seems familiarity with an L2 accent facilitates comprehension.

Two studies in Gurzynski-Weiss et al. (2017a) investigated the effects of task modality on the number of P-LRE, recasts and pronunciation accuracy. In Loewen and Isbell's (2017) study, both face-to-face and synchronized computer-mediated communication conditions generated a similar number of P-LRE, even though it was hypothesized that, in the lack of visual information, in the audio-only CMC would increase the frequency of P-LRE. Parlak and Ziegler (2017) found that only face-to-face participants noticed and were more likely to take advantage of the feedback provided than those in the synchronized computer-mediated communication condition, and overall acoustic analyses on lexical stress on 3- and 4-syllable words (i.e. pitch, intensity, and duration) showed a beneficial (albeit non-significant) effect of recasts on learners' ability to notice the correct position of stress on target words.

Given the important role of auditory, lexical and structural priming in the alignment of language forms learners produce during interaction (Trofimovich et al., 2014), Jung et al. (2017) investigated the role of task repetition in the development of L2 stress patterns through collaborative priming tasks. Their findings demonstrated that hearing and producing the same words containing the target-stress patterns repeatedly while participating in collaborative meaning-focused tasks enhanced a better control of L2 lexical stress in English and even led to delayed-learning effects. In this dissertation, tasks were designed in a way that ample tokens of lexical items could be produced throughout the 20 tasks, which would in turn aid their development of perception and production skills (Trofimovich et al., 2014). As the second research question of this dissertation deals with the effects of task complexity on segmental production, the next section is fully dedicated to this TBLT manipulation technique.

1.5.3.3. Task complexity and pronunciation

As seen in Section 1.4.4. task complexity has been found to play a very important role in L2 oral performance (CAF) and development in monologic and dialogic tasks. Whereas the Cognition Hypothesis' predictions (Robinson, 2001a, 2001b, 2005, 2007, 2011a, 2011b) have been mainly applied to learners' lexico-grammatical performance, it is under-researched how these predictions may impact accuracy of phonetic form and show through P-LRE.

On the one hand, when tasks have been designed with no specific pronunciation focus (i.e., unfocused tasks), increasing the cognitive demands of a task has been found detrimental for L2 pronunciation. Kuiken and Vedder's (2011) study investigated the effects of task complexity on linguistic performance in relation to mode. Whereas task complexity had a significantly positive effect on lexical and grammatical accuracy, learners on the most complex task (especially low-proficiency learners) made more pronunciation errors than in the simple task (consistently, albeit non-significantly), contrary to the predictions of the Cognition Hypothesis (Robinson, 2011b). This potential trade-off between lexico-grammatical and pronunciation accuracy, as a result of increased task demands, has also been shown with a negative, but non-significant correlation between a ratio of lexico-grammatical and pronunciation errors in Mora et al. (forthcoming). In Derwing et al.'s (1998) words, it is likely that "when the nature of the linguistic task necessitate[s] that attention [is] divided amongst lexical access, syntactic well-formedness, phonological accuracy, discourse organization, and so forth, speakers

[may] not [be able to] allocate enough resources to phonological concerns for there to be a noticeable transfer of segment-based skills” (p.406). Last, Crowther et al. (2018) compared global pronunciation proficiency measures of comprehensibility and accentedness and subjective specific measures for pronunciation and fluency (segmental errors, word stress errors, rhythm and speech rate) and lexis, grammar and discourse (lexical appropriateness and richness, grammatical accuracy and complexity, and discourse richness) on three tasks increasing in complexity (i.e., picture-description<IELTS<TOEFL). Task complexity was operationalized in terms of resource-directing variables, from +few elements, +spatial reasoning, +here/now, -causal reasoning, -intentional reasoning, -perspective taking, to -few elements, -spatial reasoning, -here/now, +causal reasoning, +intentional reasoning, +perspective taking. Results showed that learners’ speech was rated as significantly more strongly accented in the simple (i.e., picture-description) than the complex task (i.e., TOEFL), but no significant differences were found for comprehensibility between the simplest and most complex.

On the other hand, when tasks are designed to promote a focus-on-phonetic form, task complexity has been generally found to positively impact L2 pronunciation accuracy and lead to gains in L2 phonological development (see Table 1.8. for a summary of main findings). For instance, Solon et al. (2017) used simple and complex map tasks to elicit the pronunciation of word forms (street names) that contained target L2 vowels in an interactive map task where intermediate-level learners gave and received directions. Contrary to the Cognition Hypothesis (Robinson, 2011b), P-LRE were produced at a higher rate during the simple task (i.e., 0.70 LRE/min) than during the complex task (i.e., 0.46 LRE/min) but these differences were not significant and the effect size was relatively small. Nevertheless, vowel production accuracy generally became slightly more target-like in cognitively complex than cognitively simple tasks. Another study by Mora-Plaza et al. (2018) investigated whether increasing task complexity along ± elements and ± reasoning demands could have an effect on the occurrence of P-LRE and whether the frequency of P-LRE was related to intermediate-level learners' gains in the production of the /æ/-/ʌ/ contrast. Increased task complexity in decision-making tasks resulted in higher the occurrence of P-LRE irrespective of time-on-task, supporting the Cognition Hypothesis (Robinson, 2011b). Although Mora and Levkina (2018) did not manipulate task complexity experimentally as an independent variable, their advanced EFL learners showed significantly better perception of an L2 vowel contrast embedded in nonwords at post-test, based on a simple to complex map-task intervention. Interestingly, their findings revealed that exposing learners to simple stimuli in terms of phonological structure (1 syllable vs 3 syllables) in pronunciation tasks may be more beneficial in learning phonological contrasts than exposing them to more complex stimuli.

In an online learning-mode paradigm, Hanson (2022) manipulated task complexity to investigate whether it would promote L2 Spanish learners' accurate pronunciation of the Spanish voiceless stops /p,t,k/ and rhotic /r/. By using an information-gap task, similar to Solon et al. (2017), she found simple tasks to be more effective at directing beginner-level learners' attention to pronunciation in online learning settings extending previous findings (Baralt, 2013). Finally, only Gordon (2021) studied how increases in task complexity following the Cognition Hypothesis (Robinson, 2011b) affect suprasegmental features such as L2 lexical stress, sentence stress, rhythm, reductions, linking and intonation produced by beginner-level EFL learners. Each session followed a sequence in which the students were exposed to explicit phonetic and phonological information, engaged in controlled and guided practice exercises, and completed a communicative task following Celce-Murcia et al.'s (2010) framework. The last stage of this approach concerned communicative tasks that differed in terms of their complexity (i.e., ± elements). Only those learners who completed more complex tasks as part of their treatment, ended up being less accented and significantly more comprehensible during the post-test than in the pre-test. Gordon (2021, p.11) states that "the use of complex tasks created an environment for the learners in the Complex-Decision-Making Class to interact, negotiate meaning, and produce pronunciation focussed language-related episodes, which allowed them to automatize the phonetic and phonological forms learned in more controlled conditions during the treatment".

Overall, these studies seem to indicate that increasing task complexity along resource-directing dimensions may have the potential to draw learners' attention to phonological form during interaction and lead to more accurate realizations of L2 pronunciation targets. Nevertheless, the disparity in the results may be caused by several factors, namely, task mode (online in Hanson vs. face-to-face in Mora-Plaza et al., 2018), sample size (20 in Solon et al. vs. 36 in Mora-Plaza et al.), learners' proficiency level (beginner-level in Gordon & Hanson vs. advanced-level in Mora & Levkina), statistical models (t-tests in Solon et al. vs. linear mixed-effects models in Gordon), and generalization effects (only in Mora & Levkina). One common limitation is the lack of control of ID (e.g., attention control, anxiety) and divergence in proficiency in pairs, which may be especially relevant for P-LRE in low-proficiency learners.

The studies reviewed in Sections 1.5.3.2. and 1.5.3.3. suggest that focusing on difficulty L2 phonological features (e.g., vowel production accuracy, word stress, sentence prosody) through TBLT techniques (e.g., task complexity, task repetition) has the potential to increase L2 pronunciation accuracy and lead to more P-LRE occurrences. Nevertheless, participation and participant variables need to be considered to understand how TBPT can maximally promote gains in L2 pronunciation for everyone.

Table 1.8. Summary of five key studies on the effects of task complexity on pronunciation accuracy and P-LRE in focused dialogic tasks.

Study	**Participants**	**Task type**	**Target**	**Task complexity factor**	**Findings (complex vs. simple task)**
Solon, Long & Gurzynski-Weiss (2017)	20 English speakers in the US	Map task	Spanish /a, e, i, o, u/	-Few elements	+Pronunciation accuracy (F1: /i/, /u/; F2: /e/*; F1 & F2: /o/) -P-LRE
Mora & Levkina (2018)	81 Spanish speakers in Spain Experimental: 66 speakers Control: 15 speakers	Map task	English /iː/-/ɪ/	-Few elements	+Accurate perceptual discrimination of /iː/-/ɪ/ at post-test with simple>complex sequence* +Generalization to new contexts and speakers with simple>complex sequence *
Mora-Plaza, Mora & Gilabert (2018)	36 Spanish speakers in Spain Experimental: 18 speakers Control: 18 speakers	Decision-making task	English /æ/-/ʌ/	-Few elements +Reasoning demands	+P-LRE* +Euclidian distance between /æ/ and /ʌ/ +P-LRE + Gains in Euclidian distance*
Gordon (2021)	67 Spanish speakers in Costa Rica	Decision-making task	English prosodic features	-Few elements	+Comprehensible speech* -Accented speech
Hanson (2022)	30 English speakers in the US	Map task	Spanish /p/, /t/, /k/ and rhotics /r/	-Few elements	+Pronunciation accuracy in synchronous computer-mediated communication mode

Note: (*) indicates significance

1.5.3.4. Task conditions and learner variables

Other than task manipulations, the research agenda of TBPT cannot turn a blind eye to participation variables and learner factors that are likely to mediate the effectiveness of TBLT manipulations in driving L2 pronunciation learning.

First of all, it would be worth investigating the effect of participation variables (i.e., interactional demands of the task) on learners' negotiation of phonetic form and meaning and task outcomes. For instance, closed (i.e., one correct outcome) vs. open (i.e., multiple correct outcomes); convergent (i.e., learners need to agree to successfully complete the task) vs. divergent (i.e., learners are allowed to diverge and discuss their opinion). Another important task condition that may impact L2 pronunciation is the linguistic focus (Ellis, 2003). Tasks may be either unfocused, with no specific linguistic targets, or focused, with learners expected to use certain linguistic structures to complete the task. Loschky and Bley-Vroman's (1993) distinguished between task naturalness (i.e., features may arise, but are not necessary for the task), task utility (i.e., with the feature the task becomes easier) and task essentialness (i.e., the feature is necessary for the completion of the task) to illustrate the decision-making process in the proactive stance. To our knowledge, no TBPT studies have experimentally tested the effect of task essentialness (vs. task naturalness) in L2 pronunciation accuracy or P-LRE, but several studies have used focused tasks to investigate specific phonological features (Mora-Plaza et al., 2018; Saito, 2015; Sicola, 2008; Solon et al., 2017; the current dissertation study). An important methodological issue in TBPT is how make pronunciation elements task-essential during communicative interaction for the noticing of L1-L2 phonological differences while providing learners with fair amount of input and output practice (Mora

& Levkina, 2017). A potential solution is to build collaborative tasks flooded with target pronunciation patterns may provide learners with large practice opportunities (Loewen & Isbell, 2017; Trofimovich et al., 2014; the current dissertation study).

Secondly, learner factors involving their perceptual and productive phonological skills are likely to interact with task design features in explaining pronunciation learning outcomes. For example, if the task demands are increased along resource-directing variables (e.g., few elements) but learners lack proficiency and motivation, the initial predictions made by Robinson (2001a, 2001b) on L2 oral production will most likely not correspond, and these ID may mitigate the potential benefits of increasing task complexity on learners' linguistic outcomes. Therefore, ability variables (i.e., intelligence, aptitude, cognitive style) and affective variables (i.e., motivation, anxiety) are extremely important to on-line methodological decisions (Robinson, 2001a, 2001b, 2011b).

Section 1.5. has shed light on the low representativeness of pronunciation in the TBLT literature, despite being intrinsically related to intelligibility in L2 oral production. TBPT arises as a pedagogical approach to orient learners' attention to L2 pronunciation during communication through the use of tasks. Having reviewed previous research findings on the design and manipulation of tasks for L2 pronunciation, the study in this dissertation adopts TBLT techniques for the design of tasks and the Cognition Hypothesis for the manipulation of task complexity, and employs L2 speech measures to assess L2 phonological development. In addition, the present dissertation marks one of the first attempts to investigate whether ID play a role in mediating the impact of TBPT on L2 pronunciation performance.

1.6. Individual differences (ID) in L2 speech research

The last section of Chapter 1 explores the relative weights of learner factors in L2 speech acquisition, which may explain inter-individual variation in L2 pronunciation learning in instructional contexts. We first assess various factors that may explain variability in L2 speech acquisition, namely, experiential, psycho-social, cognitive and auditory processing. Additionally, we review previous literature on the efficacy of the elicited imitation task (EIT) as a measure of learners' L2 proficiency. Finally, we explore the relationship between WM and L2 speech perception and production and examine the role of attention control in L2 phonological performance and development.

1.6.1. Experiential, psycho-social, cognitive and auditory processing factors

Given that learner factors may interact with cognitive task complexity making it more or less difficult, it is important to consider several experiential, affective and ability variables that may mediate L2 oral, and specifically, pronunciation performance.

Decades of research in SLA has focused on uncovering the sources of ID that might explain the large inter-learner variability characteristic of L2 speech, under controlled input conditions in FL contexts (Golestani & Zatorre, 2009) and immersion contexts where exceptional learners are found (Moyer, 2014). Understanding the sources of this variability is important in order to inform instruction and assessment in L2 speaking and pronunciation, to gain a better understanding of the processes involved in L2 speech production and acquisition, and to advance current models of L2 speech learning (Mora, 2022). Moyer (2014) identified three sets of factors that could predict ultimate attainment in L2 speech learning, namely, experiential (age of onset of L2 learning, amount and quality of L2 exposure and use, learning contexts, L2 proficiency), psycho-social (motivation, personality, anxiety, beliefs, learning strategies, willingness to communicate), cognitive (WM, attention, inhibition) and general auditory processing (auditory acuity, frequency discrimination, spectral and temporal auditory motor integration, imitation ability) factors. Such factors have been investigated from a variety of theoretical approaches, leading to variability and overlap in the operationalization of the underlying predictive constructs for language learning (for reviews see Andringa & Dąbrowska, 2019).

Concerning age- and experience-related factors (i.e., experiential), factors include the extent to which the L1 and the L2 differ phonetically (Flege et al., 1995) as we know that in order for learners to establish new L2 sound categories, they need to successfully distinguish the L2 sounds from L1 sounds. In L2 immersion settings, the age at which a learner is first exposed to an L2 may determine the rate of success in learning: learners exposed to an L2 at an early age are typically better at discriminating and producing L2 segments than those who are exposed later in life, and larger amounts of L2 use and exposure to quality L2 input impact L2 speech development positively (Flege, 2016 for overviews; Muñoz, 2014 for classroom settings). Moyer's (2014) research on L2 experience concluded that what learners do outside the classroom (i.e., the L2 interlocutor characteristics, the frequency and length of L2 interactions, L2 media exposure, etc.), has an impact on how native-like their L2 speech is perceived to be. In FL contexts, experiential factors weight relatively less to L2 speech acquisition (MacKay & Fullana, 2009) due to the limited exposure and L2 use, and the lack of opportunities for meaningful language use beyond the few hours of instruction per week learners typically engage in (Muñoz, 2014). However, Suzukida and Saito (2023) found that recent L2 use outside the regular English classes was strongly associated with comprehensibility and accentedness (to a lesser extent). Therefore, outside the EFL classroom setting, English use and exposure may help learners reduce L1 phonological transfer and consequently accentedness. In sum, how much the learner can develop their pronunciation largely depends on the type of instruction (e.g. Norris & Ortega, 2000), the amount of recent classroom instruction (Saito & Hanzawa, 2016), and the amount of extra-curricular L2 learning (Muñoz, 2014).

In terms of psycho-social factors, previous research has identified characteristics related to learners' personality traits (e.g., empathy, extraversion), social affective factors (e.g., anxiety, willingness to communicate) or learner attitudes (e.g., motivation) and strategies (e.g., focus on individual words) that may have a substantial impact on L2 pronunciation and speaking (Moyer, 2014). On the one hand, one of the most widely researched affective variable in SLA is FL anxiety, which has been explored as a determinant predictor of L2 performance. Baran-Łucarz (2016) conceptualized pronunciation-specific anxiety, consisting of four main components: fear of negative evaluation, pronunciation self-efficacy and self-assessment, pronunciation self-image, and the learners' set of beliefs related to pronunciation. In fact, foreign accent is one of the most anxiety-provoking aspects of spoken language performance, according to Suzukida (2021). Research-wise, greater pronunciation anxiety has been found to be significantly associated with less comprehensibility (Saito et al., 2018) and accentedness (Suzukida & Saito, 2023). On the other hand, studies of L2 pronunciation on learners who are able to attain native-like L2 pronunciation cite motivation as one of the strongest factors impacting the learners' success. In fact, L2 motivation has been found to be an indicator of reduced foreign accent (Flege et al., 1995; Moyer, 2014) and increased comprehensibility (Saito et al., 2018) and overall perception (Sardegna & Jarosz, 2022). Saito et al. (2018) explored the links among L2 speech comprehensibility and the L2 Motivational Self-system. The authors reported that ideal L2 self (Dörnyei, 2009) may be a key factor for enhancing information processing, and helping them make the most of the available input to produce L2 comprehensible output. However, further research in the EFL setting is required to confirm the robust influence of self-guides on L2 pronunciation learning.

In terms of cognitive factors, we can distinguish language learning aptitude (i.e., the talent to learn a language) from individual variation in cognitive factors that underlie language learning talent (i.e., WM, attention, inhibitory control; speed of lexical access; phonemic coding ability). First, aptitude has been conceptualized as the following set of abilities that enhance FL learning: phonemic coding ability (noticing and analysing unfamiliar auditory information), grammatical sensitivity, inductive learning ability, and associative memory (Carroll, 1962). Although the existing aptitude tests refer to their relevance to speech learning on a broad level, LLAMA-E has been associated with segmental and suprasegmental pronunciation accuracy (Saito, 2019) and LLAMA-D has been positively related to L2 learners' development of comprehensibility and speed and breakdown fluency (Saito et al., 2019). Other researchers (Christiner & Reiterer, 2013, 2016) have also investigated the relationship of musical aptitude and L2 pronunciation attainment, with the finding that musical aptitude is correlated with higher productive and perceptive accuracy of L2 sounds (for a review see Suzukida, 2021).

Another source of ID in L2 speech learning is L2 learners' ability to auditorily process the speech input they are exposed to. Auditory processing is defined as a domain-general perceptual ability to discriminate and reproduce patterns along individual acoustic dimensions, such as pitch, formants, duration, and amplitude (Saito & Tierney, 2022). On a broader level, auditory processing can also comprise a range of neighboring abilities, such as attention to particular acoustic dimensions while ignoring others (i.e., selective attention) and the use of acoustic information for motor action (i.e., audio-motor integration). Cross-sectional investigations have found that auditory processing relates to L2 phonology (Kachlicka et al., 2019), and longitudinal investigations have revealed that learners with more precise auditory processing demonstrate gains when they receive intensive phonetic training (e.g., Saito et al., 2021b) or engage in form-focused communicative instruction (Ruan & Saito, 2023). Whereas acuity is a crucial variable for the attainment of high-level L2 speech proficiency in immersive experiences, audio-motor integration seems to be as a key factor of successful L2 speech learning especially in classroom settings where are encouraged to produce motor output based on a limited amount of input. Finally, given that Spanish listeners use duration cues for English vowel sounds, dimension-selective attention could be a key factor for successful L2 learning. Therefore, the use of a composite model of auditory processing (acuity, attention, and integration) could greatly inform the revised view aptitude-treatment interaction espoused by proponents of the Auditory Precision Hypothesis (Saito, 2023).

To conclude, psycho-social factors may interact with cognitive aptitude factors, L2 proficiency or personality traits, or experience-related factors like amount of L2 use are likely to be related to personality traits like extraversion. Also, anxiety may interfere with attention control processes and negatively affect L2 learners' oral performance. Therefore, it is important to understand the contributions of various ID to get the most complex and complete picture of what might be affecting L2 speech performance and development (Suzukida, 2021; Suzukida & Saito, 2023). This dissertation examines to what extent certain experiential factors and two specific cognitive factors, which are considered as important determinants of individual variation in L2 speech and TBLT literature, moderate L2 pronunciation performance and development.

1.6.2. L2 proficiency

Second language proficiency has been measured through different instruments (e.g., International English Language Testing System; Test of English as a Foreign Language). Ortega et al. (2002) created an EIT in four different languages (English, German, Japanese, and Spanish) as an effective measure of L2 proficiency. The simple and economical administration procedures and the flexibility in the design of task features made EIT an attractive candidate to effectively measure proficiency in both classroom and standardized assessment contexts. The EIT involves decoding a sentence (i.e., parsing the sentence syntactically and semantically), retaining its meaning, and recoding it for production (i.e., omitting or substituting those parts that are not represented in the L1 system). Failing to imitate accurately has been hypothesized to show that the structure is beyond the level of the learner's grammar (Granena, 2016).

Yan et al.'s (2016) EIT metanalysis provided supportive construct-related validity evidence for EIT as a measure of L2 proficiency. They showed that the manipulation of three task features (i.e., nature of construct, sentence length and scoring method) distinguished EIT performances across proficiency levels better. In other words, (1) EIT was especially sensitive to measure global constructs (rather than grammatical knowledge); (2) EIT with different sentence lengths would better match the ability of speakers with different proficiency levels (Kim et al., 2016); (3) EIT would discriminate proficiency-levels more accurately when a refined rating scale is used (i.e., interval rather than binary scales). Solon et al. (2019) corroborated the growing evidence that the EIT developed by Ortega et al. (2002) and extended to other languages (e.g., Wu & Ortega, 2013) indeed offered a valid, reliable, and efficient L2 proficiency assessment tool for SLA research purposes. In their study, scores on the original 30-item Spanish EIT correlated strongly with an external written proficiency measure and exhibited high internal reliability and relatively high discriminability. In a replication of Wu and Ortega (2013), McManus and Liu (2022) also found that the EIT could be used with low proficiency learners (in addition to intermediate-to-advanced L2 learners) and to predict different proficiency levels in the oral modality. In fact, good alignment has been found between EIT performances and standardized proficiency test scores (Wu et al., 2022).

Given that EIT is conducive to the retrieval of implicit knowledge (Granena, 2016), WM should not affect accuracy in EIT performance. Indeed, Kim et al. (2016) demonstrated a weak and nonsignificant correlation (*r*=.30) between WM and EIT performance that only accounted for 9% of the variance in scores, suggesting that the EIT is appropriate to test proficiency-dependent abilities without the overreliance of WM. In relation to L2 pronunciation performance,

Last, how learners' proficiency —as measured through EIT— may be related to learners' developmental success in L2 pronunciation is largely unknown. By determining L2 proficiency through a TOEFL test, Kostromitina and Kang (2021) found with an increase in learners' proficiency —measured through a TOEFL test—, the amount of high functional load consonant and vowel deviations dropped significantly. To our knowledge, no other studies have explicitly studied how EIT may predict variability in L2 pronunciation performance and learning in FL contexts.

1.6.3. Working memory

WM has been defined as "a temporary storage system under attentional control that underpins our capacity for complex thought" (Baddeley, 2007, p.1) and Baddeley's WM model states the central executive system controls cognitive processes and coordinates three slave subsystems: the phonological loop, the visuo-spatial sketch pad and the episodic buffer (see Gilabert & Muñoz, 2010). Previous research has found that a higher PSTM plays a role in L2 acquisition (e.g., Kormos & Sáfár, 2008), and correlates with more accurate L2 pronunciation (Aliaga-Garcia et al., 2011; Darcy et al., 2015). Indeed, by maintaining all the relevant pieces of information simultaneously active, higher functioning WM may facilitate the processing of rapid spoken input, allow for more precise traces of what was heard, and better storage quality. Because of its relevance to selective attention and noticing, which are assumed to affect L2 learning during negotiation of meaning, WM has a clear bearing on interaction-driven L2 learning.

Concerning the link between WM and L2 pronunciation, Darcy et al. (2015) found an L2 complex WM span measure to be related to individual L2-English phonological scores (a composite phonological score based on segmental categorization, lexical stress and phonotactics). In fact, the three strongest significant correlations with phonological processing scores that they obtained involved L2 WM: L2 complex span (r = .504), and L2 storage capacity (digits, r = .438; non-words, r = .397). Similarly, larger PSTM capacity —measured through a serial nonword recognition task— was connected to more efficient phonological processing in Mora and Darcy (2016), as it facilitated perceptual discrimination of L1-L2 phonetic differences. In fact, L2 learners with higher PSTM have an advantage in their ability to attend to relevant spectral cues and achieve target-like cue weighting (Safronova, 2016). Interestingly, in the domain of HVPT, Aliaga-Garcia, et al. (2011) found higher PSTM capacity L1-Spanish/Catalan trainees to obtain higher accuracy scores and greater perceptual accuracy gains than the lower PSTM capacity trainees in the discrimination of L2 English vowels, suggesting that PSTM may contribute significantly to the development of L2 speech perception.

In terms of production, Christiner and Reiterer (2016) conducted a study with musicians and found that WM —measured through a forward and a backward digit span—was correlated with L2 English accent imitation (*r*=.49) and English text reading (*r*=.27), suggesting that WM and attentive skills contribute significantly to aptitude for imitating and repeating unfamiliar speech material. In contrast, Hu et al. (2013) assessed PSTM through a digit span task and a non-word repetition task and found PSTM to be unrelated to L1-German advanced learners' English pronunciation accuracy in the reading of a passage in English. In sum, the forward and backward digit span tasks have been found to be weakly to moderately correlated with perception and production tasks.

In TBLT research, WM has been positively correlated with amount of modified output produced during collaborative tasks (Mackey et al., 2010). In addition, cognitive and affective learner differences may increasingly differentiate performance as task complexity is enhanced (Robinson, 2007b). Thus, differences in WM should help account for performance variance in more cognitively complex tasks. For instance, Kim et al. (2015) revealed a more significant role of WM in the complex task due to its higher cognitive demands and increased problem-solving nature. In contrast, other studies (e.g., Awwad & Tavakoli, 2019; Kormos & Trebits, 2011) found non-interaction effects between task complexity and WM in their predictions of oral performance. These contradictions may be due to the fact that, when producing a language, a number of affective and cognitive variables including attitudes, motivation, language ability, and strategies come into play, so it is possible that participants with low WM can adopt strategies to compensate in actual task performance. The current study aims at advancing current research on cognitive ID by exploring the potential contributions of WM on L2 phonological performance and development as a result of task-based EFL pronunciation instruction.

1.6.4. Attention control

Attention control plays an important role in speech processing and language comprehension and production (Miyake & Friedman, 2012) and in SLA (Segalowitz & Frenkiel-Fishman, 2005), especially at lower levels of proficiency. Attention control is especially relevant for the success of L2 phonological learning because individuals with more efficient attentional control may be able to allocate their attention to select relevant dimensions (selective attention) —while disregarding irrelevant information— in the input for making certain phonemic distinctions and switch rapidly from one dimension to another (attentional flexibility or control).

L2 learners may be often mislead by a wrong phonetic cue when perceiving differences between similar L1/L2 sounds (Flege, 1995). For example, low-level Spanish learners of English may perceive /iː/-/ɪ/ differences in terms of temporal cues and may need to refocus their attention onto a different dimension (spectral differences) to properly cue the English vowel contrast. In this case, attention control might help learners notice cross-language differences between L2 and L1 sounds and focus on relevant L2 phonetic dimensions (Mora, 2022). In fact, Iverson et al. (2005) demonstrated that manipulating the acoustic features in the signal to re-direct learners' attention towards the relevant L2 cue-weighting has led to changes in the quality of perceptual categories stored in long term memory.

Within attention control, inhibitory control is a language control mechanism that allows less interference from the L1 phonological, acoustic and articulatory properties during L2 phonological processing (Lev-Ari & Peperkamp, 2014). With respect to L2 phonological learning, Darcy et al. (2016) found that L2 learners with stronger inhibitory skills were more accurate in categorically perceiving difficult L2 vowel contrasts, but inhibitory control did not predict changes in vowel production. Similarly, Darcy and Mora (2016) revealed that learners' better inhibitory control may help them achieve greater scores in ABX discrimination. Still, further research is needed to assess to what extent inhibitory control can predict ultimate attainment in L2 pronunciation and speaking, and how mediates L2 gains from longitudinal pronunciation interventions.

In addition, both attention switching (ASW) (Mora-Plaza et al., 2022a) and auditory selective attention (ASA) are key to selectively attend to specific acoustic dimensions during speech processing and focus attentional resources on the auditory information that is relevant for language decoding processes to work efficiently (Astheimer et al., 2016). In particular, ASA allows listeners to selectively attend to a single acoustic dimension or feature during speech processing while ignoring others, enhancing learning of L2 phonological contrasts (Ou et al., 2015). Studies investigating the role of attention control in L2 phonology have shown mixed results, with positive relationships between attention control and L2 phonological processing only surfacing for specific participant groups or phonological processing tasks. For example, Kim and Hazan (2010) found ASW skills to be related to training gains in naïve L1-English speakers trained to perceive a novel Korean stop voicing contrast. Mora and Mora-Plaza (2019) trained L1-Spanish learners in the perception and production of L2-English /æ/-/ʌ/ and /iː/-/ɪ/ embedded in nonwords during 4 45-minute training sessions. A moderately strong correlation between L2 gains in the perception of the /æ/-/ʌ/ contrast (but not /iː/-/ɪ/) and ASA suggested that learners' ability to focus their attention to specific speech dimensions was related to L2 phonological acquisition. Additionally, ASW skills were related to accuracy of performance in ABX discrimination (as in Safronova, 2016). In contrast, in Mora-Plaza et al. (2022a), neither ASA nor ASW explained individual differences in training gains mainly because of the relatively low small training gains within groups. However, ASA correlated strongly with learners' T1 and T2 scores in the ABX task, indicating that ASA enhanced learners' ability to discern between the target vowels, supporting previous findings (Mora & Mora-Plaza, 2019). Contrary to these findings, Ghaffarvand Mokari and Werner (2019) found attention control (as measured through the Stroop task) to be unrelated to training gains for L1-Azerbaijani learners of L2 English.

Section 1.6. has identified some of the ID that have been most studied in L2 speech research, namely, age and experiential, psycho-social, cognitive and auditory-processing factors. Given that the current dissertation focuses on the effects of experiential and cognitive factors, we showed evidence for the use of the EIT as an appropriate measure of general L2 proficiency, and reviewed previous work on the associations between WM/selective attention and L2 speech perception and production. We hope that this dissertation sheds further light on the relationship between ID and L2 pronunciation outcomes.

Summary and research gaps

The first chapter of the current dissertation examined how second language speech is mainly characterized by L1 processing, and why noticing and attentional mechanisms are relevant to L2 learning (Section 1.1). Theoretical speech learning models were reviewed as they systematically predict cross-linguistic difficulty in the development of new categories (SLM) or re-organization of L1 categories (PAM). In addition, L2 sound perception was revealed to be often linked to production but not directly related to the updating of L2 phonolexical representations. What is clear from psycholinguistic research is that noticing and focusing processes, aided by memory and attentional resources, are fundamental for L2 phonological learning and retention. More investigations are needed to examine how directing learners' attentional resources to difficult phonetic forms can lead to better perception, production and lexical encoding of L2 sounds in FL contexts.

In particular, this thesis focuses on L2 vowels that are challenging for Catalan learners of English. Given that Catalan /i/ and /a/ approximately occupy the portion of the spectral space occupied by the English vowel contrasts /iː-/ɪ/ and /æ-ʌ/, respectively, studies have shown that Catalan learners of English assimilate English /iː/ to their existing L1 /i/ category, while /ɪ/ is mapped onto Catalan /e/ or /i/; and English /æ/ and /ʌ/ are both assimilated to Catalan /a/. In addition, research evidence shows Catalan speakers over-rely on temporal cues, given their lack of awareness of spectral differences. According to SLM, PAM-L2 and NRV speech models, the aforementioned vowel contrasts are difficult to discriminate (Section 1.2.). However, far too little attention has been paid to the role of pronunciation instruction to increase learners' attention to previously unattended spectral cues in conversational contexts.

Therefore, Section 1.3. highlighted the importance of raising learners' awareness of relevant phonetic properties of the input to facilitate phonological processing in the recognition and production of words. Whereas phonetic training and explicit instruction are relevant to inform how classroom-based instruction can promote stable, generalizable and robust L2 pronunciation gains, these methods constitute decontextualized focus on forms that do not reflect language outside the classroom. Similarly, computer-assisted pronunciation teaching promotes individualized, anxiety-free learning with varied multimodal input, but often at the expense of pedagogical values. This dissertation advocates for teaching pronunciation through form-focused instruction, which results in automaticity and generalization of speech patterns in spontaneous speech. Having identified the main challenges of pronunciation instruction (i.e., time, method and focus), little is known about approaches that offer integration of pronunciation with other language skills, focus on form within meaningful practices and prioritize aspects of pronunciation that may affect learners' comprehensibility.

One approach that may overcome some of the aforementioned challenges is TBLT. Section 1.4. aimed at reviewing the foundations of TBLT to understand how focusing on form in interactive meaningful contexts may trigger adjustments in the linguistic input and output, and promote L2 development. The definitions of "task", classification of task types, selection of tasks and task frameworks were presented to justify the task criteria that we followed in our pedagogical intervention (two-way, convergent, closed, split, information-required, dialogic, real-world, focused, task-essential tasks). Last, Robinson's (2011b) Cognition Hypothesis, the Triadic Componential Framework and SSARC model were described to understand how tasks should be designed and sequenced to raise learners' awareness of certain linguistic L2 targets. Whilst TBLT investigations have mainly focused on CAF performance and the development of grammar, lexis and, more recently, pragmatics, there is very little scientific understanding about the potentiality of task design for L2 pronunciation improvement. In addition, few empirical investigations have tested whether Robinson's (2011b) Cognition Hypothesis holds for L2 pronunciation; in other words, whether a cognitively more complex task results in phonetically more accurate speech production, as it is the case for grammatical accuracy.

As a result, Section 1.5. offered evidence that using form-focused techniques and task manipulations may promote attention to L2 pronunciation in interactive classroom-based settings. First, task-based pronunciation teaching studies considering TBLT techniques (e.g., background, context, task repetition, task complexity) for L2 pronunciation development were reviewed. Apart from a few studies, this section showed a general lack of research as well as disparity of findings on the role of task complexity on L2 phonological form and P-LRE in focused tasks. It also pointed to the need for TBPT studies that factor in learner factors to better explain L2 pronunciation development.

Consequently, the last section of Chapter 1 provided insights into the influence of experiential (age and L2 experience), psycho-social (anxiety, motivation), cognitive (aptitude, attention control) and auditory processing factors on L2 speech acquisition, especially, in FL contexts. Section 1.6. first contrasted previous literature on the EIT as an appropriate instrument to accurately discriminate L2 proficiency levels independently of WM. In addition, WM was reported to predict L2 phonological scores and efficient phonological processing, as well as play a key role in phonetic and accent imitation training. Last, previous work on inter-individual variation in attention control suggested that selective attention and attention switching are relevant for speech processing. Nevertheless, no previous study has investigated the mediating effect of experience, L2 proficiency, WM and attention control in L2 pronunciation performance and gains after a task-based pronunciation intervention.

To sum up, this dissertation seeks to bridge the gap between L2 speech learning research and TBLT research by assessing the effects of form-focused task-based instruction on L2 phonological learning, in particular, the perception, production and lexical encoding of two confusable L2 vowel contrasts (/iː-/ɪ/ and /æ-ʌ/), and the mediating effects of task complexity. To this end, sixty-three Catalan/Spanish EFL learners were asked to perform twenty dyadic decision-making tasks differing in cognitive task complexity during 7 weeks, as part of their English language classes. In order to prove the effectiveness of task-based pronunciation instruction, generalization and retention of gains, and the performance of a control group (N = 29) were considered. Finally, L2 speech learning gains were interpreted in light of experiential and cognitive learner factors.

02 OBJECTIVES AND RESEARCH QUESTIONS

CHAPTER 2. OBJECTIVES & RESEARCH QUESTIONS

This chapter summarizes the main contributions of the present study with respect to the literature reviewed in Chapter 1 (Section 2.1.), followed by the main objectives (Section 2.2.) and the description of the target L2 vowels investigated in this study (Section 2.3.). Then, the research questions that have motivated this study (Section 2.4) are listed and, subsequently, the corresponding hypothesis are formulated (Section 2.5.).

2.1. Contribution

The present study is concerned with the acquisition of four challenging English vowels /iː/, /ɪ/, /æ/, /ʌ/ by Catalan/Spanish learners of English and the potential of TBPT and task complexity in effecting gains in the perception, production and lexical encoding accuracy of the aforementioned L2 sounds. This study aims at advancing our knowledge of the field of pronunciation instruction and TBLT in a FL context. Thus, there are several important areas where this study aims at making an original contribution to:

(1) This study seeks to contribute to the literature on the effectiveness of **pronunciation instruction** for L2 segmental acquisition. Whereas in the last decade, form-focused communicative studies targeting pronunciation have gained interest in second language pronunciation research (e.g. Darcy & Rocca, 2023), explicit pronunciation interventions still outnumber form-focused communicative interventions. Most pronunciation instruction studies have assessed gains through listener-based ratings, namely, comprehensibility (Darcy et al., 2021; Gordon & Darcy, 2019) and/or accentedness (Saito, 2011), and to a lesser extent, in terms of segmental features identification (Ruan & Saito, 2023) or acoustic analyses (Saito,

2015). This study evaluates improvement in L2 pronunciation by objectively measuring learners' *perception* at the pre-lexical level, *production* (via acoustic analyses) and *lexical encoding* of L2 vowels, after a task-based form-focused communicative intervention.[34]

(2) This study intends to determine the extent to which **TBLT** principles may be applied to L2 pronunciation learning. We explore the role of task design in promoting attention to and development of L2 pronunciation, a central aspect of L2 learning and communicative competence. Few empirical studies have assessed the effects of TBLT on L2 segmentals, but those that have done so have only assessed vowel production (but not perception) through formant description (Solon et al., 2017), rather than *acoustic distances* between learners' contrastive vowels or native vs. learners' vowels (e.g., Euclidean, Mahalanobis distances) — often used in L2 speech sciences (e.g., Kartushina et al., 2015). This study investigates the effectiveness of TBLT not only on vowel production, but also on *vowel perception* and *lexical encoding* (responding to the call for further research made by Mora & Levkina, 2017 and Solon et al., 2017), as well as the production of pronunciation-focused language related episodes, as instances of L2 pronunciation development.

(3) This study aims to contribute to the growing area of research concerning **task complexity** effects on L2 pronunciation in pronunciation-focused tasks. Whilst the evidence for this relationship is inconclusive, this study aims at testing whether the benefits of increased task complexity along resource-directing variables (Robinson, 2011b) for L2 lexico-grammatical development can extent to pronunciation development, maximally *controlling for* methodological and

[34] My personal emphasis *in italics*, highlighting this study's original contributions.

learner aspects previously unattested. In addition, operationalizations of task complexity *are validated* in two different ways: expert judgements by EFL teachers and subjective self-ratings by EFL students (see Révész et al., 2016).

(4) This study considers **individual differences** as potential predictors of L2 speech performance. *None* of the TBPT studies published in academic journals (e.g., Gordon, 2021; Gurzynski-Weiss et al., 2017a) so far have considered *learner factors* in explaining variability in L2 outcomes, as it has been previously done in TBLT studies. This study was designed to offer an insight into the role of experiential and cognitive factors that have been previously explored in L2 speech research (Mora, 2022; Suzukida & Saito, 2023).

(5) This study was conducted in a foreign language **instructional context** with **adolescent** EFL learners. Whilst much empirical SLA research into phonetic training or pronunciation instruction has been done in highly controlled experimental settings, this study presents an opportunity to examine the *generalizability of findings* drawn from these investigations to high school classroom language learning, a context which is underrepresented in the TBLT (Erlam & Tolosa, 2022) and pronunciation instruction (Darcy et al., 2012a) literature.

(6) This study tried to adapt the **testing methods** which have been validated in lab-based phonetic training studies to a classroom-based pronunciation intervention. For example, we used individualized computer tests that assessed learners' perception, and learners' production was recorded through high-quality equipment. Furthermore, following L2 *phonetic training assessment methods* of L2 speech perception and production, generalization of gains to novel voices and untaught words is assessed.

(7) This study is concerned about the **retention** of L2 pronunciation gains. Whilst most pronunciation instruction studies are cross-sectional and do not tend to include delayed post-tests, this study undertakes a *longitudinal analysis* of L2 speech learning (Nagle, 2022) by evaluating learners' pronunciation gains not only immediately after the task-based intervention, but also 11 weeks after.

(8) This study aims to gain an insight into **learners' perception** about pronunciation learning (Dao, 2018; Nguyen et al., 2021). The added value of this study, compared to other TBPT studies, is that it describes learners' beliefs about pronunciation, evaluation and likeability of TBPT, and their self-reflections of improvement in L2 pronunciation and other language-related and unrelated skills (Henrichsen & Stephens, 2015).

In sum, this study explores how the beneficial effects of performing tasks are translated into the learning of L2 pronunciation in EFL instructional settings. To do so, sixty-three L1-Catalan/Spanish EFL learners received 20 TBPT lessons during 7 weeks. Thirty-one learners performed simple and 32 performed complex decision-making tasks. These two groups and a control group of 29 untaught learners were pre- and post-tested on the discrimination, production and lexical encoding of four target vowels, their interaction moves, as well as experiential and cognitive ID.

2.2. Objectives

The main aim of this study is to investigate whether TBPT can promote L2 vowel learning in an instructional EFL context.

With the aim of extending the research conducted in the TBLT field and provide evidence that form-focused pronunciation instruction can be taught communicatively, this study examines the potential of task design to draw learners' attention to challenging L2 pronunciation features during interaction, and subsequently, lead to L2 vowel learning. In particular, this study assesses the effectiveness of TBPT on L2 vowel perception, lexical encoding and production as well as learners' engagement in P-LRE during dialogic task performance. In addition, this investigation evaluates improvement in terms of generalization to novel voices and words, and retention of phonological learning.

Given the mixed findings in the literature concerning the effects of task complexity on L2 pronunciation, this dissertation set out to assess the effect of task complexity on L2 vowel learning, thus, testing Robinson's (2011b) Cognition Hypothesis predictions for segmental learning. In particular, it seeks to find out whether increasing the cognitive demands of pronunciation-focused tasks can lead to more accurate vowel discrimination, lexical encoding and production immediately after and long time after the intervention. Additionally, this study attempts to obtain conversational data which will unfold the impact of task complexity on the occurrence of P-LRE, hence, whether increased task demands help direct attention to phonetic form while communicating.

The third objective of this research is to explain learners' performance and development of L2 vowels taking into consideration experiential (past and recent English experience, L2 proficiency) and cognitive ID, namely, WM and selective attention. Drawing upon previous research on L2 speech learning in FL contexts (Christiner & Reiterer, 2013; Mora & Mora-Plaza, 2019; Suzukida & Saito, 2023), these are some of the most important learner factors that have predicted improvement in L2 speech, and may help explain the unique contribution of TBPT on L2 vowel learning once these are controlled for.

The forth objective is to offer a qualitative complement to the quantitative findings of the study by assessing learners' beliefs about learning pronunciation, their experience after the TBPT intervention and their perception of improvement. Inspired by Carlet's (2017) post-training survey, subjective impressions on the difficulty, likeability and learnability of TBPT will reinforce the quantitative results on the effectiveness of TBPT for L2 pronunciation learning, and will provide a wider and richer illustration of the learners' learning process (Henrichsen & Stephens, 2015).

Finally, the results of this study are likely to provide relevant pedagogical implications for the teaching and learning of English pronunciation and speaking skills in a FL instructed context, specifically, the EFL teaching context of the Catalan educational system. Assuming that TBPT is an effective method for teaching L2 pronunciation, it may provide learners with form-focused practice during interactive tasks and may be easily integrated within other skills, such as speaking or listening, which, in turn, may reduce the amount of time that would be allocated for teaching L2 pronunciation in isolation.

2.3. Target sounds

This dissertation deals with the perception and production of English vowels /iː/, /ɪ/, /æ/, /ʌ/ by L1-Catalan/Spanish EFL learners, which are described in terms of articulatory and acoustic features below.

According to Mott (2011), /iː/ is a front (less centralized), almost fully closed, unrounded vowel produced with lip spreading (e.g., *free, Peter*); /ɪ/ is more centralized, almost half-close, unrounded vowel produced with lip neutralization (e.g., *hill, Tilly*); /æ/ is front, between half-open and open, unrounded produced with lip spreading (e.g., *tap, magic*); and /ʌ/ is slightly forward to the centre, just below half open, unrounded produced with lip neutralization (e.g., *sun, lucky*). Concerning vowel frequencies (Table 2.1.), /iː/

has a lower F1 and higher F2 (i.e., /iː/ is closer and more front) than /ɪ/; and /æ/ has a higher F1 and higher F2 (i.e., /iː/ is more open and more front) than /ʌ/. Regarding duration, /iː/ is longer than /ɪ/ (119ms vs. 75ms); and /æ/ is longer than /ʌ/ (159ms vs. 103ms) when they occur in individual words in stressed position (Crystal & House, 1988). In addition, /iː/ is tense (i.e., produced with articulatory tension) and /ɪ/, /æ/ and /ʌ/ are lax (i.e., produced without articulatory tension), whilst /æ/ differs from /ʌ/ visually in degree of lip aperture (/æ/ is opener). Finally, /iː/ is produced with longer duration than /ɪ/. Despite British English native speakers use F1 and F2 values as a primary cue to distinguish this vowel contrast, they have also been found to rely on phonemic length (ms) when their primary cue is not available (Hillenbrand, et al., 2000).

Table 2.1. Average British English male and female values of F1 and F2 in Bark for words in isolation (Deterding, 1990) and produced in connected speech (Deterding, 1997).

	In isolation				Connected speech			
	Male		Female		Male		Female	
	F1	F2	F1	F2	F1	F2	F1	F2
iː	2.68	13.77	3.10	15.03	2.73	13.85	2.95	14.87
ɪ	3.68	12.97	4.14	13.98	3.54	12.26	3.70	13.64
æ	6.63	11.32	8.58	12.26	6.31	11.42	8.62	12.41
ʌ	6.35	9.83	7.24	10.84	5.94	10.02	7.94	11.01

The aforementioned L2 vowels were chosen due to their difficulty in acquisition, learners' inappropriate use of phonetic cues and their high functional load (see Section 1.2.4.) and, consequently, the fact that their mispronunciation may contribute to loss of intelligibility and reduction of comprehensibility (Levis, 2018). On the one hand, the English /iː-ɪ/ vowel contrast is hard to perceive and produce for Catalan learners of English because they assimilate English /iː/ and /ɪ/ to Catalan /i/ via an uncategorized-categorized assimilation pattern (Best & Tyler, 2007). Whereas English /iː/ is perceptually similar to Catalan /i/, the English vowel /ɪ/ has been found to be perceived as a poorer fit of the same L1 sound, being identified as either Catalan /i/ or /e/. In addition, Catalan EFL learners rely heavily on temporal cues to distinguish /iː-ɪ/ due to their lack of experience with spectral differences between these vowels. On the other hand, the /æ-ʌ/ English contrast is also hard to acquire (Aliaga-Garcia & Mora, 2009; Rallo-Fabra & Romero, 2012), as both English vowels are perceptually mapped onto a single L1 low central vowel category /a/ in Catalan, although /æ/ is a slightly better perceptual match for Catalan /a/ than English /ʌ/ (category-goodness assimilation) (see Cebrian 2019, 2021; Tyler, 2019, 2021 & Section 1.2.4 for further details on these cross-linguistic segmental difficulties).

Last, learning to differentiate English /iː/ from /ɪ/ and /æ/ from /ʌ/ in perception and production should be prioritized due to their high functional load (See Section 1.3.4.5.). As observed from Higgins' (2017) list of minimal pairs in non-specialist (i.e., everyday use) English contexts, the English /iː-ɪ/ contrast is present in a total of 466 pairs and the English /æ-ʌ/ contrast in 436 pairs, contrary to other contrasts (/ʊ-uː/: 18 pairs; /æ-ɑː/: 184 pairs). Learning the aforementioned target minimal pairs is expected to help them distinguish a large number of words based on these vowel contrasts effectively and enhance the comprehensibility of their speech (Munro & Derwing, 2006; Suzukida & Saito, 2022).

2.4. Research questions

This dissertation poses four main research questions (RQ), each subdivided into several subquestions.

Research question 1 (RQ1): ***Main effects of TBPT*** **(T1 - T2 - T3)**

Is TBPT effective at improving L1-Catalan/Spanish learners' perception, lexical encoding and production of L2 English vowels?

RQ1.1.: To what extent do learners improve the *discrimination* of L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/)?

RQ1.2.: To what extent do learners improve the *lexical encoding* of L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/)?

RQ1.2.1.: Are there gains in lexical encoding as tested through a forced lexical choice (FLeC) task?

RQ1.2.2.: Are there gains in lexical encoding as tested through a lexical decision (LD) task?

RQ1.3.: To what extent do learners improve the *production* of L2 vowels (/iː/, /ɪ/, /æ/, /ʌ/)?

RQ1.3.1.: Are there accuracy gains in the production of L2 vowels in isolated words?

RQ1.3.2.: Are there accuracy gains in the production of L2 vowels in words in sentence contexts?

RQ1.4.: Is there a relationship between learners' performance and gains in *perception*, *lexical encoding* and *production*?

Research question 2 (RQ2): ***Main effects of task complexity*** **(Simple - Complex)**

Does task complexity play a role in L1-Catalan/Spanish learners' perception, lexical encoding and production of L2 English vowels, as well as the occurrence of P-LRE?

RQ2.1.: What is the effect of task complexity on learners' *discrimination* of L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/)?

RQ2.2.: What is the effect of task complexity on learners' *lexical encoding* of L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/)?

RQ2.3.: What is the effect of task complexity on learners' *production* of L2 vowels (/iː/, /ɪ/, /æ/, /ʌ/)?

RQ2.4.: What is the effect of task complexity on the occurrence of *P-LRE*?

Research question 3 (RQ3): *The role of individual differences* (Experiential - Cognitive factors)

Do individual differences in L2 experiential and cognitive factors explain L1-Catalan/Spanish learners' performance and gains in English vowels?

RQ3.1.: Are there differences in terms of experiential (past and recent L2 experience, L2 proficiency) and cognitive (WM, ASA) factors between simple, complex and control groups? Are experiential factors related to cognitive factors in the experimental group?

RQ3.2.: Is there an association between experiential and cognitive ID and performance in L2 vowel discrimination, lexical encoding and production? How much unique variance in L2 vowel performance do ID explain?

RQ3.3.: Is there an association between experiential and cognitive ID and gains in L2 vowel discrimination, lexical encoding and production? How much unique variance in L2 vowel gains do ID explain?

Research question 4 (RQ4): ***Learners' perceptions of the intervention*** **(Beliefs – Likeability – Learning)**

Which were L1-Catalan/Spanish learners' perceptions of the TBPT intervention?

RQ4.1.: Which were learners' beliefs about English pronunciation, evaluation of the task-based project and perceptions of learning once the TBPT intervention ended?

RQ4.2.: Were there group differences (simple vs. complex) in terms of learners' evaluation of the TBPT intervention and perceptions of learning?

2.5. Hypotheses

Several hypotheses are formulated supported by previous research in the field of L2 speech acquisition, TBLT and individual differences in SLA.

Concerning *RQ1* on the main effects of TBPT, it can be hypothesized that a 10-hour intervention (30 minutes x 20 tasks) of form-focused pronunciation instruction is likely to enhance learners' attention to difficult L2 vowel contrasts during interactive communication, and promote L2 vowel learning, in line with Sicola (2008) and Solon et al.'s (2017) findings. Although greater L2 segmental perception gains have been found as a result of HVPT or explicit pronunciation instruction (10-15% - Carlet & Cebrian, 2019; 10-20% - Thompson, 2012) than form-focused communicative instruction (4% - Ruan & Saito, 2023), it is predicted that TBPT will lead to gains at the pre-lexical perceptual level, similar to the ones obtained from incidental HVPT (5-10% - Lim and Holt, 2011; Saito et al, 2022c). Moreover, lexical encoding gains have been found as a result of a phonological specificity treatment on L2 minimal pairs (Llompart & Reinisch, 2021) or phonetic training (Adrian & Mora, 2022; Melnik & Peperkamp, 2021). Therefore, it could be conceivably hypothesised that orienting learners' attention to L2 phonological form by making target phonological forms essential for task completion, may improve learners' sensitivity of the vowel contrast at a lexical level. Vowel productions are also predicted to become increasingly more distinct and more target-like throughout the communicative intervention (Saito, 2015) as a result of noticing L1-L2 crosslinguistic differences in their own speech and that of their interlocutors. It may also be the case that learners make use of vowel duration (Cerviño & Mora, 2009), apart from vowel quality differences, to differentiate L2 vowel contrasts.

According to PAM-L2's (Best & Tyler, 2007) systematic predictions of perceptual vowel difficulty, English vowels in the /æ/-/ʌ/ contrast (category-goodness assimilation) are expected to be more difficult to discriminate than /iː/-/ɪ/ (uncategorized-categorized assimilation), but see Cebrian (2019). Also, based on perceptual assimilation studies (Cebrian 2019, 2021), English /ʌ/ is predicted to be produced less accurately and to present more room for improvement than /æ/, which is perceptually closer to L1-Catalan /a/. Also, peripheral vowels /æ/ and /iː/ may be perceived and produced more accurately than their more centralized vowel counterparts /ʌ/ and /ɪ/, respectively, according to the NRV framework (Polka & Bohn, 2003, 2011). In addition, given the considerable length of the intervention, relative to regular phonetic training or explicit instruction studies (<5h), perception and production gains are expected to be generalizable to untaught words and voices, and retained over time (Carlet & Cebrian, 2019; Rato, 2014). In line with previous literature, learners are expected to perform significantly better in the FLeC than LD task (Kojima, 2019); however, it is currently unknown which of the two tasks would reflect more gains in the lexical encoding of L2 vowel contrasts. Furthermore, whilst eliciting words in isolation has been hypothesized to direct learners' attention to the acoustic features that distinguish the target vowels and to achieve higher levels of articulatory control than when the target words are embedded in sentences (Mora et al., 2022), improvement in production may be more noticeable in words in isolation. However, given the communicative, meaning-based nature of this intervention, we expect greater gains from vowels elicited in words in sentence contexts, in line with Hirata (2004). Finally, a weak relationship (if any) between learners' gains in perception, lexical encoding and production of L2 vowels is expected, according to previous literature on the perception-production (Casillas, 2020; Nagle, 2021), or perception-lexical encoding link (Daidone & Darcy, 2021).

Based on previous research (Solon et al., 2017; Gordon, 2021; Mora & Levkina, 2018), we hypothesize that task complexity may probably enhance learners' acquisition of L2 vowels under a task-based communicative approach (*RQ2*). Therefore, manipulating tasks along resource-directing dimensions may direct learners' attentional resources to phonetic form and, according to Robinson's (2011b) Cognition Hypothesis predictions, boost L2 pronunciation accuracy. Although manifestations of attention to form reflected through LRE have been found for grammar (Baralt, 2014) and lexis (Gilabert et al., 2009) when tasks demands increase, we may not find significant differences in terms of frequency of P-LRE as a function of task complexity, in line with Solon et al. (2017). However, learners' awareness of phonetic form after the intervention could be reflected through an increase of P-LRE from T1 to T3.

With respect to *RQ3* on the main effects of ID, experiential factors such as recent English experience outdoors are predicted to explain some variability in learners' vowel performance (in accordance with Suzukida & Saito, 2023; partly in line with Fullana & MacKay, 2003), but WM and ASA are hypothesized to better characterize learners' L2 vowel performance (Darcy et al., 2015; Mora & Mora-Plaza, 2019; Mora-Plaza et al., 2022a). On the one hand, L2 learners with high attention control and WM could be at an advantage in terms of benefiting from proactive FonF (Lee, 2021 for L2 French grammatical targets). On the other hand, we also expect learners to obtain gains after pronunciation instruction regardless of their ID, hence, it is possible that we only find weak associations between learner factors and L2 vowel gains (Mora-Plaza et al., 2022a; Suzukida & Saito, 2023). Additionally, based on Robinson's (2007b), experiential and cognitive learner differences may increasingly differentiate performance as task complexity is enhanced.

Last but not least, learners' perception of the intervention (*RQ4*) is hypothesized to be overall positive. As reported by EFL learners from Nguyen et al. (2021) and ESL learners from Henrichsen and Stephens (2015), form-focused communicative pronunciation instruction in a Vietnamese context was believed to improve their pronunciation was well as enhance L2 speaking and listening skills. We hope that this TBPT project is helpful in raising learners' awareness about the importance of L2 pronunciation for communication and leads to a general feeling of improvement. Finally, the complex group is hypothesized to report higher ratings of task difficulty and mental effort, but non-significant differences in terms of task enjoyability, relative to the simple group (Robinson, 2001a; 2007b).

03 METHODOLOGY

CHAPTER 3. METHODOLOGY

The aim of this chapter is to introduce the demographic and linguistic characteristics of the participants of the study (Section 3.1.); offer an overview of the experimental design and timeline (Section 3.2.); describe the stimuli, the intervention and testing materials and procedure, and the questionnaires (Section 3.3.); outline the overall study procedure (Section 3.4.); and present the approaches to analyses and main statistical analyses (Section 3.5.) employed to answer the research questions.

3.1. Participants

This study was initially conducted with 98 L1-Catalan/Spanish learners of L2-English who were 16-17 years old and were studying *1r Batxillerat* (Year 12) at the public secondary school *INS Vilatzara*, in Vilassar de Mar (Barcelona, Spain). Nevertheless, 6 participants had to be excluded because they missed multiple speaking sessions (≥ 8 tasks) or missed one testing time (T1, T2, T3).

In order to assess the effectiveness of TBPT and the main effects of task complexity, three intact classes were selected for convenience, as it would have been logistically impossible to randomly assign participants to different intervention groups (Mackey & Gass, 2016). One class (N = 31) conducted simple tasks (i.e., simple group: SG); one class (N = 32) conducted complex tasks (i.e., complex group: CG); and one class (N = 29) did not perform any task (i.e., control group: CTG). Therefore, 63 learners (33 males, 30 females) were part of the experimental group and 29 (14 males, 15 females) were part of the control group.

In terms of the current place of residence, most participants lived in Vilassar de Mar, where the school is, and the rest lived nearby. All learner groups took the same compulsory subjects (i.e. Catalan, Spanish, English, Maths, History and Philosophy), but differed in the optional subjects according to their career-related preferences (i.e., technology and humanities (SG), science and humanities (CG), or technology and social sciences (CTG)). The three intact classes had 1-hour EFL classes, 3 times per week at school. The English learning materials used in class aimed at an intermediate/upper-intermediate level of English. Finally, the focus of EFL instruction was typically grammar and vocabulary-centred rather than pronunciation, thus, their familiarity with the English vowel system was considered minimal. Overall, most contact with L2 English was restricted to the school, but individual differences in FL exposure outside school were present (See Appendix B).

In the remainder of the section, the most relevant demographic and linguistic characteristics of SG, CG and CTG are detailed. Finally, a comparison between the three groups of learners in terms of L2 knowledge and experience is conducted. See Appendix B.1.-B.2. for experimental/control, and Appendix B.4.-B.6. for simple/complex complete demographic and linguistic information).

In terms of demographic information, the SG (*N* = 31) was made up of 17 male and 14 female learners, whose mean age was 16.03 (*SD* = .18, *Range* = 16-17). In the CG (*N* = 32), 16 were male and 16 were female, and their mean age was 16.06 (*SD* = .24; *Range* = 16-17). Out of the 29 participants in the CTG, 14 were male and 15 were female, and their mean age was 16.07 (*SD* = .26, *Range* = 16-17). The majority of learners were right-handed, with 4 (SG), 6 (CG) and 5 (CTG) being left-handed, and only 1 ambidextrous (CTG). From the SG, two learners reported having dyslexia and one suffered a hearing impairment but was wearing a hearing aid and this allowed him to perceive sound without difficulties. From the CG, two learners reported suffering from a minor loss of audition; however, no one from the CTG reported having any speech/hearing pathologies. The aforementioned learners were all included but special caution was taken in the analyses of the results for these particular cases.

In terms of learners' mother tongue, the SG was formed by Catalan/Spanish bilinguals who differed in the degree of L1-dominance. 61.3% were Catalan-dominant and 38.7% were Spanish-dominant. Learners stated that they used Catalan around 62% of the time and Spanish 31% of their time. Similarly, CG learners used Catalan around 64% of the time and Spanish 31% of their time, and clearly differed in their degree of L1-dominance (81.3% were Catalan-dominant and 18.8% were Spanish-dominant). Similarly, the CTG's Catalan/Spanish bilinguals were 69% Catalan-dominant and 31% Spanish-dominant, and used Catalan around 62% and Spanish 38% of their time. A couple of CTG learners defined themselves as Spanish/Arabic bilinguals.[35]

As for their FL linguistic experience, all learners reported English to be their L2 but varied in AOL (*Range* = 3-10) and years of instruction (*Range* = 6-13), Table 3.1. On average, SG learners stated that they received English input from the school approximately 4 hours/week, and around 1.5 hours/week of extra-curricular lessons. In the case of CG and CTG learners, they reported receiving 3 hours/week of English input at school and around 1.5 hours/week from private language academies. Only 6 (SG), 5 (CG) and 1 (CTG) learner had obtained the Cambridge First Certificate in English at the time of testing.

[35] Despite the variability in the degree of Catalan/Spanish dominance and differences in Catalan/Spanish use, this was not expected to affect intervention outcomes, as English /iː/ and /ɪ/ are assimilated to L1-Catalan and L1-Spanish /i/, or /i/ and /e/, respectively, and English /æ/ and /ʌ/ are mapped onto a similar /a/ low vowel category in L1-Catalan and L1-Spanish (Cebrian et al., 2011).

Concerning learners' L2 language use, as expected, all groups reported speaking more English with non-native speakers (3.21 hours/week on average) than native speakers of English (1.28 hours/week on average), as shown in Table 3.1. Interestingly, learners in the SG, CG and CTG mentioned that they had similar exposure to Received Pronunciation (51.61%, 51.87%, 54.48%, respectively) and General American (48.39%, 48.13%, 45.52%, respectively) accents of English[36]. While CG and CTG learners considered themselves to sound with a greater British English than General American accent, accent-type differences were not present in the SG (Appendix B.2 & B.5.). In addition, all group learners revealed that they had more exposure and greater opportunities to produce oral English than written English (Appendix B.3. & B.6.). Finally, as observed in Table 3.1., SG, CG and CTG learners rated their own L2 proficiency as intermediate to upper-intermediate and their pronunciation as being quite accurate. Ortega et al.'s (2002) EIT showed that EFL learners were intermediate, scoring an average of 71 points out of 120 (see Table 3.1. for group comparisons). Whereas only 3 SG and 2 CTG learners had been living in an English-speaking country over 3 weeks (M = 18.33 weeks, SD = 16.62; M = 7.50 weeks, SD = 6.36, respectively), 9 CG learners reported having lived in an English-speaking country for around 10.33 weeks (SD = 14.62).

Last but not least, 14 participants from the SG commented that they were learning a third foreign language (L3), namely, German (N = 8), French (N = 5) and Italian (N = 1) and only 1 participant was studying L4 Italian. In the case of CG, 18 participants reported L3 learning, namely, German (N = 15) and French (N = 3) and 2 participants were learning Japanese (N = 1) and Russian (N = 1) as their L4. Last, 19 CTG participants

[36] While the main exposure to L2 English in school came from the teacher, who had a Southern British English accent, the reported exposure to General American may be ascribed to the learners' monthly (10.6%), weekly (20.8%) and daily (13.1%) exposure to American TV series or learners' monthly (14.6%), weekly (34.2%) and daily (30.3%) exposure to films/videos in English.

mentioned they were learning L3 German (*N* = 13), French (*N* = 5) and Arabic (*N* = 1) and only 1 person was studying L4 Japanese. SG learners had started learning their L3 when they were around 10 years old (range 4-12) and CG and CTG learners when they were around 11 years old (*Range* = 4-14; *Range* = 5-14, respectively). The two SG and CTG L4 learners had begun L4 learning at the age of 12 and the CG learner at the age of 15. To sum up, learners' linguistic profiles suggest that the great majority of learners from the present dissertation were multilingual and had great interest in the learning of foreign languages beyond L2 English.

Finally, a three-group comparison was conducted to test the hypothesis of equal population variances. On the one hand, one-way ANOVAs revealed no significant differences concerning sex, hand-dominance and mother-tongue ($p > .05$) across the three groups. On the other hand, one-way ANOVAs showed that the three groups were not significantly different concerning their L2-English experience and knowledge (Table 3.1.). To conclude, these results showed that participant groups did not differ significantly in terms of demographics or reported L2 experience, and thus, were comparable prior to the intervention.

Table 3.1. Three-group comparison (SG: simple group, CG: complex group, CTG: control group) for L2 experience and knowledge.

		SG	CG	CTG	*p*	*ηp2*
Age of onset of learning	*M*	5.74	5.53	6.14	.431	.019
	SD	2.08	1.75	1.62		
L2 instruction (years)	*M*	10.26	10.47	9.86	.453	.015
	SD	2.08	1.75	1.62		
L2 academy exposure (h/week)	*M*	1.46	1.31	1.14	.604	.011
	SD	1.25	1.26	1.19		
L2 use NNS (h/week)	*M*	4.13	2.81	2.69	.073	.057
	SD	3.69	2.33	1.53		

L2 use NS (hours per week)	*M*	1.56	1.19	1.10	.512	.015
	SD	1.80	1.71	1.37		
L2 self-estimated proficiency (1-9) [a]	*M*	6.04	5.98	5.68	.659	.009
	SD	1.71	1.53	1.68		
L2 proficiency [b] **(0-120)**	*M*	72.28	70.13	71.68	.918	.002
	SD	23.94	15.77	21.18		

[a] Averaged self-estimated ability to speak spontaneously, understand, read, write and pronounce English (1=very poor - 9=native-like)
[b] Obtained through an elicited imitation task (Ortega et al. (2002)
Note. NNS: non-native speakers, NS= native speakers, M= mean, SD= standard deviation, *p* = significance ($p < .05$)

3.2. Experimental design and timeline

In the current section, the study experimental design and timeline are summarized. To understand the procedure, the intervention and testing tasks are shortly introduced.

The main question addressed in the current study is whether a focus on phonetic form during communicative tasks improves learners' L2 vowel perception, lexical encoding and production, and whether this may depend on the cognitive complexity of the task. The two experimental groups (SG & CG) carried out a pre-test, a 7-week treatment, a post-test and a delayed post-test (11 weeks after the treatment), in order to check the amount of retention after 20 sessions of task-based pronunciation instruction. The control group received the same EFL instruction as the experimental groups at school[37], but did not participate in any task-based intervention between pre- and post-test. Examining how the control (untaught) group (CTG) performed is assumed to reveal: (1) test–retest effects (i.e. learners might have benefited because the same test was used twice) and (2) any kind of learners' improvement due to their L2 experience outside the classroom. Taken together, having a CTG would help evaluate the unique effect of pronunciation-focused tasks on L2 learners' pronunciation. For ethical reasons, CTG

[37] The three intact classes followed the same syllabus, shared the same coursebook and were instructed on the same contents.

learners carried out 10 simple and 10 complex tasks immediately after the post-test in an intensive way, that is, they completed the project in 4 weeks. In addition, these students were not informed of the existence of experimental groups in order to avoid compromising the internal validity of the experiment. Participants might have been strongly demotivated to participate if they knew that they were in a control group.

Pre- (T1), post- (T2) and delayed post-tests (T3) were identical and consisted of a battery of perceptual, lexical encoding and production tests. The ID tests (i.e. L2 proficiency, WM, ASA) were distributed along the three testing times as they were independent to the treatment. Learners were exposed to unfamiliar voices (M1, F1) in the testing in order to ensure generalization to new talkers and they were also tested on untaught words to examine whether their knowledge of the target vowels was extended to new words. These untaught tokens were similar to the taught tokens in terms of word category, consonantal context (place of articulation, voicing) and most were unknown.

Testing included perception (ABX categorical discrimination); lexical encoding (Forced Lexical Choice (FLeC) and Lexical Decision (LD)); production (Delayed Word Repetition (DWR) and Delayed Sentence Repetition (DSR)); and an oral interactive task, and happened before and after a 7-week (20 session x group) task-based treatment. Learners' level of oral proficiency (EIT), WM (forward and backward-digit span) and attention control (ASA) were tested once during the testing period (Table 3.2.). A language background questionnaire and a word familiarity questionnaire were administered online before the first day of intervention. After performing the tasks, the two experimental groups were asked to fill in a post-intervention questionnaire to assess their perception of task complexity and their general impression about the pronunciation-focused task-based intervention.

Each intervention session occupied 30-40 minutes of their regular English classes and was carried out in their EFL class altogether. In contrast, the 45 testing sessions (15 x 3 groups) involved 60 minutes (one entire English class) and took place in an isolated classroom.

Table 3.2. An overview of the pre-test/post-test/delayed post-test experimental design with three groups of participants: simple group (SG), complex group (CG) and control group (CTG).

	Pre-test	**Intervention**	**Post-test**	**Delayed post-test**
SG/CG	OI	Pre-tasks	OI	OI
	ABX		ABX	ABX
	FLeC	Simple tasks (SG)	FLeC	FLeC
	LD	Complex tasks (CG)	LD	LD
	DWR		DWR	DWR
	DSR	Post-tasks	DSR	DSR
	EI		ASA	FBDS
CTG	OI		OI	
	ABX		ABX	
	FLeC		FLeC	
	LD		LD	
	DWR		DWR	
	DSR		DSR	
	EI		ASA & FBDS	

Note. OI: Oral Interaction; ABX: ABX discrimination; FLeC: Forced Lexical Choice; LD: Lexical Decision; DWR: Delayed Word Repetition; DSR: Delayed Sentence Repetition; EI: Elicited Imitation; ASA: Auditory Selective Attention; FBDS: Forward and Backward-Digit Span

The present investigation comprised a total duration of 6 months. It started on 16 - 17 September 2019 when the author of this dissertation presented the "Our trip to London" project to the three intact classes. Two days later, the SG and CG students started doing the pre-test. The CTG group started a couple of weeks later, as their English classes coincided with the SG and CG's, thus they could not be tested at the same time. By mid-October 2019, all groups had been pre-tested. The G1 and G2 groups carried out the first interactive task in class on 2 - 3 October 2019, respectively, and the last one on 14-15 November 2019. The control group did the 20 tasks in an intensive way (4 weeks) in January 2020, after they had completed the post-test. The post-test for the 3 groups took place from mid-November until mid-December 2019. Lastly, only the SC & CG groups were tested for L2 vowel retention in the delayed post-test during the month of February 2020. (See timeline by group in Table 3.3.)

Table 3.3. Timeline of the study by group: simple group (light grey), complex group (dark grey), control group (black).

2019-2020	**1st week**	**2nd week**	**3rd week**	**4th week**	**5th week**
September			Pre-test		
			Pre-test		
October	Pre-test				
	Simple tasks				
	Complex tasks				
November	*Simple tasks*		Post-test		
	Complex tasks		Post-test		
December		Post-test			
January		*Simple/Complex tasks*			
February	Delayed post-test				
	Delayed post-test				

3.3. Materials

In this section, the materials used in the study are described. It is divided into four main subsections: First, we present the stimuli of the study, specifically, its elicitation and preparation, selection and validation, the intervention stimuli and the testing stimuli. Second, the intervention materials and procedures —pre-tasks, tasks and post-tasks— follow. Third, we describe the testing materials comprising the tasks which measure perceptual discrimination, lexical encoding, vowel production, oral interaction, target word knowledge, L2 proficiency, WM and attention control. The last subsection details three questionnaires related to learners' demographic and linguistic background, word familiarity and post-intervention perceptions.

3.3.1. Stimuli

3.3.1.1. Elicitation and preparation

A total of 3 male (M1, M2, M3) and 3 female (F1, F2, F3) native speakers of British English were recruited to record the intervention and testing stimuli, which, in turn, served as baseline data for native vs. non-native vowel production comparisons. Native speakers were asked to record words and nonwords twice; sentences once; and dialogues once, with additional repetitions when necessary. The recordings were conducted by the author of this dissertation at the *L2 Speech* laboratory. Speakers were recruited through email and were asked to fill in a language background questionnaire after being recorded (See Appendix I). They were recorded within a period of a week and received a monetary compensation for their participation. They all reported having normal vision and no speech/hearing pathologies.

The native speakers' mean age was 32 years old (*SD* = 2.09) and they were all from the United Kingdom, specifically, from the South of England, the Midlands and one from the North of England. All of them had been largely exposed to Received Pronunciation by teaching with EFL materials and were instructed to use this variety of English when recording the target sounds. At the time of the stimuli recording, five of them lived in Barcelona and one of them lived in Murcia (Spain).

All native speakers had a high level of formal education (Table 3.4.) and had been teaching English at all levels for over 10 years. F1 was also the academic coordinator of higher education and F3 was a teacher trainer. In addition, they all had enrolled courses about Languages, Linguistics, Applied Linguistics or Second Language Acquisition and had also studied English Phonetics and Phonology.

The six native speakers spoke English (L1) with their family; however, they spoke an average of 67.7% of English and 33.3% of Spanish daily. They had learned their L2 when they were 21.17 years old (*SD*=2.78) and had learned between 3 and 5 foreign languages. The foreign languages comprised L3s (Catalan, Spanish, Italian & French), L4s (French & German) and L5s (Catalan & Japanese). See a summary of the speakers' demographic and linguistic data in Table 3.4.

Table 3.4. Demographic and linguistic data of the native speakers who created the stimuli and provided baseline data for production analyses.

Speaker	**Age**	**Place of Birth (UK)**	**Education**	**L1, % use**	**L2, % use**	**L2 AOL**
M1	32	Oxford	BA + MA, CELTA	English, 60	Spanish, 40	22
M2	31	Farnborough	BA + MA	English, 80	Spanish, 20	25
M3	30	Burnley	CELTA	English, 60	Spanish, 40	18
F1	31	Rugby	BA + MA+ PhD, TESOL	English, 60	Catalan, 40	21
F2	32	Swindon	BA + MA, CELTA	English, 60	Spanish, 40	18
F3	36	Swansea	BA + MA (TESOL), CELTA	English, 80	Spanish, 20	23

Note. AOL=Age of Onset of Learning

Recordings took place in a soundproof booth in the *L2 Speech* laboratory at the University of Barcelona. The speech samples were obtained through a Marantz PMD-661 solid-state digital recorder with an external Shure SM58 voice microphone and were digitized at a 44.1 kHz sampling frequency and 16-bit quantification. Word stimuli were embedded in carrier phrases and displayed to speakers by means of a printed reading list.

In order to ensure the desired pronunciation of real words and nonwords, a rhyming carrier sentence was used, which contained a real word that rhymed with the target word/nonword, e.g. *It rhymes with "real word". I say "word/nonword"; I say "word/nonword" again*. They were instructed to read the words carefully at normal speed and on a falling intonation. The word list reading was preceded by a short practice section, consisting of a few practice items, and the researcher verified the recorded stimuli to ensure that the stimuli had been produced as intended.

The real word sentences were recorded without any carrier sentence as they contained the words which they had previously uttered. The sentences in the dialogues were recorded by the same talker with a natural pace, reflecting natural conversations. Recordings took around 60 minutes per speaker and all instances were closely monitored by the researcher in order to guarantee the desired pronunciation of each target sound. In a few instances, speakers were asked to repeat mispronounced words and nonwords.

So as to obtain the stimuli, each recording was annotated using Praat (Boersma, & Weenink, 2020). The words, nonwords and isolated sentences were segmented, extracted and saved as *wav* files. The sentences from the dialogues were also segmented and extracted, but then, they were alternated between two speakers, concatenated adding natural pauses and saved as *wav* files in the form of a dialogue. All resulting stimuli were intensity-normalized to 50 decibels.

3.3.1.2. Selection and validation

When more than two instances of each item were recorded by the aforementioned NS, a stimuli selection took place. We only selected the best tokens —in relation to clarity of articulation and absence of creakiness— on the basis of auditory judgements and spectrographic analyses in Praat (Boersma & Weenink, 2020).

Three different native British English speakers (2 females, 1 male) performed the perception tasks appearing at pre- and post-test (i.e., ABX, FLeC and LD) in order to validate the selected stimuli and provide a native speaker score as baseline data (see Table 3.5. for native speakers' demographic and linguistic information). Native speakers obtained an accuracy score of 97% in the ABX discrimination task, 98% of accuracy in the FLeC task and 95% in the LD task, and reported through a follow-up interview that the stimuli were clear and the tasks posed no difficulty in terms of understanding. They mentioned that a percentage of error may have been due to their inaccuracy in hitting one of the keys accidently. Therefore, the stimuli were considered appropriate for both perception and production tasks.

Table 3.5. Demographic and linguistic data of the native speakers who validated the stimuli in perception tasks.

Speaker	**Age**	**Place of Birth (UK)**	**Education**	**L1, % use**	**L2, % use**	**L2 AOL**
NS1	25	Cambridge	CELTA	English, 80	Spanish, 20	20
NS2	34	Bristol	CELTA	English, 90	Spanish, 10	16
NS3	30	Plymouth	CELTA	English, 70	Spanish, 30	18

Note. NS =Native Speakers, AOL=Age of Onset of Learning.

3.3.1.3. Intervention stimuli

The stimuli chosen for the task-based intervention comprised a total of 80 different consonant-vowel-consonant (CVC) words, which were carefully selected to fit in the project “Our trip to London”. Most of the target words referred to objects, food and animals, therefore, 78.75% were nouns, 11.25% were adjectives and 10% were verbs (Appendix C).

Twelve minimal pairs contained the English vowel contrast /iː-ɪ/ and twelve the English vowel contrast /æ-ʌ/. The 32 words left, which were not minimal pairs but were used during task performance, also contained the same vowels: /iː/ (8), /ɪ/ (8), /æ/ (8), /ʌ/ (8). These L2 English vowels constitute a source of confusion in perception and production for Catalan/Spanish EFL learners (Section 1.2.4. and Section 2.3). Given that phonological representations may be influenced by the orthographic input (Hayes-Harb et al., 2018; Charoy & Samuel, 2020), the effect of orthography was minimized by including words whose stressed vowel consistently matched the same letters (“i” /ɪ/, “a” /æ/, “u” /ʌ/). However, in order to provide sufficient realistic input for the performance of the task, various spellings of /iː/ had to be necessarily included, namely, “ee, ea, ei, i, e”. Words containing opaque spellings (e.g., bl**oo**d /bl**ʌ**d/) were *not* included.

Considering the Zipf-scale frequency measures provided by Subtlex-UK (Van Heuven et al., 2014), the stimuli were balanced in terms of high and low frequencies, and were distributed similarly across the four target vowels. Word familiarity or recency of word use (Darcy & Holliday, 2019) could not be initially predicted or controlled for.

Table 3.6. Intervention stimuli features.

Item	Word Type	Category	Contrast	Vowel	Syllable
Words (80) Nonwords (0)	MP words (48) Extra words (32)	Nouns (63) Adjectives (9) Verbs (8)	iː-ɪ (24) æ-ʌ (24)	iː (20) ɪ (20) æ (20) ʌ (20)	1 syllable (40) 2 syllables (40)

# PoA	# Voicing	PoA #	Voicing #	Knowledge
Bilabial (25) Labiodental (3) Alveolar (21) Palatal (10) Velar (12) Glottal (4)	Voiced (41) Unvoiced (34)	Bilabial (23) Labiodental (3) Alveolar (40) Palatal (3) Velar (11) Glottal (0)	Voiced (43) Unvoiced (37)	Unknown (33) Known (39)

Note. Brackets indicate number of tokens. MP= Minimal pair, Extra words= words which were not minimal pairs but contained the 4 target vowels, PoA= Place of articulation of the preceding or following consonant (#_beginning, _# end)

Half of the minimal pairs and extra words were monosyllabic and half were disyllabic and were preceded by bilabial, labiodental, alveolar, palatal, velar or glottal consonants. All words had a mean length of 4.77 letters (*SD*= 1.45, *Range* = 3-8). Learners knew more than half of the words presented in the intervention and were exposed to 33 unknown words, as revealed by their responses in the vocabulary knowledge scale[38]. See a summary of stimuli features in Table 3.6.

All tokens were uttered by two male speakers (M2, M3) and two female speakers (F2, F3) of British English. They were the input in the pre-task focus-on-form activities, and also in the dialogues of the pre-task listening exercises, prior to the task cycle.

3.3.1.4. Testing stimuli

The testing corpus consisted of a total of 168 CVC tokens, corresponding to 16 practice items, 16 control/filler items and 136 test items, which were perfectly balanced in terms of one and two syllables (Appendix D). Out of 168 items, 128 were words (73.43% nouns, 10.15% adjectives and 16.40% verbs) and 40 were nonwords (used in FLeC and LD tasks). In the word group, 42 minimal pairs contained the L2 vowel contrast /iː-ɪ/ and 42 minimal pairs the L2 vowel contrast /æ-ʌ/. Furthermore, 76 were other words that appeared in the intervention and testing but did not have a minimal pair and only 32 appeared in the FLeC and LD tasks. These were used to create nonwords by changing the vowel by its L2 contrastive counterpart (e.g. "kiwi" /ki:wi/>*/kɪwi/ or "jam" /ʤæm/>*/ʤʌm/) respecting the same consonantal environment. During the testing phase, learners were exposed to 80 taught words and 88 untaught words/nonwords, thus,

[38] Proper nouns (Peter, Sheila, Patrick, Cathy, Jimmy, Lily, Luster, Sunset) are excluded from this count.

tokens which had not appeared during the intervention (including 4 practice + 4 control/filler items).

The monosyllabic and disyllabic untaught minimal-pairs (excluding the nonwords) contained the target vowels surrounded by a large variety of consonantal contexts. The 1-syllable practice items (in the ABX discrimination and DWR/DSR tasks) contained the four target vowels /iː- ɪ - æ - ʌ/ surrounded by similar environments (/f_t/, /g_n/, /h_l/, /m_t/, /r_t/, /s_n/) to the test items. The 1-syllable practice trials (in the FLeC and LD tasks) had four consonant contrasts (i.e. /p/-/k/, /b/-/t/, /n/-/m/, /w/-/j/) and corresponded to the words/nonwords *pass/*kass, bird/*tird, neck/*meck, work/*yerk*. Lastly, the control minimal-pairs of the ABX discrimination tasks contained the vowel contrasts /iː-æ/ - /ɪ-ʌ/. The filler tokens of the FLeC and LD tasks consisted of words/nonwords containing the vowel contrasts /e/-/uː/, /ɔː/-/ɑː/ and were included in the words *pen, merry, horse, morning*, and the nonwords **poon, * murry, *hars, *marning*.

Overall, words and nonwords had a mean length of 4.72 letters (*SD*=1.42, *Range* = [3-8]) and there were more unknown words (95) than known words (61) because most of the untaught words (which did not appear during the intervention) had lower lexical frequencies than taught words and were unknown, as reported through the vocabulary knowledge scale questionnaire. The target words in the DSR task were the same as in the DWR task and can be read in Appendix D. See a summary of stimuli features in Table 3.7.

Finally, all tokens in the testing phase were uttered by a male (M1) and a female (F1) native speaker, who were different from the ones used during the intervention as the purpose was to examine if learners were able to generalize their knowledge of the target vowels to new talkers (Carlet & Cebrian, 2019). The stimuli in the pre-test, post-test and delayed post-test were identical.

Table 3.7. Testing stimuli features.

Item Type	Item	Word Type	Teaching	Category	Contrast	Vowel
Practice (16)	Words (128)	MP words (84)	Taught (80)	Nouns (94)	iː-ɪ (42)	iː (38)
Control (16)	Nonwords (40)	Extra words (84)	Untaught (88)	Adjectives (13)	æ-ʌ (42)	ɪ (38)
Test (136)				Verbs (21)	iː-æ (4)	æ (38)
					ɪ-ʌ (4)	ʌ (38)
					e-uː (4)	
					ɔː-ɑː (4)	
					p-k (1)	e (2)
					b-t (1)	uː (2)
					n-m (1)	ɔː (2)
					w-j (1)	ɑː (2)

Syllable	# PoA	# Voicing	PoA #	Voicing #	Knowledge
1 syllable (92)	Bilabial (47)	Voiced (86)	Bilabial (34)	Voiced (88)	Unknown (95)
2 syllables (76)	Labiodental (9)	Unvoiced (74)	Labiodental (6)	Unvoiced (80)	Known (61)
	Alveolar (52)		Alveolar (95)		
	Palatal (21)		Palatal (6)		
	Velar (23)		Velar (27)		
	Glottal (8)		Glottal (0)		

Note. Brackets indicate number of tokens. MP= Minimal pair, Extra words= words which were not minimal pairs, PoA= Place of articulation of the preceding or following consonant (#_ beginning, _# end).

3.3.2. Intervention and procedure

During 7 full weeks, participants were immersed in a project called “Our trip to London” which involved an end-of-the-course trip to the capital of England to practice their English and do multiple cultural and entertainment activities. The target words contained either the challenging L2 vowel contrast /iː- ɪ/ or /æ - ʌ/ and these were alternated across the 20 tasks (i.e., task 1 (/iː- ɪ/), task 2 (/æ - ʌ/), task 3 (/iː- ɪ/), task 4 (/æ - ʌ/), etc.). Every task contained 6 monosyllabic words (3 minimal pairs), 6 disyllabic words (3 minimal pairs) and 8 extra words which were not minimal pairs but contained the target L2 vowels (/iː, ɪ, æ, ʌ/). Each target word was repeated 5 times throughout the 20 tasks, as it has been found that repeating the same content of tasks, as well as the same procedure, helps consolidate pronunciation (Darcy et al., 2019; Jung et al., 2017).

The tasks were organized in a sequential manner and generated a logical narrative: the first ones dealt with the planning, the preparation and arrival in London and the last tasks corresponded to farewell activities, the return to the school and the creation of their own photo album. The project presented a scenario in which learners would spend 14 days in London but had been involved in the trip for around 20 days (4 preparation days + 16 in London + 2 back home), corresponding to 20 task-based classes. Having set this realistic context, they performed 20 dyadic problem-solving tasks which were always preceded by a pre-task and followed by a post-task. In the pre-task, students learned the meaning and were exposed to the pronunciation of the words uttered by native-speakers of English in isolation and in context. In the post-task, students consolidated the language which they had practiced during the task cycle through meta-communicative activities.

In order to foster students' motivation, learners were allowed to choose the classmate they wanted to work with, and every 5 tasks, they changed pairs to ensure that weaker students could benefit by listening to what more advanced students said, and the advanced ones could also improve by having to paraphrase and make themselves understood in front of weaker ones. In every class, the author of this dissertation, who adopted the role of the main EFL teacher, showed a Power Point presentation which encompassed the pre-task, task cycle and post-task activities around the same topic. I will now present each task and student-teacher roles through the adapted Task-Based Learning implementation framework (Willis, 1996; Figure 3.1.) for pronunciation, where there is a natural progression from the holistic to the specific.

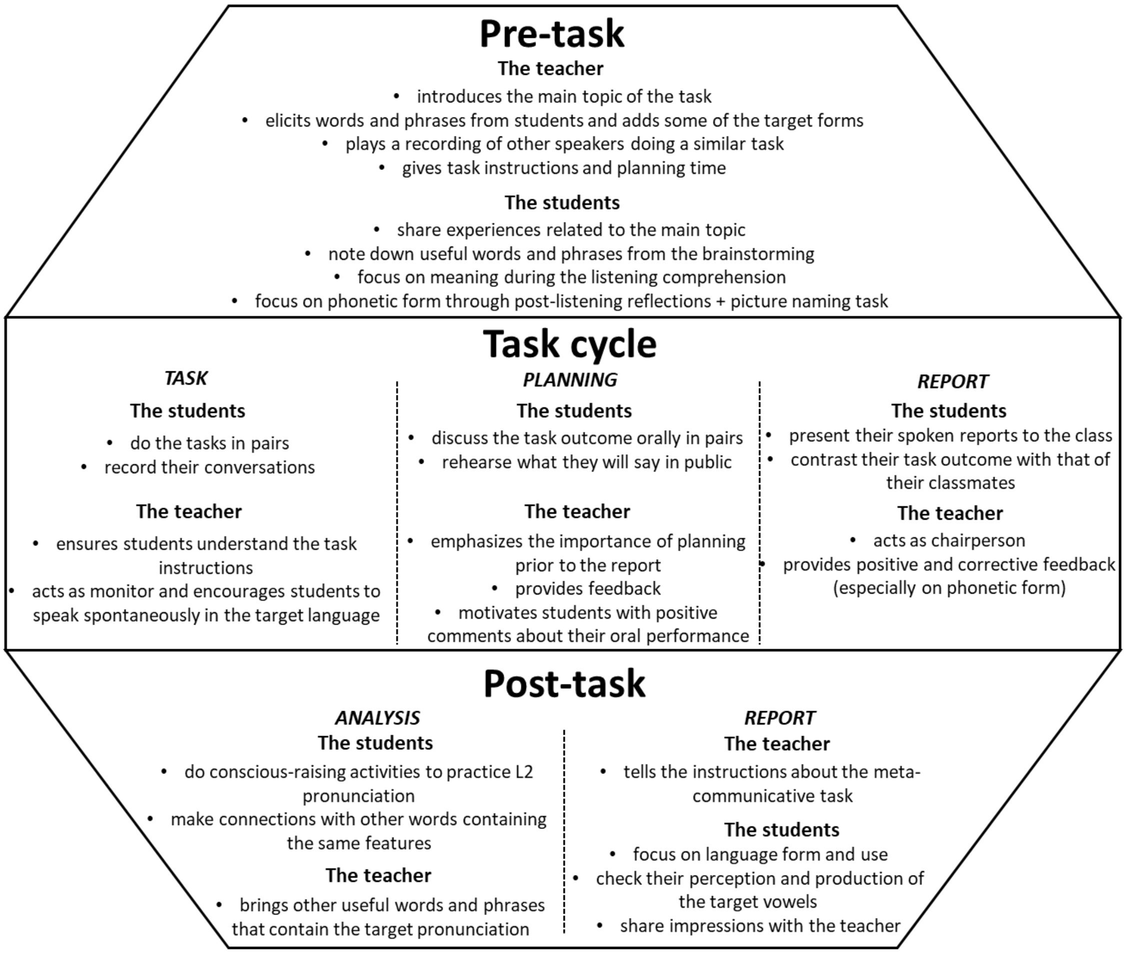


Figure 3.1. Overview of the study's pre-task, task cycle and post-task adapted from the Task-Based Learning framework (Willis, 1996) to focus on L2 pronunciation.

The study's intervention was designed around Willis' (1996) task design and methodological implementation framework of TBLT and took influences from Pica's (1984) approach because representative examples of pronunciation targets were introduced, practiced through interactive problem-solving tasks and task outcomes were exposed in the final oral report. Although the methodology used was not sequenced from controlled to spontaneous pronunciation practice (e.g., Celce-Murcia et al.'s 2010 Communicative Framework), learners took part in listening exercises which helped them discriminate minimal-pair words, and attended to both phonetic form and meaning through a picture-naming task in the pre-task phase. We opted for a more naturalistic approach were learners were exposed to and could use L2 phonetic forms incidentally in meaning-driven interactions[39]. It was in the post-task that learners' pronunciation and meaning of the words was consolidated with more meta-linguistic and communicative activities.[40]

3.3.2.1. Pre-tasks

Pre-tasks were used to introduce the class to the topic and the main task, activate topic-related words and phrases, expose them to a model of conversation and the target words, introduce the task cycle and give them planning time. Pre-tasks took place during the first *10 minutes* of class time.

To start with, the author of this dissertation —as *the main EFL teacher (**T**)*— presented the topic area of the session by showing a full-screen photo of the real-life activity (e.g., someone packing his/her clothes before going on a trip) and justified the selection of this task in the project sequence. The topics were familiar to most students and were well-integrated in the project. Then, they were asked about their experience doing that particular activity and encouraged learners to pool topic-related words and phrases they already knew. During this brainstorming activity, T helped students recall and activate words and phrases that were going to be useful during the task cycle. As T was writing them on the blackboard in the form of mind maps, T read them aloud so they

[39] In task-based interventions, incidental learning would entail brief unintentional but conscious focus on form in a meaning-driven context (Long, 1985). However, the cognitive-interactionist theory of instructed SLA underpinning TBLT embraces not just incidental focus on form, but a necessary, principled, symbiotic relationship between incidental and intentional learning (Long, 2015, 2016).

[40] The 20 tasks used in the intervention can be found here: https://drive.google.com/drive/folders/1npMcVKUovQ1iy39NsMkw1QtCyY5WJUfK?usp=sharing

were exposed to the pronunciation of the words. Once they had come up with several words and expressions, T gave out the handouts for the following activity, where learners were exposed to some of the already mentioned words in a realistic scenario.

The second task consisted of listening to a recording of a male and a female native speaker of English having a conversation which replicated the task each dyad was going to carry out during the task cycle (Appendix F). Providing “models” for learners to follow (Ellis et al., 2019) was useful for several reasons:

1) Learners were motivated to listen, as they knew they had to perform a similar task in the next phase and could rely on a model.

2) Learners understood what the task goals were and they had a fairly clear idea of the kind of meanings that might be expressed.

3) Learners learned from comprehensible input and how speakers negotiated meaning, sustained the interaction and reached a solution.

4) Learners were exposed to the pronunciation of the target words through different male and female voices and were able to clarify meanings during natural talk.

The recording was played twice. The first time, learners were asked to listen to the dialogue for overall comprehension; the second time, attention was drawn to the minimal pairs which were made task essential (e.g., *cap/cup, batter/butter*). Each listening activity was different in terms of outcome, so they were asked to circle, cross out, fill in the missing letters of words, complete the dialogue, among others (Appendix E.1.). Afterwards, students compared their answers with their pairs and T and students discussed the correct outcome, which was also shown on screen through a Power Point presentation (Figure 3.2.).

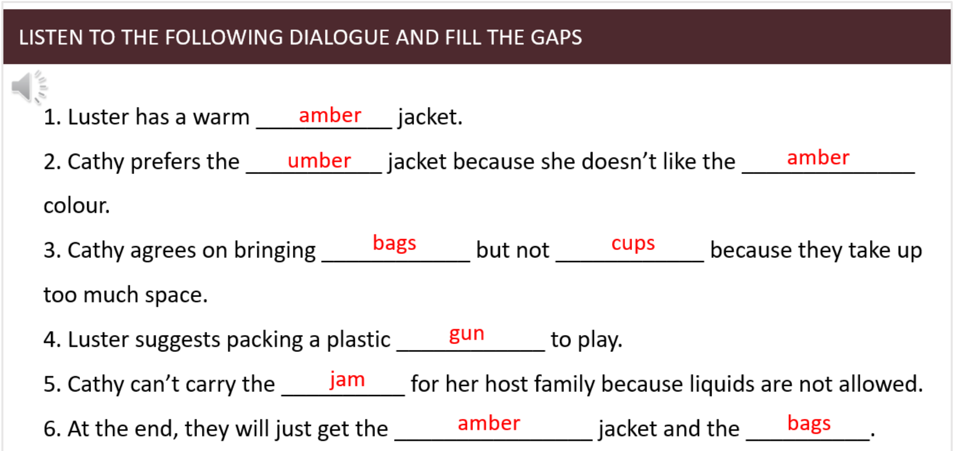


Figure 3.2. Example of the pre-task listening comprehension with the key in red.

Having had a first experience with the target words, the last task of the pre-task phase required a more controlled exposure and production of the target words. In order to do so, T showed an illustration of the target object on screen and elicited the word. As a form of feedback, they were asked to listen to the pronunciation of the word uttered by a male/female native speaker of English (different from the ones heard in the listening comprehension task) and produce it afterwards. The same procedure was applied to the 20 tasks but, so as for students not to become very repetitive and monotonous, the way of eliciting the target words was different from task to task. For instance, T asked individual students, rows of students or half of the group or T asked for volunteers to participate and point at other students to join. Overall, this activity actively involved all learners, gave them relevant exposure and, above all, created interest in doing a task on this topic.

Finally, T gave clear instructions about the communicative task, showing an example of the task performance and setting an approximate time for completion. Then, T handed out a "task pack" which contained the written instructions of the task, the conditions and target items (in the format of flashcards) for student A and student B, and a game board which the pair shared (Figure 3.3.)

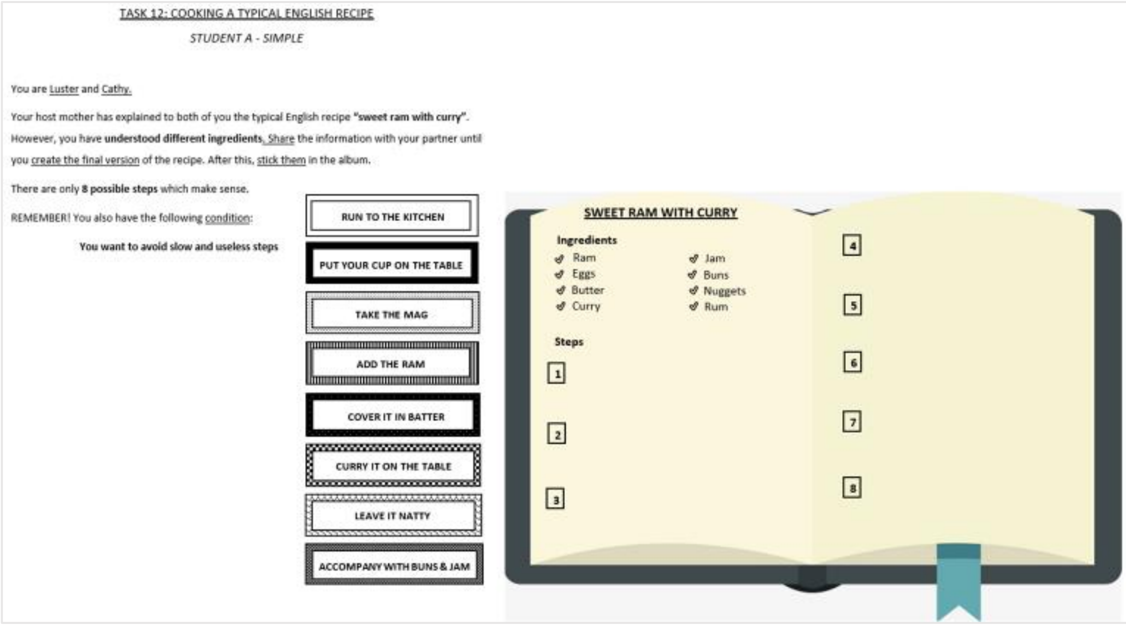


Figure 3.3. Example of the "task pack" for Student A performing the simple version of *the recipe* task.

T asked them to read the written instructions by themselves again and ask any questions that they had. They were given around 2-minute planning time to look at their cards and conditions and think about the language they were going to use to perform their tasks. In the case of a communication breakdown, learners were encouraged to exaggerate their pronunciation.

3.3.2.2. Tasks

The task cycle offered learners the chance to use the linguistic resources they had recently been exposed to carry out the task, and then improve the target phonological form under teacher guidance while planning their reports of the task (*~20 minutes*). As the task cycle was related to "a holistic experience of language in use" (Willis, 1996, p. 40), learners were hypothesized to have increased their L2 fluency and confidence in themselves as communicators, and used the available linguistic resources to solve the task.

Despite the quality of L2 input from their peers not being the richest, learners were exposed to the phonological and lexical form of the words in the pre-task, and the task cycle was a vital opportunity for learners to use those L2 phonological forms spontaneously to achieve the goals of the task. If inaccuracies were produced, learners had the chance to revise and consolidate the target language during the post-task. Overall, the task cycle had a duration of 15 to 20 minutes.

Adopting the definitions by Pica et al. (1993), the twenty problem-solving tasks were two-way, split, close and convergent because the two interlocutors always had different information which they had to share in order to reach one possible solution. In addition, learners were not able to solve the task if they did not produce the L2 phonological contrasts (/iː- ɪ/ or /æ - ʌ/) distinctively, as the target words were made task-essential (Loschky & Bley-Vroman, 1993). In other words, success depended on being intelligible to their partners, hence, on producing distinguishable L2 minimal pairs. Despite performing these tasks in a classroom setting, these communicative tasks mirrored the kind of activities that these students would be doing in London the following semester.

All tasks involved two mental operations: information-sharing and decision-making but simple tasks differed from complex tasks in terms of cognitive complexity (Robinson, 2001a, 2005, 2011b). Along resource-directing dimensions, complex decision-making tasks were manipulated by increasing task complexity through ± reasoning demands, but maintaining the same number of lexical elements. Reasoning demands were understood as the task component which made learners reason about certain actions and justify their choices (Robinson, 2007b). CG tasks contained more conditions (6 condition per dyad) than SG tasks (3 condition per dyad) that learners had to respect prior to reaching a solution (Appendix E.2.).

The perceived cognitive demands posed by the tasks (i.e., mental effort and difficulty) were independently assessed on 7-point Likert-scales (*1=very low mental effort/extremely easy; 7=very high mental effort/extremely difficult*) by 10 experienced EFL language teachers (Révész et al., 2016; Révész & Gurzynski-Weiss, 2016), the 92 experimental learners from the present study (as also did Gilabert et al., 2009; Révész, 2011; Révész et al., 2016, 2022; Robinson 2007b) and a group of pilot learners (N = 26) that shared the same demographic and linguistic profile as the experimental group. Following Révész et al.'s (2016) recommendation for the independent validation of task complexity, EFL teachers and pilot learners were asked to critically evaluate a subset of tasks for complexity prior to this study's intervention. In contrast, the experimental learners evaluated them all after completing the 20 tasks from this study, given the little time available at the end of each task cycle.

Table 3.8. Self-perceptions of mental effort and difficulty in simple and complex-decision making tasks by EFL teachers, pilot and experimental learners.

	Simple decision-making tasks			Complex decision-making tasks				
	M	*SD*	*95% CI*	*M*	*SD*	*95% CI*	*p*	ηp^2
MENTAL EFFORT								
EFL teachers	2.60	0.94	[2.16-3.04]	5.65	1.04	[5.16-6.14]	<.001	.71
Pilot learners	3.36	1.12	[2.61-4.12]	5.09	1.04	[4.39-5.79]	.001	.41
Experimental learners	2.94	1.18	[2.50-3.37]	5.06	1.24	[4.61-4.51]	<.001	.44
DIFFICULTY								
EFL teachers	2.70	0.86	[2.30-3.10]	5.75	0.91	[5.32-6.18]	<.001	.75
Pilot learners	3.55	1.44	[2.58-4.51]	5.27	1.01	[4.59-5.95]	.004	.34
Experimental learners	3.03	1.19	[2.59-3.47]	5.16	0.95	[4.81-5.50]	<.001	.49

One-way ANOVAs showed that language teachers, pilot learners and experimental learners perceived complex tasks to involve significantly more mental effort and be more difficult than simple tasks (Table 3.8.). Resource-dispersing dimensions (i.e., planning time, prior knowledge of the task and target items) were controlled for during the task cycle.

While learners were performing the communicative tasks, their regular EFL teacher and T monitored the conversations from the distance, only engaging in teacher-student language-related episodes in case of communication breakdown. As it was a student-led phase, we tried not to stand too close to the groups so they could freely engage in the conversation and:

- Make sure that all pairs were doing the right task and were clear about the objectives.
- Encourage students to take part although they were not very proficient.
- Avoid correcting errors of form unless they caused an important communication breakdown.
- Observe if any groups switched to their mother tongue repeatedly.
- Identify which students seemed to talk more and give support to those who struggled to speak.
- Control the time and give 1-minute warning before the end of the task.

Once all the groups had finished performing the tasks, T briefly commented on one or two points of interest which T had taken notes about while monitoring, and provided general feedback about learners' task performance. Finally, each pair of learners was asked to send T via WhatsApp Messenger or e-mail the audio file of the task which each pair had recorded with one of their mobile phones.

In the planning stage, dyads were asked to orally discuss the outcome of the task and rehearse what they would say in front of the class. T highlighted the importance of the planning stage before the oral presentation of the report. In addition, T helped them express what they wanted to say, providing feedback on the meaning and pronunciation of the words if needed.[41] T was always very positive about learners' use of the target forms and creative use of the language and motivated them before the report.

Lastly, in the report stage, some pairs presented the outcome of the task to their classmates orally and contrasted their results with the rest of the class. According to Willis (1996), the report enhanced the process of planning, drafting and rehearsing. T acted as a chairperson to organize the oral presentations, gave some feedback on content and mainly on the phonetic form of the words, and emphasized the great improvement they were making after each session.

3.3.2.3. Post-tasks

The last phase in the framework was the post-task/language focus, which lasted between *5 and 10 minutes*. At this point, learners had already worked with the language and processed its meaning and pronunciation, so they were ready to consolidate it through an activity that required a focus on phonetic form. According to Willis and Willis (2007), this practice helped learners make sense of the language they had experienced; generalize the pronunciation of the target words to other words inside and outside the class; and build on their motivation and self-confidence when speaking a foreign language. In case of time constraints, the analysis and practice stages were naturally combined in a single activity.

In the analysis stage, learners examined and discussed the meaning and pronunciation of the target words needed during the task cycle through conscious-raising

[41] During the task cycle, which was a purely learner-centred phase, individualized feedback could not be provided. It was only during the planning and report stages that, globally, feedback (mainly in the form of recasts) was provided to help students express themselves more confidently. Still, it was mainly during the post-task phase that learners could consolidate the meaning and pronunciation of the target words.

activities and we compared them to other words which sounded in a similar way. T hung on the blackboard a flashcard containing the phonetic transcription of the two L2 vowels that they had been practicing during that session, together with a drawing of the mouth pronouncing the vowels and a couple of examples, so that we could use it as a complement to the conscious-raising activities if needed.

Afterwards, in the practice stage, learners carried out a meta-communicative task which focused on language form and use (Figure 3.4.). Learners were able to check if they could accurately perceive and produce the L2 vowel contrasts (/iː-ɪ/ or /æ-ʌ/) accurately. These tasks always included the target items used during the task cycle and ranged from dictations and creating dialogues to drawing objects, inventing a rap or playing Bingo (Appendix E.3.). At the end of the class, learners were encouraged to share their feelings of like and dislike and impressions concerning the task.

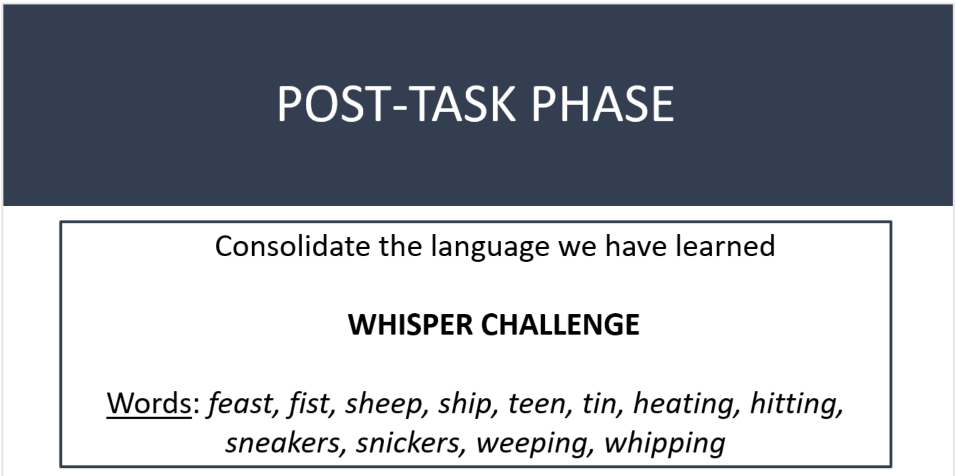


Figure 3.4. Example of a post-task activity to consolidate the phonetic form and meaning of words containing the /iː- ɪ/ contrast.

3.3.3. Testing and procedure

Before and after the 7-week task-based treatment, we assessed learners' improvement in the perception, lexical encoding and production of the L2 target vowel contrasts (/iː- ɪ/ or

/æ - ʌ/) through individual computer-based tests which could last from 3 up to 12 minutes. Prior to the intervention, all tests were previously piloted with a similar group of learners.

Learners were equipped with Beyerdynamic DT-770 closed (noise-cancelling) headphones for the perceptual discrimination task, two lexical encoding tasks differing in difficulty, the ASA and the WM task. The stimuli of the delayed word repetition task, the delayed sentence repetition task and the L2 oral proficiency task were administered through Beyerdynamic DT-990 open headphones and the participants' speech was recorded through Marantz PMD-661 mkii and Marantz PMD-661 mkiii solid-state digital recorders with external Shure SM58 voice microphones, and was digitized at a 44.1 kHz sampling frequency and 16-bit quantification. Lastly, learners' interactive task was recorded through the researcher's mobile phone and 3 Tascam DR-05X recorders.

3.3.3.1. Perceptual discrimination

In order to assess learners' perception of L2 vowel contrasts at a pre-lexical level, a speeded categorical ABX discrimination test (Melnik-Leroy et al., 2022; Mora & Mora-Plaza, 2019; Tyler et al., 2014) was administered via the *DMDX* software (Forster & Forster, 2003) on a laptop computer (*Total task duration*=8-9 minutes). Trials were created by combining words into ABX triads with a 500ms inter-stimulus-interval (e.g. A=*cap*-B=*cup*-X=*cap*). Participants were instructed to decide, as accurately and as fast as they could, whether the last word in the triad contained the same stressed vowel as the first (A) or the second word (B) by selecting a key labelled as A (Alt) or B (Alt gr) on the computer keyboard.

A, B and X were produced by two speakers (M1, F1) who had not appeared during the intervention to ensure that participants made a decision based on the phonological categorization of the stimuli while disregarding indexical phonetic variability between words coming from the familiarity with the speakers' voices. A and B always belonged to the same speaker and X was uttered by the other speaker, therefore, the only two possible sequences were M1-M1-F1 and F1-F1-M1. These speakers' voices appeared the same number of times across the four possible orders (ABA, ABB, BAB, BAA).

The test consisted of a total of 120 trials: 96 experimental trials and 8 practice trials testing the target contrasts (/iː- ɪ/ and /æ - ʌ/), and 16 control trials testing vowel contrasts (/iː- æ/ and /ɪ - ʌ/) that were not expected to pose perceptual difficulty in the discrimination of the minimal pairs: *bleak-black, cheating-chatting, pin-pun,* and *fizzy-fuzzy*. Within each trial block, i.e. ABA, there were 30 pairs of contrasting items, 15 of which were monosyllabic and 15 of which were disyllabic. Furthermore, 12 target minimal word pairs were taught and 12 were untaught (excluding the 2 practice + 4 control trials). Untaught words, which were mostly unknown according to the vocabulary knowledge scale questionnaire, were included to test for generalization to new word items, that is, to observe whether the task-based phonetic intervention was effective in modifying sensitivity to the target vowel contrasts regardless of the lexical item (Appendix D.1.).

Before doing the test, learners performed 8 practice trials during which they received visual feedback for error and response latency. The rest of the trials were presented in fully randomized order. If a participant made no response within 2500 milliseconds, the next trial was initiated. The response latencies in milliseconds measured from the onset of the third word in the triad were used as a measure of speed. Both accuracy and speed measures were meant to reflect lexical encoding of the vowels being tested.

3.3.3.2. Lexical encoding

The lexical encoding of L2 vowel contrasts (i.e., perceptual sensitivity of L2 vowel contrasts in lexical contexts) was measured through a fairly innovative task paradigm, namely, forced-lexical choice (FLeC) (Daidone, 2020; Kojima, 2019), and a recurrently used lexical-decision (LD) task (Darcy & Thomas, 2018; Mora & Mora-Plaza, 2019; Llompart, 2021a, 2021b). In the LD task, even though they hear only one stimulus, to accept the stimulus as a word or to reject it as a nonword, participants might activate more competitors than two. For instance, upon hearing the nonword **cuthy*, learners might activate *Cathy, coffee*, and so on before they decided that **cuthy* is a nonword. Whereas the LD task makes an open-ended question, the FLeC asks participants to choose from one of two possible competitors. The FLeC task is used to complement the LD task as we assume the FLeC task would reduce participants' cognitive load (hence, would be less demanding) by presenting two options to choose from and, thus, we expect higher accuracy overall. Both FLeC and LD task were administered via the *DMDX* software (Forster & Forster, 2003) on a laptop computer.

3.3.3.2.1. Forced lexical choice (FLeC)

The FLeC task illustrated learners' perceptual sensitivity to the L2-English contrasts /iː-ɪ/ and /æ- ʌ/ in a lexical context, hence, their ability to encode these phonological contrasts in their mental lexicon with precision. Learners were asked that they would be hearing real English words and invented words -that in some cases could sound similar to English words- and that their task was to decide whether what they heard was an existing English word by pressing a key on the computer keyboard. Participants were asked to hit "1" (Alt) when they thought the first one in a given pair was the English word and to hit "2" (Alt

Gr) when they believed the second one was the English word (*Total task duration*=3-4 minutes).

The FLeC task was created following an AX paradigm, where A and X belonged to two different speakers, thus, the only possible sequences were M1-F1 and F1-M1. These voices appeared the same number of times across the two possible orders (AX and XA). The task consisted of a total of 40 trials: 32 test trials testing the target contrasts (/iː-ɪ/ and /æ - ʌ/) through minimal pairs, and 4 practice and 4 filler trials which were pairs formed by a different set of consonant and vowels, respectively. The 32 minimal pairs were 16 monosyllabic and 16 disyllabic items containing the target vowels /iː/ (8), /ɪ/ (8), /æ/ (8), /ʌ/ (8). Each pair was formed by an English word (e.g. *kiwi* /ˈkiːwi/*)* and a nonword created by swapping the two target vowels (e.g. /ˈkɪwi/). Additionally, learners were exposed to 4 filler word-nonword pairs (i.e. *pen-*poon, merry-*murry, horse-*hars, morning-*marning*) containing the English contrasts /e/-/uː/, /ɔː/-/ɑː/ that were expected to be easy to distinguish for L1-Catalan/Spanish learners. The inclusion of fillers served to provide a baseline measure of performance with easy L2 phonological contrasts and allowed for a controlled observation of the effects that the difficult L2 phonological contrasts (i.e., /iː- ɪ/ and /æ- ʌ/) had on word recognition (Appendix D.2.).

Four pairs of practice trials involving easy consonant contrasts (i.e. /p/-/k/, /b/-/t/, /n/-/m/, /w/-/j/) were used to familiarize learners with the task procedure as learners received visual feedback regarding their accuracy and response time (i.e. *pass-*kass, bird-*tird, neck-*meck, work-*yerk*). All trials, except for the practice, were presented in fully randomized order. If a participant made no response within 2500 milliseconds, the next trial was initiated. The response latencies in milliseconds were taken from the onset of the auditory presentation of the word.

Accuracy on all items was screened by native speakers to see if nonwords were perceived as words and vice versa. Despite the fact that the tokens "hum, tird and laster" were actual words, they were used as the nonword counterpart in the task because they were completely unknown by the learners, as shown in the results of the word familiarity questionnaire. This applies to the LD task, described below.

We calculated an overall accuracy and RT rate which corresponded to the correct identification of real words out of the two items (Kojima, 2019) for every participant, testing time (pre-test, post-test and delayed post-test)[42] and vowel (/iː/, /ɪ/, /æ/, /ʌ/). Preliminary analyses included the comparison between test and filler trials, but only test trials were part of the main analyses.

3.3.3.2.2. Lexical decision (LD)

The LD task was used to assess L1-Catalan learners' phonolexical encoding of the L2-English contrasts /iː- ɪ/ and /æ - ʌ/ in a lexical context, reflecting the extent to which L2 learners had precisely encoded these phonological contrasts lexically. Participants were asked to decide, as accurately and as fast as possible, whether a sequence of sounds presented auditorily (as spoken by M1 or F1) constituted an English word or an invented word by pressing a green key as "real word" (Alt gr) or a red key as "invented word" (Alt) on the computer keyboard. They were warned that invented words could sound similar to real English words (*Total task duration*=5-6 minutes).

The LD task contained 80 trials (i.e., 40 word and 40 nonword trials). Test trials (*N*=64) consisted of 32 monosyllabic and 32 disyllabic items, containing the L2 target vowels /iː/ (*N*=16), /ɪ/ (*N*=16), /æ/ (*N*=16) and /ʌ/ (*N*=16). Each block of 16 was formed

[42] Initially, we calculated an overall error rate (to make it more comparable to the LD measure) but both error and accuracy rate represented mirror images of each other. For the sake of clarity, the selected FLeC accuracy measure was overall accuracy rate and reaction times on correct responses.

by 8 English words (e.g., kiwi, pin, jacket, drum → /ˈkiːwi/ /ˈpɪn/ /ˈʤækɪt/ /ˈdrʌm/) and 8 nonwords with systematic mispronunciations created by substituting the target test vowels by their contrasting counterparts (e.g. */ˈkɪwi/ */ˈpiːn/ */ˈʤʌkɪt/ */ˈdræm/). Learners were also exposed to 8 filler trials that did not pose any difficulty to provide a baseline for performance in the lexical decision task in the absence of confusable L2 phones. Fillers consisted of 4 monosyllabic and 4 disyllabic words (i.e., *pen*, *merry*, *horse*, *morning*→ /ˈpen/ /ˈmeri/ /ˈhɔːs/ /ˈmɔːnɪŋ/) and 4 monosyllabic and 4 disyllabic nonwords (i.e., */ˈpuːn/ */ˈmuːri/ */ˈhɑːs/ */ˈmɑːnɪŋ/), that contained two non-confusable phones, thus, /e/-/uː/, /ɔː/-/ɑː/ (Appendix D.2.)

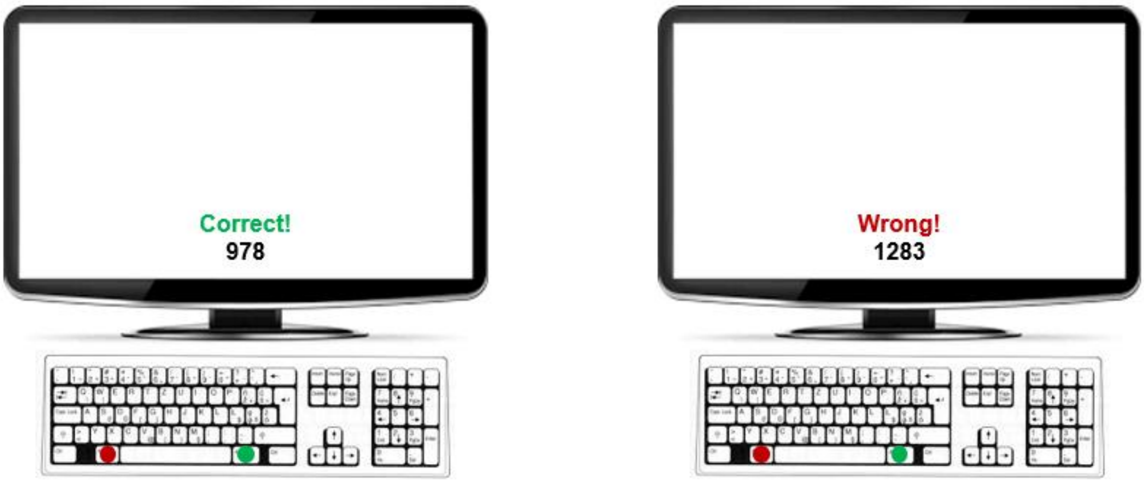


Figure 3.5. Feedback responses (accuracy and reaction time) in the LD task.

In order to get used to the LD procedure, learners encountered 8 practice word and nonword trials involving easy consonantal contrasts (i.e. /p/-/k/, /b/-/t/, /n/-/m/, /w/-/j/) found in *pass, bird, neck, work, and *kass, *tird, *meck, *yerk*. Learners received visual feedback for error and response latency (Figure 3.5.). All trials, except for the practice ones, were presented in fully randomized order so they were not affected by the lexical type (word/nonword) or the voice (male/female). If a participant made no response within 2500 milliseconds, the next trial was initiated. Response latencies in milliseconds were taken from the onset of the auditory presentation of the word.

Native-like sensitivity to the /iː-ɪ/ and /æ-ʌ/ contrasts in the lexicon would be reflected in correctly identifying both test words and nonwords. I calculated, for every participant and testing time (pre-test, post-test and delayed post-test), a global accuracy rate: average accurate rate and RT scores per vowel (including both words and nonwords) separately for test and filler trials, as well as an individual measure of perceptual sensitivity: nonword rejection rate, hence, correct identification of nonwords as nonwords (Amengual, 2016; Darcy & Thomas, 2019; Llompart, 2021a, 2021b).

Despite the fact that the words in the two lexical tasks were chosen in order to be familiar to L2 learners, a vocabulary knowledge scale (Section 3.3.4.2.) determined learners' familiarity with the words appearing the lexical tasks (test and filler items).

3.3.3.3. Vowel production

Learners produced the L2 target vowels through delayed word repetition (DWR; Munro & Derwing, 2008) and delayed sentence repetition (DSR; Mora et al., 2022) tasks, which were administered in *DMDX* (Forster & Forster, 2003) on a laptop computer. Word and sentence repetition tasks have been previously used in communicative approaches to assess gains in pronunciation instruction (Elliott, 1995, 1997). In addition, in the use of controlled tasks, which are designed to avoid the communicative pressure or the need to retrieve grammar, vocabulary or speak pragmatically appropriate, speakers can direct more attention to their pronunciation and have been found to show improvement after pronunciation instruction (Darcy & Rocca, 2023; Mora et al., 2022).

3.3.3.3.1. Word repetition (DWR)

The purpose of the DWR was to determine whether experimental participants would produce the target vowels /iː-ɪ/ and /æ-ʌ/ more contrastively at post-test and at delayed post-test, and assess whether L2 vowel quality would become closer to the native speakers' values across testing times. Learners were instructed to repeat words in isolation after a *beep sound*. Participants heard the stimulus (e.g., /lɪvə/) followed by a 1500-millisecond pause before a tone signal (200ms) prompted them to repeat it. They had 2000 milliseconds to repeat the stimulus before the next one was presented (Figure 3.6.; *Total task duration*=5-6 minutes). This delayed elicitation procedure, previously used to investigate L2 segmental production (Nagle, 2021), was used to elicit the L2 sounds items within a phonological (rather than acoustic or phonetic) processing mode, thus avoiding direct imitation from sensory memory (Werker & Logan, 1985). Learners' productions should have reflected their phonolexical representations. In order to test for generalization effects, the testing stimuli comprised taught and untaught words. To ensure generalization to new voices, the testing stimuli were produced by 2 speakers (M1, F1), which participants had not been exposed to during the intervention.

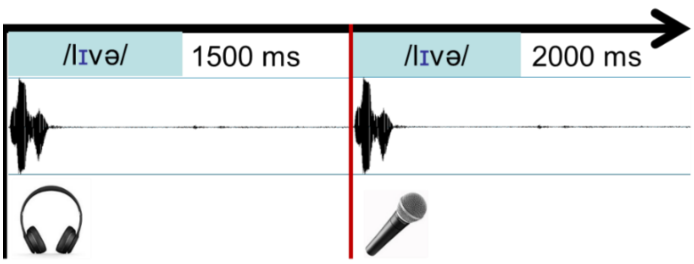


Figure 3.6. Delayed word repetition procedure

The test consisted of a total of 68 trials, i.e., 64 test and 4 practice trials. The testing stimuli were monosyllabic (*N*=32) and disyllabic (*N*=32) and belonged to 48 words coming from minimal pairs (i.e., /iː-ɪ/ and /æ-ʌ/) and 16 words which did not have a contrasting counterpart but which contained the 4 target vowels and had appeared during the intervention (Appendix D.3.). Also, 24 minimal-pair words were taught and 24 were untaught, meaning that learners had not practiced their form and meaning during the intervention and most of them were unknown to them (Table 3.9.).

Table 3.9. DWR test items' distribution by Type (minimal pairs/extra words) and Word Type (taught/untaught).

	Taught				Untaught			
	/iː/	**/ɪ/**	**/æ/**	**/ʌ/**	**/iː/**	**/ɪ/**	**/æ/**	**/ʌ/**
Minimal Pairs	6	6	6	6	6	6	6	6
Extra words	4	4	4	4				

An introduction to the production task was given to the participants beforehand in the form of 4 words that did not have a contrasting pair (i.e., *feet, hill, ran, sun*) and had not appeared during the intervention but allowed participants to get familiarized with the task procedure. All words, except for the practice ones, were distributed into two randomized blocks with 32 stimuli each and a break in between. The first block included words which contained the L2 vowels /iː/ and /ɪ/, and the second one /æ/ and /ʌ/.

3.3.3.3.2. Sentence repetition (DSR)

DSR tasks (e.g., Flege et al., 1995; Piske et al., 2001; Trofimovich & Baker, 2006) consist of an elicitation method which eliminates the effect of reading and orthography, but produces controlled, usually fluent, speech. The DSR was used to examine whether learners were able to produce the L2 English vowels in the contrasts more distinctively and more accurately at post-test and delayed post-test and whether they were able to transfer the pronunciation of L2 words produced in insolation (DWR) to the context of a meaningful sentence. If they were able to accurately produce the L2 target words included in the sentences, this meant they could produce L2 vowels precisely in a meaningful context (*Total task duration*=11-12 minutes).

Learners were asked to (1) read the sentence appearing in standard orthography on the computer screen for 3000 ms (i.e., *The **cat** is hidden*), (2) listen to the sentence over the headphones, (3) repeat the sentence from memory after a sound signal occurring 1500ms after the offset of the sentence stimulus (Figure 3.7.). In order to test for generalization effects, the testing stimuli comprised taught and untaught words as well as untaught voices (M1, F1), which participants had not been exposed to during the intervention.

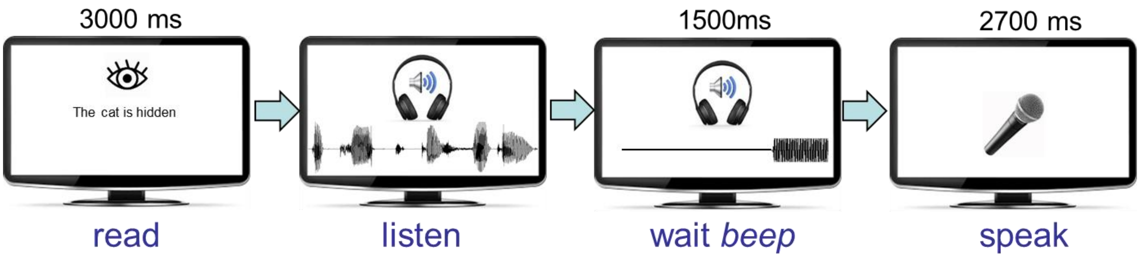


Figure 3.7. Delayed sentence repetition procedure.

In terms of stimuli, the DSR was identical to the DWR task but, instead of producing words, learners produced sentences containing those L2 target words. Learners were exposed to 64 test sentences and 4 practice sentences, which were 4 words long. They were always formed by the determiner/pronoun *THE/THEY* + the *TARGET WORD* containing /iː - ɪ - æ - ʌ/ + *VERB* + *OBJECT* (e.g., The *bin* is empty). The testing words in the sentence were either part of minimal pairs (*N*=48) or other words (*N*=16), which did not have a contrasting counterpart but which contained the L2 vowels (/iː- ɪ - æ - ʌ/). Half of the 48 target MP words were taught and half were untaught (Appendix D.4.).

Before starting the test trials, learners practiced uttering some sentences, which contained the same words (i.e. *feet, hill, ran, sun*) as the practice trials in the DWR task. All sentences, except for the practice ones, were distributed into two randomized blocks with 32 stimuli each and a break in between. The first block included words which contained the L2 vowels /iː/ and /ɪ/, and the second one /æ/ and /ʌ/.

Once recorded and stored digitally on the recorders' hard disk, the recording for all subjects' words (DWR) and sentences (DSR) in the pre-, post- and delayed post-test were transferred to a laptop and acoustically analysed through *Praat* (Boersma & Weenink, 2020).

We decided to use objective measures of production accuracy (i.e., acoustic analyses) as they have been found to be a valuable and more sensitive tool for the assessment of training effects than listener-based judgements (Delvaux et al., 2013; Saito & Plonsky, 2019). Acoustic measures of vowel pronunciation improvement typically involve computing pronunciation accuracy scores based on the qualitative distance between contrastive vowels, i.e., how much distinct vowel qualities have become through training (Melnik-Leroy et al., 2022), a measure of vowel *distinctiveness*. They may also involve computing the distance between learners' vowel productions and vowel target spaces in F1/F2 space, hence, how much learners' vowel qualities approximate those of native speakers (Kartushina et al. 2015, 2022), a measure of vowel *nativelikeness*. We used Mahalanobis distances of *nativelikeness* and *distinctiveness* (Mora, 2021), which compute the distances in standard deviations between a point and the centroid of a distribution (Kartushina et al. 2015, 2022; Melnik-Leroy et al., 2022). That is, Mahalanobis distances were used to calculate a measure of vowel *distinctiveness* (i.e., every token of a vowel (/æ/ or /iː/) and the centroid of the distribution of the tokens of the other vowel in the contrast (/ʌ/ or / ɪ/) and vice versa), so a larger distance meant less of an overlap between the two vowels (Melnik-Leroy et al., 2022). In addition, we calculated the distance between NS' and learners' productions of each target vowel produced in the same phonetic context (vowel *nativelikeness*), so a smaller distance meant a more target-like production (Kartushina et al., 2015). The reason for preferring Mahalanobis to Euclidean distances (Flege et al., 1997; Iverson & Evans, 2007) is that the former takes into consideration not only the centroid location, but also the spread and orientation of the reference distribution, thus reflecting token variability. As a secondary measure, we obtained a measure of vowel duration and duration ratio in milliseconds (ms) between the target contrastive L2 vowels (i.e., /iː-ɪ/ and /æ-ʌ/) produced by L2 learners.

3.3.3.4. Oral interaction

A dyadic interactive L2 task, which mirrored the type of tasks learners were performing during the intervention, was given to learners in the three testing times (*Total task duration*=5-7 minutes). The SG performed a simple decision-making task, and the CG performed a complex decision-making task. Participants' conversations were audio-recorded.

Following Swain and Lapkin's (1995) understanding of LRE as segments of discourse in which interlocutors topicalize linguistic items (i.e., talk about and question the language they are producing) either because of a focus on linguistic accuracy or a communication breakdown, we operationalized P-LRE as instances where learners focused on phonetic form. In particular, we only annotated negotiations (i.e., confirmation checks, clarification requests, comprehension checks, recasts) that were centred around the target phonological forms of the study (i.e., English vowels /iː/, /ɪ/, /æ/, /ʌ/).

Example. Travel task

A: Hi Lily! So, which are your preferences for our destination?
B: Oslo because there is the biggest *[ʃip] museum
A: <u>Sheep [ʃiːp] or ship [ʃɪp]?</u>

B: <u>Sheep [ʃiːp]</u>
A: Oh, I also like Oslo because I love animals.

Once a P-LRE was identified, it was secondarily annotated in a Praat (Boersma & Weenink, 2020) TextGrid. P-LREs started with a negotiation move (e.g., a confirmation check) and ended when the communication breakdown or metalinguistic conversation was over. 25% of these annotations were double-checked and fully agreed with a trained applied linguist (see Section 3.5.1. for inter-rater analyses). A Praat script was used to extract the frequency and duration of P-LRE.

To control for the differences in time on task between the simple and complex tasks, a rate of frequency (number of P-LRE/total duration of the audio) and duration (duration of P-LRE/total duration of the audio) of P-LRE production by dyad per minute was calculated, and the mean of both rates of P-LRE production between the simple and complex tasks was statistically compared over time. The dyads were the same across the three testing times.

3.3.3.5. Target word assessment

Immediately after the TBPT intervention, learners in the experimental group carried out a pen-and-paper vocabulary test in class to assess whether they knew the words they had been exposed to during the task-based treatment (Appendix G). They were asked to translate the target L2 words into Catalan/Spanish or make a drawing which would represent that concept, if they could not come up with the L1 translation (*Total task duration*=20-25 minutes). A measure of accuracy was obtained to assess the proportion of correctly translated words (%) after the treatment.

3.3.3.6. L2 oral proficiency

Overall L2 proficiency was assessed through an EIT (*Total task duration*= 7-8 minutes) constructed on a Power Point presentation, which was originally designed by Ortega et al. (2002) for a cross linguistic study on syntactic complexity measures.

The test included 30 sentences in English ranging from 7 to 19 syllables that increased in grammatical and lexical complexity. These included high frequency vocabulary items, a range of syntactic complexity, and typical grammatical features known to challenge instructed learners. The sentences were produced by a female native speaker of English (Speech rate: M=4.58, SD=.58; Sentence duration: M=3.02, SD=.74) and were presented auditorily over headphones for delayed repetition. Participants were instructed to repeat each sentence as accurately as they could after a 250ms *beep* signal, which occurred 2000ms after the sentence ended. Participants had 6.8 seconds to repeat the sentence after the beep. The learners' productions were recorded onto a digital recorder and assessed for accuracy following Ortega et al.'s (2002) rubric, where each sentence received a score from 0 to 4 as a function of how much of it was repeated and the type of inaccuracies and missing unrepeated material. Individual scores could range from 0 to 120 points.

3.3.3.7. Working memory

In order to obtain an individual measure of PSTM and complex WM capacity, learners carried out two digit-span tasks in Inquisit 5 Lab (Draine, 1999). The forward-digit span test provided a measure of learners' PSTM and the backward-digit span test provided a measure of complex WM (*Total test duration:* 5 minutes each), both suitable for adolescent FL learners (Jarvis & Gathercole, 2003 for a review). Nonverbal tasks have been frequently used in L2 research (Darcy et al., 2015; Kormos & Sáfar, 2008) and their main advantage is that we can obtain measures of cognitive ID that are language independent both in terms of the materials used and the participants tested (Juffs & Harrington, 2011). We used a forward-digit span task to assess storage capacity in the phonological short-term memory. The test consisted of 17 trials, 2 practice and 15 test trials. The number of digits in a sequence started at two in the first trial and increased until the participants failed to recall correctly two consecutive sequences with the same length. Digits appeared one by one on the computer screen with a 60-ms pause between items. If students typed the digits in the correct order of at least one set, they could go to the next span.

The backward-digit span required both storage of information as well as manipulation of that information (involving both the central executive and the phonological loop) and could be considered a complex WM task (Montero Pérez, 2020; Juffs & Harrington, 2011). The task consisted of 17 trials, 2 practice and 15 test trials. In each trial, participants were visually presented with a random sequence of digits, which appeared one by one on the computer screen with a 60-ms pause between items. Then they were asked to recall the sequence of digits in the reverse order in which they had appeared by typing the answer into a presented textbox. The number of digits in a sequence started at two in the first trial and increased until the participants failed to recall correctly two consecutive sequences with the same length. If a consecutive error occurred, the participant moved back down to a lower level, starting over.

We obtained a mean span (MS) with the scores on each one of the tasks giving an estimate of the score a participant would obtain 50% of the time on the basis of overall performance during 15 trials (Brunfaut et al., 2021). Instead of using a two-error maximum length measure in the analyses (Kormos & Sáfár, 2008), we used MS because

Woods et al. (2011) showed that MS showed reduced variance, improved test–retest reliability, and obtained higher correlations with the results of other neuropsychological test results relative to a two-error maximum length measure. The MS metric also enhanced the sensitivity of forward versus backward span comparisons, and elucidated changes in digit span performance with age and education level. In addition, metrics based on this approach provide useful information that is present in performance variability around maximal span.

3.3.3.8. Auditory selective attention

Individual measures of auditory attention control were obtained from the ASA test (Humes et al., 2006), which was carried out in *Inquisit 5 Lab* software (Draine, 1999). The ASA test was based on a single-talker competition paradigm. The test consisted of 40 trials of pairs of different sentences presented simultaneously, target vs. competitor, spoken by a male voice and a female voice respectively (e.g. male: *Ready CHARLIE go to BLUE SIX now*; female: *Ready TIGER go to RED EIGHT now*). The sentences were normalized for duration (1700ms). In every trial, a call signal (e.g. *TIGER*), appearing on the screen previous to the auditory presentation of the sentence, cued the voice participants had to attend to before the sentence for correctly identifying 1 of 4 colours and 1 of 8 digits visually presented on the screen (Figure 3.8.; *Total task duration*: 7-8 minutes). Learners were first presented with the procedure through a Power Point presentation, then, they were exposed to 8 practice trials and, finally, they went through the 32 test trials. Individual ASA scores were computed by adding up all correctly identified colours (32 points) and digits (32 points) up to a maximum score of 64.

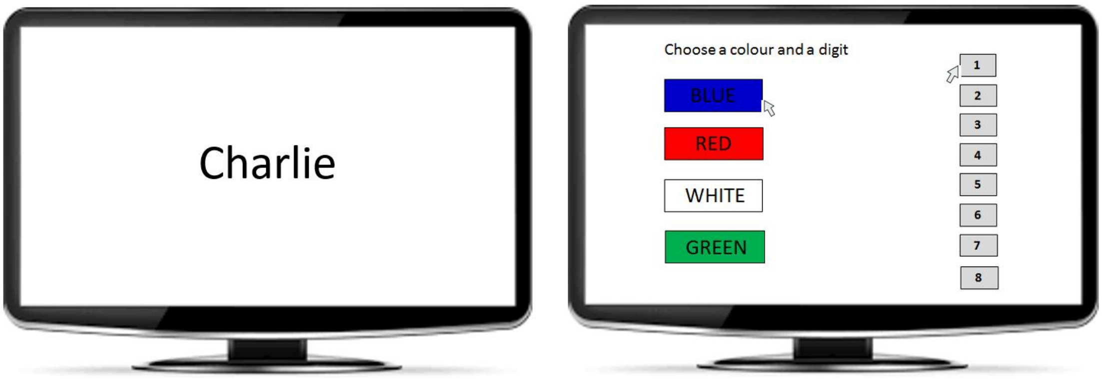


Figure 3.8. Auditory selective attention task in the Inquisit 5 software.

3.3.4. Questionnaires

The purpose of this section is to present the questionnaires used in the study to gather information about the participants' demographic and linguistic characteristics, their word familiarity and their impressions of the task-based pronunciation-teaching intervention.

The online demographic and linguistic background questionnaire was created to obtain learners' personal and linguistic information, centred on their English learning experience. The word familiarity questionnaire was used to gain insights into participants' knowledge of L2 words before prior to intervention. Lastly, learners expressed their perceptions and evaluation about the tasks carried out during the intervention through an online post-intervention questionnaire.

3.3.4.1. Demographic and linguistic background

A large body of research exists on the effect of individual variables on the acquisition of foreign language speech (Piske et al., 2001 for a review). Factors such as age of acquisition, amount of L2 experience, L2 use, L1 use and quality of the L2 input, among others, have been widely studied, and evaluated in relation to L2 learning outcomes. A questionnaire was used to target these variables as precisely as possible in order to gather detailed information about their L2 experience. The language background questionnaire (LBQ)[43] was created in the online platform Google Forms and was sent to the participants prior to the pre-test. The LBQ link was sent to the informants via email or WhatsApp. The responses were gathered and stored in a separate file on the online Google server from where the researcher downloaded them into an Excel spreadsheet.

The LBQ consisted of three parts (Appendix H): the first section inquired about general demographic information; the second section presented questions related to the learners' language profile and L1 and L2 language use; and the final section dealt with learners' English learning experience. Concerning the first part of the questionnaire, participants were inquired about basic demographic data (age, sex) and hand dominance. Additionally, participants were asked if they had been diagnosed with any speech/hearing problems in order to pay special attention to those participants who responded affirmatively.

The second section of the questionnaire was designed to obtain detailed information about the participant's language profile. With this aim, questions about languages spoken daily (*Q1*), their mother tongue (*Q2*), their home language (*Q3*) and dominance in other languages were formulated (*Q4.1*-*Q4.5*). The last three questions were related to their daily use of Catalan, Spanish and English (*Q5.1*-*Q5.3*).

The last section of the questionnaire targeted participants' L2 English learning experience. Questions targeted their age of onset of learning (*Q6*), number of weekly hours of English instruction in different contexts (*Batxillerat*, and extracurricular) (*Q7*, *Q8*) as well as English use with their family (*Q15*) and time spent in English speaking

[43] The LBQ was adapted from the pen-and-paper questionnaire employed in studies by the GRAL group and can be found in http://www.ubgral.com/

countries (*Q17.1, Q17.2*). Additionally, *Q10* and *Q11* were formulated to obtain information about their weekly use of English with non-native speakers and native speakers of English, respectively, as well as the amount of time doing activities in English (*Q14*). Learners were also asked about the % of exposure (*Q12*) and production (*Q13*) of RP vs General American English. Finally, they evaluated their command of English for each one of the 5 competences (*Q16*).

The data obtained from the questionnaire was categorical (e.g., sex, hand-dominance), ordinal (self-estimated language proficiency, frequency of exposure to English, etc.) and scale (age, nº of years studied L2, etc.). A summary of participants' demographic and linguistic information can be found in Appendix B. The calculation of spoken and written input / output (*Q14*) variables employed as predictors for RQ3 analyses are discussed next. The *Spoken L2 input* variable was obtained from summing the responses of the statements "Watching English language television", "Listening to songs in English" and "Watching videos or movies in English". The *Written L2 input* variable was computed as the sum of the responses regarding "Reading newspapers/magazines in English" and "Reading books in English". Moreover, the *Spoken L2 output* measure came from the sum of responses in "Speaking English with native or fluent speakers" and "Speaking English with non-native speakers". Finally, *Written L2 output* was obtained from the last statement in *Q14*: "Writing emails/letters in English". Learners' self-perceived L2 proficiency (*Q16)* was obtained by computing the average of responses regarding the 5 competences (i.e. reading, listening, speaking, writing, pronunciation).

In order to answer RQ3 on the effects of experiential factors on L2 vowel performance and learning, participants' *past* and *recent* L2 learning experience were selected as predictors (Kissling, 2014; Muñoz, 2014; Suzukida & Saito, 2023). On the one hand, *past* L2 learning experience was operationalized as years of instruction. On the other hand, *recent* L2 learning experience was divided into (a) inside classroom experience (operationalized as the average number of hours per week of L2 learning in the school and language academy) and (b) outside classroom experience (operationalized as the average time spent receiving L2 spoken and written input and producing L2 spoken and written output [*1= never, 5= every day*]). The Cronbach's alpha value of each construct for *recent* outside L2 learning experience indicated a relatively high level of internal consistency (α = .85).

3.3.4.2. Word familiarity

Learners' familiarity with the experiment words was extracted from an adapted version of the Vocabulary Knowledge Scale (VKS) (Wesche & Paribakht, 1996) and presented to the students prior to testing through a link to the online platform Google Forms (see a sample in Appendix J). The questionnaire exposed the learners to several words/nonwords (e.g. *bean*) and they had to indicate how familiar they were with those by selecting 1 of 4 options (Figure 3.9.). Responses were gathered and stored in a separate file which was downloaded in Excel format. Then, they were labelled with categorical values (1= I've never seen this word before; 4= I know what this word means and I can use it in a sentence)

☐ I know what this word MEANS and I can USE it in a sentence

☐ I know what this word MEANS, but I'm NOT SURE how to USE it

☐ I've SEEN this word before, but I DON'T KNOW what it MEANS

☐ I've NEVER SEEN this word before

Figure 3.9. Sample of a 4-option response in the vocabulary knowledge scale questionnaire.

On the one hand, the VKS questionnaire was used to exclude any words from the FLeC and LD tasks which they were unfamiliar with to ensure that responses would be a reliable reflection of learners' phonolexical knowledge. On the other hand, the VKS questionnaire was used to classify the words that students knew initially from the ones they were not so familiar with; we computed the total sum by word, run the median and classified the highest numbers as unknown and the rest as already known tokens. This variable was incorporated in the perception and production datasets, but it is outside of the scope of this thesis to explore the effect of word familiarity on L2 vowel learning.

3.3.4.3. Post-intervention perceptions

The qualitative part of this study stems from learners' perceptions of the task-based project "A trip to London". Right after the performance of the last task, learners in the experimental group were asked to fill in an online questionnaire in Google Forms about (1) their beliefs about English pronunciation, (2) their evaluation of the tasks and the overall project and (3) their perceptions of improvement (Appendix K).

To start with, learners were asked about their believes regarding the importance of learning English pronunciation before (*Q4*) and after (*Q5*) doing the tasks on 7-point Likert scales (*1= "not important at all", 7= "extremely important"*) as well as the number of hours per month they would like to dedicate to L2 pronunciation in class (*Q6*), once the project finalized.

The second section of the questionnaire focused on learners' opinions concerning the pre-tasks, tasks and post-tasks. First, learners shared their perceptions about the pre-tasks, namely, whether they found them enjoyable (*Q10*) and whether they had helped them learn the meaning (*Q7*) and pronunciation (*Q8*) of the words they encountered in the task cycle. Also, they were asked whether they found the recordings of the listening comprehension exercise difficult to understand (*Q9*). Concerning the tasks, learners were inquired about the difficulty (*Q12*) and degree of mental effort (*Q11*) they had to put to solve the tasks, as well as the source (*Q13*) of mental effort (i.e., task conditions, pronunciation or both). In addition, learners were asked to rate how interesting/ enjoyable (*Q14*) and realistic (*Q17*) they considered the tasks to be; select which tasks they enjoyed the most and the least (*Q15*); and state the reason (*Q16*) that would make a task more enjoyable (i.e. the topic, the images/drawings, the text or the difficulty). We also elicited information about their opinions regarding the presence of images in the flashcards (*Q18*) and we posed an open question regarding the kind of strategies they used when they had problems communicating with their classmates (*Q18*). This section finished with questions regarding the post-task, namely, whether they had helped learners revise the meaning (*Q20)* and the pronunciation (*Q21*) of the target words, and whether they considered the post-tasks to be interesting and enjoyable (*Q22*).

To finish with, learners were asked to assess difficulty in the pronunciation of the target vowels (*Q23*) they had been exposed to during the tasks, and self-assess their improvement in pronunciation (*Q24*) on a 7-point Likert scale. In addition, they had to indicate to what extent (i.e., from 1= "*no improvement at all"* to 7= "*a lot of improvement"*) they felt they had improved the pronunciation of each one of the target vowels (*Q25*-*Q28*). The post-intervention questionnaire concluded with two open-ended questions about learners' general opinion concerning the likes and dislikes of the TBPT project (*Q29*) and evaluation (*Q30*) of their learning process. Open-ended questions were included as they have been found to permit greater freedom of expression, yielding illustrative quotes and suggestions that could have not been anticipated by the researcher (Dörnyei & Taguchi, 2009).

In order to answer RQ4 on learners' perceptions about the TBPT intervention, categorical responses were given a numerical value and responses from ordinal questions were defined (e.g., *1= strongly disagree, 2= disagree, 3= somehow disagree,* etc.). Percentages were used to express the frequency of learners who gave a particular answer. The protocol of coding the open-ended questionnaire items was done by the author of this dissertation and an additional EFL teacher. In order to avoid the harmful effects of rater subjectivity, the themes were ''data-driven'' (i.e. inductive category coding) and processed by means of systematic content analysis (Dörnyei & Taguchi, 2009) in Microsoft Excel[44]. This was achieved through a stepwise process that involved six main phases, adapted from Braun and Clarke (2013, 2021):

1- *Data familiarization and writing familiarization notes*: Taking each learner response, distinct content elements or key points were highlighted in bold. Could a response be assigned to several themes, an asterisk (*) was placed next to it.
2- *Systematic data coding*: Each text unit was assigned a different colour depending on the theme they belonged.
3- *Generating initial themes from coded data*: Themes were grouped together by colour and provisional names were allocated. Responses which were marked with an asterisk (*) were duplicated, highlighting in bold the key ideas representing each theme.
4- *Developing and reviewing themes*: Themes were redefined and agreed with an experienced EFL teacher (Figure 3.10.). Afterwards, an applied linguist assigned each one of the different responses to themes. Inter-coder agreement coefficients

[44] For larger qualitative datasets, a qualitative data analysis software (e.g., NVivo) should be used to systematically code and generate themes and a thematic map.

were calculated to observe the extent to which the two coders agreed (see Section 3.5.1. for inter-coder reliability results).

5- *Refining and naming themes and subthemes*: Coding decisions and theme/subtheme naming were discussed at length between the author of this dissertation and the aforementioned applied linguist, and where there as any disagreement, adjustments were considered until full consensus was reached. Thematic maps were used to aid visualize the different categories (Figure 3.11.)

6- *Data analysis and writing of the report*: The themes and subthemes obtained in Phase 5 were numerically coded to be treated as quantitative data and expressed in percentages (Appendix M). In addition, some of the most relevant responses highlighted in Phase 1 were quoted in the discussion for the purpose of exemplification.

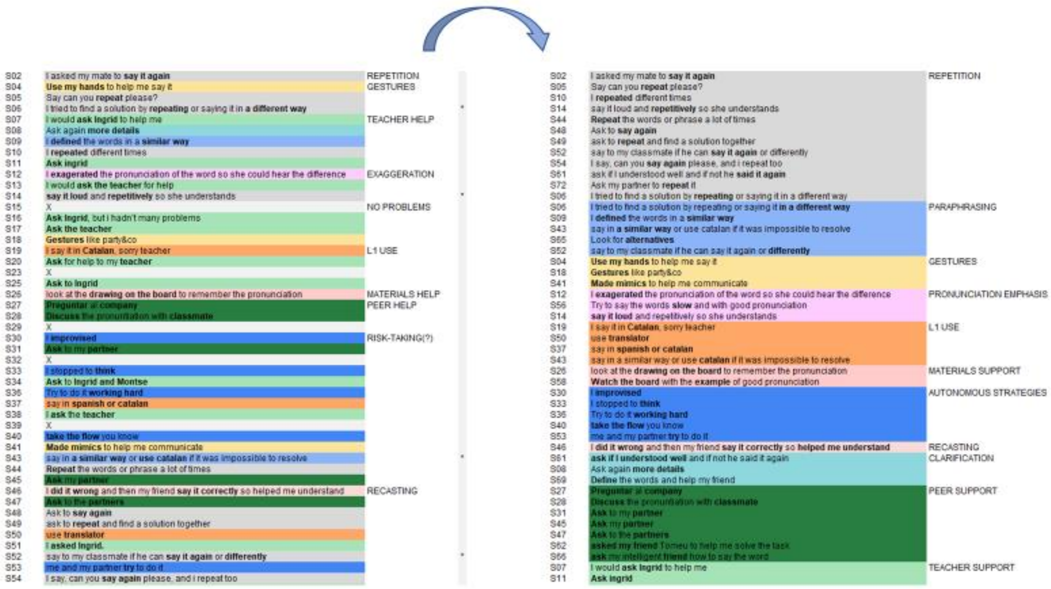


Figure 3.10. Visual representation of phases 1 to 4 of the coding of open-ended responses.

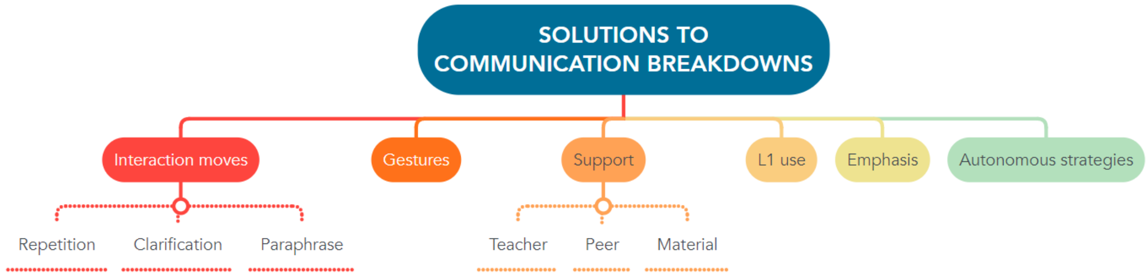


Figure 3.11. Example of a thematic map.

3.4. General procedure

This section outlines the overall procedure of the intervention and testing phases of the study. Six months before the start of the pedagogical intervention, the head of the public school, language department coordinators and the two English teachers responsible for teaching L2 English in *Batxillerat* were contacted. This study was presented as part of the state-funded RecerCaixa project "Habilitats orals per a les professions del futur: un programa d'intervenció en el currículum de secundària i batxillerat a Catalunya" (RecerCaixa 2017ACUP 00249), which aims at improving learners' L1 Catalan and L2 English oral abilities in secondary schools in Catalunya. The rector of the University of Barcelona and the head of the INS Vilatzara school signed an agreement of collaboration, once the study had been approved by the University of Barcelona Bioethics Board (Institutional Review Board – IRB00003099). Afterwards, a detailed information sheet was provided to the head of the school, department coordinators and English teacher containing the objective of the project, the type of tasks, the dates and duration of the intervention and testing phase, and the equipment and spaces needed (Appendix A.1.). The school agreed to participate under the premise that I would dedicate the first part of their English lesson to do communicative activities focused on English pronunciation and the English teachers would dedicate the second-half to practice the rest of linguistic competences (i.e., writing, reading, use of English, etc.). Students would get a 15% of their overall English mark in the first trimester if they completed the pre-test, intervention and post-test successfully, and a 5% in the second trimester if they participated in the delayed post-test.

During the first week of the course, learners were introduced to the project "A trip to London" through a Power Point presentation in the English class. This pedagogical intervention was presented as part of the aforementioned state-funded project on L1/L2 oral abilities. The objectives and methodology of the project were presented to the learners. Finally, learners read the privacy statement and were asked to sign a consent form (Appendix A.3.) and write down their name, surnames, and other contact details. The students' parents were also informed about the study through a formal letter (Appendix A.2.). The Association of Students' Fathers and Mothers of the school (A.M.P.A) gave green light to carry out this pedagogical intervention as part of their English course. Before starting the project, the participants were instructed to fill in an online language background questionnaire and a word familiarity questionnaire during the first week of the course at home.

3.4.1. Intervention

The intervention phase took place in the usual English classroom and was carried out with the whole group of students three times per week. Learners dedicated the first 20-30 minutes of their English class to practice reading, listening, writing and/or use of English with their usual teacher and the last 30-40 minutes to do oral tasks with the author of this dissertation. Their usual English teacher helped me check class assistance and we closely monitored the students' interactions. Learners carried out the pre-task individually and, in the task cycle, they were instructed to join desks in a way that each pair/group of 3 would be facing each other. Learners performed the post-task in pairs or small groups but did not change seats.

3.4.2. Testing

The pre-test, post-test and delayed post-test took place in a separate room outside the main building, which was usually used for school meetings and examinations. The room was big enough to test 8 participants at once in 8 different laptops separated by >2 meters. Due to the large number of students, communication between the researcher and students was facilitated by a WhatsApp group created especially for this study. Several procedures and/or any changes on the schedule (i.e. due to strikes or class cancelations) were immediately notified via WhatsApp. At the end of each day, all absent students were contacted and a make-up session was arranged.

The testing was made possible by the assistance of three other researchers. For every 8 participants, one researcher was in charge of 4 students and another researcher guided the remaining 4. Within each group of 4 students, a pair carried out the interaction task together at the beginning and, whereas one other student was doing the perception tasks (i.e., ABX, FLeC, LD), the one left was performing the production tasks (i.e. DWR, DSR). This procedure was used to maximally avoid student distractions and sound interference across tasks. The ID tests (i.e., EIT, WM, ASA) took place either after the oral interaction task at the beginning or after the perception/production tasks, so that L2 vowel perception and production could always be tested sequentially. Researchers ticked the tasks which were performed and the recording number by participant in an Excel spreadsheet. Learners who were being tested missed an hour of their English class but they could catch up with the content by checking their Moodle, as the English teacher always uploaded the materials and task instructions of the missing class. After the delayed post-test, participants were informed that the experiment's results would be shared with them at the beginning of the following academic year (September 2020).

3.5. Data analysis

This section comprises (1) prior analyses to the main statistical analyses and (2) the main statistical tests that were conducted to answer the research questions of the study.

3.5.1. Approach to analyses

Before dealing with the primary inferential analyses of the perception and production tasks, the VKS questionnaire revealed that participants were generally very familiar with the words (M = 3.76, SD = 0.71) appearing in the FLeC/LD tasks on a 4-point scale, where "1" meant no knowledge of the word and "4" indicated that the word was well-known. Since there was no particular word from the FLeC/LD task that was overall unfamiliar to the learners and could affect the interpretability of the results, the pre-selected words and nonwords were used as the stimuli of the FLeC/LD task (as other studies did, e.g., Simonchyk & Darcy, 2021). Only the pair "run-ran" was excluded from FLeC analyses as the author of this dissertation made an initial mistake selecting the past form of the verb "to run" as a nonword.

All recordings were checked for absence of noise (e.g. coughs, sneezes, etc.), recording failures, and productions that differed from the target (i.e. words that did not correspond to the target words). In terms of acoustic analyses for L2 vowel production, vowel quality measurements (f0, F1, F2) were extracted from a 10-millisecond window by manually placing a cursor at the midpoint of the steady-state portion of the target vowels in Praat (Boersma & Weenink, 2020). To minimize age, gender and vocal tract size effects, frequency values were then converted from Hertz (Hz) to the psycho-acoustical scale Bark (B),[45] and then a Bark-distance normalisation procedure (Syrdal & Gopal, 1986) was used to provide speaker-independent estimates of vowel quality.[46] The difference in Bark between F1 and *f*0 (B1-B0) estimated vowel height, whereas the difference between F2 and F1 (B2-B1) estimated vowel frontness (Bohn & Flege, 1990). Mahalanobis distances of *distinctiveness* and *nativelikeness* (Kartushina et al., 2015, 2022; Melnik-Leroy et al., 2022) were calculated from Bark-normalized frequencies. In terms of vowel duration, we manually selected the portion from the onset to the offset of vocal cord vibration, and obtained the duration in milliseconds (ms) through a Praat script.

The last preliminary steps before conducting the primary analyses involved screening the perception (ABX, FLeC, LD) and production (DWR, DSR) data. When response latencies in perception and lexical decision tasks are reported, these correspond to reaction times (RT) screened for accuracy (only including correct responses) and extreme values (2.5 standard deviations below or above the by-subject by-item type (test vs. practice item) and by-time mean). A total of 2.7%, 1.4% and 0.8% of the datapoints were discarded in the ABX, FLeC and LD datasets, respectively, mainly coming from extreme values that did not reflect learners' accurate performance.

In addition, to minimize measurement errors or extreme values in vowel production, *f*0, F1 and F2 values were screened by replacing values above or below 2.5

[45] Vowel frequencies (Hz) were converted to Bark (B) using the formula Zi = 26.81/(1+1960/ Fi) -0.53, where Fi is the frequency value in Hz for a given formant i and Z the frequency in Bark (Traunmüller, 1997).

[46] Euclidean distances (SD scores) were calculated by means of the following formula, where Va and Vb are the two vowels for which the Euclidean distance is calculated:

$$\sqrt{((VaB2-B1)-(VbB2-B1))^2+((VaB1-B0)-(VbB1-B0))^2}$$

standard deviations from the mean within each participant's mean value for that vowel in the same testing time. Outliers more than 2.5 standard deviations from the by-talker by-vowel by-time mean were discarded (2.9% and 3.2% of the datapoints in the DWR and DSR datasets, respectively). The exact same procedure was followed for vowel duration. Milliseconds (ms) above 2.5 standard deviations from the by-talker by-vowel by-time mean were discarded (0.7% and 1.5% of the datapoints in the DWR and DSR datasets, respectively).

Lastly, inter-rater reliability analyses were high (α = .92) for the identification of P-LRE from 25% of the interactions of SG and CG learners. The author of this dissertation and the rater discussed differences in P-LRE identification until strong consensus was reached. Furthermore, inter-coder reliability (Cronbach's alpha intra-class correlation coefficients) was moderate-to-high for the coding of all open-ended responses belonging to *Q18* "What did you do if you had PROBLEMS communicating with your classmates?" (α = .90), *Q19* "What do you think of the IMAGES/DRAWINGS in the flashcards?" (α = .98), *Q29.1* "What did you most LIKE about doing the project?" (α = .84), *Q29.2* "What did you most DISLIKE about doing the project?" (α = .98), and *Q30* "What do you think you have IMPROVED after doing the tasks?" (α = .73).

3.5.2. Statistical analyses

3.5.2.1. Perception, lexical encoding and production

In order to answer RQ1 and RQ2 about the main effects of TBPT and task complexity on vowel perception, lexical encoding and vowel production, generalized linear mixed-effects models (GLMM) were conducted in SPSS 27.[47] When the dependent variable was categorical (e.g., accuracy), binary logistic regressions were applied to the data. Instead, when the dependent variable was continuous (e.g., reaction times and Mahalanobis distances), either identity link or gamma regressions were conducted. Gamma link function was used for gamma distributions (i.e., all positive values clearly skewed to the right in the distribution) when, after normalizing the Pearson residuals from the original variable through various mathematical operations (i.e., Sqrt, Log10), the Akaike's Information Criterion (AIC)[48] was still lower, the gaussian distribution had a better bell-shape and the model showed a better fit using the gamma link function. Fixed-effects structures were defined for each one of the models. Random-effects structures for all analyses in this study were chosen by a model fitting procedure (comparing AIC estimators across models) and random slopes were only included if they improved the model's fit (i.e., AIC decreased). Finally, Bonferroni adjustments were used for pairwise contrasts, and all parameter estimates were placed in Appendix L.

Accuracy, Mahalanobis distances of vowel *distinctiveness* gains were calculated by, first, aggregating the by-item into a by-subject datafile and, second, subtracting T1 scores from T2 or T3 scores. With respect to reaction times and Mahalanobis distances of vowel *nativelikeness*, T2 and T3 were subtracted from T1 because scores decreased over time, hence, negative values were avoided[49]. The vowel duration ratio (DR) was calculated by dividing the total milliseconds (ms) of the tense (/iː/, /æ/) by the lax (/ɪ/, /ʌ/) vowels. Finally, Spearman-rank correlations were conducted (1) to assess the association

[47] Linear mixed models were preferred to repeated measures ANOVAs because they can take into account the variability ascribed to random factors (e.g., subject, item) in addition to assess the main effects of the independent variables (e.g., accuracy). The statistical analyses conducted in the present dissertation were consulted with a statistician and were conducted following the recommendations in Larson-Hall (2016).
[48] AIC is an estimator of the relative amount of information lost by a particular model.
[49] Raw gains (rather than residualized gains) were used because the groups' performance at T1 was rather homogenous. However, follow-up analyses could compare whether the correlations between perception and production tasks' gains (ABX, FLeC, LD, DWR, DSR) would vary in strength as a function of using residualized gains, which would control for learners' initial level of performance.

between vowel *distinctiveness* and *nativelikeness* in the vowel production tasks, and (2) to examine the relationship in performance/gains between different tasks (ABX, FLeC, LD, DWR, DSR). To answer RQ2.4, frequency and duration ratios of P-LRE were submitted to linear mixed models (LMM).

3.5.2.2. ID

Prior to conducting analyses on the role of ID in L2 vowel performance and gains, to ensure the comparability of the dependent variables which were measured through different scaling systems, experiential factors (i.e., past and recent experience learning L2 English) were converted to z-scores.

Initially, one-way ANOVAs were conducted to determine whether ID scores differed as a function of group (i.e., SG, CG, CTG). So as to provide an answer to RQ3 concerning the relationship between ID, and between L2 experiential and cognitive factors and learners' performance and gains in English vowels, either non-parametric (Spearman-rank) or parametric (Pearson) correlations were employed. Standard multiple regressions were conducted to determine how much unique variance experiential and cognitive ID explained in L2 vowel perception, lexical encoding and production, as well as their overall contribution to each one of the dependent measures.

3.5.2.3. Learners' perceptions of the intervention

Responses from closed questions were first aggregated into a by-subjects dataset in SPSS, and percentages were calculated by each one of the questions. Responses from each one of the open-ended questions were coded into themes, expressed in percentages and analysed in Microsoft Excel.

In order to answer RQ4.1. about learners' beliefs about pronunciation, opinions of the TBPT project and perception of learning, frequencies were used and illustrated in either bar or pie charts. In terms of SG vs. CG perceptions' comparison (RQ4.2.), non-parametric Mann-Whitney U-tests were conducted on ordinal questions and Chi-square tests for independence on categorical questions.

Summary

The third chapter of the present dissertation described the participants of the current study, the experimental design, the stimuli and materials used for the intervention and testing, and the general study procedure. A section was dedicated to prior analyses before the main analyses and the statistical tests applied to answer the main research questions.

The objective of this study was to examine whether a focus on phonetic form during interactive task would improve L2 learners' vowels, and whether this could depend on task complexity. Ninety-two EFL Catalan/Spanish learners of English (63 experimental, 29 control) took part in the current study. These belonged to three intact classes corresponding to learners who engaged in simple-decision making tasks (*N*=31; SG), complex-decision making tasks (*N*=32; CG), or did not engage in any tasks but were pre- and post-tested (*N*=29; CTG). The three groups were comparable in terms of demographic and L2 experience.

SG and CG carried out a pre-test, a 7-week task-based intervention, an immediate post-test and a delayed post-test (11 weeks after). These learners performed 20 dyadic problem-solving tasks that were two-way, split, close, convergent and task-essential, meaning that learners had to produce the target L2 vowels (/iː- ɪ/ or /æ - ʌ/) embedded in minimal pairs distinctively to solve the task. Tasks were always preceded by a pre-task (listening comprehension and word elicitation) and followed by a post-task (consolidation games), and differed in cognitive complexity (-reasoning demands – SG; +reasoning demands – CG). The perceived cognitive demands imposed by the tasks were assessed by EFL language teachers, pilot learners and the experimental learners of this study.

A battery of tests assessed learners' perception (ABX categorical discrimination), lexical encoding (FLeC and LD), production (DWR and DSR), oral interaction (focus-on-form communicative task), and individual differences in L2 proficiency (EIT), WM (forward and backward-digit span) and attention control (ASA). A total of six English native speakers (3 male and 3 female) were recruited to record the stimuli of the intervention (80 tokens) and testing (168 tokens) materials and provide baseline data for production analyses. Three extra native speakers were used to validate the target stimuli and provide baseline data for perception tests.

Before conducting the main statistical analyses, several steps were taken: (1) the target words from the FLeC/LD task were checked for familiarity, (2) preliminary data cleaning for production tests involved deleting noises and transform frequency values from Hertz to Bark and apply a Bark-distance normalization procedure, as well as screen perceptual and production data for accuracy errors and extreme values, and (3) inter-rater reliability analyses were conducted for the occurrence of P-LRE and the classification of codes obtained from open-ended responses into themes.

Finally, statistical analyses for RQ1 and RQ2 involved conducting GLMM in SPSS 27 for each one of the models. Fixed and random effects were selected with the purpose of obtaining the best model's fit (i.e., normally-distributed residuals and lowest AIC value). Accuracy and reaction time gains were calculated in order to compare gains across groups and between tasks. Pearson/Spearman-rank correlations were used for between-task performance/gain comparisons, as well as to answer RQ3 about the relationship between experiential/cognitive factors and learners' performance and gains in vowel acquisition. Standard multiple regressions were conducted to determine how much variance ID explained in L2 perception, lexical encoding and production. Last, learners' perceptions of the TBPT intervention were analysed descriptively through percentages and between-group differences were tested through Mann-Whitney U-tests and Chi-square tests for independence.

04

RESULTS

CHAPTER 4. RESULTS

In this chapter, the analyses and results of the four main research questions are presented. The results are presented in four different sections corresponding to each one of the main research questions this study examined: the effectiveness of TBPT for L2 vowel learning (Section 4.1.); the effects of task complexity on L2 vowel learning and occurrence of pronunciation-based language-related episodes (Section 4.2.); the contributions of individual differences (i.e., L2 proficiency, WM, attention control) to L2 vowel learning (Section 4.3.) and learners' opinions after the TBPT intervention (Section 4.4.).

4.1. Effectiveness of TBPT

This section examines whether TBPT leads to improvement in L2 English vowels (RQ1), specifically, in the perceptual discrimination (RQ1.1), lexical encoding (RQ1.2) and production (RQ1.3) of L2 English vowels produced by the experimental group (simple + complex[50]) – relative to the control group. In order to address these research questions, the effects of *Contrast/Vowel* (/iː/, /ɪ/, /æ/, /ʌ/), *Word Type* (taught/untaught) and *Retention* of learning (T2-T3) are assessed[51]. Generalization to novel voices is assumed if improvement is found for the experimental group from pre- to post-test, as the stimuli voices in the tests were different from the voices that learners were exposed to during the TBPT intervention. Additionally, the relationship between perception, lexical encoding and production performance and gains is explored (RQ1.4).

[50] Data from simple and complex groups (SG/CG) are addressed together in RQ1 under the name of "experimental group".

[51] Recall that the control group did not do a delayed post-test because no changes were expected after the immediate post-test if these learners were not part of the TBPT intervention.

Prior to assessing gains in perception and production, it is important to note that the target word assessment results after the intervention revealed that learners in both experimental groups knew well over 85% of the target words after the 20-session task-based intervention, and there was no particular word that was known by less than 50% of the class. As a result, these words were included in the main analyses of perception and production.

4.1.1. Perceptual discrimination

This subsection intends to show the results concerning perceptual discrimination of the target contrastive vowels after TBPT in terms of accuracy and reaction time (RT), how this may vary as a function of *Contrast* (/iː-ɪ/, /æ-ʌ/) and *Word Type* (taught/untaught), and whether improvement was *retained* after a period of 11 weeks. For the subsequent analyses, 3 experimental subjects out of 63 were unselected due to inappropriate attitude during one of the task performances, thus, the data from 60 experimental + 29 control participants was entered.

The overall effectiveness of the TBPT intervention (RQ1) was assessed by fitting the accuracy of test trials[52] in the ABX task to a GLMM with a binary logistic regression function with **accuracy** (*0 = incorrect, 1 = correct*) as the categorical dependent variable, and *Group* (experimental/control), *Time* (T1/T2[53]), and their interactions as predictors[54]. The random-effects structure included random intercepts for *Subject* and *Item.* A random

[52] As expected, learners were significantly more accurate (M = .91, SD = .28 vs M = .69, SD = .46) and faster (M = 949.78, SD = 290.96 vs. M = 1028.77, SD = 312.96) in the control -which did not pose any difficulty in discrimination- than in test trials, $p < .05$. *Order* of stimuli presentation (AB, BA) was initially included as a random effect in the ABX analyses but, given that the covariance structure showed no significant effect of *Order* and complexified the model, we decided not to include it in the forthcoming analyses.

[53] T1= pre-test, T2=immediate post-test, T3= delayed post-test (11 weeks after the intervention)

[54] *Contrast* was not included as a factor in this first set of analyses because the main aim was to compare overall pre-test/post-test accuracy and RT differences between experimental and control groups.

slope for *Time* over *Subject* was not included because its inclusion resulted in non-convergence, thus, in uncertainty of the model validity.

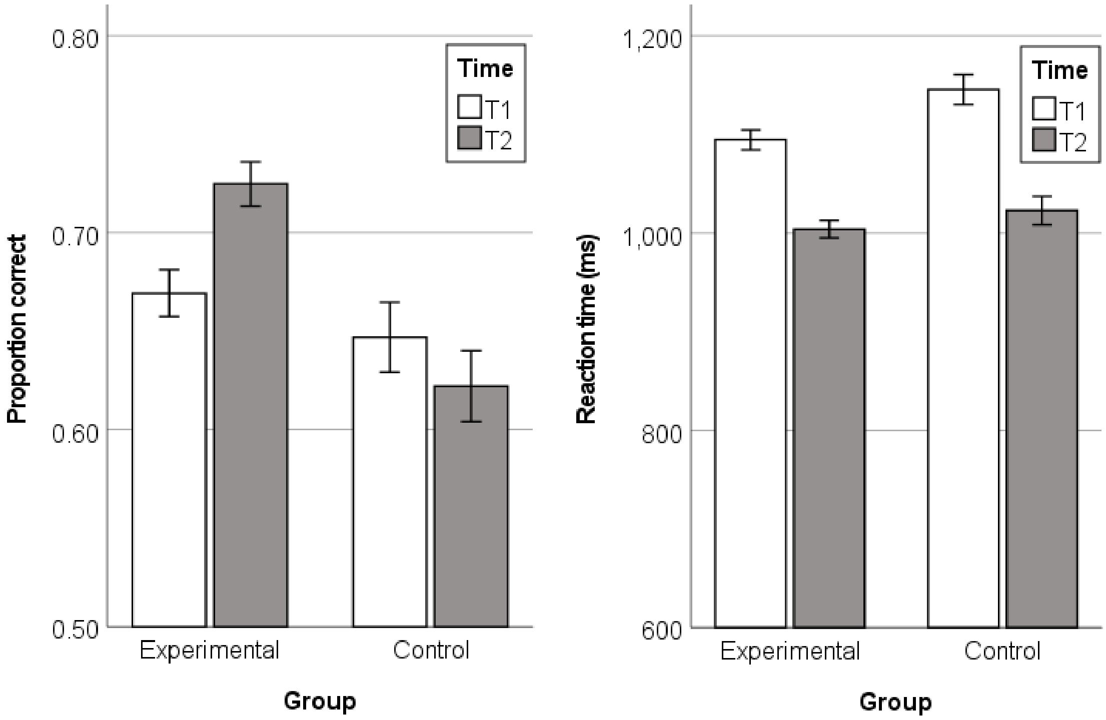


Figure 4.1. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the ABX task as a function of Group (experimental/control) and Time (T1 in white/T2 in grey). Error bars indicate 95% confidence intervals.

The analyses on discrimination accuracy yielded a significant main effect of *Group* (F [1, 17756] = 8.91, p = .003), as the experimental group obtained significantly higher accuracy scores than the control group, and a significant main effect of *Time* (F [1, 17756] = 6.41, p = .011) because overall learners' scores were higher at than T1. Crucially, *Group* interacted with *Time* (F [1, 17756] = 34.33, p < .001) because, whereas the experimental group improved the discrimination of L2 contrasts significantly (t [17756] = −7.05, p < .001) from T1 to T2, the control group did not make any improvement, in fact, their performance decreased from T1 to T2 (Figure 4.1., left panel). Whereas the experimental group improved 6% on the discrimination of English vowel contrasts, the amount of gain evidenced by the control group was below 0% (Table 4.1.). Additionally, the non-significant differences (t [17756] = −.941, p = .347) between experimental and control groups at T1 suggested they were comparable before the pedagogical intervention (Appendix L.1. for parameter estimates).

A GLMM with a gamma linking function was fit with the **RT** of correct test trials as the continuous dependent variable and *Group* (simple/complex), *Time* (T1/T2) as well as their interactions as predictors. Random intercepts by *Subject* and *Item* were included in the random-effects structure. Finally, a random slope for *Time* over *Subject* was not included because it did not improve the model's fit. The model revealed a significant main effect of *Time* (F [1, 11638] = 393.22, p < .001) as, overall, learners were faster at discriminating the L2 vowel contrast at T2 compared to T1 but not of *Group* (F [1, 11638] = .98, p = .322). Pairwise contrasts (Bonferroni-adjusted) showed that the *Group* x *Time* interaction (F [1, 11638] = 7.88, p = .005) arose because at T1 participants in the control group were slightly slower than the experimental group (Figure 4.1., right panel) but this group difference did not reach significance (t [11638] = −1.41, p = .156). Surprisingly, the control group obtained greater gains (108 ms faster) than the experimental group (91 ms faster) in the discrimination of /iː-ɪ/ and /æ-ʌ/ (Table 4.1.), but such differences did not reach significance (Appendix L.2. for parameter estimates). However, the experimental group was overall faster at T2.

Within RQ 1.1. we asked whether the variability of the dependent variable (accuracy/RT) could depend on the discrimination of each L2 vowel contrast.

Table 4.1. Accuracy (proportion correct) and RT (in milliseconds) by Group (experimental/ control), Contrast (/iː-ɪ/, /æ-ʌ/), Time (T1/T2) and gains (T2-T1) in the ABX task.

		/iː-ɪ/					
		T1		T2		*gains*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.63	0.13	0.67	0.13	0.05	0.15
	Control	0.62	0.14	0.58	0.14	-0.04	0.15
RT (ms)	Experimental	1128.67	171.29	1033.63	191.63	95.03	179.66
	Control	1171.20	166.87	1056.39	187.57	114.81	152.38

		/æ-ʌ/					
		T1		T2		*gains*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.71	0.13	0.78	0.12	0.07	0.16
	Control	0.67	0.14	0.67	0.17	-0.01	0.13
RT (ms)	Experimental	1075.88	178.55	988.80	165.78	87.08	152.38
	Control	1127.89	159.99	1027.14	192.59	100.75	207.50

Note. M = Mean, SD = Standard Deviation

Once the experimental group was selected, **accuracy** scores were submitted to a GLMM with *Time* (T1/T2), *Contrast* (/iː-ɪ/, /æ-ʌ/) as well as their interactions as fixed effects (with *Subject* and *Item* random intercepts). The model, whose parameter estimates are shown in Appendix L.3., rendered a significant effect of *Time* (F [1, 12188] = 55.87, p < .001) and *Contrast* (F [1, 12188] = 14.94, p < .001) but no significant *Time* x *Contrast* interaction (F [1, 12188] = 3.58, p = .058). These results suggest that, overall, learners were more accurate at discriminating the vowels in the /æ-ʌ/ (M = .75, SD = .13, *95% CI* [.73, .76]) than the /iː-ɪ/ (M = .65, SD = .13, *95% CI* [.63, .67]) contrast but learners in the experimental group improved the discrimination of the two L2 vowel contrasts to a similar extent after the TBPT intervention.

Applying the same GLMM model to the RT data, the model with **RT** as the continuous dependent variable showed very similar results to accuracy, hence, a significant effect of *Time* (F [1, 8186] = 256.04, p < .001) and *Contrast* (F [1, 8186] = 13.29, p < .001) but no significant *Time* x *Contrast* interaction (F [1, 8186] = 1.16, p = .281). Despite being faster at discriminating the contrasting vowels in /æ-ʌ/ (M = 1032.33, SD = 177.38, *95% CI* [1010.32, 1054.34]) than /iː-ɪ/ (M = 1081.15, SD = 187.52, *95% CI* [1057.88, 1104.41]), speed of discrimination improved significantly for both contrasts after the intervention (see Appendix L.4. for parameter estimates).

In addition, we explored whether improvement could be generalized to words that had not been part of the TBPT intervention. For this analysis, data were submitted to a GLMM with **accuracy** as the categorical dependent variable and *Time* (T1/T2), *Word Type* (taught/untaught) as well as their interactions as fixed effects. The random-effects structure included random intercepts for *Subject* and *Item*. *Contrast* (/iː-ɪ/, /æ-ʌ/) was not included as it did not predict covariance in the dependent variable (z = .65, p = .51). The accuracy analyses uncovered a significant main effect of *Time* (F [1, 12188] = 53.88, p < .001) but not of *Word Type* (F [1, 12188] = .00, p = .986) suggesting non-significant differences between taught and untaught words. A significant disordinal *Time* x *Word Type* interaction (F [1, 12188] = 5.18, p = .023) showed that learners were more accurate at discriminating words that they had been taught than untaught at T2 (Figure 4.2., left panel). This difference did not reach significance (t [12188] = .603, p = .547), suggesting that learners were able to generalize L2 vowel discrimination to words they had not been taught (See parameter estimates in Appendix L.5).

In addition, **RT** were submitted to a GLMM with *Time* (T1-T2), *Word Type* (taught/untaught) as well as their interactions as predictors, and random intercepts for *Subject* and *Item*. *Contrast* (/iː-ɪ/, /æ-ʌ/) was not included in the random-effect structure as it did not predict covariance in RT (z = .65, p = .51). Similar to the accuracy results, only *Time* (F [1, 8186] = 244.42, p < .001) emerged as a significant main effect, whereas *Word Type* (F [1, 8186] = 244.42, p < .001) and *Time x Word Type* (F [1, 8186] = 1.00, p = .316) did not. Bonferroni-adjusted pairwise contrasts indicated no significant differences between taught and untaught words at T2 in terms of speed of L2 contrast discrimination (t [8186] = −21.71, p = .129), hence, learners were similarly fast at discriminating L2 contrastive vowels embedded in novel and taught words. See Figure 4.2., right panel, and Appendix L.6. for parameter estimates.

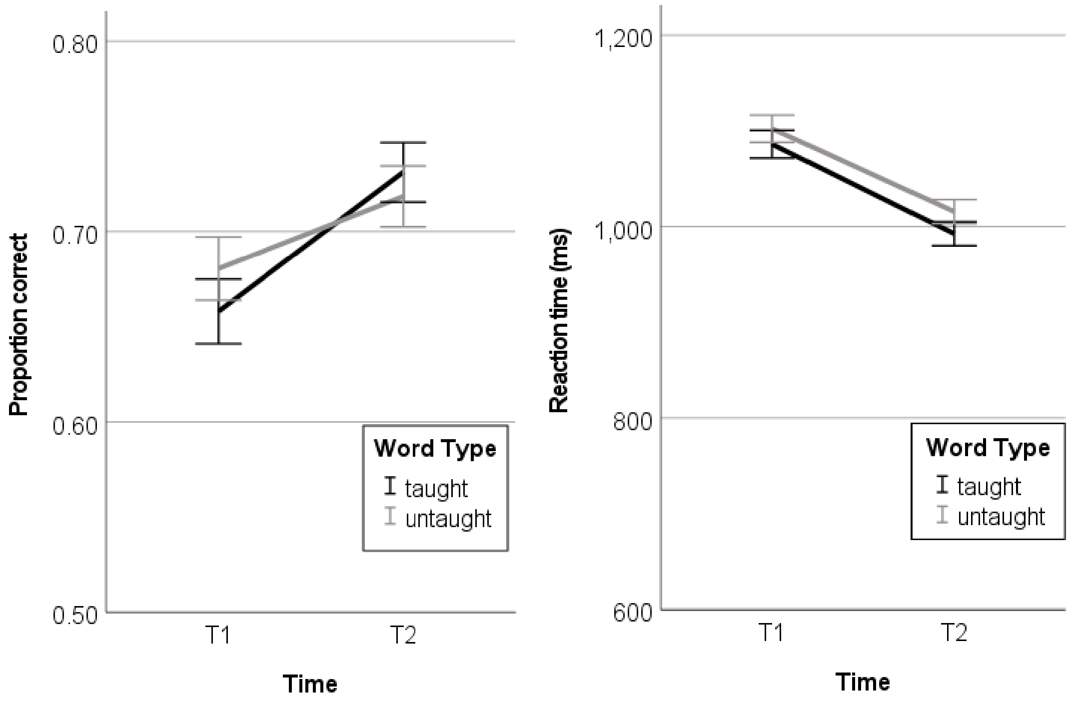


Figure 4.2. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the ABX task as a function of Word Type (taught/untaught) and Time (T1/T2). Error bars indicate 95% confidence intervals.

Finally, effectiveness of TBPT cannot be fully ensured without considering retention effects (T3) which, in the present study, were assessed 11 weeks after learners took part in the TBPT intervention. **Accuracy** scores were fitted to a GLMM with *Time* (T1/T2/T3) as predictor, and *Subject* and *Item* as random intercepts. The model, whose parameter estimates are shown in Appendix L.7., revealed a significant main effect of *Time* (F [2, 17949] = 45.84, p < .001). Figure 4.3. (left panel) depicts that learners' improvement in L2 vowel discrimination was kept from T2 to T3, as no significant differences were shown in the Bonferroni-adjusted pairwise contrasts (t [17949] = −1.79, p = .073).

Following the same GLMM structure, the model with **RT** revealed very similar results to accuracy, hence, significant main effects of *Time* (F [2, 12411] = 356.20, p < .001). Bonferroni-corrected pairwise contrasts showed significant differences between T2 and T3 (*t* [12411] = 11.07, *p*< .001) in the expected direction, thus, learners became increasingly faster at discriminating L2 contrastive vowels over time (see Figure 4.3., right panel, and parameter estimates in Appendix L.8.). Overall, the experimental group gained 7% in accuracy (5% in /iː-ɪ/ and 8.5% in /æ-ʌ/) and 151ms in speed (155ms in /iː-ɪ/ and 147ms in /æ-ʌ/) in the discrimination of L2 contrastive vowels between T1 (pre-test) and T3 (delayed post-test).

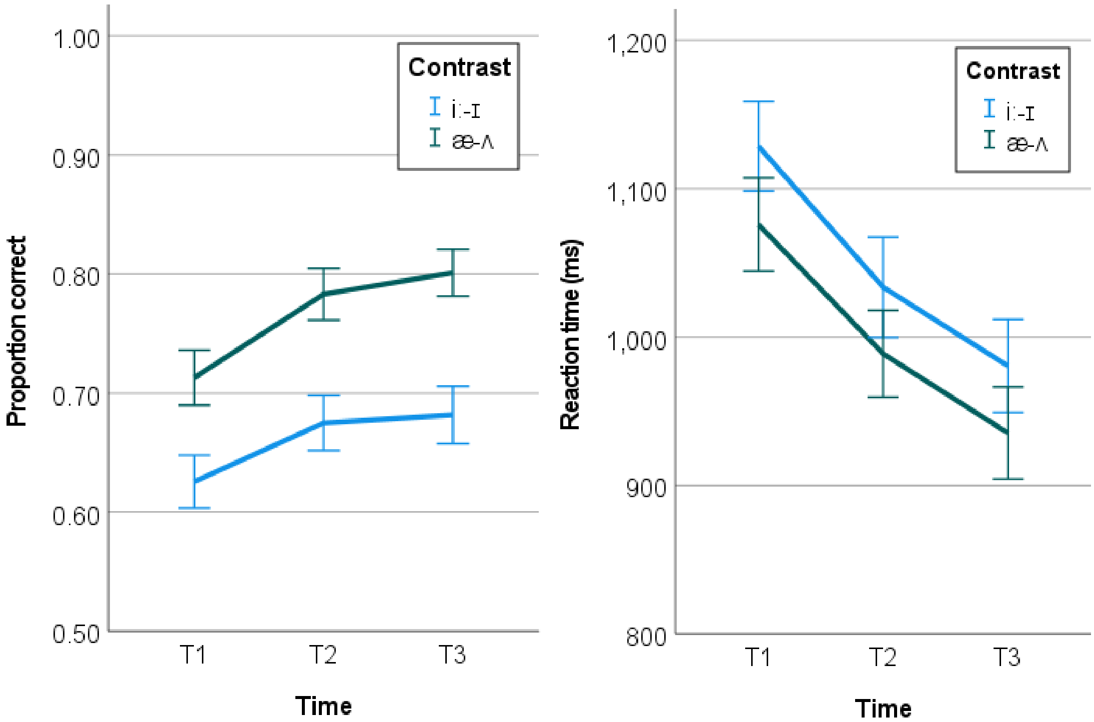


Figure 4.3. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the ABX task as a function of Time (T1/T2/T3) and Contrast (/iː-ɪ/, /æ-ʌ/). Error bars indicate 95% confidence intervals.

Summary of results:

The experimental group improved significantly in pre-lexical discrimination from T1 to T2 whereas the control group did not. Whereas both experimental and control groups became faster from T1 to T2, the experimental group's reaction times were lower at T2 than those from the control group. The experimental group became more accurate and faster in the discrimination of both L2 vowel contrasts, especially, /æ-ʌ/. Overall, improvement in vowel discrimination generalized to untaught words and was retained 11 weeks after the intervention.

4.1.2. Lexical encoding

This subsection presents the results about the lexical encoding of L2 vowel contrasts from two different tasks, namely, FLeC and LD tasks. Contrary to the pre-lexical ABX discrimination task (Section 4.1.1.), these tasks tap into the perception of L2 vowels at the lexical level, in other words, tap into learners' phonolexical representations. As assessed through a word familiarity questionnaire (Section 3.3.4.2.), learners were overall familiar with the words appearing in the FLeC and LD tasks, so no trials were removed from their datasets. Given the wide range of L2 proficiency levels of our participants, both tasks were employed and later compared to determine which one better reflected improvement in the lexical encoding of the target vowel contrasts.

To do so, we assessed learners' lexical encoding in terms of accuracy and RT, and explored how this may vary as a function of *Contrast* (/iː-ɪ/, /æ-ʌ/), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) and whether improvement was *retained* after a period of 11 weeks.[55]. Data from 91 participants was included in the FLeC task, and from 89 participants in the LD task. Excluded experimental subjects were due to task misunderstanding, distractions or inappropriate class attitudes.

4.1.2.1. FLeC

Before directly assessing the effects of *Group*, *Time* and *Vowel* on performance with /iː/-/ɪ/ and /æ/-/ʌ/ trials, a preliminary analysis was conducted in order to ascertain that the

[55] Generalization to novel words was not tested in the FLeC and LD tasks. However, provided that the words included in the FLeC and LD tasks were not "task essential" in the TBPT intervention (i.e., they could or could not use them to solve the tasks), we could assume that pre- to post-test improvement in the lexical encoding of these words suggests generalization of L2 vowel learning. However, given this limitation, further studies should include "untaught" words in the lexical decision tasks.

L2 learners in the present study indeed exhibited enhanced difficulties with trials containing the “difficult” /iː-ɪ/ and /æ-ʌ/ contrasts in comparison to the remaining filler trials containing “easy” L2 contrasts /e-uː/ and /ɔː-ɑː/. This FLeC task did not include as many filler trials as those reported in previous literature (Kojima, 2019), which is an important limitation of this experimental paradigm (See Limitations Section 6.2.). Still, assessing the expected differences between experimental and filler trials should provide evidence for the reliability of the task to detect difficulties in the lexical encoding of difficult L2 phonological contrasts.

Pre-test data were submitted (i.e., responses to filler and test trials) to a GLMM with a binary logistic regression function with **accuracy** (*0 = incorrect, 1 = correct*) as the categorical dependent variable, and *Item Type* (test/filler), as fixed effects. The random-effects structure for this model included random intercepts for *Subject* and *Item*. The model revealed significant main effects of *Item Type* (F [1, 2390] = 55.18, p < .001), showing that learners were less accurate identifying English words in trials containing /iː-ɪ/ and /æ-ʌ/ minimal pairs (M = .64, SD = .09, *95% CI* [.62, .66]) than with filler trials /e-uː/ and /ɔː-ɑː/ (M = .91, SD = .17, *95% CI* [.88, .95]). See Appendix L.9.

Parallel analyses on pre-test data were conducted on **RT** only for correct trials. A GLMM was fit with RT as the continuous dependent variable and *Item Type* (test/filler), as predictor, and random intercepts by *Subject* and *Item*. The model (parameter estimates in Appendix L.10) unfolded significant main effects of *Item Type* (F [1, 1445] = 71.06, p < .001), suggesting that learners were slower in trials containing the target difficult contrast (M = 1726.91, SD = 152.82, *95% CI* [1695.26, 1758.56]) than filler trials containing easy contrasts (M = 1550.48, SD = 207.18, *95% CI* [1507.58, 1593.39]). Overall, these analyses showed evidence of a more unreliable lexical encoding of the difficult L2 vowel contrasts, relative to the easier L2 vowel contrasts.

In order to assess the effectiveness of TBPT on the phonolexical encoding of difficult L2 sounds (RQ1.2), the **accuracy** of test trials was fitted to a GLMM with a binary logistic regression function. Accuracy (*0 = incorrect, 1 = correct*) was the categorical dependent variable, and *Group* (experimental/control), *Time* (T1/T2), and their interactions constituted the fixed effects. The random-effects structure included random intercepts for *Subject* and *Item,* but a random slope for *Time* over *Subject* could not be included due to non-convergence. The model revealed a significant main effect of *Group* (F [1, 4412] = 12.00, p < .001), given that the experimental group obtained significantly higher accuracy scores than the control group. The model also showed a significant main effect of *Time* (F [1, 4412] = 22.72, p < .001) because, in general, overall learners' scores were higher at T2 than T1. However, *Group* interacted with *Time* (F [1, 4412] = 12.72, p < .001) indicating that the effect of *Time* depended on *Group*.

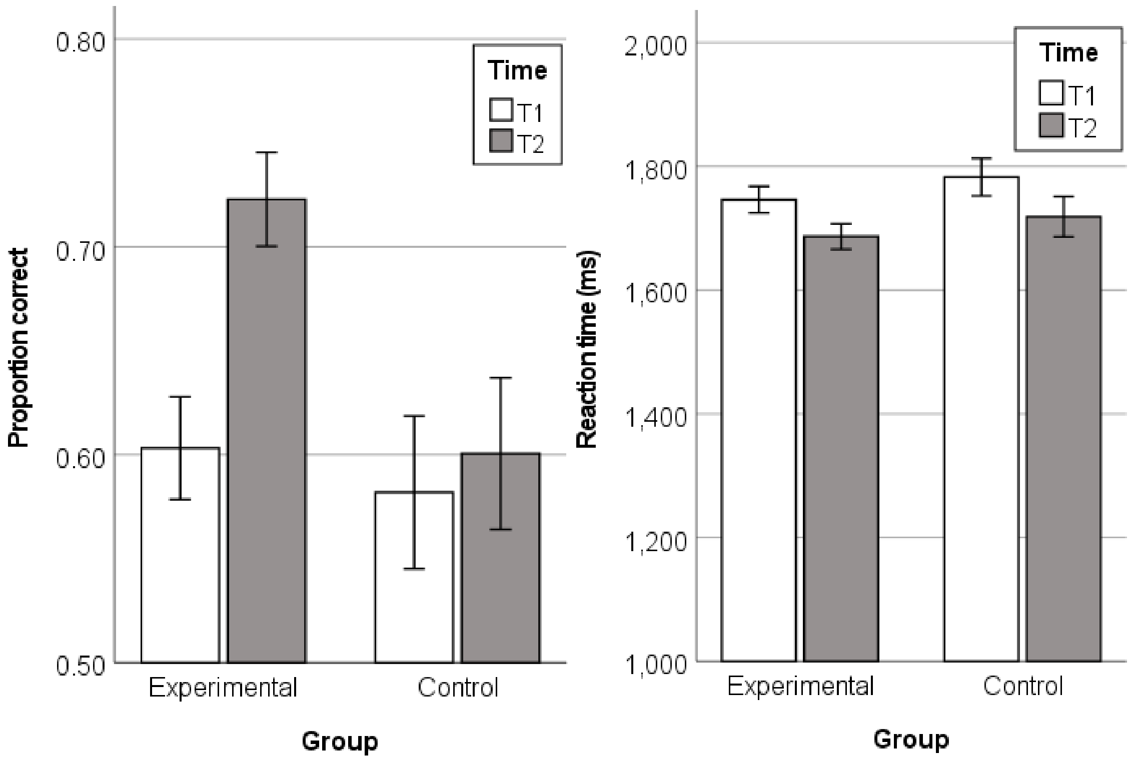


Figure 4.4. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the FLeC task as a function of Group (experimental/control) and Time (T1 in white/T2 in grey). Error bars indicate 95% confidence intervals.

Pairwise contrasts (Bonferroni-adjusted) showed that whereas the experimental group got significantly better at the phonolexical encoding of L2 contrasts from T1 to T2 (t [4412] = −6.88, $p < .001$), the control group's improvement from T1 to T2 was not significant (t [4412] = −.73, $p = .461$) (Figure 4.4., left panel). In terms of gains, whereas the experimental group gained a total of 12% in the lexical encoding of the target vowel contrasts from T1 to T2, the control group only obtained an overall 0.2% gain (Table 4.2.). In addition, experimental and control groups were initially comparable as Bonferroni-adjusted pairwise contrasts showed non-significant differences (t [4412] = .79, $p = .426$) between groups at pre-test (See Appendix L.11. for parameter estimates).

Mirroring the previous GLMM structure, **RT** of correct test responses were submitted to the model as a continuous dependent variable with a gamma linking function and *Group* (experimental/control), *Time* (T1/T2) as well as their interactions as predictors. Random intercepts by *Subject* and *Item* were included in the random-effects structure but a random slope for *Time* over *Subject* was finally not included because it did not show significance in the covariate parameters of the model and did not improve the model's fit. In this case, the model revealed a significant main effect of *Time* (F [1, 2761] = 29.75, $p < .001$), given that, in general, learners were faster at responding which was the English word at T2 than at T1. Nevertheless, the main effects of *Group* did not reach significance (F [1, 2761] = .93, $p = .334$), possibly indicating minimal differences between experimental and control groups (Figure 4.4., right panel). Despite the non-significant *Group*Time* interaction (F [1, 2761] = .10, $p = .748$), Bonferroni-adjusted pairwise contrasts showed that experimental learners were faster (albeit non-significantly) than control learners at T2 but both groups' responses were equally fast at T1 (t [2761] = −1.01, $p = .312$). The control group obtained slightly greater gains in speed

(59 ms) than the experimental group (51 ms) (See Table 4.2., & Appendix L.12. for parameter estimates.

Within RQ 1.2., we explored whether (a) each L2 vowel contrast and (b) each one of the L2 target vowels could explain variability in the dependent variable (accuracy/RT), [56] hence, whether learners' accuracy and speed of response could be affected by the presented L2 vowel contrast, but also the target vowel the English word contained. Knowing that the control group did not improve the lexical encoding of L2 contrasts over time, only the experimental group is selected henceforth.

In terms of (a) the L2 vowel contrasts, **accuracy** scores were submitted to a GLMM with *Time* (T1/T2), *Contrast* (/iː-ɪ/, /æ-ʌ/) and their interactions as fixed effects, and *Subject* and *Item* as random intercepts, but no random slope for *Time* over *Subject*. The model revealed significant main effects of *Time* (F [1, 3020] = 50.70, p < .001) but not of *Contrast* (F [1, 3020] = .13, p = .712) and no significant *Time* x *Contrast* interaction (F [1, 1953] = 1.39, p = .237) because learners were significantly more accurate at the lexical encoding of both L2 vowel contrasts from T1 (/iː-ɪ/: M =.63, SE = .04; /æ-ʌ/: M =.58, SE = .05) to T2 (/iː-ɪ/: M =.73, SE = .04; /æ-ʌ/: M =.74, SE = .04) (see parameter estimates in Appendix L.13. and Figure 4.5.).

When the same GLMM was applied to **RT** of correct test responses the model, whose parameter estimates are shown in Appendix L.14., results revealed significant main effects of *Time* (F [1, 1953] = 20.55, p < .001) but non-significant main effects of *Contrast* (F [1, 1953] = .029, p = .864) nor a significant *Time* x *Contrast* interaction (F [1, 1953] = 1.39, p = .237). Bonferroni-adjusted pairwise contrasts showed that learners became significantly faster in the correct identification of L2 words containing the two

[56] In the LD and FLeC tasks, we decided to assess L2 vowel differences to observe whether there could be any asymmetries ascribed to vowel acoustics.

vowel contrasts, from T1 (/iː-ɪ/: *M* = 1759.47, *SE* = 41.79; /æ-ʌ/: *M* =1734.70, *SE* = 41.53) to T2 (/iː-ɪ/: *M* = 1682.15, *SE* = 39.63; /æ-ʌ/: *M* =1689.63, *SE* = 39.82) (Figure 4.5.).

Table 4.2. Accuracy (proportion correct) and RT (in milliseconds) by Group (experimental/ control), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2) and gains (T2-T1) in the FLeC task.

		/iː/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.58	0.20	0.66	0.19	0.08	0.28
	Control	0.61	0.21	0.53	0.21	-0.08	0.20
RT (ms)	Experimental	1737.01	234.44	1737.40	223.04	-4.19	279.70
	Control	1765.24	191.65	1729.50	263.35	35.74	325.40
		/ɪ/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.67	0.23	0.78	0.18	0.10	0.19
	Control	0.63	0.23	0.67	0.18	0.04	0.30
RT (ms)	Experimental	1780.91	232.58	1647.16	224.58	138.46	238.74
	Control	1799.79	178.48	1757.07	182.49	42.72	194.51
		/æ/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.62	0.20	0.72	0.21	0.10	0.25
	Control	0.61	0.23	0.66	0.17	0.05	0.23
RT (ms)	Experimental	1714.54	261.10	1703.19	229.52	11.35	273.24
	Control	1778.06	205.16	1649.13	182.17	128.92	224.89
		/ʌ/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.53	0.21	0.73	0.21	0.20	0.28
	Control	0.48	0.18	0.55	0.25	0.07	0.29
RT (ms)	Experimental	1776.42	202.40	1715.87	264.47	56.67	319.66
	Control	1773.72	274.33	1713.09	232.12	28.29	289.85

Note. M = Mean, SD = Standard Deviation

In terms of each one of the vowels (b), **accuracy** scores were submitted to a GLMM with *Time* (T1/T2), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) as well as their interactions as fixed effects. The random-factor structure comprised *Subject* and *Item* as intercepts. Convergence errors appeared when the random-factor structure included a random slope for *Time* over *Subject*, thus, it was not included. The model revealed significant main effects of *Time* (*F* [1, 3016] = 50.22, *p* < .001) but not of *Vowel* (*F* [3, 3016] = .88, *p* =

.450) and no significant *Time* x *Vowel* interaction (F [3, 3016] = 1.96, p = .117) (see parameter estimates in Appendix L.15.). Pairwise contrasts (Bonferroni-corrected) showed that learners' accuracy was significantly higher at T2 than T1 with all vowels: /iː/ (t [3016] = −2.37, p = .018), /ɪ/: (t [3016] = −3.07, p = .002), /æ/: (t [3016] = −2.91, p = .004), /ʌ/: (t [3016] = −5.43, p < .001), but especially /ɪ/ and /ʌ/ (Table 4.2).

In addition, **RT** of correct test responses were fitted to a GLMM with *Time* (T1/T2), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) as well as their interactions as predictors (only *Subject* and *Time* as intercepts). The model, whose parameter estimates are shown in Appendix L.16., rendered significant main effects of *Time* (F [1, 1949] = 20.11, p < .001) but non-significant main effects of *Vowel* (F [3, 1949] = .233, p = .873). Crucially, the *Time* x *Vowel* interaction was significant (F [3, 1949] = 3.54, p = .014) because, as the Bonferroni-adjusted pairwise contrasts showed, learners became significantly faster at identifying English words containing /ɪ/ (t [1949] = 4.96, p < .001) and /ʌ/ (t [1949] = 2.24, p = .025) from T1 to T2 (Table 4.2), but not those having /iː/ and /æ/ (p > .05).

Finally, in order to assess whether improvement in the phonolexical encoding of difficult L2 phonological contrasts remained over time, retention effects (T3) were assessed 11 weeks after learners finished the TBPT intervention. For this analysis, data were submitted to a GLMM with **accuracy** as the categorical dependent variable and *Time* (T1/T2/T3), as fixed factor. *Subject* and *Item* were random intercepts. The model revealed a significant main effect of *Time* (F [2, 4509] = 64.38, p < .001) (see parameter estimates in Appendix L.17.). As shown in Figure 4.5. (left panel), not only did learners retain their L2 phonolexical knowledge at T3, but they also improved significantly (t [4509] = −3.87, p < .001) from T2 to T3.

Parallel analyses were conducted with **RT** of correct responses. Maintaining the same GLMM structure, the model, whose parameter estimates can be shown in Appendix

L.18., rendered significant main effects of *Time* (*F* [2, 2980] = 14.29, *p* < .001). Bonferroni-adjusted pairwise contrasts showed no significant differences between T2 and T3 (*t* [2980] = -.07, *p*= .944) indicating equal speed of response with the target L2 vowel contrasts 11 weeks after the intervention (see Figure 4.5., right panel).

To sum up, the experimental group gained around 18% in accuracy (16.5% in /iː-ɪ/ and 19.5% in /æ-ʌ/) and 51ms in speed (59ms in /iː-ɪ/ and 44ms in /æ-ʌ/) in the phonolexical encoding of L2 contrasts between pre-test and delayed post-test.

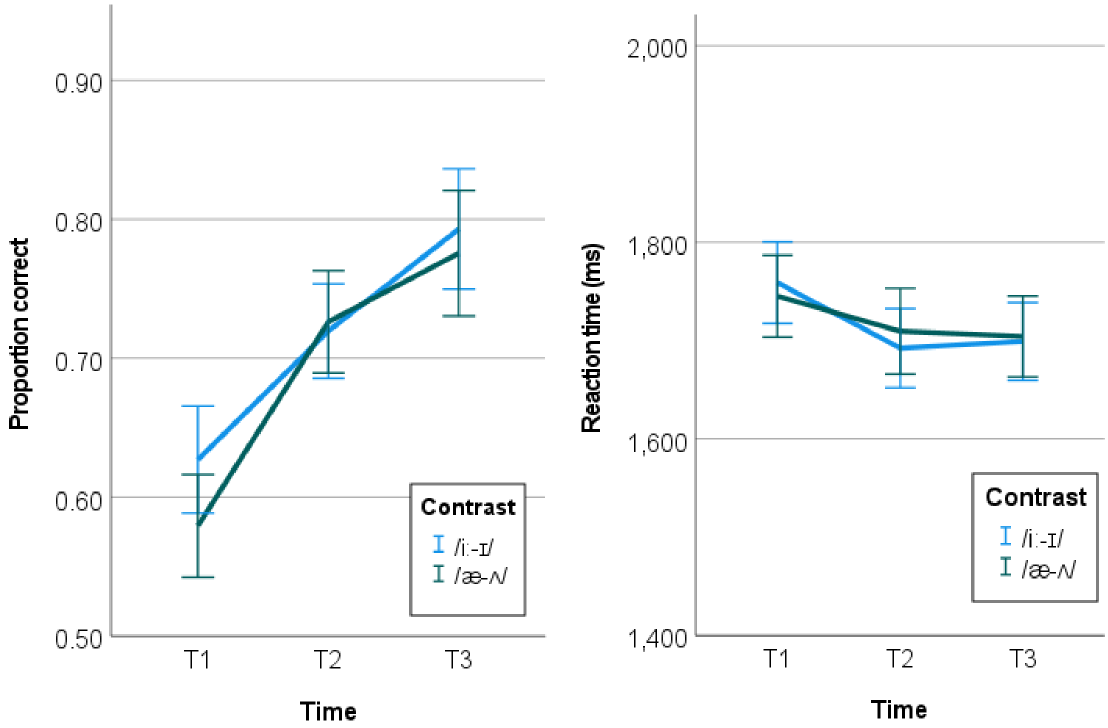


Figure 4.5. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the FLeC task as a function of Time (T1/T2/T3) and Contrast (/iː-ɪ/, /æ-ʌ/). Error bars indicate 95% confidence intervals.

4.1.2.2. LD

In addition to the FLeC task, learners also conducted a more demanding version which also tapped into the lexical encoding of difficult L2 vowel contrasts[57]. First of all, in accordance with previous studies (Amengual, 2016; Darcy & Thomas, 2019; Llompart, 2021a, 2021b), only responses to nonwords (i.e., a nonword rejection accuracy measure) were included in all forthcoming analyses.[58] In other words, we only considered nonword trials that learners had recognised as invented words.

Before providing an answer to RQ1.2.2. on the effectiveness of TBPT in improving learners' phonolexical encoding of L2 vowel contrasts, preliminary analyses were conducted to confirm that learners could perfectly encode L2 phonolexical contrasts that did not pose any difficulties to Catalan leaners of English (i.e., filler words containing /e/-/uː/ and /ɔː/-/ɑː/) but had difficulties with test trials containing the challenging /iː/-/ɪ/ and /æ/-/ʌ/ contrasts (i.e., the target contrasts of this study). Despite the limited number of filler trials (see Limitations Section 6.2.), greater accuracy in filler than test trials might indicate that the task was effective at detecting difficulties in the lexical encoding of challenging L2 vowel contrasts.

Pre-test nonword rejection **accuracy** scores (*0 = incorrect, 1 = correct*) were submitted to a GLMM with a binary logistic regression function, with *Item Type* (test/filler), as predictor. The random-effects structure included random intercepts for *Subject*. *Item* was not included in the random-effects structure as it did not predict covariance in the dependent variable (z = .70, p = .48). The model, whose parameter

[57] Were gains to be found in the more demanding version of the FLeC task, it could be hypothesized that improvement in the lexical encoding of L2 contrasts from L2 intermediate learners can be captured through a cognitively demanding task.
[58] Accuracy with real words was significantly higher (65% correct; SD = 47.90) than accuracy with nonwords (36% correct; SD = 48.10), so a nonword rejection rate was chosen to better capture changes in lexical encoding.

estimates are shown in Appendix L.19., rendered significant main effects of *Item Type* (*F* [1, 4498] = 80.68, *p* < .001) because learners were significantly more accurate at rejecting nonwords belonging to "easy" filler trials (*M* = .91, *SD* = .19, *95% CI* [.87, .95]) than "difficult" test trials (*M* = .41, *SD* = .15, *95% CI* [.38, .44]).

In addition, the same analyses were conducted for **RT** only for correct trials (i.e., nonword trials recognized as nonwords). RT were submitted to a GLMM with *Item Type* (test/filler), as predictor, and random intercepts by *Subject* and *Item.* The model revealed significant main effects of *Item Type* (*F* [1, 787] = 5.89, *p* = .015) indicating that learners were significantly slower in the identification of nonwords that contained the target /iː/-/ɪ/ and /æ/-/ʌ/ contrasts (*M* = 1492.77, *SD* = 112.58, *95% CI* [1469.19, 1516.35]) than those nonwords that contained the /e/-/uː/ and /ɔː/-/ɑː/ contrasts (*M* = 1456.50, *SD* = 223.94, *95% CI* [1408.20, 1504.80]). See Appendix L.20. for the parameter estimates. In sum, these analyses provided compelling evidence that it is nonword rejection of test trials (i.e., nonwords containing /iː/, /ɪ/, /æ/, /ʌ/) that learners struggled with, rather than the identification of nonwords containing easy contrasts, in line with the findings from the FLeC task.

In order to assess the effectiveness of TBPT on the lexical encoding of challenging L2 vowel contrasts, the **accuracy** of nonword test trials in the LD task were fitted to a GLMM with a logistic linking function with accuracy as the categorical dependent variable, and *Group* (experimental/control), *Time* (T1/T2), and their interactions as fixed factors. The random-effects structure included random intercepts for *Subject* and a random slope for *Time* over *Subject*. Item was not included as a random intercept because it did not predict covariance in the dependent variable (*z* = .63, *p* = .52).

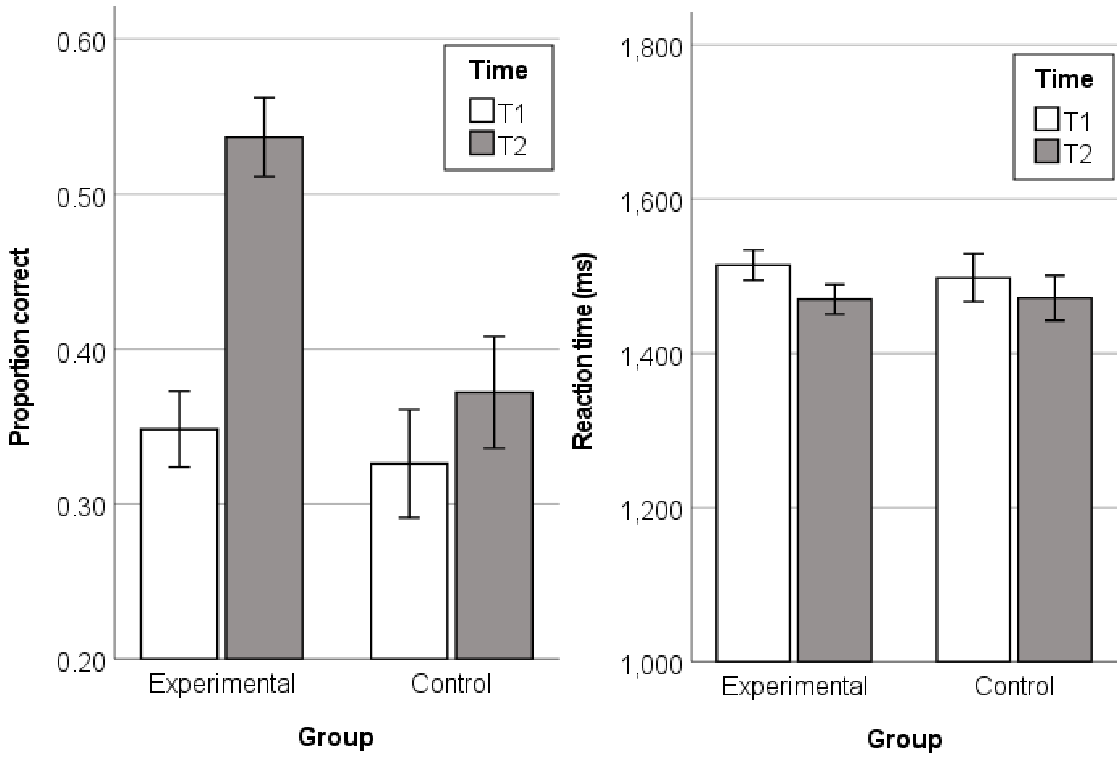


Figure 4.6. Proportion of correct nonword rejection (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the LD task as a function of Group (experimental/control) and Time (T1 in white/T2 in grey). Error bars indicate 95% confidence intervals.

The analyses on lexical encoding yielded significant main effects of *Group* (*F* [1, 4316] = 7.13, *p* = .008) given that the experimental group was significantly more accurate than the control group, and significant main effects of *Time* (*F* [1, 4316] = 7.13, *p* = .008) because, overall, learners' accuracy scores were higher at T2 than T1. The *Group* x *Time* interaction also reached significance because, whilst the experimental group significantly improved in the lexical encoding of L2 contrasts (*t* [4316] = −6.68, *p* < .001) from T1 to T2, the control group's improvement from T1 to T2 was small (*t* [4316] = −1.06, *p* = .289). See Figure 4.6. (left panel) and parameter estimates in Appendix L.21. The experimental group gained 18.8% from T1 to T2 in the lexical encoding of the target L2 contrasts, whereas the control group only obtained a 0.05% gain. Since no differences existed between groups (experimental/control) at T1 (*t* [4316] = .41, *p* = .676), changes in accuracy produced at T2 are likely to be due to the effects of the TBPT intervention.

In addition, **RT** of correct nonword test trials in the LD task were fitted to a GLMM with a gamma linking function and *Group* (experimental/control), *Time* (T1/T2), and their interactions, as predictors. The random-effects structure included random intercepts for *Subject* and a random slope for *Time* over *Subject,* as it improved the model's fit. *Item* was not included as a random intercept because including it resulted in non-convergence. The model, whose parameter estimates are provided in Appendix L.22., rendered significant main effects of *Time* (F [1, 1569] = 6.12, p = .013), because all learners had become faster at rejecting the nonword from T1 to T2. However, neither *Group* (F [1, 1569] = .23, p = .625) nor *Group x Time* interactions were significant (F [1, 1569] = .19, p = .657). Overall, RT differences between experimental and control groups were very small. Nevertheless, pairwise contrasts (Bonferroni-adjusted) revealed that the experimental group (but not the control group: t [1569] = 1.22, p = .220) was significantly faster (t [1569] = 2.58, p = .010) from T1 to T2 at nonword rejection (Figure 4.6., right panel). In fact, the experimental group obtained greater gains in speed of nonword rejection (30 ms) than the control group (16 ms) for both L2 target contrasts. Both groups produced similar RT at T1, which made them comparable for the present study.

As part of RQ 1.2., the differential effect of (a) L2 vowel contrast and (b) each L2 target vowel on nonword rejection accuracy and RT was explored. Given that the control group did not show changes in lexical encoding, only the experimental group is included in the analyses henceforth.

As for differences in the lexical encoding of L2 vowel contrasts, **accuracy** scores were submitted to a GLMM with accuracy as the categorical variable. In this case, *Time* (T1/T2), *Contrast* (/iː-ɪ/, /æ-ʌ/) and their interactions were entered as fixed effects. The random-factor structure included an intercept for *Subject* and a random slope for *Time* over *Subject* (z = 3.89, p < .001). The model showed significant main effects of *Time* (F

[1, 2924] = 28.85, $p < .001$) and *Contrast* (F [1, 2924] = 16.89, $p < .001$), but no significant *Time* x *Contrast* interaction (F [1, 2924] = 3.38, $p = .066$). Bonferroni-adjusted pairwise contrasts showed that learners were generally more accurate at rejecting nonwords containing /æ-ʌ/ than /iː-ɪ/ at T2 (t [2924] = −4.12, $p < .001$) but not at T1 (t [2924] = −1.64, $p = .099$). However, significant improvement happened in the lexical encoding of both contrasts from T1 (/iː-ɪ/: M =.32, SE = .03; /æ-ʌ/: M =.36, SE = .04) to T2 (/iː-ɪ/: M =.51, SE = .04; /æ-ʌ/: M =.63, SE = .04) (see Figure 4.7., and parameter estimates in Appendix L.23.).

Applying the same GLMM to **RT** of correct nonword test trials, the model rendered significant main effects of *Time* (F [1, 1087] = 10.08, $p < .001$), but non-significant main effect of *Contrast* (F [1, 1087] = 1.79, $p = .181$), nor a significant *Time x Contrast* interaction (F [1, 1087] = 1.00, $p = .316$). Interestingly though, Bonferroni-adjusted pairwise contrasts indicated that learners became significantly faster at rejecting nonwords from T1 to T2, when the tested contrast was /æ-ʌ/ (t [1087] = 3.11, $p = .002$), relative to /iː-ɪ/. (see Figure 4.7., and parameter estimates in Appendix L.24.).

Table 4.3. Accuracy (proportion correct nonword rejection) and RT (in milliseconds) by Group (experimental/control), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2) and gains (T2-T1) in the LD task.

		/iː/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.32	0.18	0.48	0.30	0.16	0.32
	Control	0.30	0.19	0.38	0.17	0.08	0.17
RT (ms)	Experimental	1533.02	195.72	1496.14	171.58	36.06	238.78
	Control	1488.52	201.37	1503.09	200.12	26.79	223.84
		/ɪ/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.34	0.21	0.50	0.35	0.17	0.36
	Control	0.28	0.18	0.24	0.18	-0.05	0.20
RT (ms)	Experimental	1507.12	178.38	1483.32	218.52	30.74	280.32
	Control	1530.56	203.91	1544.57	203.76	0.57	211.70

		/æ/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.45	0.21	0.64	0.30	0.19	0.31
	Control	0.42	0.19	0.55	0.23	0.13	0.25
RT (ms)	Experimental	1496.83	177.24	1453.66	190.73	46.66	226.07
	Control	1478.43	176.14	1427.05	194.71	44.20	218.51
		/ʌ/					
		T1		T2		*gain*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Accuracy	Experimental	0.29	0.18	0.52	0.35	0.23	0.37
	Control	0.30	0.19	0.33	0.21	0.02	0.24
RT (ms)	Experimental	1541.26	163.31	1517.60	209.71	6.12	194.76
	Control	1504.92	184.65	1501.04	192.36	-6.54	238.61

Note. M = Mean, SD = Standard Deviation

Accuracy scores were submitted to a GLMM with accuracy as the categorical variable. *Time* (T1/T2), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) as well as their interactions were introduced as fixed effects. The random-factor structure included an intercept for *Subject* and a random slope for *Time* over *Subject* (z = 3.90, p < .001). The model, whose parameter estimates can be shown in Appendix L.25., revealed significant main effects of *Time* (F [1, 2920] = 28.86, p < .001) and *Vowel* (F [3, 2920] = 16.51, p < .001); however, no significant *Time* x *Vowel* interaction (F [3, 2920] = 1.34, p = .258) was found. Bonferroni-adjusted pairwise contrasts indicated that learners' accuracy was significantly higher at T2 than T1 with all vowels (Table 4.3.): /iː/ (t [2920] = −3.39, p = .001), /ɪ/: (t [2920] = −3.73, p < .001), /æ/: (t [2920] = −4.87, p < .001), /ʌ/: (t [2920] = −5.33, p < .001).

Parallel analyses were conducted with **RT** of correct nonword test trials. *Group* (experimental/control) and *Time* (T1/T2) were fitted to a GLMM, as predictors. The random-effects structure included random intercepts for *Subject* but no random slope for *Time* over *Subject,* as the random effect covariance appeared to be non-significant (z = 1.41, p = .157), hence, it did not improve the model's fit. The model yielded significant main effects of *Time* (F [1, 1083] = 9.72, p = .002), but non-significant main effects of

Vowel (*F* [3, 1083] = 2.27, *p* = .078) nor a *Time x Vowel* interaction (*F* [3, 1083] = .60, *p* = .615). See parameter estimates in Appendix L.26. Bonferroni-adjusted pairwise contrasts indicated no significant differences between target vowels at T1 or T2 ($p < .05$) but it seems correct rejection of nonwords containing /æ/ (*t* [1083] = −2.85, *p* = .004) was the fastest from T1 to T2 (Table 4.3.).

Finally, in order to determine whether the TBPT intervention was effective in the lexical encoding of the target L2 contrasts, retention effects (T3) were tested. **Accuracy** scores were submitted to a GLMM with *Time* (T1/T2/T3) as the main predictor. The random-factor structure included an intercept for *Subject* and a random slope for *Time* over *Subject* ($z = 5.65$, $p < .001$). A random slope for *Vowel* over *Subject* was not included because it did not improve the model's fit. The model, whose parameter estimates are presented in Appendix L.27., rendered main effects of *Time* (*F* [2, 4365] = 33.17, $p < .001$). Figure 4.7. (left panel) illustrates learners' improvement in the lexical encoding of L2 contrasts did not stop from T2 to T3, which was confirmed by Bonferroni-adjusted pairwise contrasts (*t* [4365] = −3.25, *p* = .001).

Finally, **RT** of correct nonword rejection responses were submitted to a GLMM with *Time* (T1/T2/T3) as the fixed-effects factor. The random-effects structure only included an intercept for *Subject* because a random slope for *Time* over *Subject* did not improve the model's fit. The model revealed significant main effects of *Time* (*F* [2, 1694] = 12.82, $p < .001$); however, pairwise contrasts (Bonferroni-adjusted) showed no significant differences between T2 and T3 (*t* [1694] = 1.81, *p*= .069). The fact that learners' speed remained stable in time means that learners were as fast as T2 at rejecting nonwords in the LD task (see Figure 4.7., right panel; Appendix L.28. for parameter estimates).

In short, the experimental group significantly gained around 29.5% in accuracy (27.5% in /iː-ɪ/ and 32% in /æ-ʌ/) and 58ms in speed (44ms in /iː-ɪ/ and 72ms in /æ-ʌ/) in the phonolexical encoding of L2 contrasts between pre-test and delayed post-test.

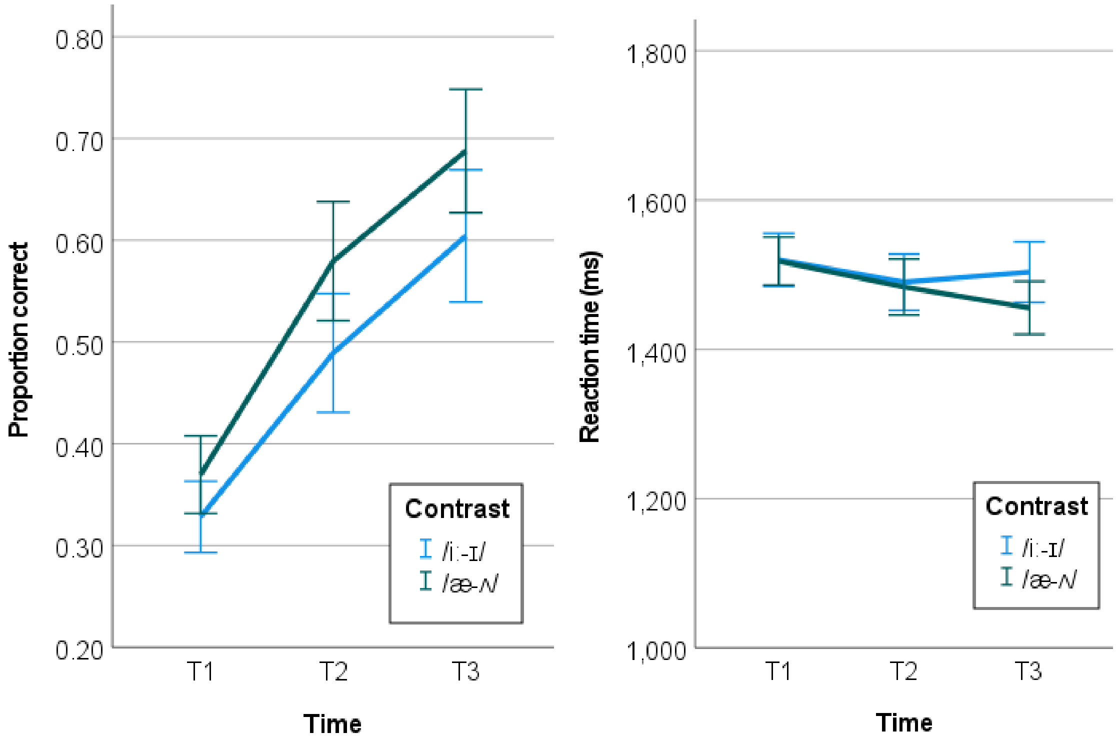


Figure 4.7. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the LD task as a function of Time (T1/T2/T3) and Contrast (/iː-ɪ/, /æ-ʌ/). Error bars indicate 95% confidence intervals.

Summary of results:

In both FLeC and LD tasks, the experimental group, but not the control group, got significantly more accurate in the lexical encoding of L2 vowel contrasts from T1 to T2. Both experimental and control groups also became faster at identifying words and rejecting nonwords in general. In addition, improvement in lexical encoding happened with both L2 vowel contrasts and across all L2 vowels. Accuracy gains were especially greater in the lexical encoding of /æ-ʌ/ and when the words in the FLeC and LD tasks contained vowel /ɪ/ and /ʌ/. Last, accuracy and RT gains were retained 11 months after the experimental group finished the intervention.

Having assessed the effectiveness of the TBPT intervention in the lexical encoding of L2 vowel contrasts under two different lexical decision tasks differing in complexity, the present study concludes that both tasks are sensitive at capturing gains in lexical encoding. Globally, higher scores were obtained in the FLeC than LD task (Kojima, 2019) and overall greater gains were obtained in the LD than FLeC tasks, yet learners were able to show improvement in the lexical encoding of L2 contrasts in both tasks.

4.1.3. Production

After examining learners' perception of L2 vowel contrasts, this subsection focuses on the results concerning the production of the target L2 vowel contrasts embedded in words (i.e., produced in isolation) and sentences (i.e., produced in meaningful contexts) coming from two different tests: a delayed word repetition (DWR) and a delayed sentence repetition (DSR) test. To assess learners' productions before and after the intervention, on the one hand, (1) formant frequencies of learners' vowel qualities were compared across the three times (T1-T2-T3), and we obtained (2) vowel quality Mahalanobis distances of *distinctiveness* (i.e., how distinct the quality of non-native vowels in each target contrast was) and *nativelikeness* (i.e., how much non-native vowel qualities approximate those of native speakers). Correlational analyses were conducted to determine (3) a relationship between vowel quality *distinctiveness* and *nativelikeness* performance and gains. On the other hand, duration (ms) and a duration ratio of learners' contrastive vowels was compared across T1-T2-T3. For vowel quality, we assessed improvement (T1-T2), how this may vary as a function of *Contrast* (/iː-ɪ/, /æ-ʌ/) or *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/), generalization to novel words (taught/untaught) and whether phonological learning was maintained after 11 weeks (i.e., T3). Generalization to novel voices was assumed if improvement was found for experimental groups at post-test. Finally, one CTG subject had to be excluded from the main production analyses due to inappropriate behaviour during DWR and DSR task performance (*TotalN* = 91).

4.1.3.1. DWR

4.1.3.1.1. Vowel quality

RQ1.3.1. asked about the effects of TBPT on L2 vowel quality changes in words elicited in isolation. To answer this question, we looked at (1) F1 and F2[59] vowel description in Bark for learners and native speakers, (2) vowel *distinctiveness* and *nativelikeness* and (3) the relationship between these two aforementioned measures.

Vowel height and frontness (B1-B2 comparison)

On the one hand, initial observations from Figure 4.8., suggest that, the experimental group's production of English /iː/ and /ɪ/ became more distinct because learners produced more centralized (i.e., lower B2) and especially lower (i.e., higher B1) productions of /ɪ/ at T3 than at T1. Instead, the control group's production of /iː/ became lower, almost undistinguishable from /ɪ/, from T1 to T2 because B1 differences were minimal. Despite making more distinction of these confusable vowels at T1, their vowel production was not very accurate. Instead, both /iː/ and /ɪ/ became more targetlike (i.e., closer to the native speaker B1 and B2 values) from T1 to T3 for experimental learners.

On the other hand, both experimental and control groups separated the English /æ-ʌ/ vowels across testing times. The control group's /ʌ/ production became slightly higher (i.e., lower B1) but less targetlike from T1 to T2. In contrast, the experimental learners' productions of /æ/ became more front (i.e., higher B2) and the productions of /ʌ/ became more centralized (i.e., lower B2) and higher (i.e., lower B1) from T1 to T3, approximating the native speaker reference points (Figure 4.8.).

[59] F1 refers to height (i.e., how high/low the tongue is from the palate when producing a monophthong) so the lower the F1 value, the higher the tongue position. F2 refers to frontness (i.e., how front/back the tongue is in the oral cavity when producing a monophthong) so the higher the F2 value, the further forward the tongue position.

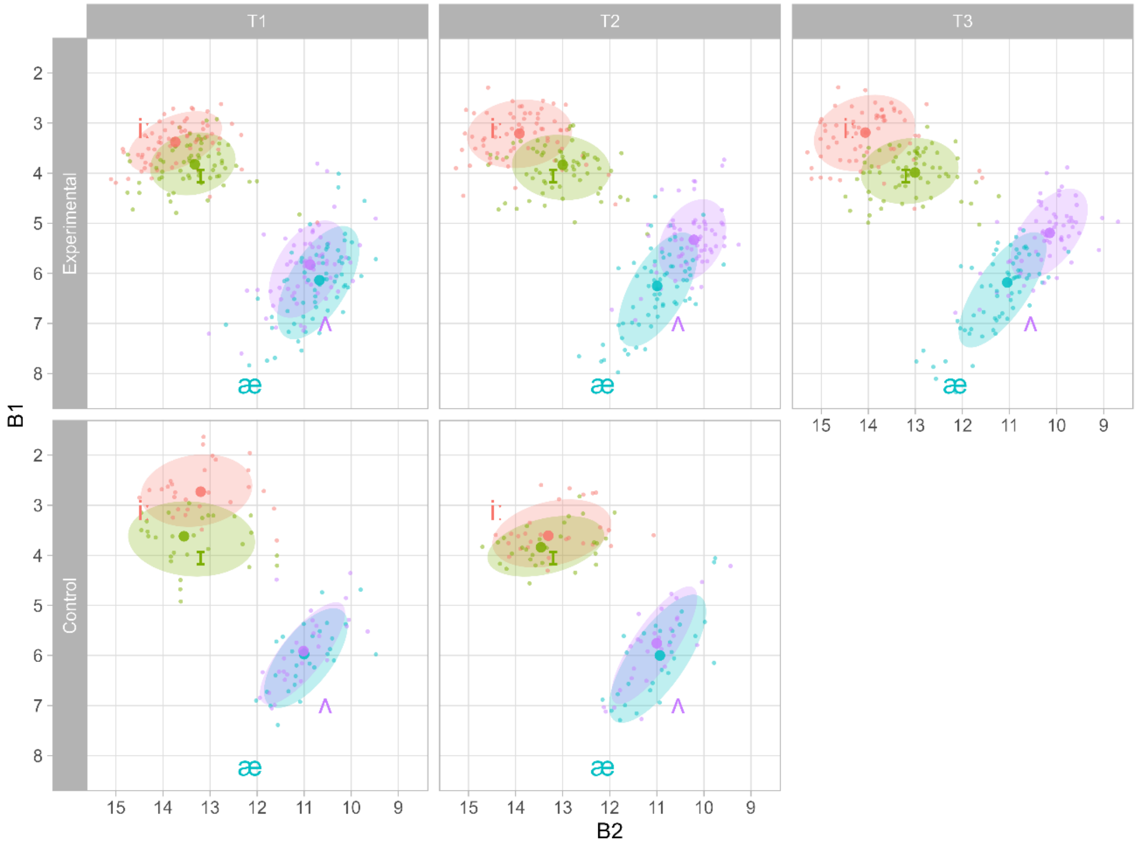


Figure 4.8. Learners' mean B1-B2 formant values (big dots) and mean B1-B2 formant values for each learner (small dots) for English vowels /iː/, /ɪ/, /æ/, /ʌ/ in word contexts. Filled ellipses represent 75% confidence intervals. Vowel transcriptions indicate mean B1-B2 formant values for native speakers of English. Graph organized by Time (T1/T2/T3) and Group (experimental/control).

Vowel distinctiveness and nativelikeness

On the one hand, the overall effectiveness of the TBPT intervention on vowel production (RQ1.3) was assessed by fitting log-transformed Mahalanobis distance scores between contrastive vowels (/iː-ɪ/ and /æ-ʌ/) of test trials[60] in the DWR task (i.e., ***vowel distinctiveness***) to a GLMM with distance as the continuous dependent variable, and *Group* (experimental/control), *Time* (T1/T2), and their interactions as predictors. The random-effects structure included random intercepts for *Subject* and *Item*. A random

[60] Practice trials were checked for learners' understanding of the task but were eliminated in all subsequent production analyses.

slope for *Time* over *Subject* could not be included because its inclusion resulted in non-convergence.

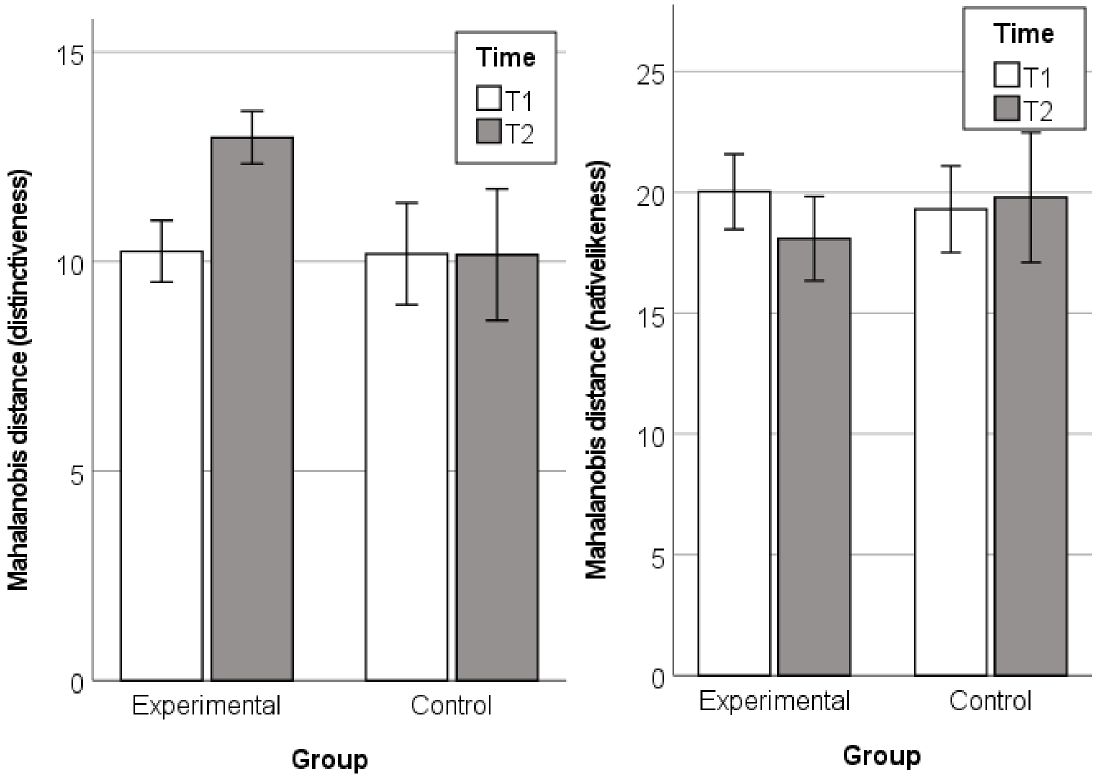


Figure 4.9. Mahalanobis distances (*distinctiveness*) on the left panel, and Mahalanobis distances (*nativelikeness*) on the right panel, produced in words in isolation (DWR). Graphs organized by *Group* (experimental, control) and *Time* (T1 in white/T2 in grey). Error bars show 95% confidence intervals.

The model rendered non-significant main effects of *Group* (*F* [1, 8684] = .63, *p* = .426), but significant main effects of *Time* (*F* [1, 8684] = 7.60, *p* = .006) because, in general, learners' contrastive vowels became significantly more distinct from T1 to T2. Crucially, a significant *Group* x *Time* interaction (*F* [1, 8684] = 6.40, *p* = .011) revealed that, whereas the experimental group produced vowels significantly more separated (*t* [8684] = −5.04, *p* < .001) from T1 to T2 (Figure 4.9., left panel), none of the control group's contrasting vowels distinguished significantly from T1 to T2. As seen in the means of Table 4.4, the experimental group (2.7 SD[61]) gained more than the control group (0.1 SD)

[61] Mahalanobis distances are measured in Standard Deviations (SD).

in the separation of vowels /æ-ʌ/ and /iː-ɪ/. Furthermore, experimental and control groups were shown to produce similar Mahalanobis distances between contrastive vowels at T1 differences (t [8684] = .108, p = .914), suggesting they were comparable before the pedagogical intervention (Appendix L.29. for parameter estimates).

On the one hand, the effectiveness of the TBPT intervention on vowel production (RQ1.3) was assessed by fitting the log-transformed Mahalanobis distance between non-native and native vowels (/iː/, /ɪ/, /æ/, /ʌ/) of test trials in the DWR task (i.e., ***vowel nativelikeness***) to a GLMM with distance as the continuous dependent variable, and *Group* (experimental/control), *Time* (T1/T2), and their interactions as fixed factors, and *Subject* and *Item* as random intercepts. The model, whose parameter estimates can be shown in Appendix L.30., yielded non-significant effects of *Group* (F [1, 8684] = .18, p = .672), *Time* (F [1, 8684] = .19, p = .660) nor a significant *Group x Time* interaction (F [1, 8684] = 2.76, p = .097). Nevertheless, Bonferroni-adjusted pairwise contrasts revealed that, overall, the experimental group (but not the control group) significantly shortened Mahalanobis distances with respect to native speakers' vowels (i.e., became more accurate) between testing times (t [8684] = 1.98, p = .047; see Figure 4.9., right panel). Overall, the experimental group's vowel productions got significantly closer to native speaker vowel productions (2 SD) than the control group's (-0.4 SD) (see mean gains in Table 4.5.). Last, experimental and control groups were comparable at T1 in terms of vowel nativelikeness: t [8684] = .25, p = .805).

Within RQ 1.3.1., we explored potential differences in vowel *distinctiveness* and *nativelikeness* as a function of L2 vowel contrast or L2 vowel, respectively. Only the experimental group was included in the forthcoming analyses.

Table 4.4. Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) by Group (experimental/control), Contrast (/iː-ɪ/, /æ-ʌ/), Time (T1/T2) and gains (T2-T1) in the DWR task.

	/iː-ɪ/								
	T1			T2			*Gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	11.65	6.97	13.43	13.28	9.57	11.34	1.63	2.60	13.06
Control	8.93	6.36	8.83	6.73	3.81	9.15	-2.20	-2.55	9.99
	/æ-ʌ/								
	T1			T2			*Gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	8.84	6.72	8.33	12.64	9.61	9.04	3.81	2.89	9.28
Control	11.40	6.37	15.42	13.48	4.50	26.50	2.08	-1.87	20.35

Note. M = Mean, Mdn = Median, SD = Standard Deviation

Table 4.5. Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) by Group (experimental/control), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2) and gains (T2-T1) in the DWR task.

	/iː/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	16.64	16.40	7.41	15.05	13.49	8.64	1.58	2.91	7.72
Control	20.68	19.22	10.81	18.82	14.63	12.73	1.86	4.59	11.45
	/ɪ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	31.21	16.27	66.89	34.17	15.27	60.63	-2.96	1.00	68.08
Control	22.76	13.42	30.17	20.90	10.30	44.86	1.86	3.12	53.57
	/æ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	21.37	16.65	20.06	15.78	11.22	13.76	5.60	5.43	19.72
Control	17.75	15.31	9.87	15.86	13.97	8.04	1.89	1.34	10.31
	/ʌ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	10.91	10.53	4.74	7.35	5.71	4.90	3.56	4.82	5.38
Control	16.21	9.62	16.26	23.60	9.79	41.64	-7.39	-0.17	31.82

Note. M = Mean, Mdn = Median, SD = Standard Deviation

Vowel ***distinctiveness*** scores were submitted to a GLMM with *Time* (T1/T2), *Contrast* (/iː-ɪ/, /æ-ʌ/) as well as their interactions as fixed effects. In the random-effects structure, *Subject* and *Item* were random intercepts. The model revealed significant main effects of *Time* (F [1, 6044] = 32.77, p < .001), but not of *Contrast* (F [1, 6044] = .33, p = .564), indicating that vowel *distinctiveness* was similar for both L2 vowel contrasts. Interestingly, the *Time x Contrast* interaction (F [1, 6044] = 5.22, p = .022) reached significance. Pairwise contrasts (Bonferroni-corrected) indicated that the task-based intervention helped learners produce less overlap between /iː/ and /ɪ/ (t [6044] = -2.59, p = .010) and, especially, /æ/-/ʌ/ (t [6044] = -5.71, p < .001). See Table 4.4. and Appendix L.31. for parameter estimates.

Parallel analyses were conducted for vowel ***nativelikeness,*** where a GLMM with *Time* (T1/T2), *Vowel* (/iː, ɪ, æ, ʌ/) and their interactions as fixed effects (*Subject* and *Item* as random intercepts), showed significant effects of *Time* (F [1, 6040] = 4.01, p = .045), *Vowel* (F [3, 6040] = 78.85, p < .001) and a *Time x Vowel* (F [3, 6040] = 2.68, p = .045) interaction. On the one hand, the main effect of *Vowel* showed significant main differences across all vowels (p < .05); on the other hand, the interaction showed that improvement (T1-T2) depended on the target vowels. Bonferroni-adjusted pairwise contrasts indicated that only learners' productions of /æ/ (t [6040] = 2.65, p = .008) and /ʌ/ (t [6040] = 1.95, p = .049) significantly approximated the values of native speakers' vowel productions at T2. Nevertheless, learners' productions of /iː/ (t [6040] = .883, p = .377) and /ɪ/ (t [6040] = -1.12, p = .262) did not get significantly more target-like from T1 to T2.[62] See Table 4.5. and Appendix L.32. for parameter estimates.

[62] When the GLMM included *Contrast* (/iː-ɪ/ and /æ-ʌ/), instead of *Vowel*, as predictors, there were significant effects of *Contrast* (F [1, 6044] = 84.89, p < .001) and a *Time x Contrast* interaction (F [1, 6044] = 5.43, p = .020). Nevertheless, Bonferroni-adjusted pairwise contrasts indicated that the vowel contrast /æ-ʌ/ became significantly more target-like from T1 to T2 (t [6044] = 3.03, p = .002), whereas learners' productions of /iː-ɪ/ did not change from T1 (t [6044] = -.16, p = .868) to T2.

In order to assess whether improvement in vowel *distinctiveness* and *nativelikeness* happened in different lexical contexts, we explored whether improvement could be generalized to words that had not been part of the TBPT intervention (i.e., untaught). The experimental group data were submitted to a GLMM with vowel ***distinctiveness*** as the dependent variable and *Time* (T1/T2), *Word Type* (taught/untaught) as well as their interactions as fixed effects. The random-effects structure included random intercepts for *Subject* and *Item, and* a random slope for *Contrast* over *Item,* which improved the model's fit. The model (parameter estimates in Appendix L.33.), yielded significant main effects of *Time* (*F* [1, 6044] = 32.88, $p < .001$); however, neither the main effects of *Word Type* (*F* [1, 6044] = 1.25, $p = .262$) nor the *Time x Word Type* interaction (*F* [1, 6044] = .31, $p = .573$) reached significance. Overall, learners increased the distance between L2 confusable vowels in both taught and untaught words. The L2 vowel learning happening with taught (*t* [6044] = -4.56, $p < .001$) extended to untaught (*t* [6044] = -3.78, $p < .001$) words across testing times (T1-T2). See Figure 4.10., left panel.

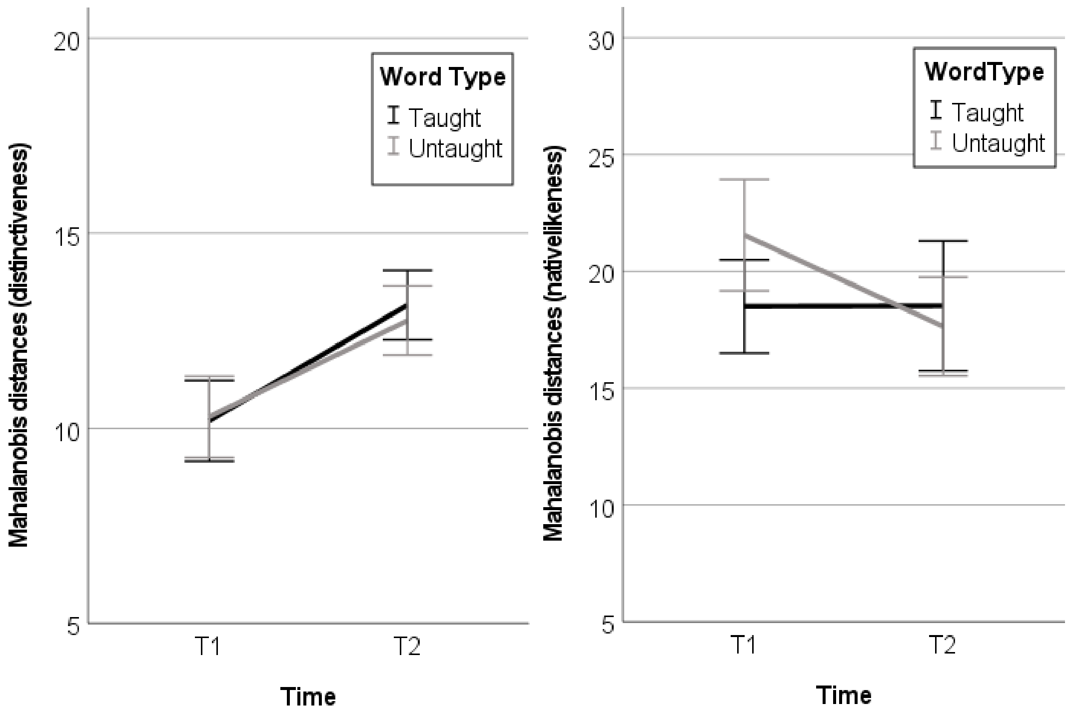


Figure 4.10. Mahalanobis distances between non-native contrastive vowels (left panel) and Mahalanobis distances between native and non-native vowels (right panel) in the DWR task as a function of Word Type (taught/untaught) and Time (T1/T2). Error bars indicate 95% confidence intervals.

As for ***nativelikeness,*** parallel analyses were carried out with *Time* (T1/T2), *Word Type* (taught/untaught) and their interactions as predictors. The random structure included *Subject* and *Item,* and a random slope for *Vowel* over *Item,* which resulted in a better model fit. The model, whose parameter estimates are shown in Appendix L.34., revealed significant main effects of *Time* (F [1, 6040] = 5.23, p = .022), non-significant main effects of *Word Type* (F [1, 6040] = .51, p = .473) and a *Time x Word Type* interaction (F [1, 6040] = 4.47, p = .035). Whereas overall there were no significant differences between taught and untaught words for *nativelikeness*, the significant disordinal interaction (Figure 4.10., right panel) illustrated that at T1 learners' produced vowels which were significantly less target-like in untaught than taught words (t [6040] = -1.99, p = .046). However, improvement (i.e., greater approximation to native vowels) in *nativelikeness* happened in both kind of words.

Finally, it is relevant for this study to consider whether improvement in vowel *distinctiveness* and *nativelikeness* was kept 11 weeks after the TBPT intervention (i.e. retention effects; T2-T3). To do so, ***distinctiveness*** distance scores were submitted to a GLMM with *Time* (T1/T2/T3) as predictor. The random structure included *Subject* and *Item* as random intercepts, and a random slope for *Contrast* over *Item* (z = 6.60, p < .001). The model yielded significant main effects of *Time* (F [2, 9069] = 27.39, p < .001), see parameter estimates in Appendix L.35. As illustrated in Figure 4.11. (left panel), learners produced contrastive vowels with greater distinctiveness from T1 to T2 (t [9069] = -5.35, p < .001), and they kept separating the target contrasting vowels from T2 to T3 (t [9069] = -1.98, p = .048).

In terms of how close the non-native vowel productions were of native speakers (vowel ***nativelikeness***) at T3, a similar model with *Time* (T1/T2/T3) as fixed effect, and *Subject* and *Item* as random intercept and a random slope for *Vowel* over *Item* (z = 10.17, p < .001) was conducted. The model, whose parameter estimates are shown in Appendix L.36., rendered significant main effects of *Time* (F [2, 9069] = 13.46, p < .001). From Figure 4.11. (right panel), it can be observed that not only were leaners' vowels produced in a more native-like manner from T1 to T2 (t [9069] = 2.33, p = .020), but they kept reducing the distance with the native speaker values at T3 (t [9069] = 2.94, p = .007), indicating robust retention effects and gains over time.

Overall, the experimental group gained 3.75 SD in Mahalanobis distance in vowel *distinctiveness* (i.e., separation between non-native contrastive vowels) and 4.75 SD in Mahalanobis distance in *nativelikeness* (i.e., approximation to native speaker values).

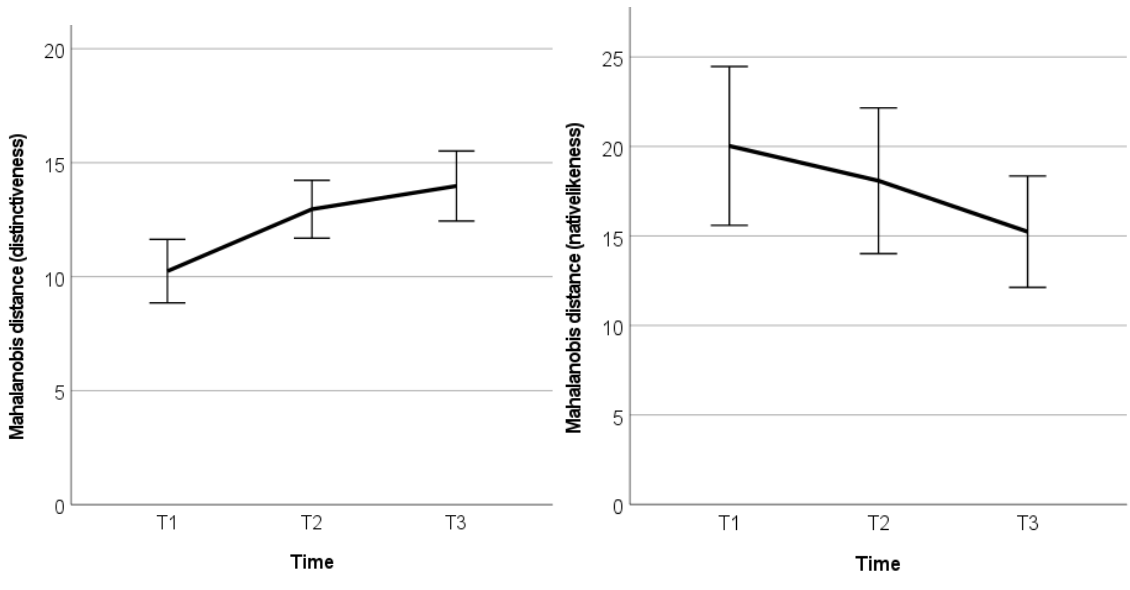


Figure 4.11. Mahalanobis distances between non-native contrastive vowels (left panel) and Mahalanobis distances between native and non-native vowels (right panel) in the DWR task as a function of Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

Correlations between vowel distinctiveness and nativelikeness

In order to assess overall vowel pronunciation improvement from the experimental group (i.e., producing the target confusable vowels more distinctively and in a more native-like manner), two-tailed Spearman rank-order correlations were run to examine the relationship between vowel *nativelikeness* (i.e., how close non-native vowel productions were of native vowel productions) and *distinctiveness* (i.e., how distant the contrastive vowels were from each other) at pre-test (T1), post-test (T2) and delayed post-test (T3) of the experimental group. As illustrated in Figure 4.12., negative correlations indicated that larger non-native vowel distances between /iː/ and /ɪ/ or /æ/ and /ʌ/ (i.e., less overlap between contrastive vowels) corresponded to smaller distances between non-native and native vowel productions (i.e., greater approximation to native speaker values). This negative relationship was significant and weak-to-moderate at T2 and T3. Interestingly, T1-T3 gains in distinctiveness were moderately correlated with gains in nativelikeness when vowels were embedded in words produced in isolation (r_s = .330, n = 63, p = .008). This suggests that learners who produced more distinct vowel qualities for each contrast after the TBPT intervention, also produced vowels that were more target-like.

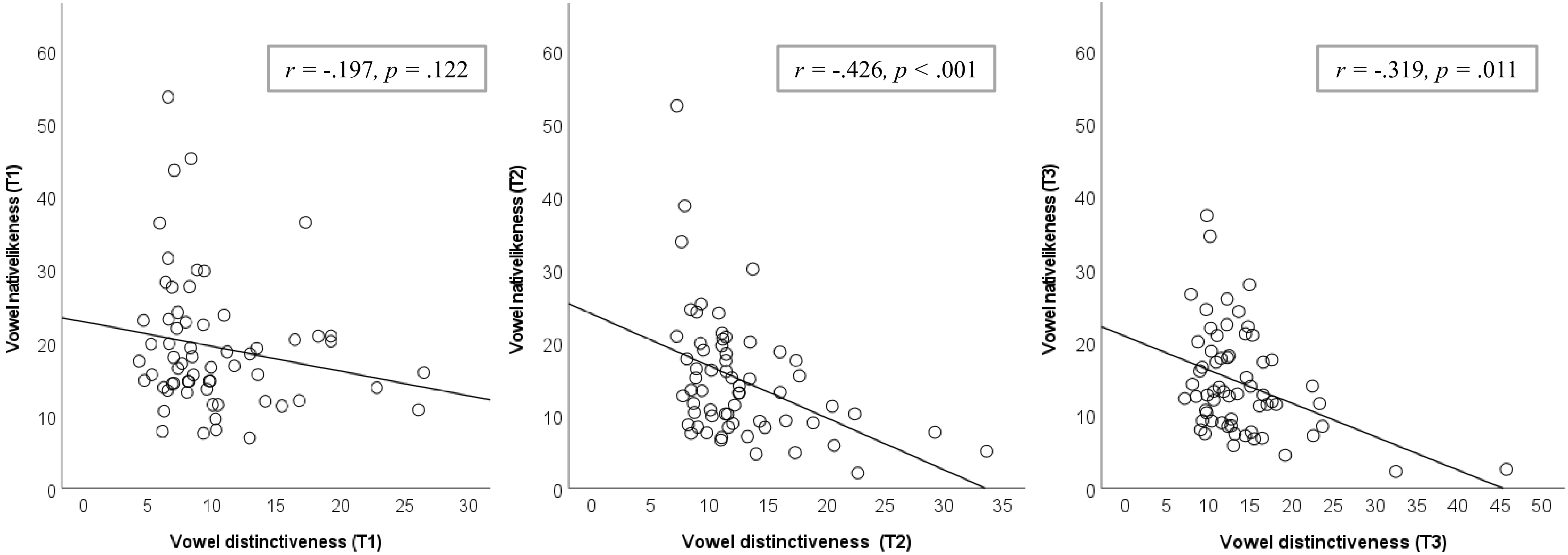


Figure 4.12. Two-tailed Spearman rank-order correlations between vowel nativelikeness (y axis) and distinctiveness (x axis) by Time (T1/T2/T3) from words produced in isolation (DWR).

4.1.3.1.2. Vowel quantity

Given that Catalan learners of English often rely on vowel quantity (i.e., duration) to distinguish similar L2 vowels, apart from changes in vowel quality assessed through Mahalanobis distances (Section 4.1.3.1.1), we observed whether learners' vowel durations changed over time as a result of the TBPT intervention.

Before assessing learners' Duration Ratio (DR) per contrast (/iː-ɪ/ - /æ-ʌ/), we first observed through two-tailed paired sample t-tests that experimental learners' duration of /iː/ was significantly larger than /ɪ/ (t [125] = 15.23, p < .001), and the duration of /æ/ was significantly larger than /ʌ/ (t [125] = 23.94, p < .001). Similarly, two-tailed paired sample t-tests with the control group showed that learners' duration of /iː/ was significantly larger than /ɪ/ (t [55] = 11.30, p < .001), and the duration of /æ/ was significantly larger than /ʌ/ (t [55] = 11.51, p < .001) (See Table 4.6.)

Table 4.6. Duration and duration ratio by Group (experimental/control) and Time (T1/T2) in the DWR task.

		Duration (ms)				Duration ratio (ms)			
		T1		T2		T1		T2	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experimental	/iː/	111.56	13.58	121.80	13.11	1.15	0.14	1.24	0.14
	/ɪ/	97.45	11.42	99.15	11.24				
	/æ/	134.02	15.55	145.23	16.11	1.20	0.10	1.24	0.12
	/ʌ/	111.83	12.97	117.67	13.98				
Control	/iː/	112.97	13.93	115.59	12.09	1.16	0.12	1.18	0.12
	/ɪ/	98.17	14.82	98.98	13.62				
	/æ/	135.20	17.06	138.91	14.43	1.19	0.14	1.20	0.12
	/ʌ/	114.61	14.09	116.82	11.76				

Note. M = Mean, SD = Standard Deviation

Secondly, we submitted DR scores to LMM where *Group*, *Time* and their interactions were submitted as fixed effects, and the random-effects structure included an intercept for *Subject*. The first model with DR for /iː-ɪ/ yielded non-significant main effects of *Group* (F [1, 178] = 1.38, p = .242) but significant effects of *Time* (F [1, 178]

= 5.85, p = .017), and the *Group* x *Time* interaction did not reach significance (F [1, 178] = 2.41, p = .122) However, Bonferroni-adjusted pairwise contrasts showed that the experimental group's DR significantly increased from T1 to T2 (t [178] = -.89, p < .001) (See Appendix L.37.). The second model with DR for /æ-ʌ/ revealed non-significant main effects of *Group* (F [1, 178] = 2.64, p = .106), *Time* (F [1, 178] = 1.47, p = .226), nor a significant *Group* x *Time* interaction (F [1, 178] = .62, p = .432). Overall, these results showed that the duration distance between /iː/ and /ɪ/ significantly increased between T1 and T2 for the experimental group, but this did not happen between /æ/ and /ʌ/. See Table 4.6. and Appendix L.38.

4.1.3.2. DSR

4.1.3.2.1. Vowel quality

In RQ1.3.2., we sought to investigate the effects of TBPT on L2 vowel quality changes in words in sentence contexts. To answer this question, we looked at (1) F1 and F2 vowel description in Bark for learners and native speakers, (2) vowel *distinctiveness* and *nativelikeness* and (3) the relationship between these two aforementioned measures.

Vowel height and frontness (B1-B2 comparison)

First of all, Figure 4.13. illustrates how the experimental group's production of English /iː/ and /ɪ/ became less similar over time because they produced higher (i.e., lower B1) and slightly more fronted (i.e., higher B2) /iː/, and lower (i.e., higher B1) and more centralized (i.e., lower B2) /ɪ/ from T1 to T3. In contrast, the control group's vowels were produced in a very similar manner from T1 to T2, only /ɪ/ became higher (i.e., lower B1) and less target-like in comparison to native speakers' production of /ɪ/. However, /iː/ and

/ɪ/ became more targetlike (i.e., closer to the native speaker B1 and B2 values) from T1 to T3 for experimental learners.

Lastly, the control group made slightly more distinctions between English /æ-ʌ/ at T1 than the experimental group; however, their /æ/ production became less targetlike (i.e., closer to their /ʌ/ production and more distant to native speakers' production of /æ/) at T2. Nevertheless, the experimental group's productions of /æ/ became slightly lower (i.e., higher B1) and more front (i.e., higher B2), and /ʌ/ became higher (i.e., lower B1) and more centralized (i.e., lower B2) from T1 to T3, closer to the native speaker B1-B2 values (Figure 4.13.).

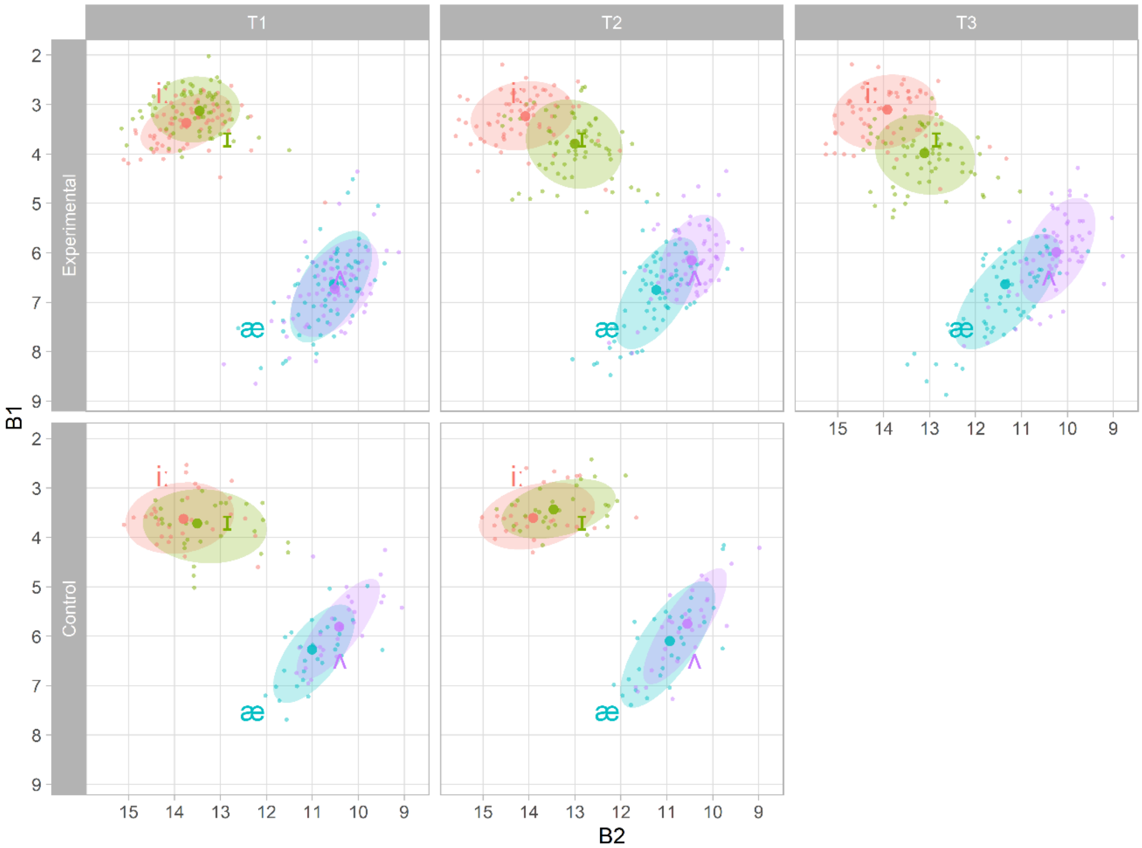


Figure 4.13. Learners' mean B1-B2 formant values (big dots) and mean B1-B2 formant values for each learner (small dots) for English vowels /i:/, /ɪ/, /æ/, /ʌ/ in sentence contexts. Filled ellipses represent 75% confidence intervals. Vowel transcriptions indicate mean B1-B2 formant values for native speakers of English. Graph organized by Time (T1/T2/T3) and Group (experimental/control).

Vowel distinctiveness and nativelikeness

First of all, we assessed whether the TBPT had had an effect on vowel production (RQ1.3.) by entering the log-transformed Mahalanobis distance between contrastive vowels (/iː-ɪ/ and /æ-ʌ/) of test trials in the DSR task (i.e., ***vowel distinctiveness***) to a GLMM. Distance between non-native contrastive vowels was declared the continuous dependent variable, and *Group* (experimental/control), *Time* (T1/T2), and their interactions as predictors. The random-effects structure included random intercepts for *Subject* and *Item.* A random slope for *Time* over *Subject* led to non-convergence and was excluded from the model.

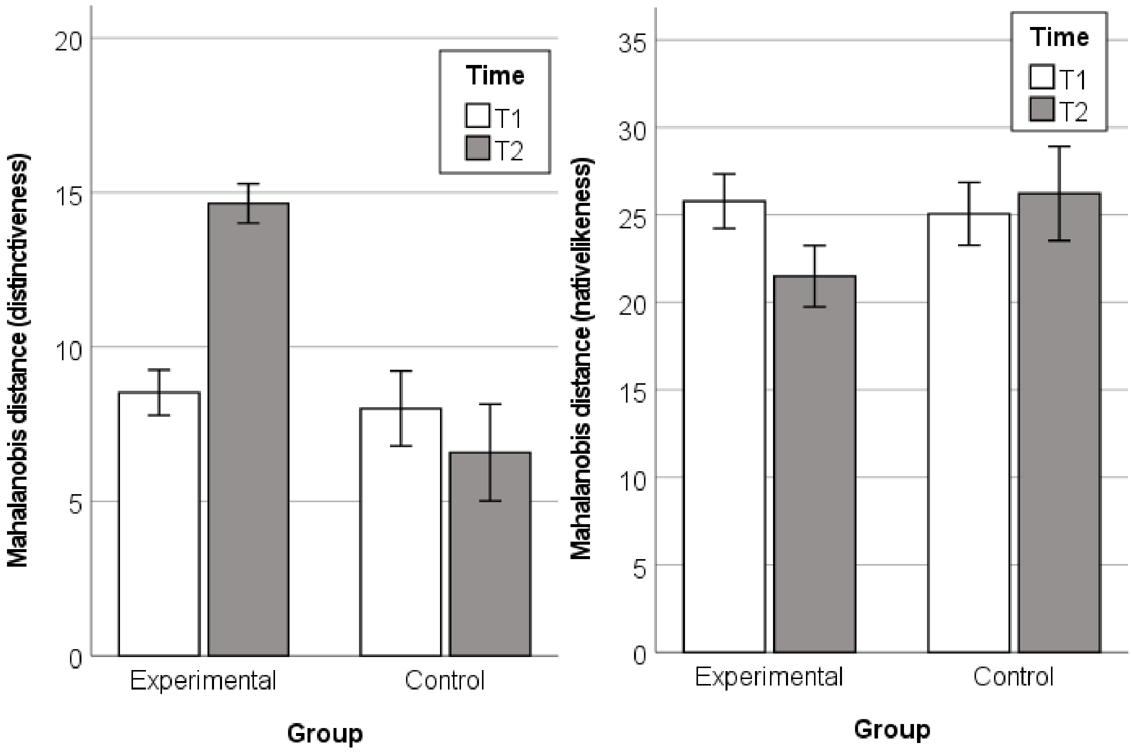


Figure 4.14. Mahalanobis distances (*distinctiveness*) on the left panel, and Mahalanobis distances (*nativelikeness*) on the right panel, produced in words in sentence contexts (DSR). Graphs organized by *Group* (experimental, control) and *Time* (T1 in white/T2 in grey). Error bars show 95% confidence intervals.

The GLMM yielded significant main effects of *Group* (*F* [1, 8684] = 4.94, *p* = .026), because the experimental group distinguished L2 vowels embedded in sentences to a greater extent than the control group, and significant main effects of *Time* (F [1, 8684] = 21.25, $p < .001$). Interestingly, the *Group* x *Time* interaction (F [1, 8684] = 49.01, $p < .001$) showed that vowel distinctiveness improvement depended on the group. Pairwise contrasts (Bonferroni-adjusted) revealed that, whereas the experimental group produced vowels significantly more distinct (t [8684] = −11.07, $p < .001$) from T1 to T2 (Figure 4.14., left panel), the control group's distinction of contrastive vowels in sentence contexts slightly overall worsened from T1 to T2 (t [8684] = 1.40, $p = .161$). Table 4.7. shows that, whereas the experimental group increased the distance between the contrastive vowels in both contrasts (/iː-ɪ/ and /æ-ʌ/) (6.2 SD mean gain), the control did not gain at all (-1.4 SD mean gain) in vowel distinctiveness. In addition, experimental and control groups were comparable before the TBPT treatment as they produced similar Mahalanobis distances between contrastive vowels at T1 (t [8684] = .38, $p = .703$). See Appendix L.39. for parameter estimates.

The second way to observe whether the TBPT intervention had changed learners' vowel productions was by fitting the log-transformed Mahalanobis distance between non-native and native vowels (/iː/, /ɪ/, /æ/, /ʌ/) of test trials in the DSR task (i.e., ***vowel nativelikeness***) to a GLMM with distance as the continuous dependent variable, and *Group* (experimental/control), *Time* (T1/T2), and their interactions as fixed factors. *Subject* and *Item* were the random intercepts. A random slope for *Time* over *Subject* was not included because it did not improve the model's fit. The model, whose parameter estimates can be shown in Appendix L.40., revealed non-significant effects of *Group* (F [1, 8684] = 1.08, $p = .297$) or *Time* (F [1, 8684] = 1.37, $p = .241$). Nevertheless, a significant disordinal *Group x Time* interaction (F [1, 8684] = 9.19, $p = .002$) showed that groups' Mahalanobis distance of nativelikeness went in opposite directions. Bonferroni-adjusted pairwise contrasts uncovered that, whereas the experimental group's vowels significantly approximated those of native speakers from T1 to T2 (Figure 4.14., right panel), hence, became significantly more target-like (t [8684] = 3.95, p < .001), the control group's Mahalanobis distances with respect to native speakers' vowels became larger from to T2, albeit non-significantly (t [8684] = -1.09, p = .273). Overall, whereas the experimental group's vowel productions got more target-like over time, hence, closer to native speaker values (4.3 SD), the control group's vowels became overall less target-like (-1.2 SD), as seen from mean gains in Table 4.7. Bonferroni-adjusted pairwise contrasts also revealed that both experimental and control groups were comparable before the TBPT intervention (t [8684] = .42, p = .671).

As part of RQ 1.3.2., we analysed potential differences in vowel *distinctiveness* and *nativelikeness* as a function of L2 vowel contrast or L2 vowel, respectively. Only the experimental group was included in the forthcoming analyses.

Table 4.7. Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) by Group (experimental/control), Contrast (/iː-ɪ/, /æ-ʌ/), Time (T1/T2) and gains (T2-T1) in the DSR task.

	/iː-ɪ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	9.93	5.44	13.42	16.57	12.95	11.56	6.64	7.51	13.19
Control	6.75	4.18	5.63	4.67	1.76	8.00	-2.08	-2.42	8.00
	/æ-ʌ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	7.11	5.19	8.33	12.72	9.57	9.18	5.61	4.38	9.29
Control	9.22	4.19	13.12	8.42	.55	22.22	-0.80	-3.64	15.22

Note. M = Mean, Mdn= Median, SD = Standard Deviation

Table 4.8. Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) by Group (experimental/control), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2) and gains (T2-T1) in the DSR task.

	/iː/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	22.39	22.15	9.51	17.90	15.51	9.76	4.49	6.64	8.25
Control	26.43	24.97	11.55	26.57	23.38	15.25	-0.14	1.59	13.63
	/ɪ/								
	T1			T2			*Gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	36.96	22.02	69.44	36.76	17.75	60.65	0.20	4.27	71.05
Control	28.51	19.17	32.21	33.65	23.05	40.22	-5.14	-3.88	58.29
	/æ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	27.12	22.40	23.44	20.85	15.97	15.10	6.27	6.43	20.49
Control	23.50	21.06	10.85	23.61	21.72	8.50	-0.11	-0.66	11.05
	/ʌ/								
	T1			T2			*gain*		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Experimental	16.66	16.28	7.88	10.49	9.76	5.27	6.17	6.52	6.03
Control	21.96	15.37	20.33	21.36	11.97	45.25	0.60	3.40	40.88

Note. M = Mean, Mdn= Median, SD = Standard Deviation

We submitted vowel ***distinctiveness*** distance to a GLMM with *Time* (T1/T2), *Contrast* (/iː-ɪ/, /æ-ʌ/) as well as their interactions as predictors. *Subject* and *Item* were included as random intercepts in the random-effects structure. The model, whose parameters are found in Appendix L.41., rendered significant main effects of *Time* (*F* [1, 6044] = 156.23, *p* < .001) and *Contrast* (*F* [1, 6044] = 6.33, *p* = .012) because overall, experimental learners produced greater distance between /iː/ and /ɪ/ (*M* = 13.25, *SD* = 12.93, *95% CI* [11.64, 14.85) than /æ/ and /ʌ/ (*M* = 9.91, *SD* = 9.18, *95% CI* [8.77, 11.05). The *Time x Contrast* interaction (*F* [1, 6044] = 1.27, *p* = .259) did not reach significance because the task-based intervention helped learners distinguish confusing vowels embedded in both target contrasts in similar ways (Table 4.7.).

As for vowel ***nativelikeness***, Mahalanobis distances were submitted to a GLMM with *Time* (T1/T2), *Vowel* (/iː, ɪ, æ, ʌ/) and their interactions as predictors. The random-effects structure included *Subject* and *Item* as random intercepts. The model, whose parameter estimates are shown in Appendix L.42., yielded significant main effects of *Time* (F [1, 6040] = 15.61, p < .001) and *Vowel* (F [3, 6040] = 78.14, p < .001) because all vowels presented different Mahalanobis distances with respect to native speakers' values (p < .05). Despite a non-significant *Time x Vowel* (F [3, 6040] = 1.63, p = .179) interaction, Bonferroni-adjusted pairwise comparisons revealed learners' productions of /iː/ (t [6040] = 2.10, p = .035), /æ/ (t [6040] = 2.93, p = .003) and /ʌ/ (t [6040] = 2.87, p = .004) significantly approximated the values of native speakers' vowel productions at T2. However, learners' productions of /ɪ/ (t [6040] = .25, p = .807) did not get significantly more accurate from T1 to T2[63] (Table 4.8.).

The third set of GLMM involved assessing improvement in vowel *distinctiveness* and *nativelikeness* in new lexical contexts. In other words, we were interested in observing whether improvement from pre- to post-test from the experimental group could be generalized to words that had not been part of the TBPT intervention (i.e., untaught). To do so, Mahalanobis distances between contrastive vowels (i.e., vowel ***distinctiveness***) were fitted to a GLMM with distance as the dependent variable and *Time* (T1/T2), *Word Type* (taught/untaught) as well as their interactions as predictors. The random-effects structure included random intercepts for *Subject* and *Item*, *and* a random slope for *Contrast* over *Item*. The model rendered significant main effects of *Time* (F [1, 6040] =

[63] When the GLMM included *Contrast* (/iː/-/ɪ/ and /æ/-/ʌ/), instead of *Vowel*, as fixed effects, there were significant effects of *Contrast* (F [1, 6040] = 81.11, p < .001) but a non-significant *Time x Contrast* interaction (F [1, 6040] = 3.08, p = .020). Interestingly, though, Bonferroni-adjusted pairwise contrasts indicated that the vowel contrast /æ/-/ʌ/ became significantly more accurate from T1 to T2 (t [6040] = 4.02, p < .001), whilst learners' productions of /iː/- /ɪ/ did not approximate the native-like model significantly (t [6040] = 1.63, p = .102) at T2.

156.81, $p < .001$), but not of *Word Type* (F [1, 6044] = 1.08, $p = .298$) or a *Time x Word Type* interaction (F [1, 6044] = .49, $p = .484$). Bonferroni-adjusted pairwise contrasts showed that learners' improvement in vowel distinctiveness from T1 to T2 happened when both taught (t [6044] = -9.61, $p < .001$) and untaught (t [6044] = -8.63, $p < .001$) words were part of sentences (Figure 4.15., left panel), indicating generalization of production gains over time. See parameter estimates in Appendix L.43.

Parallel analyses were conducted for vowel ***nativelikeness*****.** Again, Mahalanobis distances between native and non-native target vowels were submitted as the dependent variable, and *Time* (T1/T2), *Word Type* (taught/untaught) and their interactions as fixed factors. *Subject* and *Item* were included as random intercepts, and we added a random slope for *Vowel* over *Item.* The model, whose parameter estimates are presented in Appendix L.44., revealed significant main effects of *Time* (F [1, 6044] = 21.74, $p < .001$); however, non-significant main effects of *Word Type* (F [1, 6044] = .45, $p = .500$) and a *Time x Word Type* interaction (F [1, 6044] = 4.49, $p = .034$). These results suggested that learners' overall vowel accuracy was similar in taught and untaught words. Although learners became more target-like in vowels embedded both in taught (t [6044] = 2.01, $p = .046$) and untaught (t [6044] = 4.81, $p < .001$) between testing times (T1-T2), the disordinal interaction shown in Figure 4.15. (right panel) depicts that at T1 learners produced larger distances with respect to native speaker values with untaught words than taught words (t [6044] = 1.96, $p = .049$), but after the TBPT intervention, differences between taught and untaught words disappeared (t [6044] = 1.00, $p = .317$).

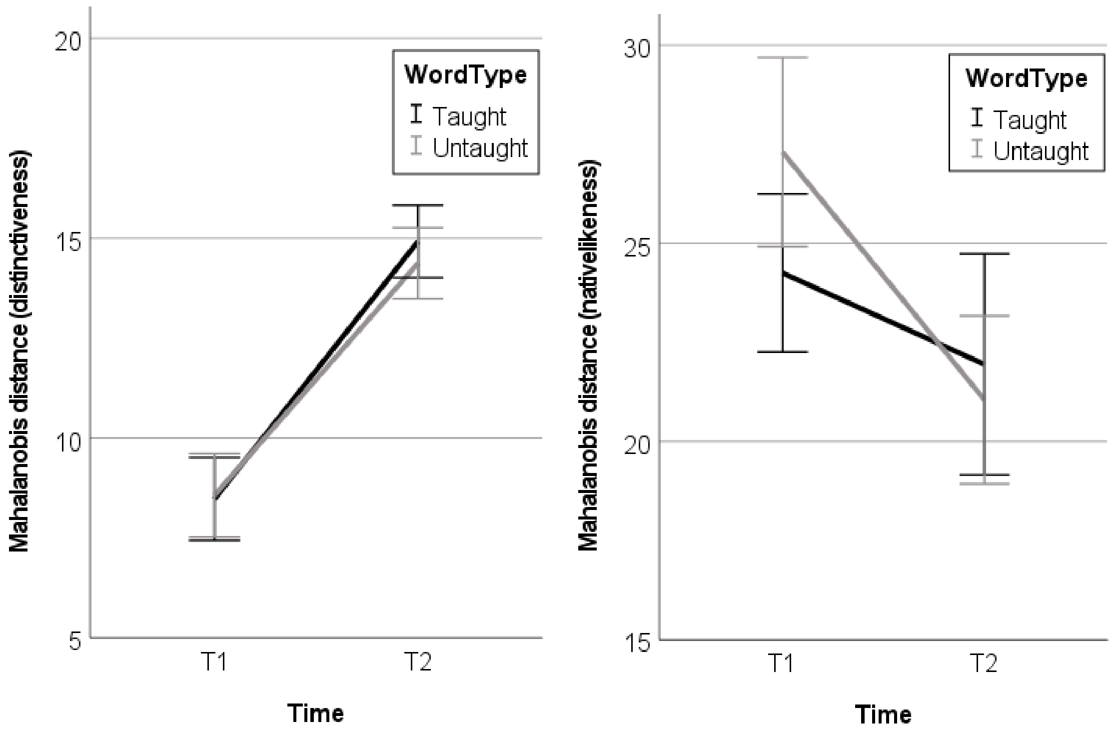


Figure 4.15. Mahalanobis distances between non-native contrastive vowels (left panel) and Mahalanobis distances between native and non-native vowels (right panel) in the DSR task as a function of Word Type (taught/untaught) and Time (T1/T2). Error bars indicate 95% confidence intervals.

Last but not least, the present study investigated whether improvement in vowel *distinctiveness* and *nativelikeness* was maintained 11 weeks after the TBPT intervention (i.e. retention effects; T2-T3). To do so, Mahalanobis distances between contrasting vowels (i.e., vowel ***distinctiveness***) were submitted to a GLMM with *Time* (T1/T2/T3) as fixed factor. We also included Subject and Item as random intercepts and a random slope for Contrast over Item (z = 6.70, p < .001). The model, whose parameter estimates are shown in Appendix L.45., rendered significant main effects of *Time* (F [2, 9069] = 78.16, p < .001). Figure 4.16. (left panel) illustrates that TBPT was effective at producing distinct confusing vowels from T1 to T2 (t [9069] = -11.79, < .001), and despite the relative decline from T2 to T3, the effects were still maintained (t [9069] = 1.89, p = .058).

Taking vowel ***nativelikeness*** as the dependent variable, *Time* (T1/T2/T3) was entered as fixed effect. The random-structure included *Subject* and *Item* as random intercepts and a random slope for *Vowel* over *Item* (z = 10.33, p < .001). The model yielded significant main effects of *Time* (F [2, 9069] = 16.77, p < .001), see parameter estimates in Appendix L.46. Figure 4.16. (right panel) depicts significant improvement from T1 to T2, hence, significantly reduced distance with respect to native speakers' vowels (t [9069] = 4.39, p < .001). From T2 to T3, learners kept producing vowels in a more native-like manner (albeit non-significantly: t [9069] = 1.18, p = .234). In sum, these findings show how gains were retained over time, suggesting the TBPT intervention was effective.

To sum up, the experimental group gained 5.2 SD in Mahalanobis distance in vowel *distinctiveness* (i.e., separation between non-native contrastive vowels) and 5.6 SD in Mahalanobis distance in *nativelikeness* (i.e., approximation to native speaker values) from pre-test to delayed post-test.

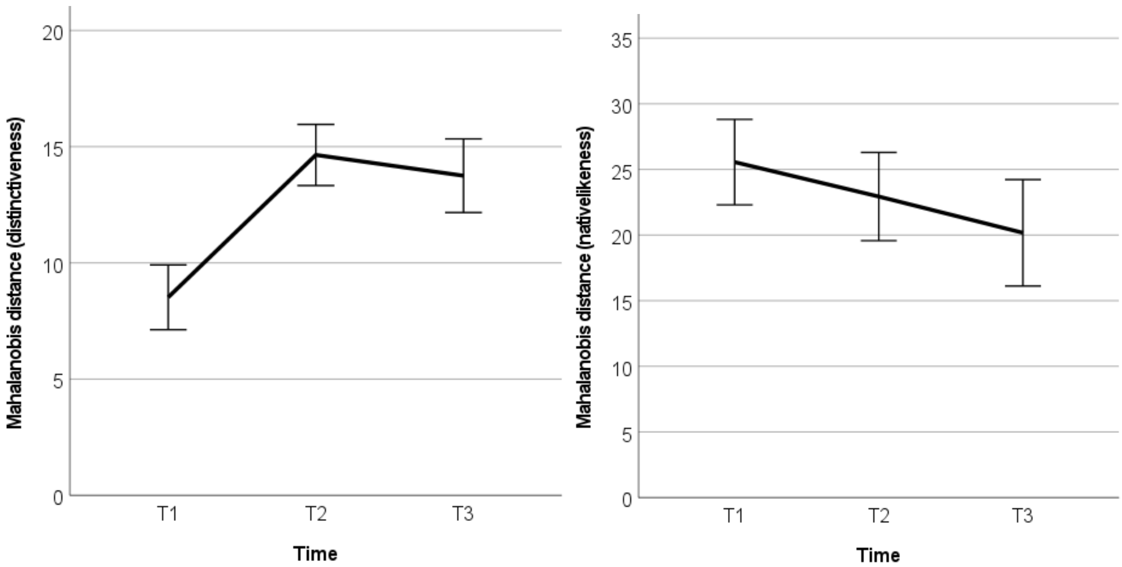


Figure 4.16. Mahalanobis distances between non-native contrastive vowels (left panel) and Mahalanobis distances between native and non-native vowels (right panel) in the DSR task as a function of Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

Correlations between vowel distinctiveness and nativelikeness

In order to assess the relationship between vowel *nativelikeness* (i.e., how close non-native vowel productions were of native vowel productions) and *distinctiveness* (i.e., how distant the contrastive vowels were from each other) at pre-test (T1), post-test (T2) and delayed post-test (T3) of the experimental group, two-tailed Spearman rank-order correlations were conducted. Figure 4.17. depicts negative correlations between the two Mahalanobis distance measures, suggesting that the larger the non-native vowel distances between /iː/ and /ɪ/ or /æ/ and /ʌ/ (i.e., less overlap between confusing vowels), the smaller the distances between non-native and native vowel productions (i.e., more-target like productions). Whereas the strength of this association was weak at T1, after the treatment it was slightly stronger, from weak to almost moderate. Crucially, as in the DWR task, T1-T3 gains in distinctiveness were moderately correlated with gains in nativelikeness (r_s = .429, n = 63, p < .001). This finding indicates that, in general, learners who learned to make a spectral distinction between the confusable vowels after the TBPT intervention, also learned to do so in a nativelike manner.

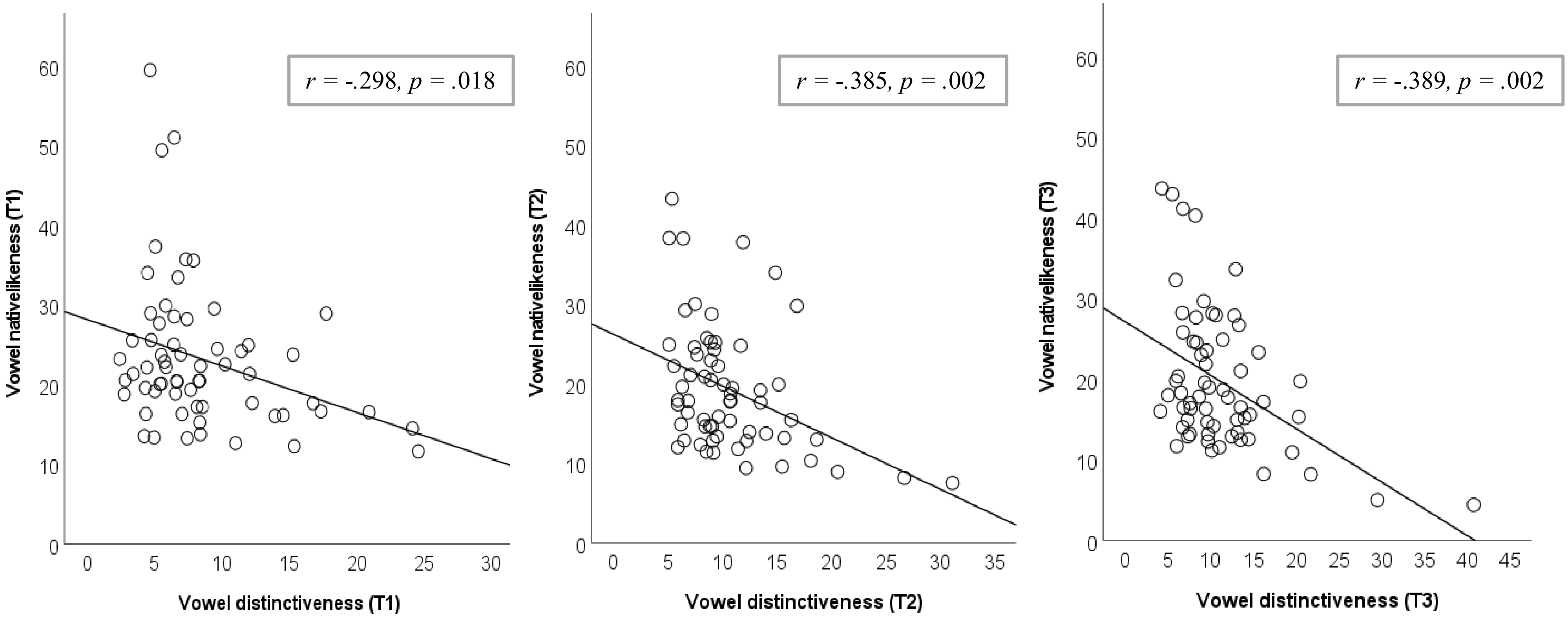


Figure 4.17. Two-tailed Spearman rank-order correlations between vowel nativelikeness (y axis) and distinctiveness (x axis) by Time (T1/T2/T3) from words produced in sentence contexts (DSR).

4.1.3.2.2. Vowel quantity

Having observed learners' changes in vowel quality (Section 4.1.3.2.1.), this short subsection is dedicated to examine whether vowel quality changes were accompanied by changes in vowel quantity (duration) as well.

Prior to analysing learners' DR per contrast (/iː-ɪ/ - /æ-ʌ/), two-tailed paired sample t-tests showed that experimental learners' duration of /iː/ was significantly larger than /ɪ/ (t [125] = 14.34, $p < .001$), and the duration of /æ/ was significantly larger than /ʌ/ (t [125] = 23.24, $p < .001$). As for the control group, learners' duration of /iː/ was significantly larger than /ɪ/ (t [55] = 12.30, $p < .001$), and the duration of /æ/ was significantly larger than /ʌ/ (t [55] = 11.82, $p < .001$) (See Table 4.9.)

Table 4.9. Duration and duration ratio by Group (simple/control) and Time (T1/T2) in the DSR task.

		Duration (ms)				Duration ratio (ms)			
		T1		T2		T1		T2	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experimental	/iː/	101.57	13.45	101.89	13.07	1.17	0.15	1.25	0.17
	/ɪ/	87.47	11.46	82.48	11.15				
	/æ/	124.03	15.49	126.22	16.11	1.22	0.11	1.26	0.14
	/ʌ/	101.84	12.98	101.01	14.01				
Control	/iː/	102.47	13.52	104.09	11.29	1.18	0.13	1.20	0.14
	/ɪ/	87.67	14.70	87.48	12.42				
	/æ/	124.70	17.43	127.41	15.53	1.21	0.16	1.22	0.14
	/ʌ/	104.11	13.20	105.32	12.77				

Note. M = Mean, SD = Standard Deviation

Secondly, the DR was submitted to LMM where *Group*, *Time* and their interactions were submitted as fixed effects, and the random-effects structure included an intercept for *Subject*. The first model with DR for /iː-ɪ/ yielded non-significant main effects of *Group* (F [1, 178] = .44, $p = .505$) but significant effects of *Time* (F [1, 178] =

4.04, p = .046). The *Group* x *Time* interaction did not reach significance (F [1, 178] = 1.14, p = .286). Yet, experimental learners' DR significantly increased from T1 to T2 (t [178] = -.07, p = .006) (Appendix L.47.). In terms of DR for /æ-ʌ/, LMM yielded non-significant main effects of *Group* (F [1, 178] = 1.79, p = .182), *Time* (F [1, 178] = 1.10, p = .296), nor a significant *Group* x *Time* interaction (F [1, 178] = .32, p = .568). In sum, despite the DR increasing from T1 to T2 for the experimental group, the DR was maintained over time for the low vowels /æ-ʌ/. See Table 4.9. and Appendix L.48.

Summary of results:

Learners' productions of L2 vowels in words in isolation (DWR) and sentence contexts (DSR) show that the target confusable vowels became more distinct and more target-like from T1 to T2 for the experimental group, but not the control group. In both tasks, only the experimental learners showed gains in vowel distinctiveness and nativelikeness. In the case of words in isolation, learners produced especially more distinct and accurate productions of L2 vowels /æ/-/ʌ/. In the case of words in sentences, L2 vowels /æ/-/ʌ/ (and especially /iː/-/ɪ/) became significantly more distinct across times but vowels /æ/, /ʌ/ and /iː/ were more accurately produced from T1 to T2. In both word and sentence contexts, L2 production learning generalized to untaught words and was retained 11 weeks after the intervention. Interestingly, more distinctiveness between contrastive vowels was associated with more target-like productions in the DWR and DSR tasks.

To conclude, having assessed learners' production of L2 vowels under two different tasks (DWR and DSR), this study concludes that the TBPT intervention was equally beneficial for L2 vowel production accuracy when they were produced in words in isolation and surrounded by the meaningful context of a sentence. Descriptively, although L2 vowels were produced more accurately (closer to the native speaker model) in isolation than in sentences, overall T1-T3 gains in vowel distinctiveness and vowel nativelikeness were greater when L2 vowels were produced in sentence contexts.

4.1.4. Discrimination - lexical encoding - production comparisons

The last section within 4.1. compares the overall performance and raw gains from T1 to T3 obtained by the experimental group in each one of the perceptual and production tests, namely, ABX discrimination, FLeC, LD, DWR and DSR.[64] Out of 63 experimental learners, only 56 could be included in the forthcoming analyses because they contained data from the three testing times (i.e., T1-T2-T3) of all tasks.

Two-tailed Spearman-rank correlations revealed moderate associations between learners' accuracy performance in ABX perceptual discrimination and lexical encoding as assessed through FLeC (r_s = .42, n = 56, p = .001) and LD (r_s = .38, n = 56, p = .003) tasks. In addition, speed of ABX discrimination was also moderately related to speed of lexical decision in both FLeC (r_s = .58, n = 56, p = .001) and LD (r_s = .41, n = 56, p = .001) tasks. Therefore, learners who were overall more accurate and faster at perceptual discrimination appeared to be also more accurate and faster at accepting words (FLeC) and rejecting nonwords (LD) containing the target vowels. As expected, both lexical decision tasks appeared to be strongly correlated for accuracy (r_s = .62, n = 56, p = .001) and RT (r_s = .65, n = 56, p = .001). Similarly, performance in vowel distinctiveness and nativelikeness were significantly and moderately correlated between the DWR and DSR production tasks (r_s = .54, n = 56, p = .001; r_s = .46, n = 56, p = .001). However, neither of the production task performances was related to the perception/lexical encoding tasks. Additionally, no significant correlations appeared between perceptual and production gain data.

[64] For the current analyses, the data was averaged across all tested vowel tokens (taught and untaught). As for the LD task, only gains in nonword rejection were averaged.

4.2. Task complexity effects

In the second section of the results, we investigated whether task complexity led to differential effects in the perceptual discrimination (RQ2.1), lexical encoding (RQ2.2) and production (RQ2.3) of L2 English vowels. To do so, *Group* (simple/complex), *Time* (T1/T2/T3) and *Contrast/Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) were declared predictors in the mixed-effects models, and gains by group were computed. In addition, we assessed whether the occurrence of pronunciation-based language-related episodes was associated with task complexity (RQ2.4) and whether their occurrence changed over time.

4.2.1. Perceptual discrimination

This subsection explores to what extent task complexity leads to differential effects in the perceptual discrimination of the target L2 vowel contrasts in terms of accuracy and RT over time and how this vary as a function of *Contrast* (/iː-ɪ/, /æ-ʌ/). In the subsequent analyses, only the experimental group was selected, hence data from 60 experimental participants was entered (i.e., SG: *N*= 31; CG: *N*= 29).

Task complexity effects on the improvement of the TBPT intervention by each pair of vowels was assessed by fitting the accuracy of test trials in the ABX task to a GLMM with a binary logistic regression function with **accuracy** (*0 = incorrect, 1 = correct*) as the categorical dependent variable, and *Group* (simple/complex), *Time* (T1/T2/T3), *Contrast* (/iː-ɪ/, /æ-ʌ/) and their interactions as fixed effects. The random-effects structure included random intercepts for *Subject* and *Item.* A random slope for *Time* over *Subject* was excluded because it did not improve the model's fit. The model, whose parameter estimates are shown in Appendix L.49., yielded non-significant main effects of *Group* (*F* [1, 17940] = .72, *p* = .396) but a significant main effect of *Time* (*F* [2, 17940] = 49.57, *p* < .001) and *Contrast* (*F* [1, 17940] = 17.08, *p* < .001). Crucially, the *Group x Time* (*F* [2, 17940] = 3.15, *p* = .043) and *Group x Time x Contrast* (*F* [5, 17940] = 4.29, *p* < .001) interactions reached significance. On the one hand, Bonferroni-adjusted pairwise contrasts revealed that both SG (*t* [17940] = −6.00, *p* < .001) and CG (*t* [17940] = −4.26, *p* < .001) improved significantly in the discrimination of L2 vowels from T1 to T2. Nevertheless, whereas the CG kept improving from T2 to T3 (*t* [17940] = −3.10, *p* = .002), the SG did not (*t* [17940] = .36, *p* = .719). See Figure 4.18., left panel. On the other hand, Bonferroni-adjusted comparisons from the triple interaction revealed that, whereas the SG and CG improved in both vowel contrasts equally from T1 to T2 (*p* < .05), the CG improvement from T2 to T3 mainly came from the distinction of /æ-ʌ/ (*t* [17940] = -3.51, *p* = .001). Whereas the SG improved 6.5% on vowel discrimination, the amount of gain evidenced by the CG was 8% (see Table 4.10.). Additionally, SG and CG were comparable before the TBPT intervention because pairwise comparisons (Bonferroni-adjusted) showed no significant differences (*t* [17940] = −.82, *p* = .408) between groups at T1.

The **RT** of correct test trials was submitted to a GLMM with a gamma linking function with *Group* (simple/complex), *Time* (T1/T2/T3), *Contrast* (/iː-ɪ/, /æ-ʌ/) and their interactions as fixed effects. Random intercepts by *Subject* and *Item* were included in the random-effects structure. A random slope for *Time* over *Subject* was not included because it did not improve the model’s fit. The model revealed non-significant main effects of *Group* (*F* [1, 12401] = .04, *p* = .841) but significant main effects of *Time* (*F* [2, 12401] = 330.96, *p* < .001) and *Contrast* (*F* [1, 12401] = 10.79, *p* < .001). Interestingly, the *Group x Time* interaction (*F* [2, 12401] = 13.67, *p* < .001) reached significance, whilst the *Group x Time x Contrast* (*F* [5, 12401] = .44, *p* = .818) interaction did not.

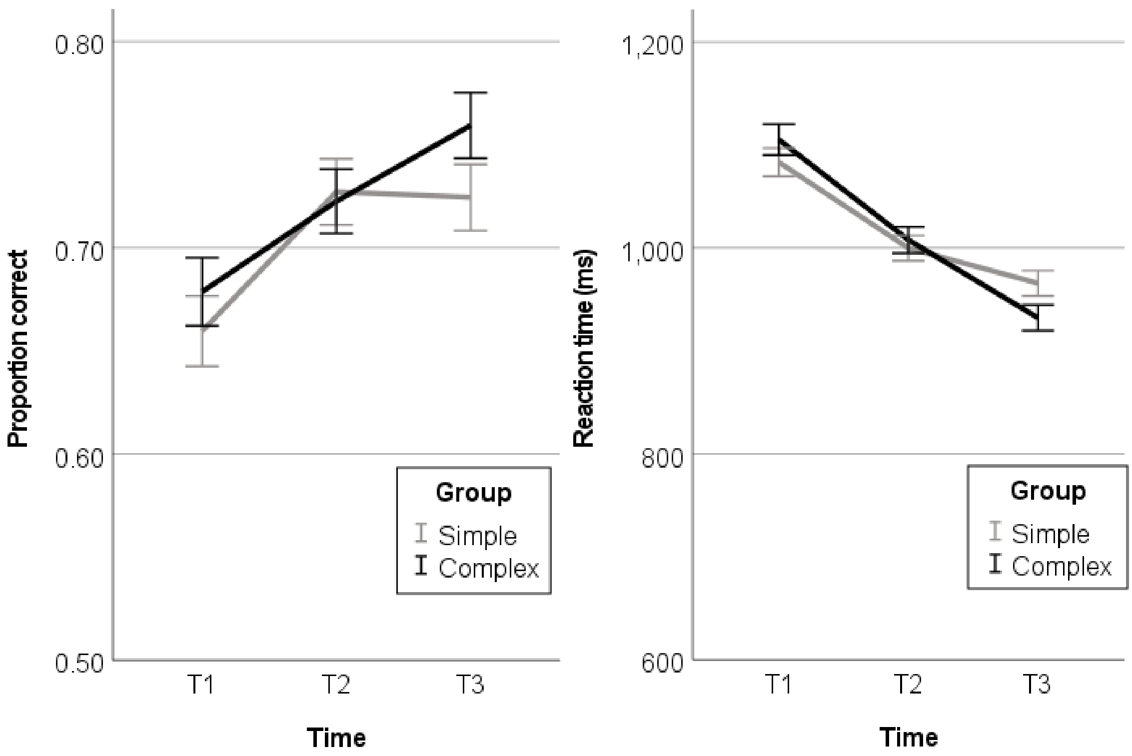


Figure 4.18. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the ABX task as a function of Group (simple in grey/complex in black) and Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

Bonferroni-adjusted pairwise contrasts showed that both SG (t [12401] = 10.21, p < .001) and CG (t [12401] = 10.46, p < .001) became increasingly faster from T1 to T2, but the CG became faster than the SG from T2 to T3, albeit non-significantly (see Figure 4.18., right panel, and Appendix L.50. for parameter estimates). Importantly, improvement in RT happened equally with both target contrasts. Overall, the CG obtained greater gains in speed of response (188 ms) than the SG (117 ms) in the discrimination of /iː-ɪ/ and /æ-ʌ/ (Table 4.10.). Both groups were comparable at T1 (t [12401] = -.26, p = .792).

Table 4.10. Accuracy (proportion correct) and RT (in milliseconds) by Group (simple/complex), Contrast (/iː-ɪ/, /æ-ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the ABX task.

			Accuracy				RT (ms)			
			T1	T2	T3	*gain*	T1	T2	T3	*gain*
Simple	/iː-ɪ/	*M*	0.60	0.66	0.66	0.06	1119.00	1018.46	993.73	125.27
		SD	0.10	0.09	0.10	0.09	140.50	163.66	172.17	154.47
	/æ-ʌ/	*M*	0.71	0.80	0.79	0.07	1063.38	987.22	954.08	109.30
		SD	0.11	0.10	0.10	0.11	134.15	156.45	185.15	153.62
Complex	/iː-ɪ/	*M*	0.65	0.69	0.70	0.05	1138.96	1047.98	968.33	186.59
		SD	0.12	0.13	0.13	0.10	178.58	208.24	166.00	155.57
	/æ-ʌ/	*M*	0.71	0.77	0.82	0.11	1090.35	991.85	915.68	189.79
		SD	0.13	0.12	0.08	0.11	202.56	164.72	144.60	155.46

Note. M = Mean, SD = Standard Deviation

Summary of results:

Both SG and CG became significantly more accurate and faster in the discrimination of the two target L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/) from T1 to T2. However, only the CG kept discriminating the contrastive vowels at T3, especially those in the /æ-ʌ/ contrast. Overall gains in accuracy and speed were overall larger in the CG than the SG.

4.2.2. Lexical encoding

4.2.2.1. FLeC

This subsection aims at assessing task complexity effects in the lexical encoding of English contrasts (/iː-ɪ/, /æ-ʌ/), measured in terms of accuracy scores and RT over time. In the following analyses, only the experimental group (*N*=62) was selected (SG: *N*= 31; CG: *N*= 31).

So as to assess the effects of task complexity on the lexical encoding of L2 sounds (RQ2.2), the **accuracy** of test trials (*0 = incorrect*, *1 = correct*) was fitted to a GLMM with a binary logistic regression function with *Group* (simple/complex), *Time* (T1/T2/T3), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) and their interactions as predictors. *Subject* and *Item* were included as random intercepts, and a random slope of *Time* over *Subject* was also included because it improved the model's fit and explained significant covariance in the dependent variable (z = 2.94, p = .003). The model, whose parameter estimates can be found in Appendix L.51, unfolded significant main effects of *Group* (F [1, 4488] = 20.25, p < .001) and *Time* (F [2, 4488] = 47.57, p < .001), but non-significant main effects of *Vowel* (F [3, 4488] = 1.19, p = .310). As for the interactions, the *Group x Time* (F [2, 4488] = 24.55, p < .001) reached significance whereas the *Group x Time x Vowel* (F [15, 4488] = 1.33, p = .173) interaction did not. The main effects of *Group* revealed that, overall, the accuracy scores were significantly higher for the CG than the SG. In fact, Bonferroni-adjusted pairwise contrasts from the *Group x Time* interaction showed that, on the one hand, both groups improved significantly from T1 to T2 (t [4488] = −3.21, p = .004). However, only the CG kept being more accurate at T3 (t [4488] = −5.69, p < .001) whereas the SG did not (t [4488] = .86, p = .388). See Figure 4.19., left panel. Interestingly, pairwise contrasts from the triple interaction showed that the SG only improved significantly in the identification of words containing /ʌ/: t [4488] = −4.18, p < .001) from T1 to T2 but identification of words containing the L2 vowel contrasts did not improve from T2 to T3; nevertheless, the CG improved significantly from T1 to T2 in the identification of English words containing all vowels (/iː/: t [4488] = −2.97, p = .003; /ɪ/: t [4488] = −2.33, p = .019; /æ/: t [4488] = −2.12, p = .034; and /ʌ/: t [4488] = −3.21, p = .003); especially, from T2 to T3 (p < .01).[65] Overall, the SG gained 7.5% and the CG 28.5% in accuracy from T1 to T3 (Table 4.11.). Last, both groups were comparable before the TBPT intervention because, as revealed by Bonferroni-adjusted pairwise comparisons, there were no significant differences between groups at T1 (t [4488] = −.28, p = .776).

[65] When the GLMM included *Contrast*, instead of *Vowel*, Bonferroni-adjusted pairwise contrasts showed that the SG only improved the lexical encoding of /æ-ʌ/ significantly between T1 and T2 (p < .001), but not /iː-ɪ/ (p = .615), and their accuracy scores for /iː-ɪ/ and /æ-ʌ/ decreased (albeit non-significantly) from T2 to T3. Instead, the CG significantly improved in the lexical encoding of both L2 vowel contrasts from T1 to T2 (p < .001) and T2 to T3 (p < .001).

The second way of evaluating whether task complexity had an impact on the lexical encoding of L2 vowels was by submitting **RT** of correct test responses to a GLMM with a gamma linking function and *Group* (simple/complex), *Time* (T1/T2/T3), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/), and their interactions as fixed-effects. The random-effects structure only included *Subject* and *Item* as intercepts because a random slope for *Time* over *Subject* resulted in uncertainty of the model's fit. Fixed effects from the model rendered significant main effects of *Time* (F [2, 2959] = 11.53, $p < .001$), *Vowel* (F [3, 2959] = 4.91, $p = .002$) and a significant *Group x Time* (F [2, 2959] = 3.91, $p = .020$) interaction; however, no significant main effects of *Group* (F [1, 2959] = .49, $p = .483$) nor a *Group x Time x Vowel* (F [15, 2959] = 1.22, $p = .245$) interaction arose.

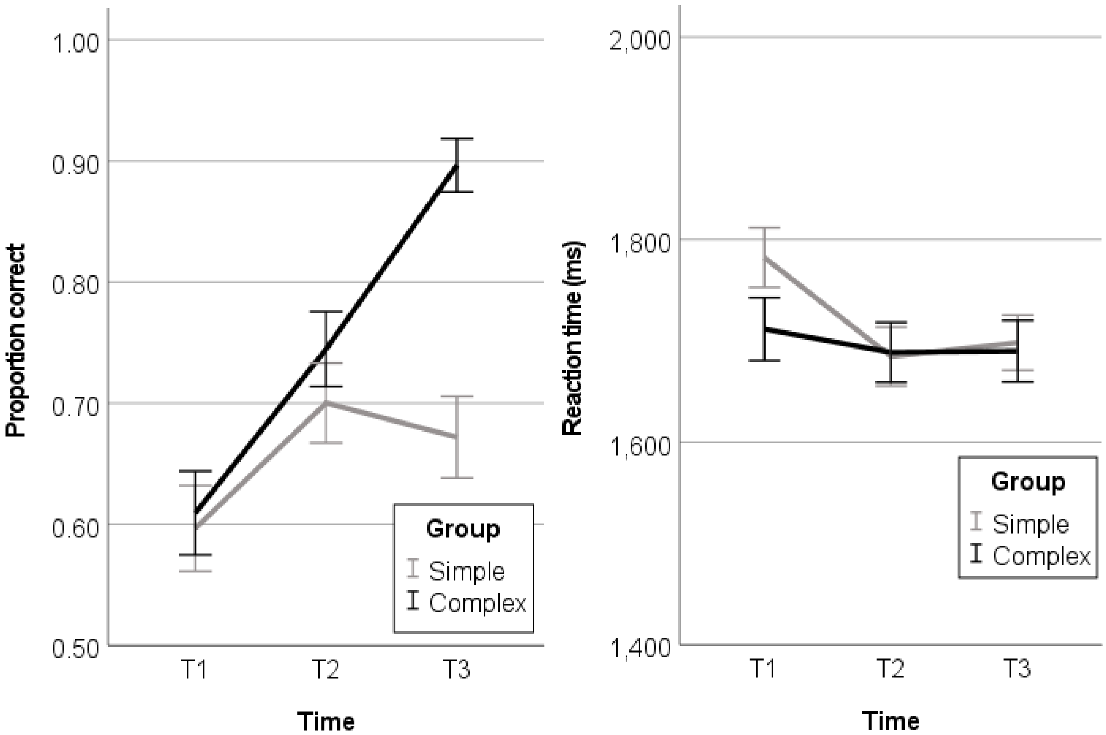


Figure 4.19. Proportion of correct responses (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the FLeC task as a function of Group (simple in grey/complex in black) and Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

Pairwise contrasts (Bonferroni-adjusted) revealed that only the SG (t [2959] = 4.75, $p < .001$) became significantly faster from T1 to T2, because its RT were higher than the CG at T1, despite non-significantly, *t* [2959] = 1.68, *p* = .092. However, from T2 to T3, no significant differences arose for the SG or CG (*t* [2959] = -.08, *p* = .932) (Figure 4.19., right panel). Crucially, pairwise contrasts (Bonferroni-adjusted) showed that both groups had only reduced their RT in the identification of English words containing /ɪ/ from T1 to T2 (SG: *t* [2959] = 4.04, *p* < .001; CG: *t* [2959] = 2.63, *p* = .026)[66]. Therefore, despite initial differences between groups, both mainly improved in speed when words contained /ɪ/. Overall, the SG obtained greater gains (85.8 ms) than the CG (72.9 ms) in the lexical encoding of L2 vowels (Table 4.11.), but the CG was overall faster at T3. See parameter estimates in Appendix L.52.

Table 4.11. Accuracy (proportion correct) and RT (in milliseconds) by Group (simple/complex), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the FLeC task.

			Accuracy				RT (ms)			
			T1	T2	T3	*gain*	T1	T2	T3	*gain*
Simple	/iː/	*M*	0.62	0.63	0.64	0.02	1786.85	1720.43	1745.77	41.08
		SD	0.18	0.18	0.22	0.24	210.03	212.24	173.29	195.64
	/ɪ/	*M*	0.68	0.77	0.74	0.05	1820.34	1643.59	1661.24	159.10
		SD	0.22	0.19	0.26	0.18	218.90	201.08	220.62	250.26
	/æ/	*M*	0.61	0.70	0.68	0.06	1744.76	1718.80	1728.57	16.19
		SD	0.18	0.21	0.18	0.26	251.60	254.86	210.27	332.77
	/ʌ/	*M*	0.47	0.69	0.63	0.17	1796.95	1697.03	1672.64	126.84
		SD	0.21	0.22	0.27	0.30	194.01	264.12	234.14	263.66
Complex	/iː/	*M*	0.54	0.70	0.85	0.31	1687.17	1753.85	1732.73	-58.35
		SD	0.21	0.19	0.25	0.32	250.01	235.23	243.23	294.35
	/ɪ/	*M*	0.66	0.78	0.94	0.28	1741.48	1650.63	1656.57	90.70
		SD	0.24	0.18	0.12	0.25	242.61	248.43	240.51	247.24
	/æ/	*M*	0.64	0.74	0.89	0.24	1685.27	1688.08	1705.92	-16.88
		SD	0.21	0.22	0.22	0.30	270.71	204.99	230.35	269.74
	/ʌ/	*M*	0.60	0.77	0.90	0.31	1757.18	1734.11	1709.08	55.09
		SD	0.20	0.19	0.22	0.33	211.21	267.73	253.66	318.69

Note. M = Mean, SD = Standard Deviation

[66] When the GLMM included *Contrast*, instead of *Vowel*, Bonferroni-adjusted pairwise contrasts showed that the SG became significantly faster from T1 to T2 in the lexical encoding of /iː-ɪ/ (*p* < .001) because this group was slower than the CG at T1. Nevertheless, this group difference disappeared at T2 and at T3.

4.2.2.2. LD

This subsection explores task complexity effects in the lexical encoding of English contrasts (/iː-ɪ/, /æ-ʌ/) (RQ 2.2), measured in terms of nonword rejection accuracy scores and RT of correctly rejected nonwords over time. For the subsequent analyses, only the experimental group (*N*= 60) was selected (SG: *N*= 31; CG: *N*= 29).

To start with, nonword rejection **accuracy** scores (*0 = incorrect*, *1 = correct*) were submitted to a GLMM with a binary logistic regression function. The fixed-effects structure included *Group* (simple/complex), *Time* (T1/T2/T3), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/) and their interactions as predictors. The random-effects structure included *Subject* and *Item* as intercepts, and a random slope of *Time* over *Subject* because it improved the model's fit and explained significant covariance ($z = 5.56$, $p < .001$). The model, whose parameter estimates are shown in Appendix L.53, yielded significant main effects of *Group* (F [1, 4344] = 6.82, $p = .009$), *Time* (F [2, 4488] = 47.57, $p < .001$) and *Vowel* (F [3, 4344] = 26.21, $p < .001$). The main effect of *Group* indicated that, overall, the CG had higher scores in nonword rejection accuracy than the SG. Bonferroni-adjusted pairwise contrasts from the *Group x Time* interaction (F [2, 4344] = 5.98, $p = .003$) showed that, whereas the SG significantly improved in nonword rejection accuracy from T1 to T2 (t [4344] = -2.54, $p = .022$), they did not improve from T2 to T3. Instead, the CG significantly improved from T1 to T2 (t [4344] = -4.96, $p < .001$), as well as from T2 to T3 (t [4344] = -3.44, $p = .001$). See Figure 4.20., left panel. Finally, the *Group x Time x Vowel* interaction (F [15, 4344] = .53, $p = .922$) did not reach significance but, interestingly, Bonferroni-adjusted pairwise contrasts revealed that, from T1 to T2, the SG only improved the nonword rejection from words containing /ʌ/: t [4344] = −2.95, $p = .006$). Instead, the CG improved the nonword rejection of words containing /iː/: t [4344] = −3.45, $p = .001$; /ɪ/: t [4344] = −3.43, $p = .001$; /æ/: t [4344] = −4.70, $p < .001$; and /ʌ/: t

[4344] = −4.43, $p < .001$); especially, from T2 to T3 ($p < .01$).[67] In total, the SG gained 19.75% and the CG 40% in accuracy from T1 to T3 in the LD task (Table 4.12.). Last, SG and CG were comparable before the TBPT intervention (t [4344] = −.14, $p = .888$).

Parallel analyses were conducted with RT data. **RT** of correct test responses were fitted to a GLMM with a gamma linking function and *Group* (simple/complex), *Time* (T1/T2/T3), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/), and their interactions as predictors. *Subject* and *Item* were inserted as random intercepts. The model (Appendix L.54.) did not show significant main effects of *Group* (F [1, 1673] = .051, $p = .821$) but revealed significant main effects of *Time* (F [2, 1673] = 11.28, $p < .001$), *Vowel* (F [3, 1673] = 10.19, $p < .001$) and a *Group x Time* interaction (F [2, 1673] = 3.49, $p = .031$).

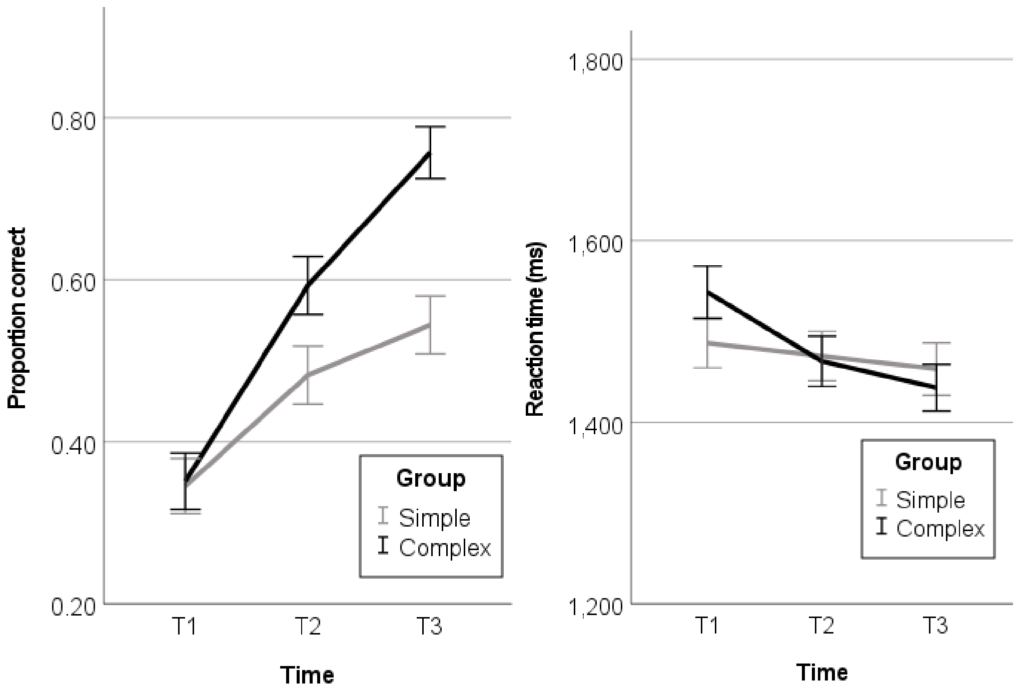


Figure 4.20. Proportion of correct nonword rejection (left panel) and reaction times (in milliseconds) for correct responses (right panel) in the LD task as a function of Group (simple in grey/complex in black) and Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

[67] Similar to the FLeC results, when the GLMM with accuracy as dependent variable included *Contrast*, instead of *Vowel*, Bonferroni-adjusted pairwise contrasts revealed that the SG only improved the lexical encoding of /æ-ʌ/ significantly between T1 and T2 ($p < .001$), but not /iː-ɪ/ ($p = .133$), and their accuracy scores for /iː-ɪ/ and /æ-ʌ/ did not increase significantly from T2 to T3. Instead, the CG significantly improved in the lexical encoding of both L2 vowel contrasts from T1 to T2 ($p < .001$) and T2 to T3 ($p < .001$).

Overall, SG and CG seemed comparable in terms of speed of response; however, Bonferroni-adjusted pairwise contrasts uncovered non-significant differences between T1 and T2 for the SG (t [1673] = .60, p = .623); and a significant decrease in reaction times between T1 and T2 for the CG (t [1673] = 2.79, p = .011), mainly because at T1 they produced longer RT (Figure 4.20., right panel). No RT differences arose for both SG and CG groups (p > .05) at T3. A triple non-significant interaction for *Group x Time x Vowel* (t [1673] = .69, p = .791) suggested that there were no significant differences between groups over time as a function of vowel[68]. Overall, the CG obtained greater gains (92.5 ms) than the SG (36.9 ms) in the lexical encoding of L2 vowel contrasts (Table 4.12.), and both groups were comparable at T1 (t [1673] = -1.35, p = .174).

Table 4.12. Accuracy (proportion correct) and RT (in milliseconds) by Group (simple/complex), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the LD task.

			Accuracy				RT (ms)			
			T1	T2	T3	*gain*	T1	T2	T3	*gain*
Simple	/iː/	*M*	0.32	0.41	0.48	0.15	1506.12	1489.17	1492.40	38.09
		SD	0.17	0.24	0.31	0.31	199.35	166.87	252.30	294.37
	/ɪ/	*M*	0.32	0.45	0.50	0.18	1492.10	1499.62	1523.44	-13.92
		SD	0.20	0.32	0.34	0.41	174.12	153.50	214.42	282.85
	/æ/	*M*	0.46	0.59	0.68	0.22	1460.67	1441.26	1478.79	7.80
		SD	0.22	0.29	0.30	0.28	163.43	181.79	205.23	214.75
	/ʌ/	*M*	0.28	0.48	0.51	0.24	1521.60	1520.48	1472.62	64.89
		SD	0.18	0.29	0.35	0.40	148.59	223.70	208.62	203.57
Complex	/iː/	*M*	0.32	0.54	0.72	0.40	1563.01	1503.11	1475.36	125.50
		SD	0.19	0.34	0.35	0.37	190.94	178.95	199.43	279.02
	/ɪ/	*M*	0.35	0.56	0.74	0.38	1523.29	1468.90	1526.26	23.16
		SD	0.23	0.37	0.37	0.36	184.89	265.50	195.74	238.78
	/æ/	*M*	0.44	0.69	0.82	0.37	1532.99	1466.92	1377.24	157.84
		SD	0.20	0.30	0.28	0.30	185.83	202.22	137.92	182.94
	/ʌ/	*M*	0.30	0.56	0.75	0.45	1560.92	1513.85	1494.13	63.78
		SD	0.19	0.40	0.35	0.36	177.43	194.84	171.82	200.64

Note. M = Mean, SD = Standard Deviation

[68] When *Contrast* substituted *Vowel* in the GLMM with RT, pairwise contrasts revealed a significant decrease in speed between T1 and T3 in /æ-ʌ/ for the CG (but not the SG), probably due to the slower RT at T1, but in general, the CG was faster at the lexical encoding of L2 vowel contrasts than the SG.

Summary of results:

In terms of accuracy, results from both FLeC and LD tasks revealed that both SG and CG significantly improved the lexical encoding of L2 vowel contrasts from T1 to T2. Whereas the CG improved significantly in the two L2 vowel contrasts (and all vowels), the SG mainly improved in /æ-ʌ/, and in the identification of words/rejection of nonwords containing /ʌ/. Only the CG kept being significantly more accurate at the lexical encoding of the two L2 vowel contrasts at T3, and overall, gained more than the SG. In terms of RT, both SG and CG became faster at lexically encoding L2 sounds but T1-T2 improvement in response speed happened significantly for the SG in the FLeC task and CG in the LD task, mainly due to differences at T1. RT were very similar from T2 to T3.

4.2.3. Production

4.2.3.1. DWR

This subsection explores task complexity effects on the production of English vowels /iː/, /ɪ/, /æ/ and /ʌ/ in words in isolation, measured in terms of vowel quality (B1-B2 descriptions, Mahalanobis distances between contrastive non-native vowels -*distinctiveness*-, and Mahalanobis distances between learner and native speaker vowels -*nativelikeness*-) and vowel quantity. In the next set of analyses, the experimental group (*N*=63) was selected (SG: *N*= 31; CG: *N*= 32).

In general, Figure 4.21. illustrates that confusable vowels got more contrastive and nativelike from T1 to T3 for both groups. First of all, the SG learned to distinguish English /iː/ from /ɪ/ from T1 to T2 by producing /iː/ further forward (i.e., higher B2) and slightly higher (i.e., lower B1) and by centralizing (i.e., lower B2) and lowering vowel /ɪ/ (i.e., higher B1). At T3 learners slightly lowered vowel /ɪ/. The SG's productions of /iː/, especially, became closer to the native model from T1 to T3. As for the CG, English /iː/ was produced with a higher (i.e., lower B1) and more front (i.e., higher B2) tongue position and /ɪ/ became more centralized (i.e., lower B2) from T1 to T3. Moreover, learners' productions became more accurate (i.e., closer to native speakers' /iː-ɪ/ B1-B2 values) across testing times.

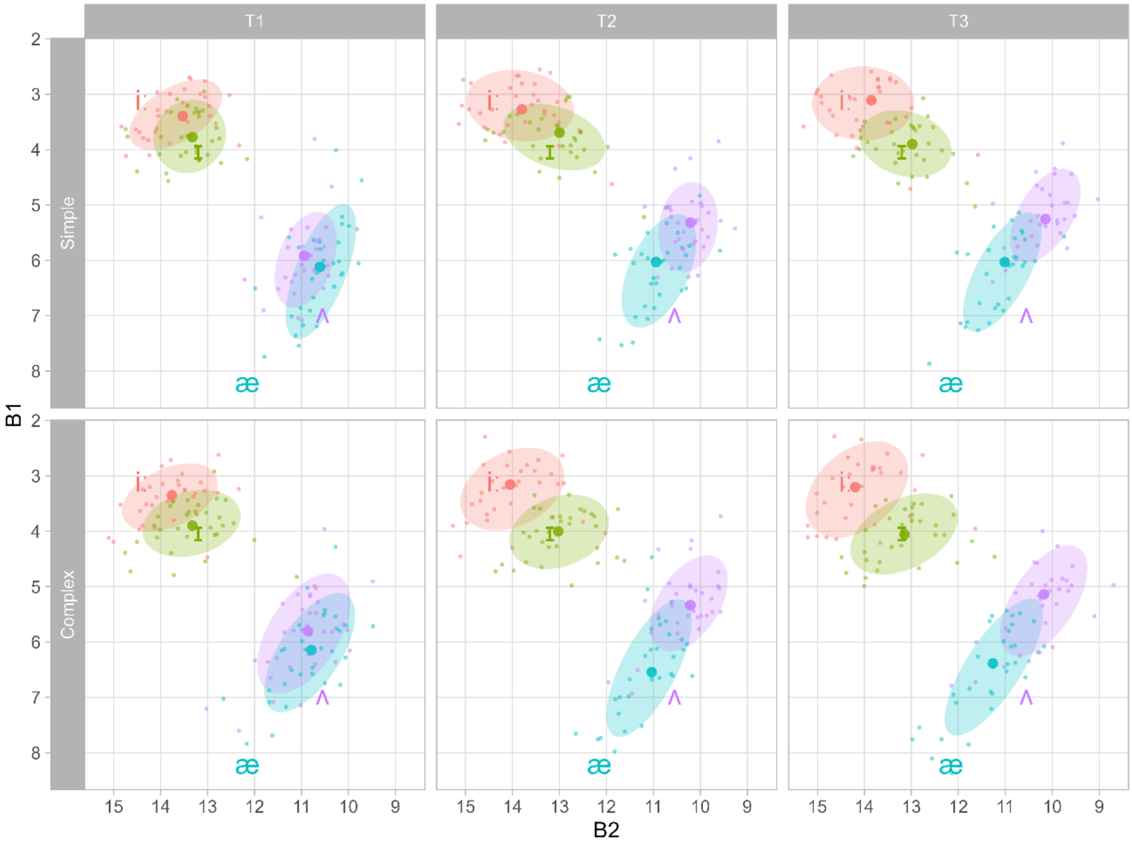


Figure 4.21. Learners' mean B1-B2 formant values (big dots) and mean B1-B2 formant values for each learner (small dots) for English vowels /iː/, /ɪ/, /æ/, /ʌ/ in word contexts. Filled ellipses represent 75% confidence intervals. Vowel transcriptions indicate mean B1-B2 formant values for native speakers of English. Graph organized by Time (T1/T2/T3) and Group (simple/complex).

Secondly, whereas the SG's production of /æ/ became further forward (i.e., higher B2) from T1 to T2, and /ʌ/ became more centralized (i.e., lower B2) and higher (i.e., lower B1), minimal formant changes happened from T2 to T3. With respect to native speaker's values, the distance was still considerable but they approximated vowels in the expected direction. Instead, the CG's production of /æ/ became substantially lower (i.e., higher B1) and more front (i.e., higher B2), and /ʌ/ productions became considerably higher (i.e., lower B1) and more centralized (i.e., lower B2) from T1 to T3. As depicted in Figure 4.21., learners' distance between native and learner vowels got significantly reduced from T1 to T3.

Table 4.13. Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) by Group (simple/complex), Contrast (/iː-ɪ/, /æ-ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the DWR task.

			Vowel distinctiveness			
			T1	T2	T3	*gain*
Simple	/iː-ɪ/	*M*	11.19	12.17	13.97	2.78
		Mdn	7.04	9.50	10.56	3.52
		SD	10.97	7.34	11.36	13.49
	/æ-ʌ/	*M*	8.93	13.26	12.16	3.23
		Mdn	7.46	10.16	10.08	2.62
		SD	6.59	9.66	6.54	6.63
Complex	/iː-ɪ/	*M*	12.10	14.37	16.45	4.35
		Mdn	6.33	9.68	9.82	3.49
		SD	15.52	13.17	15.46	18.41
	/æ-ʌ/	*M*	8.75	12.04	13.28	4.54
		Mdn	5.61	9.47	10.87	5.26
		SD	9.79	8.44	7.53	9.32

Note. M = Mean, Mdn= Median, SD = Standard Deviation

Concerning vowel ***distinctiveness***, Mahalanobis distances between contrastive vowels were fitted to a GLMM with an identity linking function. *Group* (simple/complex), *Time* (T1/T2/T3), *Contrast* (/iː-/ɪ/, /æ-ʌ/), and their interactions were entered as predictors. The random-effects structure included *Subject* and *Item* as random intercepts. The model, whose parameter estimates are shown in Appendix L.55., yielded significant main effects of *Time* (F [2, 9060] = 26.11, p < .001) and *Contrast* (F [1, 9060] = 22.92, p < .001), but non-significant main effects of *Group* (F [1, 9060] = 1.19, p = .274) or a *Group x Time* (F [2, 9060] = .95, p = .387) interaction. Despite the non-significant interaction, Bonferroni-adjusted pairwise contrasts revealed that the SG improved the distinctiveness of confusable vowels from T1 to T2, but not so much from T2 to T3 (t [9060] = -.53, p = .591). Instead, the CG made less of an overlap between vowels from T1 to T2 (t [9060] = -3.69, p < .001), and kept increasing the vocalic distance from T2 to T3 (t [9060] = -2.14, p = .032). See Figure 4.22., left panel. The *Group x Time* *x Contrast* (*F* [5, 9060] = 2.54, *p* = .026) interaction reached significance because, from T1 to T2, the SG got significantly better at producing distinct vowels from the /æ-ʌ/ contrast (*t* [9060] = -4.26, *p* < .001), but not the /iː-ɪ/ contrast, *t* [9060] = -.95, *p* = .338. However, the CG got significantly better at distinguishing the confusable vowels from the /iː-ɪ/ (*t* [9060] = -2.28, *p* = .049) and /æ-ʌ/ (*t* [9060] = -3.20, *p* = .003) contrasts. Overall, the CG obtained greater gains (4.5 SD) than the SG (3.0 SD) in the separation of L2 confusing vowels (Table 4.13.). Last, both groups were comparable before the intervention (*t* [9060] = -.85, *p* = .393).

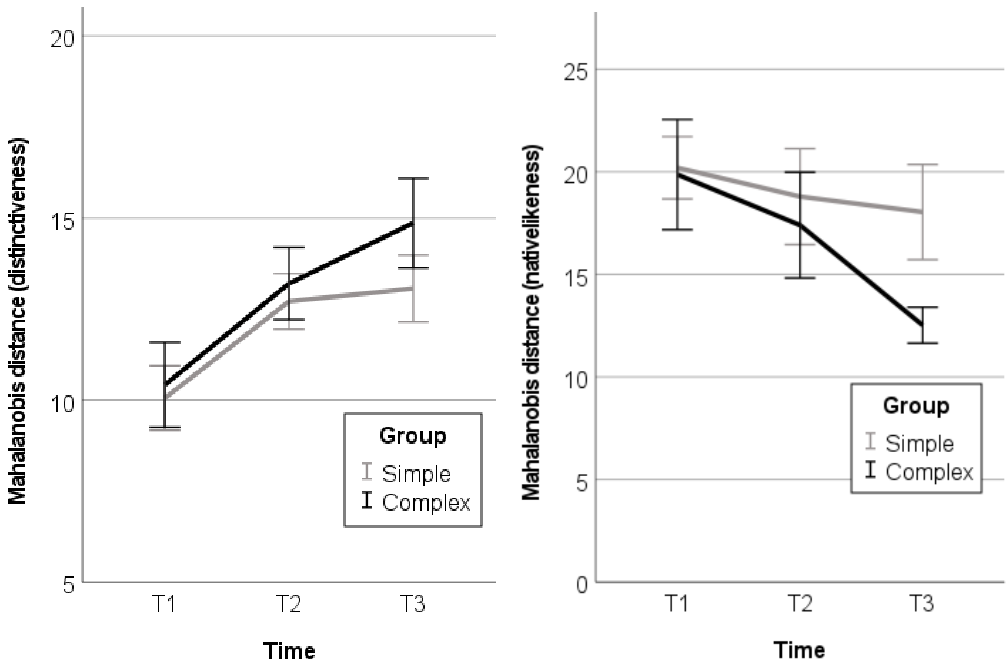


Figure 4.22. Mahalanobis distances (*distinctiveness*) on the left panel, Mahalanobis distances (*nativelikeness*) on the right panel, produced in words in isolation (DWR). Graphs organized by Group (simple in grey/complex in black) and Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

In terms of vowel ***nativelikeness***, Mahalanobis distances between learner and native vowels were fitted to a GLMM with an identity linking function. *Group* (simple/complex), *Time* (T1/T2/T3), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/), and their interactions were entered as fixed effects. *Subject* and *Item* were included as random intercepts. The model yielded non-significant main effects of *Group* (*F* [1, 9048] = 1.53, p = .215), but significant main effects of *Time* (*F* [2, 9048] = 11.62, p < .001), *Vowel* (*F* [3, 9048] = 11.62, p < .001) and significant *Group x Time* (*F* [2, 9048] = 4.63, p = .010) and *Group x Time x Vowel* (*F* [15, 9048] = 4.12, p < .001) interactions. Despite apparent non-group differences, the *Group x Time* interaction revealed that, the SG decreased the distance with native speaker vowels from T1 to T2 and from T2 to T3 but these differences did not reach significance (*t* [9048] = .94, p = .689; *t* [9048] = .41, p = .700). In contrast, the CG's vowels got significantly more target-like from T1 to T2 (*t* [9048] = 2.28, p = .022), and from T2 to T3 (*t* [9048] = 3.30, p = .002). See Figure 4.22., right panel. The triple interaction reached significance because, from T1 to T2, the SG got more accurate at producing /iː/, /æ/, /ʌ/ (but not /ɪ/); instead, all CG's vowel productions were more native-like at T2 (especially /æ/ and /ʌ/, p < .05). In general, the CG became closer to native-speaker values than the SG at T3 and obtained overall greater gains (SG: 2.2 SD, CG: 7.4 SD) (but see Table 4.14. for differences among vowels). Last but not least, SG and CG obtained similar nativelikeness distances at T1 (*t* [9048] = .10, p = .913), which made them comparable before the intervention. See parameter estimates in Appendix L.56.

Table 4.14. Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) by Group (simple/complex), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the DWR task.

			Vowel nativelikeness			
			T1	T2	T3	*gain*
Simple	/iː/	*M*	19.07	17.12	16.06	3.01
		Mdn	18.39	16.37	13.16	5.23
		SD	6.72	10.53	8.35	7.09
	/ɪ/	*M*	26.58	33.65	29.95	-3.37
		Mdn	19.17	17.64	12.12	7.05
		SD	24.32	52.13	65.89	62.06
	/æ/	*M*	22.92	17.76	18.92	4.00
		Mdn	17.03	12.99	15.39	1.64
		SD	20.86	15.75	14.12	19.99
	/ʌ/	*M*	12.24	6.65	7.24	4.99
		Mdn	12.00	5.68	6.23	5.77
		SD	4.10	4.36	3.92	3.55

Complex	/iː/	*M*	14.28	13.05	13.69	0.59
		Mdn	12.75	12.59	12.99	-0.24
		SD	7.38	5.79	5.45	6.51
	/ɪ/	*M*	35.70	34.68	14.73	20.96
		Mdn	15.75	15.58	9.35	6.40
		SD	91.29	68.71	12.12	92.67
	/æ/	*M*	19.88	13.86	13.29	6.59
		Mdn	15.13	10.03	11.40	3.73
		SD	19.46	11.44	8.78	16.73
	/ʌ/	*M*	9.62	8.02	8.38	1.25
		Mdn	9.28	6.59	8.17	1.11
		SD	5.02	5.35	4.22	4.97

Note. M = Mean, Mdn= Median, SD = Standard Deviation

In terms of vowel ***duration***, before assessing learners' DR per contrast (/iː-ɪ/ - /æ-ʌ/), two-tailed paired sample t-tests showed that the SG' duration of /iː/ was significantly larger than /ɪ/ (t [95] = 15.17, p < .001) and the duration of /æ/ was significantly larger than /ʌ/ (t [95] = 19.34, p < .001). Likewise, paired sample t-tests with the CG group showed that learners' duration of /iː/ was significantly larger than /ɪ/ (t [95] = 13.34, p < .001), and the duration of /æ/ was significantly larger than /ʌ/ (t [95] = 20.87, p < .001) (Table 4.15.).

Table 4.15. Duration and duration ratio by Group (simple/complex) and Time (T1/T2/T3) in the DWR task.

		Duration (ms)						Duration ratio (ms)					
		T1		T2		T3		T1		T2		T3	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Simple	/iː/	113.89	13.35	120.76	13.58	135.03	18.77	1.18	0.14	1.25	0.15	1.25	0.14
	/ɪ/	97.50	13.66	97.49	11.88	108.91	14.37						
	/æ/	135.53	12.39	144.89	16.74	161.40	21.44	1.21	0.09	1.24	0.13	1.27	0.14
	/ʌ/	112.11	11.22	117.68	15.69	127.80	16.26						
Complex	/iː/	109.30	13.45	122.81	12.68	126.42	17.86	1.13	0.13	1.23	0.14	1.29	0.17
	/ɪ/	97.41	9.58	100.75	10.25	98.51	10.44						
	/æ/	132.55	16.88	145.55	15.38	148.41	15.47	1.19	0.11	1.24	0.10	1.24	0.11
	/ʌ/	111.55	13.41	117.66	12.22	120.46	14.68						

Note. M = Mean, SD = Standard Deviation

We submitted DR scores to LMM. *Group*, *Time* and their interactions were submitted as fixed effects, and the random-effects structure included an intercept for

Subject. The first model with DR for /iː-ɪ/ revealed non-significant main effects of *Group* (*F* [1, 183] = .26, *p* = .607) but significant effects of *Time* (*F* [2, 183] = 10.75, *p* < .001), and the *Group* x *Time* interaction did not reach significance (*F* [2, 183] = 1.65, *p* = .194) (Appendix L.57.). The second model with DR for /æ-ʌ/ revealed non-significant main effects of *Group* (*F* [1, 183] = .94, *p* = .333), *Time* (*F* [2, 183] = 3.29, *p* = .039), nor a significant *Group* x *Time* interaction (*F* [2, 183] = .31, *p* = .730). Bonferroni-adjusted pairwise contrasts revealed that both SG and CG increased the duration between /iː-ɪ/, but they did not do so for vowels /æ-ʌ/. See Table 4.15. and Appendix L.58.

4.2.3.2. DSR

In this subsection, we assess task complexity effects on the production of English vowels /iː/, /ɪ/, /æ/ and /ʌ/ in words in sentence contexts, measured in terms of vowel quality (B1-B2 descriptions, vowel *distinctiveness*, and vowel *nativelikeness*) and vowel quantity. Forthcoming analyses only included the experimental group (*N*=63) (SG: *N*= 31; CG: *N*= 32).

Overall, both groups learned to produce L2 confusing vowels more distinctively and accurately from T1 and T2 (Figure 4.23.). On the one hand, the SG's separation of /iː-ɪ/ from T1 to T2 was because they produced /iː/ further forward (i.e., higher B2) and slightly higher (i.e., lower B1) and /ɪ/ became more centralized (i.e., lower B2) from T1 to T2. While this contrast productions were more targetlike from T1 to T2, at T3, learners' production of /iː/ became further way (i.e., lower B2) from the native speakers' model and /ɪ/ slightly lower (i.e., higher B1); hence, vowel distinctiveness remained similar from T2 to T3. Concerning the CG, the TBPT intervention clearly helped learners produce more distinct realizations of /iː-ɪ/. As depicted in Figure 4.23., English productions of /iː/ became higher (i.e., lower B1) and more front (i.e., higher B2) and /ɪ/ more centralized (i.e., lower B2) and lower (i.e., higher B1) from T1 to T3. At the same time, their productions approached native speakers' /iː-ɪ/ B1-B2 values.

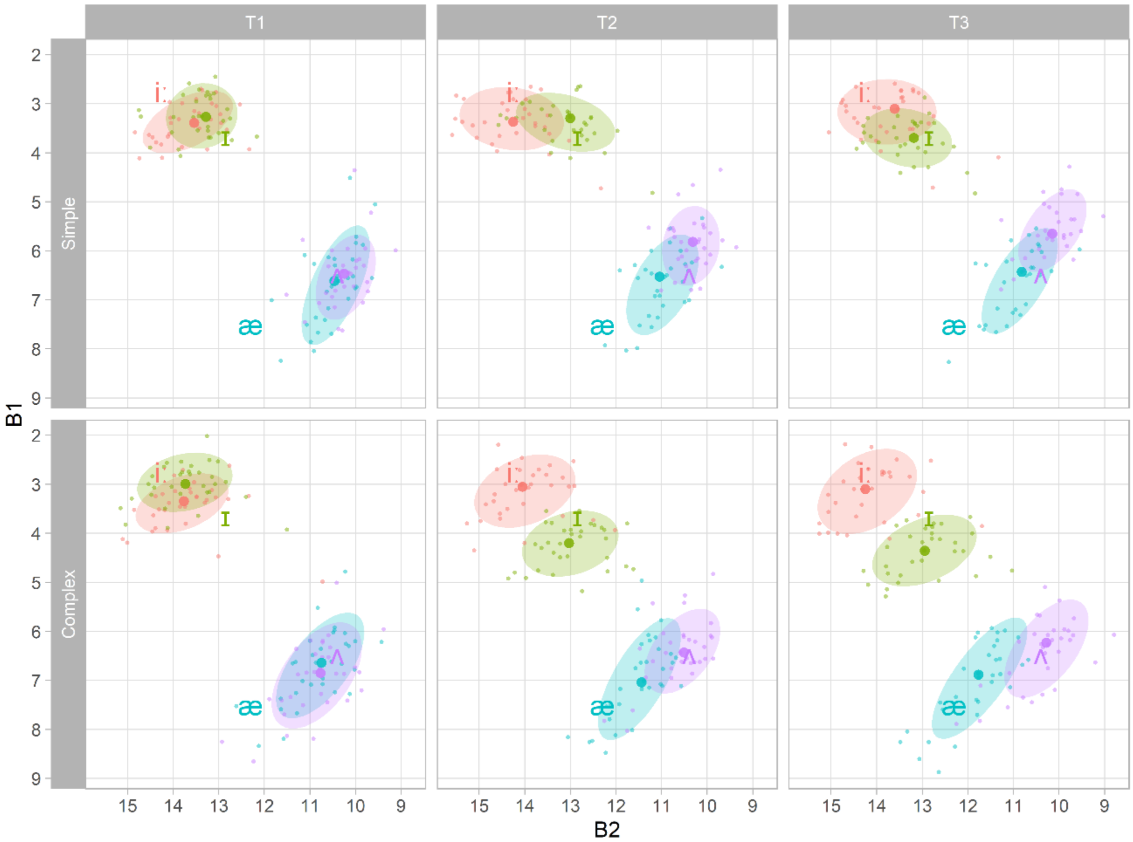


Figure 4.23. Learners' mean B1-B2 formant values (big dots) and mean B1-B2 formant values for each learner (small dots) for English vowels /iː/, /ɪ/, /æ/, /ʌ/ in sentence contexts. Filled ellipses represent 75% confidence intervals. Vowel transcriptions indicate mean B1-B2 formant values for native speakers of English. Graph organized by Time (T1/T2/T3) and Group (simple/complex).

On the other hand, the SG's production of /æ/ became more front (i.e., higher B2), but not lower, and /ʌ/ became higher (i.e., lower B1) and slightly more centralized (i.e., lower B2) from T1 to T2. However, learners' productions did not get more accurate or more distinct from T2 to T3 because the realizations of /æ/ were less front. Contrastively, the CG's production of /æ/ became considerably lower (i.e., higher B1) and especially more front (i.e., higher B2), and /ʌ/ productions became higher (i.e., lower B1) and more centralized (i.e., lower B2) from T1 to T3. In addition, the CG's /æ/ and /ʌ/ changed in the direction of the native speakers' productions, as shown in Figure 4.23.

Table 4.16. Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) by Group (simple/complex), Contrast (/iː-ɪ/, /æ-ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the DSR task.

			Vowel distinctiveness			
			T1	T2	T3	*gain*
Simple	/iː-ɪ/	*M*	9.66	15.34	11.99	2.33
		Mdn	5.51	12.67	8.58	3.07
		SD	8.11	6.44	9.25	12.19
	/æ-ʌ/	*M*	7.40	13.43	10.18	2.78
		Mdn	5.93	10.32	8.10	2.17
		SD	5.22	8.55	7.64	7.66
Complex	/iː-ɪ/	*M*	10.20	17.77	19.42	9.22
		Mdn	4.40	13.44	12.79	8.39
		SD	13.71	15.22	17.44	16.11
	/æ-ʌ/	*M*	6.83	12.03	13.25	6.42
		Mdn	3.68	9.18	10.84	7.16
		SD	8.43	8.66	6.22	8.40

Note. M = Mean, Mdn= Median, SD = Standard Deviation

In terms of vowel ***distinctiveness,*** we submitted Mahalanobis distances between contrastive vowels to a GLMM with an identity linking function. The fixed-effects structure included *Group* (simple/complex), *Time* (T1/T2/T3), *Contrast* (/iː-ɪ/, /æ-ʌ/), and their interactions; the random-effects structure included *Subject* and *Item* as intercepts. The model, whose parameter estimates can be found in Appendix L.59., rendered significant main effects of *Group* (F [1, 9060] = 20.82, p < .001), *Time* (F [2, 9060] = 82.08, p < .001), *Contrast* (F [1, 9060] = 70.98, p < .001), as well as significant *Group x Time* (F [2, 9060] = 15.85, p < .001) and *Group x Time x Contrast* (F [5, 9060] = 3.61, p = .003) interactions. In general, the CG separated contrasting vowels to a larger extent than the SG. Bonferroni-adjusted pairwise contrasts showed that both SG (t [9060] = -7.97, p < .001) and CG (t [9060] = -8.84, p < .001) improved the distinctiveness of L2 contrastive vowels from T1 to T2. Nevertheless, while the CG's vowels became more distinct at T3 (t [9060] = -1.98, p = .047), the SG significantly reduced the distance between contrasting vowels at T3 (t [9060] = 4.49, p < .001). See Figure 4.24., left panel. Pairwise comparisons (Bonferroni-adjusted) also showed that learners' improvement in vowel *distinctiveness* happened with both contrasts (/iː-ɪ/ and /æ-ʌ/) from T1 to T2 ($p < .001$) for both groups; and the SG got significantly worse at separating L2 contrastive vowels from both contrasts, while the CG got better (albeit non-significantly) with both contrasts ($p > .05$) at T3. Overall, the CG who obtained the largest gains (7.8 SD, vs. 2.5SD from the SG) (Table 4.16.). Finally, both groups were shown to perform similarly before the TBPT intervention (t [9060] = .019, p = .985).

In terms of vowel ***nativelikeness***, Mahalanobis distances between native and non-native vowels were submitted to a GLMM with *Group* (simple/complex), *Time* (T1/T2/T3), *Vowel* (/iː/, /ɪ/, /æ/, /ʌ/), and their interactions as fixed effects. *Subject* and *Item* were included as random intercepts. The model rendered non-significant main effects of *Group* (F [1, 9048] = 2.32, p = .127) but significant main effects of *Time* (F [2, 9048] = 13.90, p < .001), *Vowel* (F [3, 9048] = 107.61, p < .001) and significant *Group x Time* (F [2, 9048] = 5.68, p = .003) and *Group x Time x Vowel* (F [15, 9048] = 1.77, p = .032) interactions.

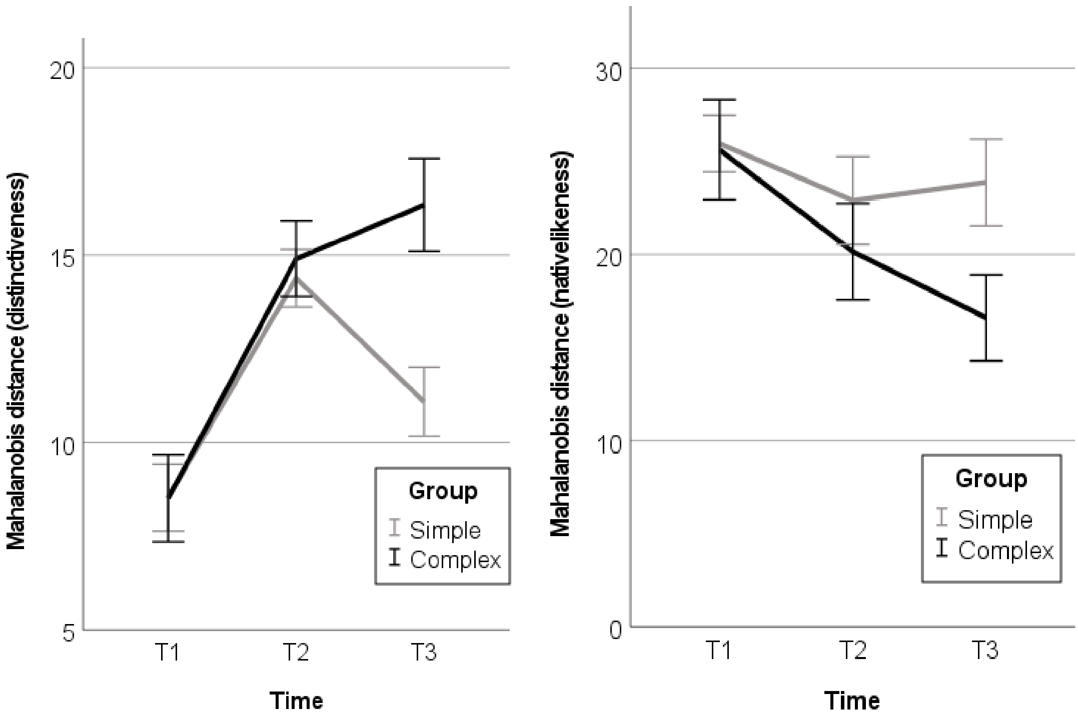


Figure 4.24. Mahalanobis distances (*distinctiveness*) on the left panel and Mahalanobis distances (*nativelikeness*) on the right panel, produced in words in sentences (DSR). Graphs organized by

Group (simple in grey/complex in black) and Time (T1/T2/T3). Error bars indicate 95% confidence intervals.

Bonferroni-adjusted pairwise contrasts from the *Group x Time* interaction showed that, whereas both SG and CG improved from T1 to T2, only the CG improved significantly: t [9048] = 3.77, p < .001 (vs. SG: t [9048] = 1.90, p = .172). From T2 to T3, the SG produced slightly less target-like vowels (t [9048] = -.66, p = .508); however, the CG kept producing more target-like vowels at T3 (t [9048] = 2.13, p = .033). See Figure 4.24., right panel.

Table 4.17. Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) and vowel duration by Group (simple/complex), Vowel (/iː/, /ɪ/, /æ/, /ʌ/), Time (T1/T2/T3) and gains (T3-T1) in the DSR task.

			Vowel nativelikeness			
			T1	T2	T3	*gain*
Simple	/iː/	*M*	24.82	22.24	21.97	2.85
		Mdn	24.14	21.12	18.91	5.23
		SD	8.42	10.98	10.66	6.22
	/ɪ/	*M*	32.33	38.40	35.70	-3.37
		Mdn	24.92	22.39	17.87	7.05
		SD	28.22	54.11	52.99	70.16
	/æ/	*M*	28.67	22.53	24.80	3.86
		Mdn	22.78	17.74	21.14	1.64
		SD	19.46	16.76	17.29	18.00
	/ʌ/	*M*	17.99	8.40	12.99	5.00
		Mdn	17.75	7.43	11.98	5.77
		SD	5.12	5.26	4.52	6.05
Complex	/iː/	*M*	20.03	13.69	12.49	7.54
		Mdn	18.50	13.09	11.74	6.76
		SD	10.48	6.56	6.22	6.77
	/ɪ/	*M*	41.45	35.18	28.04	13.41
		Mdn	21.49	14.08	11.05	10.44
		SD	81.11	69.51	53.99	104.66
	/æ/	*M*	25.63	19.22	18.63	7.00
		Mdn	20.88	14.53	15.15	5.73
		SD	20.03	15.01	11.05	14.09
	/ʌ/	*M*	15.37	12.52	7.23	8.14
		Mdn	15.03	11.09	6.92	8.11
		SD	7.30	6.33	3.24	5.18

Note. M = Mean, Mdn= Median, SD = Standard Deviation

Further pairwise comparisons revealed that only the SG's realizations of /ʌ/ (t [9048] = 2.98, p = .009) became more accurate from T1 to T2; instead, the CG produced significantly more accurate productions of /iː/, /æ/, /ɪ/ (and almost /ʌ/) from T1 to T2 (p < .05). Overall, the CG obtained greater accuracy gains (9.0 SD) than the SG (2.1 SD) (Table 4.17.). Last, non-significant differences were found between T1 and T2 (t [9048] = .18, p = .852). See parameter estimates in Appendix L.60.

Finally, in terms of **vowel duration** (ms), two-tailed paired sample t-tests showed that the SG' duration of /iː/ was significantly larger than /ɪ/ (t [95] = 14.09, p < .001) and the duration of /æ/ was significantly larger than /ʌ/ (t [95] = 18.64, p < .001). Similarly, paired sample t-tests with the CG revealed that learners' duration of /iː/ was significantly larger than /ɪ/ (t [95] = 12.41, p < .001), and the duration of /æ/ was significantly larger than /ʌ/ (t [95] = 20.26, p < .001) (See Table 4.18.)

Table 4.18. Duration and duration ratio by Group (simple/complex) and Time (T1/T2/T3) in the DSR task.

		Duration (ms)						Duration ratio (ms)					
		T1		T2		T3		T1		T2		T3	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Simple	/iː/	102.89	9.44	101.77	13.89	105.10	16.08	1.20	0.16	1.26	0.18	1.30	0.20
	/ɪ/	86.50	13.26	81.74	11.79	82.16	10.96						
	/æ/	124.53	14.39	126.80	10.91	132.47	11.92	1.24	0.10	1.26	0.15	1.34	0.19
	/ʌ/	101.11	12.76	101.93	15.30	99.87	16.39						
Complex	/iː/	100.30	13.33	102.02	12.46	105.74	17.92	1.14	0.14	1.24	0.17	1.31	0.21
	/ɪ/	88.41	11.53	83.20	10.62	81.01	10.85						
	/æ/	123.55	16.70	125.66	15.55	128.73	15.98	1.21	0.13	1.26	0.13	1.28	0.14
	/ʌ/	102.55	13.36	100.11	12.81	101.78	14.22						

Note. M = Mean, SD = Standard Deviation

DR scores were submitted to LMM. The fixed-effects structure included *Group*, *Time* and their interactions, and the random-effects structure included an intercept for *Subject*. The first model with DR for /iː-ɪ/ revealed non-significant main effects of *Group* (F [1, 183] = .71, p = .398) but significant effects of *Time* (F [2, 183] = 8.91, p < .001), and the *Group* x *Time* interaction did not reach significance (*F* [2, 183] = .82, *p* = .438) (Appendix L.61.). When applying the same LMM for /æ-ʌ/, results revealed non-significant main effects of *Group* (*F* [1, 183] = 1.94, *p* = .165) but significant effects of *Time* (*F* [2, 183] = 5.69, *p* = .004), and the *Group* x *Time* interaction did not reach significance (*F* [2, 183] = .98, *p* = .376). It can be concluded that both SG and CG's duration ratio increased over time, especially for /iː-ɪ/, but no differences were found between groups. See Table 4.18. and Appendix L.62.

Summary of results:

Globally, learners' productions of /iː/ became more fronted and higher, /ɪ/ more centralized and lower, /æ/ more fronted and lower, and /ʌ/ more centralized and higher from T1 to T3, but these changes depended on the group. Concerning the words produced in isolation, confusable vowels became more distinct from T1 to T2 for the SG and CG. Whereas the CG separated the vowels of the two L2 vowel contrasts significantly, the SG mainly separated vowels /æ-ʌ/. Only the CG kept separating the confusable vowels at T3. In terms of vowel accuracy, non-native vowels became closer to native vowels for both groups from T1 to T2 and T2 to T3, but such improvement in vowel nativelikeness only reached significance for the CG who produced all vowels in a more target-like manner over time. Concerning vowels produced in sentence contexts, both SG and CG produced significantly more distinct vowels from T1 to T2, but only the CG kept improving significantly in vowel distinctiveness at T3 with both L2 vowel contrasts. As for vowel accuracy, the CG improved significantly from T1 to T2 and T2 to T3 with almost all vowels, whereas the SG improvement did not reach significance. Globally, in both DWR and DSR tasks, the CG gained more in vowel distinctiveness and nativelikeness than the SG, but both groups relied on vowel duration to separate /iː-ɪ/ over time.

4.2.4. P-LRE

This subsection examines the extent to which task complexity influenced the occurrence and duration of pronunciation-focused language-related episodes (P-LREs) over time. The conversation data was obtained from 31 dyads (SG: *N*= 15; CG: *N*=16). Data from one learner from the SG had to be excluded because, due to odd numbers, performed the interactive task with the researcher in charge. To start with, there was a total of 282 P-LREs across all observed interactions (SG: *N* = 111; CG: *N* = 171), with an average of

3.03 P-LREs per dyad (*SD* = 1.95; *95% CI*: 2.62-3.43) and a range from 0 to 7. In terms of P-LRE duration, the average length of each P-LRE was 11.92 seconds (*SD* = 4.49; *95% CI*: 10.89, 12.94), and differed across groups (SG: *M* = 10.07; *SD* = 3.63; *95% CI*: 8.84, 11.30; CG: *M* = 13.58; *SD* = 4.57; *95% CI*: 12.12, 15.04).

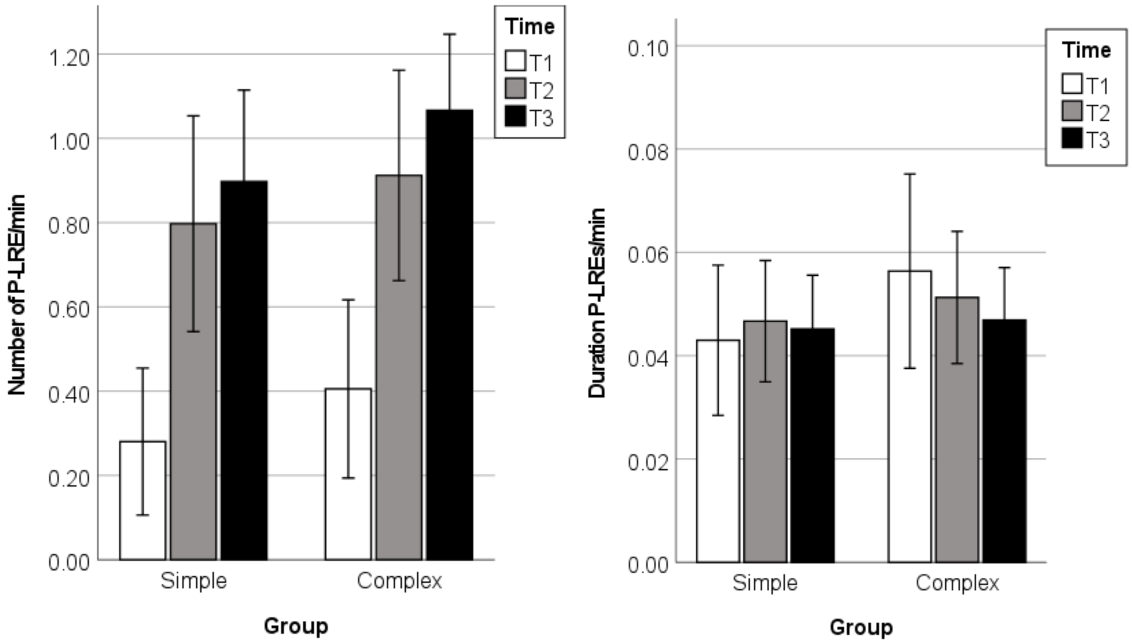


Figure 4.25. Number of P-LRE per minute (left panel) and duration of P-LRE per minute (right panel) in the interactive task as a function of Group (simple/complex) and Time (T1 in white/T2 in grey/T3 in black). Error bars indicate 95% confidence intervals.

On the one hand, the number of P-LRE per minute were submitted to LMM with *Group* (simple/complex), *Time* (T1/T2/T3), and their interactions as fixed effects. The random-effects structure included random intercepts for *Pair*. The model, whose parameter estimates are shown in Appendix L.63., yielded significant main effects of *Time* (*F* (2, 87) = 22.17, *p* < .001) but not of *Group* (*F* (1, 87) = 2.69, *p* = .104) or a *Group x Time* (*F* (1, 87) = .04, *p* = .960) interaction. Therefore, incidence of P-LRE increased significantly from T1 (*M* = .63; *SD* = .22; *95% CI*: .51, .74) to T2 (*M* = .91; *SD* = .41; *95% CI*: .75, 1.07) to T3 (*M* = 1.01; *SD* = .32; *95% CI*: .89, 1.13). Although the number of P-LRE was higher in the CG than the SG in each time (Figure 4.25., left panel), the ratio (P-LRE/min) did not reveal significant differences between groups.

On the other hand, the duration of P-LRE per minute was submitted to linear mixed models with *Group* (simple/complex), *Time* (T1/T2/T3), and their interactions as predictors, and a random intercept for *Pair*. The model, whose parameter estimates can be shown in Appendix L.64., rendered non-significant main effects of *Group* (F (1, 70) = 2.29, p = .134), *Time* (F (2, 70) = .38, p = .683) or a *Group x Time* interaction (F (2, 70) = .74, p = .481). Although the CG spent more time on the P-LREs at T1 than the SG, these learners were engaging in P-LREs of similar duration at T3, and the error bars indicate a lot of individual variability, especially at T1 (Figure 4.25., right panel).

4.3. The role of ID

The third section of the results aims to (1) compare experiential (L2 past and recent L2 learning experience, L2 proficiency) and cognitive factors (WM and attention control) across groups (SG, CG and CTG) and the relationship between such factors for the experimental groups (RQ3.1.); (2) establish associations between experiential and cognitive ID and L2 vowel perception, lexical encoding and production performance (RQ3.2.); and (3) investigate how experiential and cognitive ID mediate gains in L2 speech development (RQ3.3.).

4.3.1. Experiential and cognitive factors across groups

Table 4.19. first presents the descriptive statistics of experiential factors (see Section 3.3.4.1. and Section 3.3.3.6., for operationalization and calculation of the variables) and cognitive factors (see Section 3.3.3.7., 3.3.3.8. for operationalization and calculation of WM and ASA, respectively) by group. In addition, after confirming the null hypothesis

of equal population variances through Levene's test, one-way ANOVAs were conducted to examine possible differences across groups.

Table 4.19. Three-group comparison (simple [SG], complex [CG], and control [CTG]) for experiential and cognitive factors.

		SG	CG	CTG	*p*	*ηp2*
Experiential factors						
Past English learning experience (years)	*N*	31	32	29	.431	.019
	M	10.26	10.47	9.86		
	SD	2.08	1.75	1.62		
	95% CI	9.49-11.02	9.83-11.10	9.25-10.48		
Recent English learning inside the classroom (hours x week)	*N*	31	32	29	.104	.050
	M	2.62	2.23	2.22		
	SD	1.06	.68	.62		
	95% CI	2.22-3.01	1.98-2.48	1.99-2.46		
Recent English learning outside the classroom (frequency: 1=never; 5=every day)	*N*	31	32	29	.432	.019
	M	2.54	2.64	2.39		
	SD	.73	.75	.75		
	95% CI	2.28-2.81	2.36-2.91	2.10-2.68		
L2 Proficiency (1-120)	*N*	29	30	27	.918	.002
	M	72.28	70.13	71.68		
	SD	23.94	15.77	21.18		
	95% CI	63.17-81.38	64.24-76.02	63.29-80.06		
Cognitive factors						
Working memory -Forward DS (Mean Span)	*N*	31	30	24	.803	.005
	M	6.52	6.7	6.58		
	SD	1.06	.88	1.12		
	95% CI	6.13-6.91	6.36-7.03	6.11-7.06		
Working memory - Backward DS (Mean Span)	*N*	31	30	24	.368	.024
	M	5.49	5.76	5.32		
	SD	1.12	1.15	1.12		
	95% CI	5.08-5.90	5.33-6.19	4.85-5.80		
Auditory selective attention (0-64)	*N*	30	28	26	.033	.137
	M	37.93	44.43	36.73		
	SD	7.58	8.6	9.62		
	95% CI	35.10-40.77	41.09-47.76	32.84-40.50		

Note. N = Number, M = Mean, SD = Standard Deviation, CI = Confidence intervals

In terms of experiential factors, the three groups showed no significant differences for past English learning experience and recent English learning experience inside and outside the classroom. Whereas the SG appeared to have greater English learning inside the classroom, the CG reported having greater past English learning experience and recent learning experience outside the classroom; however, differences between these two experimental groups did not reach significance. In addition, L2 proficiency appeared to be slightly higher in the SG than the CG and CTG. In terms of cognitive factors, learners in the CG seemed to have greater WM and ASA. Only the CG appeared to have a significantly better ASA than the CTG.

Given that there were overall no significant differences in terms of experiential and cognitive factors across groups, two-tailed Spearman-rank correlations between experiential and cognitive factors were jointly conducted for the experimental groups ($N = 56$: $SG = 28$, $CG = 28$), which were also selected for RQ4.2. and RQ4.3.

Table 4.20. Two-tailed Spearman-rank correlations between experiential and cognitive factors of experimental and control groups.

Predictor	1	2	3	4	5	6	7
1. Past English learning experience	—	.09	.07	.04	.10	.05	.12
2. Recent English learning inside the classroom	.09	—	.07	.34*	.31*	.11	.00
3. Recent English learning outside the classroom	.07	.07	—	.33*	.36**	.26*	.37**
4. L2 Proficiency	.04	.34*	.33*	—	.50**	.20	.35**
5. Working memory- Forward digit span	.10	.31*	.36**	.50**	—	.61**	.33*
6. Working memory- Backward digit span	.05	.11	.26*	.20	.61**	—	.40**
7. Auditory selective attention	.12	.00	.37**	.35**	.33*	.40**	—

Note. * $p < .05$ (2-tailed); ** $p < .001$ (2-tailed); N=56

As observed in Table 4.20., learners' experience learning English in the *past* (i.e., years of instruction) was not associated with any other experiential or cognitive factor. However, the greater their *recent* English learning experience inside and outside the classroom, the greater their English proficiency (i.e., greater scores in the EIT) and also PSTM. Interestingly, their recent experience with L2 English (inside and outside the classroom) was positively (albeit weakly) related to complex WM, and ASA. Finally, L2 proficiency was moderately associated with PSTM and weakly to ASA, and ASA was weakly related to PSTM and complex WM.

4.3.2. ID and L2 speech performance

First of all, we were interested in assessing the relationship between experiential/cognitive ID and averaged performance (T1/T2/T3) in L2 vowel perception, lexical encoding and production of the experimental group only (*N*=56).[69] Pearson correlations, reported in Table 4.21., showed that learners with greater recent English learning experience outside the classroom obtained greater overall accuracy in the lexical encoding of L2 vowels (FLeC). In addition, L2 proficiency and PSTM were weakly correlated with vowel discrimination (ABX) accuracy and moderately correlated with lexical encoding (FLeC) and vowel nativelikeness (i.e., more target-like vowel productions) in the production of words (DWR) and sentences (DSR) task. Interestingly, complex WM (backward digit-span) was positively and moderately correlated with accuracy in vowel discrimination (ABX), lexical encoding (FLeC and LD) as well as vowel nativelikeness (i.e., greater approximation with native speaker vowel productions) in the DWR and DSR task. Last but not least, learners with greater ASA appeared to be significantly more accurate in speed and accuracy of vowel discrimination (ABX) and lexical encoding (FLeC and LD). The relationship between ASA and perception performance was the strongest among all[70].

Finally, standard multiple regressions were conducted to assess the overall predictive ability of cognitive ID on L2 vowel performance as well as the unique contribution of each one of the ID (i.e., proficiency, PSTM, complex WM and ASA) to L2 vowel performance (i.e., ABX accuracy, FLeC accuracy, LD accuracy, vowel

[69] Initially, correlations between performance in the perception and production tasks were initially conducted separately for each time (T1, T2 & T3), but given that the correlations were similar, we only reported the results of the averaged three times.
[70] Partial correlations controlling for L2 proficiency were also conducted. Although the strength of the correlations became overall slightly weaker, all significant associations were maintained except for PSTM and vowel perception, lexical encoding and vowel nativelikeness. Therefore, when L2 proficiency was taken into account, PSTM stopped predicting variation in L2 vowel performance.

nativelikeness in DWR and DSR).[71] As shown in Appendix L.65., for ABX discrimination accuracy, proficiency, PSTM, complex WM and ASA explained 38.5% of variance and complex WM made the largest unique contribution (β = .41; 10.30% of variance), although ASA also made a statistically significant contribution (β = .34). In terms of lexical encoding, ID accounted for 41% and 25% of the variance in the FLeC and LD tasks. With respect to FLeC, ASA made the largest unique contribution (β = .38; 10.11% of variance), followed by complex WM (β = .28); however, only ASA (β = .40; 10.25% of variance) arose as a significant contributor to the performance in the LD task. In terms of vowel production, ID explained significant variance in the production of target-like vowels in the DWR (52%) and DSR (39%) task. In both cases, complex WM made the largest unique contribution (DWR: β = -.73, 30.50% of variance; DSR: β = -.65, 26.01% of variance) but PSTM also contributed significantly to performance in the production of vowels embedded in words (β = -.62) and sentences (β = -.47). On average, 39% of L2 vowel performance from the experimental group (T1/T2/T3) was explained by ID, namely, proficiency, PSTM, complex WM and ASA and only 6% by L2 self-reported experience-related factors.

---

[71] This subset of variables was selected as they were found to be the most significantly and strongly correlated with L2 vowel performance (see Table 4.21.).
Correlations and collinearity diagnostics showed absence of multicollinearity (correlations <.70; Tolerance value < .10; Variance inflation factor [VIF] value > 10)

Table 4.21. Two-tailed Pearson correlations between experiential and cognitive individual differences and L2 vowel perception (ABX discrimination), lexical encoding (Forced lexical choice [FLeC] and Lexical decision [LD]), and production (Delayed word repetition [DWR] and Delayed sentence repetition [DSR]) T1/T2/T3 averaged scores of the experimental groups.

Predictor	ABX		FLeC		LD		DWR		DSR	
	Accuracy	RT	Accuracy	RT	Accuracy	RT	Vowel distinctiveness	Vowel nativelikeness	Vowel distinctiveness	Vowel nativelikeness
Past English learning experience	.07	-.20	.00	-.33*	.11	-.11	-.07	-.05	-.06	-.06
Recent English learning inside the classroom	.16	-.02	.09	.05	-.01	-.12	.21	-.05	.21	-.04
Recent English learning outside the classroom	.18	-.16	.52**	-.07	.24	-.07	-.13	-.11	-.19	-.08
L2 Proficiency	.33*	-.06	.36**	-.03	.24	-.10	.03	-.39**	.02	-.37**
PSTM-Forward DS	.30*	-.06	.38**	.10	.23	.02	-.03	-.34*	-.08	-.27*
Complex WM-Backward DS	.52**	-.27*	.48**	.17	.39**	.09	.28*	-.39**	-.26	-.37**
Auditory selective attention	.55**	-.37**	.57**	-.12	.52**	-.24	.20	.04	-.19	.01

Note. * p < .05 (2-tailed); ** p < .001 (2-tailed). *RT*= Reaction time; *DS*= Digit span. *N*=56

4.3.3. ID and L2 speech gains

To start with, in order to investigate the relationship between experiential and cognitive ID and L2 vowel perception, lexical encoding and production gains[72] immediately after the TBPT intervention, only SG (N = 28) and CG (N = 28) were selected. Pearson correlations, conducted separately by task (Table 4.22.), showed that overall the CG's gains correlated significantly more with individual factors (and associations were overall stronger) than the SG's gains. With respect to the SG, the greater the learners' past experience with English, the greater speed gains in vowel discrimination and the greater the gains in vowel nativelikeness in the DWR task. PSTM was positively related to speed gains in lexical encoding (LD) and those learners with higher complex WM were better and faster at vowel discrimination and lexical encoding (FLeC), respectively. No other significant correlations arose. In the case of the CG, recent L2 learning experience inside the classroom was related to gains in vowel discrimination and speed in lexical encoding, and those learners who had greater experience outside the classroom were faster in ABX discrimination. Interestingly, learners with higher proficiency and PSTM obtained the greatest gains in vowel nativelikeness, hence, became more accurate at the production of target vowels in isolated words. The complex WM measure was moderately correlated with vowel distinctiveness (DWR and DSR task) and vowel nativelikeness (DSR task) and, finally, learners with higher ASA obtained greater gains in the distinctiveness of vowels embedded in words.

[72] L2 vowel gains used for these analyses were calculated by subtracting T1 to T2, hence, raw gains obtained right after the intervention.

Table 4.22. Two-tailed Pearson correlations between experiential and cognitive individual differences and L2 vowel perception (ABX discrimination), lexical encoding (Forced lexical choice and Lexical decision), and production (Delayed word repetition and Delayed sentence repetition) T1-T2 gains of the simple (SG) and complex (CG) groups.

	ABX		FLeC		LD		DWR		DSR	
	Accuracy	RT	Accuracy	RT	Accuracy	RT	Vowel distinctiveness	Vowel nativelikeness	Vowel distinctiveness	Vowel nativelikeness
SG										
Past English learning experience	-.07	.45*	-.03	.07	-.12	.12	-.17	.42*	-.20	-.30
Recent English learning inside the classroom	.01	.19	-.01	.18	.17	-.06	-.15	.06	-.17	.00
Recent English learning outside the classroom	.15	-.09	-.13	.24	-.25	-.20	-.01	-.14	-.02	-.17
Proficiency	.20	.02	.06	.22	.21	.29	.16	-.13	.16	-.11
Working memory- Forward DS	.13	-.27	-.12	.01	.37	.52**	.35	.08	.35	.09
Working memory- Backward DS	.43*	-.30	-.14	.43*	.08	.12	.06	.21	.09	.10
Auditory selective attention	-.02	.04	.04	.26	-.04	.12	-.18	.03	-.13	-.07
CG										
Past English learning experience	-.12	.06	.12	.16	-.25	-.36	.30	.07	.25	-.04
Recent English learning inside the classroom	.39*	.30	-.28	.43*	-.15	.25	.07	-.24	-.17	-.09
Recent English learning outside the classroom	.13	.55**	.09	-.07	.02	.01	-.04	-.12	-.20	-.19
Proficiency	.12	-.02	.13	.32	.13	.19	-.06	.45*	-.09	-.34
Working memory- Forward DS	.10	-.29	.01	.22	-.19	.05	-.01	.45*	-.25	-.32
Working memory- Backward DS	.22	-.27	.32	.33	.11	.13	.52**	.33	.41*	.47*
Auditory selective attention	.01	-.25	.27	.27	-.10	.01	.38*	.03	.34	.04

Note. * p < .05 (2-tailed); ** p < .001 (2-tailed). *RT*= Reaction time; *DS*= Digit span. Simple: N=28; Complex: N=28

Despite the fact ID seemed to be more strongly related with gains in L2 vowel acquisition for the group who was instructed with complex tasks, the correlations were relatively weak and not consistent across groups, which means these results should be treated with caution. In addition, when the same Pearson correlations were conducted with the joint experimental group, correlations were spurious and weak ($r < .30$) and only the complex WM measure explained significant variance in accuracy gains in ABX discrimination ($\beta = .44$), speed gains in lexical encoding ($\beta = .52$), and more target-like vowels ($\beta = .46$; $\beta = .51$) in the production of words in isolation and sentence contexts, respectively. In total, only 7% of the variance in L2 vowel gains was explained by experiential and cognitive ID.

4.4. Learners' perceptions of the intervention

The last section of the results explores learners' perceptions of the TBPT intervention under three overarching categories, namely, learners' beliefs about pronunciation, learners' evaluation of the tasks and the overall project, and learners' sense of performance and improvement after the intervention. To do so, only the most relevant data of closed and open-ended questions from the post-intervention questionnaire (refer back to Section 3.3.4.3 for more details) will be commented. General results and reflections that apply to both intervention groups will be presented first (Section 4.4.1.), followed by a between-group comparison (Section 4.4.2.) of the most relevant questions regarding the evaluation of the project and self-perception of improvement. All results from closed questions and individual responses for open-ended questions can be found in Appendix M.

4.4.1. Beliefs, likeability and learning

This section presents the results of learners' beliefs about pronunciation (*Q4-Q6*), their opinions of the task-based project (*Q7-Q22, Q29*) and their perceptions of pronunciation performance and learning (*Q23-Q28, Q30*) after the TBPT intervention. Concerning the open-ended questions (*Q18, Q19, Q29, Q30*), a few representative comments have been selected in order to illustrate the learners' general opinions regarding the intervention received. All the remaining comments can be seen in Appendix M.

First of all, over 50% of the learners considered that practicing pronunciation in class prior to the intervention was slightly important, whereas their beliefs about the importance of working on pronunciation in class changed after participating in the tasks. Over 80% of the participants considered pronunciation to be either important or extremely important and the great majority considered that it should be instructed at least 3-4h per month, with 16% preferring 4-5h and 13% 2-3h.

Secondly, in terms of the project evaluation, over 85% of the learners reported that the pre-tasks had been useful to learn the meaning of the words and, especially, their pronunciation (92%). Only a minor percentage of the learners stated that the pre-tasks had not helped them learn either the meaning (6.5%) or the pronunciation (8%) of the words used in the task. Whereas the majority found the pre-tasks interesting and enjoyable (65%), a 19% did not show a clear position and around 16% of the learners did not enjoy this phase of the task framework. It may have been that the pre-task listening comprehension was too easy (71.5%) and did not mean a big challenge for them, especially as the sessions went by. Still, around 19% of learners found the listening task difficult to understand.

Concerning the tasks, there was a lot of variability in learners' perception of the mental effort and difficulty in task performance. Whereas 42% reported the tasks posed high mental effort, 38% reported low levels of mental effort, whereas 21% was unsure. Likewise, 47% of the learners considered the tasks to have been difficult, whilst the other half were unsure or perceived them to be easy. Interestingly, the source of difficulty seems to have stemmed from the task conditions, as stated by almost half of the learners, followed by pronunciation of the words or both the task conditions and pronunciation. Overall, 89% of the learners found the tasks interesting and enjoyable, with a very small percentage of task dislike, due to the visual materials of the tasks (24%), the topic of doing activities in London (19%), or both images and topic (17.5%). In fact, when inquired about their opinion of the drawings appearing in the flashcards, 10% of the learners' responses referred to their appealing design, 49% indicated they generated a positive reaction (i.e., they were funny and motivated them) and also helped achieve task success (14%). Learners also believed they aided the learning of vocabulary (16%) and, to a lesser extent, pronunciation (4%). Still, 7% of the responses indicated that images could be less "childish" and some might have been unnecessary.

Selected comments about the images' opinion

> They where [were] funny and motivate you to make the task (S38)
>
> They are perfect for each activity to help solve the tasks (S34)
>
> I think that images were a good idea because helped me remember the meaning (S02)
>
> I remember the sounds with the images (S32)
>
> they were ok but some for children (S33)

An additional 13% found it was both the topic and difficulty of the tasks that made them enjoyable. When asked which were the tasks that learners had enjoyed the least and the most, less than 7% of the learners indicated that they disliked any task, whilst the "history museum task" was the task they disliked the most (11%). Nevertheless, over 40% of the learners reported that they (really) enjoyed doing all tasks, especially task 20 "photos in the album", followed by task 19 "photos in the website", task 13 "clothes", task 17 "souvenirs", task 14 "roleplay party", task 9 "shopping centre" and task 8 "escape room". In fact, over 75% of the learners believed the tasks were realistic and could have taken place in London. Finally, learners were inquired about their strategies when they encountered problems communicating with their classmates. 39% of learners' responses revealed that they mainly asked for support, especially to the teacher (22%) or to their peers (15%), but they also tried to solve the communicative breakdowns by using interactional moves such as repetition (7%), paraphrasing (7%), clarification (13%), or making some kind of emphasis on pronunciation (4%). Other strategies included the use of gestures (4%), code-switching to the L1 (6%), or self-resolution through autonomous strategies such as improvising or thinking hard (7%).

Selected comments about communication problem-solving strategies

> I would ask the teacher for help. (S13)
>
> ask my intelligent friend how to say the word. (S66)
>
> Look at the drawing on the board to remember the pronunciation. (S26)
>
> I tried to find a solution by repeating or saying it in a different way. (S06)
>
> Ask again more details. (S08)
>
> I defined the words in a similar way (S09)
>
> I exagerated [exaggerated] the pronunciation of the word so she could hear the difference. (S12)
>
> Use my hands to help me say it. (S04)

As far as post-tasks are concerned, learners' agreement about the usefulness to remember the meaning and pronunciation of the target words ranged widely. On the one hand, almost 50% of the learners believed the post-task did not help them revise the meaning of the target words, whilst 25% believed it did, and the remaining 25% were unsure. Conversely, the majority of the learners (65%) did think that the post-tasks were useful to revise the target words' pronunciation, but almost 30% did not think so. In general, though, learners expressed post-task enjoyability, with over 60% of the learners agreeing that the post-tasks had been interesting and enjoyable. The last open-ended question dealt with the learners' likes and dislikes about the project. On the one hand, learners' likeability responses were diverse and were grouped into multiple themes and subthemes. The greatest number of responses referred to the overall enjoyability of the tasks (41%) as well as the impact on learning (25%). The 15% of responses highlighted the likeability of the project in general, whereas others evidenced that learners enjoyed the overall procedure (15%) including recording (7%) and working with small groups (8%). Regarding the tasks, responses manifested their likeability with respect to the different stages of the task framework (report, planning and post-tasks) (6%), the difficulty and challenges they posed on learners (5%), and the wide diversity (6%) and originality (9%) of tasks. In fact, learners commented on the lack of speaking practice in schools and the advantages of not following their book. In terms of learning, 15% of the responses emphasized they liked learning pronunciation in class and, interestingly, 6% of the responses evidenced that learners liked practicing pronunciation while communicating. Others referred to overall learning (4%), and in terms of vocabulary (4%), speaking and fluency (2%).

Selected comments about the likeability of the project

> I really enjoyed the project cause I think the topic was great and that made the project different and enjoyable. (S54)
>
> I was motivated to go to english class to do the tasks (S68)
>
> I liked that I have to make efforts to solve the task and think how to do it. (S39)
>
> many many different super original tasks. (S06)
>
> I loved doing the recording in class. I love this project. (S13)
>
> I loved it that we changed pairs and I could learn from other people. (S17)
>
> practice pronunciation when speaking, not repeating the teacher. (S28)
>
> I like practicing speaking because we don't do it. Always study from the book. (S53)
>
> My language was better after speaking with my classmates (S62)

Despite 61 % of the learners not highlighting any dislikes about the project, others manifested their aversion towards the testing phase of the experiment (9% of responses) and the repetitive nature of the content of the tasks (9% of responses). In addition, 9% of the responses referred to disinclination to the pre-task listening comprehension, also expressed by 15% of the learners in Question 10. For other learners, the difficulty of the tasks (5%), the fact of presenting their work in front of others (3%) and recording the tasks (2%) was not enjoyable.

Selected comments about the dislikeability of the project

> I disliked how repetitive it became to say the same words. (S05)
>
> I dislike the tests outside. (S29)
>
> doing the listening before the task. (S49)
>
> i [I] don't like to speak in front of the class. (S30)
>
> I finish too fast because the task is easy. (S19)
>
> I disliked some of the other classmates' attitude. (S08)

The last section of the post-intervention questionnaire was concerned with learners' perceptions of pronunciation performance and learning once the TBPT project was over. First, less than 5% of the learners indicated that the pronunciation of the four target vowels (/iː/, /ɪ/, /æ/, /ʌ/) was difficult or extremely difficult after the intervention. Overall, 14% of learners found /ɪ/ to be the most difficult and 93.6% /æ/ the easiest to pronounce. In addition, learners rated /iː/ as easier to pronounce than /ʌ/. In terms of overall pronunciation improvement, learners generally agreed (86%) that their pronunciation had improved after the TBPT intervention. As for specific vowel improvement, learners' self-perception of improvement for the English vowel /iː/ was considerable (i.e., around 80% claimed they felt they had improved from somewhat to a lot, whereas 20% believed they had improved little or nothing). Likewise, 80% of the learners reported a feeling of improvement in the pronunciation /ɪ/, whilst around 20% did not. However, only 13-14% of the learners manifested having improved a lot in the pronunciation of /iː/ and /ɪ/. Interestingly, 25% and 38% of the learners revealed they had improved A LOT in the pronunciation of /æ/ and /ʌ/, respectively. An average of 87% of the learners considered their improvement in the pronunciation of /æ/ to range from somewhat to a lot, whilst only 14% indicating little or no improvement, and an average of 95% of the learners believed their pronunciation of /ʌ/ had improved significantly (i.e., ranging from somewhat to a lot). Comments from the last open-ended question of the post-intervention questionnaire revealed that, apart from the general feeling of improvement in pronunciation (55%), in particular, the differentiation of the target vowel contrasts (10%), 17% of the responses referred to the fact learners had learned many new words and they had become more fluent after the intervention (8%). Interestingly, not only had their listening and communicative (4%) skills got better, but their problem-solving skills (2%) were also reported to have improved. In fact, in terms of affective factors, they expressed having gained self-confidence and reduced shyness (7%) and feeling more motivated (2%) when speaking in English.

Selected comments about learners' perception of learning

> I know how to pronounce better and I have learned many new words. (S72)
>
> I've learned lots of new words and I've learned to pronounce words that are written different but sound "the same". (S67)
>
> I don't spend so much time thinking about the word I want to say, I just jump. (S17)
>
> Understanding people in the listenings [listening comprehension task] with different pronunciation. (S32)
>
> I think that I have improved my pronunciation and my solving-things skills. (S64)
>
> I am more motivated to speak English now. (S15)
>
> My pronunciation and relationship with my classmates. (S68)
>
> The fact of communicating in English. I felt very confident and comfortable in the last classes (S27)

4.4.2. Simple vs. complex group perceptions

This section aims at comparing the results of (1) learners' evaluation of the project and (2) their perception of performance and improvement in pronunciation as a function of the type of intervention they received, namely, simple or complex decision-making tasks. To do so, only the questions' responses which were believed task complexity could have had an impact on (i.e., task mental effort, difficulty, enjoyability, problem resolution, pronunciation performance and overall improvement) were selected for group comparison.

Given that the ordinal data obtained from Likert-scale responses was not normally distributed, Mann-Whitney U tests were conducted to explore the effects of group (*simple/complex*) on learners' perception of mental effort during task performance, task difficulty and task enjoyability. Mann-Whitney U Test revealed significant differences in both mental effort (*U* = 873, *z* = 5.27, *p* < .001, *r* = .66) and difficulty (*U* = 900, *z* = 5.68, *p* < .001, *r* = .71) between SG and CG, suggesting that the CG perceived the tasks as significantly more difficult and requiring more mental effort than SG. Nevertheless, both SG and CG rated tasks as equally interesting and enjoyable (*U* = 601, *z* = 1.51, *p* = .129, *r* = .19), despite the CG's rating being slightly higher than the SG's (Table 4.23.).

Table 4.23. Descriptive statistics for learners' ratings of mental effort, task difficulty and enjoyment as a function of group (simple [SG]/complex [CG]).

	SG (*N*=31)			CG (*N*=32)		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*
Mental effort	2.94	3.00	1.18	5.06	5.00	1.24
Difficulty	3.03	3.00	1.20	5.16	5.00	0.95
Enjoyability	5.58	6.00	1.03	5.94	6.00	1.05

Note. 7-point Likert scales: Mental effort (*1=very low*; *7=very high*); difficulty (*1=extremely easy; 7=extremely difficult*): enjoyability (*1=strongly disagree; 7=strongly agree*). M= Mean, Mdn= Median, SD=Standard deviation

With respect to group differences regarding learners' task difficulty reasons (i.e., task conditions, pronunciation, task conditions + pronunciation) and task enjoyability (i.e., topic, images, target words, difficulty + multiple combinations), Chi-square test for independence were applied. On the one hand, this test indicated a significant association between *Group* (simple/complex) and *Task difficulty reasons*, $\chi2$ (2, *n* = 63) = 10.36, *p* = .006, *phi* = .40, because whereas a 66% of learners in the CG voted for task conditions (i.e., what learners had to consider to complete the task) for the reason of difficulty, only 26% did so from the SG. Instead twice the percentage of learners from the SG (vs. CG) considered vowel pronunciation to cause difficulty and only an average of 20% of both groups indicated that it was both the conditions and pronunciation which caused difficulty in the resolution of the task. On the other hand, Chi-square test for independence showed no significant association between *Group* (simple/complex) and *Task enjoyability reasons*, $\chi2$ (9, *n* = 63) = 9.40, *p* = .401, *phi* = .38. Interestingly, 9.5% of the CG learners (vs. 3% of the simple) indicated that difficulty made the tasks enjoyable, especially when combined with the topic (19% of CG vs. 7% of SG learners stated so). Instead, the images were the features that, according to the SG, made the tasks more enjoyable, as reported by 36% of learners (vs. only 13% of learners from the CG). The learners' evaluation group-comparison finishes with the comments regarding learners' strategies when problems in communication arose. Having assigned a numerical value to each one of the themes, responses concerning the use of interaction moves were greater in the CG (34%) than SG (21%); however, the SG reported using gestures (6.1%) and emphasis (6.1%) more than the CG (2.8% and 2.8%, respectively).

Concerning learners' perception of performance and improvement in pronunciation, Mann-Whitney U tests were used to investigate the effects of *Group* (simple/complex) on learners' difficulty in the pronunciation of the target vowels (/iː/, /ɪ/, /æ/, /ʌ/), overall sense of pronunciation improvement, improvement of each one of the target vowels, and their general perception of L2 improvement (via open-ended response group-comparison).

Table 4.24. Descriptive statistics for learners' ratings of pronunciation difficulty and improvement of each of the target vowels (/iː/, /ɪ/, /æ/, /ʌ/), and overall pronunciation improvement as a function of group (simple [SG]/complex [CG]).

	SG			CG		
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*
/iː/ pronunciation difficulty	2.71	3.00	1.10	2.16	2.00	1.19
/ɪ/ pronunciation difficulty	3.74	4.00	1.29	2.72	3.00	1.22
/æ/ pronunciation difficulty	2.13	2.00	1.06	1.53	1.00	0.80
/ʌ/ pronunciation difficulty	3.16	3.00	1.32	2.09	2.00	1.17
pronunciation improvement	5.35	5.00	1.08	5.81	6.00	1.15
/iː/ improvement	4.03	4.00	1.66	5.50	6.00	1.11
/ɪ/ improvement	4.48	5.00	1.63	5.38	5.50	1.26
/æ/ improvement	4.61	5.00	1.84	5.88	6.00	1.07
/ʌ/ improvement	5.87	6.00	1.18	5.91	6.00	1.20

Note. 7-point Likert scales: pronunciation difficulty (*1=extremely easy*; *7=extremely difficult*); pronunciation improvement (*1=strongly disagree*; *7=strongly agree);* improvement (*1=not at all*; *7=a lot*). M= Mean, Mdn= Median, SD=Standard deviation

Mann-Whitney U Tests revealed significant between-group differences in pronunciation difficulty for /ɪ/ (U = 287, z = -2.94, p = .003, r = .37), /æ/ (U = 334, z = -2.41, p = .016, r = .30), /ʌ/ (U = 280, z = -3.05, p = .002, r = .38) and almost /iː/ (U = 358, z = -1.95, p = .050, r = .24), because, the CG perceived them to be overall easier to pronounce than the SG, especially vowels /ɪ/ and /ʌ/ (Table 4.24.). In terms of overall pronunciation improvement, the CG's responses indicated slightly greater perceived improvement than the SG's but Mann-Whitney U Tests revealed no significant group-differences (U = 617, z = 1.73, p = .083, r = .21) (Table 4.24.). Considering each individual target vowel, the CG reported greater improvement in /iː/ (U = 757, z = 3.65, p < .001, r = .46), /ɪ/ (U = 644, z = 2.09, p = .036, r = .26), /æ/ (U = 698, z = 2.84, p = .004, r = .35) than the SG but not /ʌ/ (U = 513, z = .25, p = .806, r = .03). In fact, /ʌ/ was the vowel both groups reported they had improved the most, followed by /æ/, /ɪ/ and /iː/, as shown in Table 4.24. The group-comparison regarding learners' perception of improvement ends with an analysis of the open-ended comments concerning their overall sense of improvement after the TBPT project. Overall, very subtle differences arose between groups' responses. In general, the CG reported having improved in pronunciation (57%) vocabulary (18.5%) and all English skills (4.1%) in general more than the SG (51%, 14.5% and 2.4%, respectively). Responses referring to problem-solving strategy improvement only came from the CG. Nevertheless, learners in the SG (14.5%) reported having improved in motivation, self-confidence and social skills to a greater extent than the CG (6.1%).

Summary

The fourth chapter of the present dissertation was divided into four main sections corresponding to each one of the main research questions of this study. Specifically, the results section provided an answer to the effectiveness of TBPT for L2 vowel perception, lexical encoding and production; the effects of task complexity on L2 vowel learning and occurrence of pronunciation-based language-related episodes; the role of ID (i.e., L2 proficiency, WM, attention control) in L2 vowel learning; and finally, learners' perceptions about the TBPT intervention.

The results concerning RQ1 (*Is TBPT effective at improving L1-Catalan learners' acquisition of L2 English vowels?)* revealed that, overall, teaching L2 pronunciation through tasks was beneficial for the development of challenging L2 vowels. In terms of L2 vowel perception, whereas the control group did not become more accurate at discriminating vowels at post-test, the experimental group gained over 7% in accuracy and reduced their response times across testing times, especially in the discrimination of English /æ-ʌ/. In addition, learners' discrimination of L2 vowel contrasts was generalized to untaught tokens and retained 11 weeks after the treatment. Concerning the lexical encoding of L2 vowel contrasts, both FLeC and LD tasks revealed that the experimental group improved significantly in accuracy and speed from T1-T2, whilst the control group did not. Improvement by the experimental group was reflected with both L2 vowel contrasts, but learners' gains were higher in the identification of words/rejection of nonwords formed by the lax vowels /ɪ/ and /ʌ/. Crucially, T1-T3 comparisons showed learning was retained and, overall, they gained 18% in accuracy (51 ms in speed) and 29.5% in accuracy (58ms in speed) in the FLeC and LD tasks, respectively. When comparing both lexical decision tasks, despite obtaining greater accuracy scores in the FLeC task, gains were larger in the LD task.

Regarding L2 vowel production in words in isolation (DWR task), only the experimental group's confusing vowels got more separated and more target-like from T1 to T2, especially /æ/ - /ʌ/. In addition, learners' improvement in vowel *distinctiveness* and *nativelikeness* generalized to untaught words and was retained over 2 months after the treatment. Interestingly, less overlap between contrastive vowels was moderately correlated with more accurate productions, and gains in vowel *distinctiveness* were positively associated with gains in vowel *nativelikeness*. Last, experimental learners' (but not control) duration ratio between contrastive vowels only increased between the high vowels /iː/-/ɪ/. Regarding L2 vowel production in words in sentence contexts (DSR task), a very similar picture emerged. The experimental group (but not the control) learned to make less overlap between confusable vowels and approximated those of native speakers after the TBPT intervention. All vowels (except for /ɪ/) became more target-like from T1 to T2. Learning in vowel *distinctiveness* and *nativelikeness* was generalized to untaught words and retained 2 months after the treatment. Performance and gains in vowel *distinctiveness* were moderately associated with performance and gains in vowel *nativelikeness.* Nevertheless, in the case of L2 vowels in sentence contexts, neither experimental nor control groups increased overall vowel duration over time. Yet, the duration ratio for /iː/ - /ɪ/ increased significantly over time for the experimental group, as also found for words produced in isolation. When comparing both production tasks, despite obtaining greater overall scores in the DWR task, gains were greater in the DSR task for vowel *distinctiveness* and *nativelikeness*. Last but not least, comparisons across perception, lexical encoding and production task performance revealed that accuracy and speed were moderately associated between the FLeC and LD tasks, and vowel *distinctiveness* and vowel *nativelikeness* were correlated between the DWR and DSR task. Moderate correlations were also found between ABX discrimination and lexical decision accuracy, but gains in one dimension did not reflect gains in the other. Production gains were also disassociated with perception/lexical encoding gains.

The results concerning RQ2 (*Does task complexity play a role in L1-Catalan learners' acquisition of English vowels, as well as the occurrence of pronunciation-focused language-related episodes?)* showed that learners who engaged in complex decision-making tasks obtained overall greater gains in English vowel acquisition from T1 to T3. In terms of discrimination accuracy and speed, both SG and CG improved significantly from T1 to T2 but the complex group kept getting significantly better and faster at T3 with both vowel contrasts, especially /æ-ʌ/. In terms of the lexical encoding of vowels, results from the FLeC task showed that, whereas the SG improved in the lexical encoding of /æ-ʌ/ from T1 to T2, the CG became more accurate with the two target contrasts from T1 to T2 and kept increasing their accuracy at T3. Overall the CG yielded larger gains than the SG in accuracy and was overall faster than the SG. As for lexical encoding gains from the LD task, both groups also improved from T1 to T2 but, as in the FLeC task, the SG only improved the lexical encoding of /æ-ʌ/ whereas the CG was more accurate at both L2 vowel contrasts, especially from T2 to T3. Overall, the CG obtained greater gains than the SG in the lexical encoding of L2 vowels. Vowel production results somehow echoed perception results. When vowels were produced in words in isolation, the SG significantly produced more distance between /æ/ and /ʌ/ from T1 to T2 but stopped separating the vowels significantly from T2 to T3. Instead the CG produced significantly more distinct vowels from both contrasts from T1 to T2 (and T2 to T3) and gained more than the SG. In terms of vowel *nativelikeness*, only the CG produced all vowels in a more target-like manner immediately after the intervention and at T3, whereas the SG only got better with some of the vowels. When vowels were part of words in sentence contexts, similar results emerged. Both experimental groups improved the *distinctiveness* of vowels from T1 to T2, but only the CG's vowel productions got significantly more distinct and more nativelike from T1 to T2 as well as from T2 to T3. In the case of word and sentence production, both SG and CG increased the duration ratio of /iː-ɪ/ across time but differences across time did not reach significance between /æ/ and /ʌ/. Finally, CG learners engaged in P-LRE more often in each one of the three times (T1, T2, T3) than SG; however, when time-on-task was considered, no significant differences arose between groups over time.

The results from RQ3 (*Do individual differences in L2 experiential and cognitive factors explain L1-Catalan learners' performance and gains in English vowels?)* showed that, whereas L2 proficiency, PSTM, complex WM and ASA explained around 40% of variance in L2 vowel performance, only around 7% of variance explained vowel gains after the TBPT intervention. In contrast, experiential factors explained very little about learners' performance and gains in L2 vowel perception, lexical encoding and production. Firstly, SG, CG and CTG showed to be similar in terms of experiential and cognitive factors. Recent experience learning L2 English inside and outside the classroom was positively correlated with greater proficiency and PSTM, but only experience outside the classroom was associated with the complex WM measure and ASA. Secondly, when T1 to T3 performance was averaged for the experimental group, L2 proficiency, working memory (PSTM + complex WM) and ASA were positively associated with greater accuracy in vowel discrimination, lexical encoding and more target-like production of target vowels in the DWR and DSR tasks. In fact, complex WM explained greater performance in vowel discrimination and production, and ASA explained greater performance in lexical encoding, followed by vowel discrimination. Last but not least, only the complex WM predicted variance in L2 vowel gains (mainly vowel nativelikeness) after the TBPT intervention. Associations between experiential/cognitive

ID and L2 vowel gains, separately by group (SG vs. CG) were overall weak, although a greater number of correlations appeared for the group who had been taught with complex tasks.

Lastly, RQ4 (*Which were L1-Catalan learners' perceptions of the TBPT intervention?)* inquired about learners' beliefs about pronunciation, likeability of the project and perceived performance and learning after completing the intervention, and we compared the SG and CG perceptions. In general, results showed that learners' awareness about the importance of practicing pronunciation in class increased considerably after having completed the intervention. Concerning learners' opinions about the enjoyability of the project, they reported finding the pre-tasks useful to learn both the meaning and pronunciation. Almost 90% of the learners found the tasks interesting and enjoyable, suggesting that the images in the flashcards were appealing, useful for task success and overall learning. When struggling to solve the task, learners asked for teacher and peer support or engaged in interaction moves, used gestures or switched back to their L1. Furthermore, almost half of the learners stated that the post-tasks had been useful to revise the meaning of words, but especially to consolidate pronunciation. In general, learners' overall enjoyability of the project was ascribed to the great variety of tasks and the task procedure, and their impact on learning (pronunciation, vocabulary and speaking). In terms of learning perceptions, there was an overall agreement that learners' pronunciation had improved after the intervention, finding the pronunciation vowels rather easy. Specifically, they believed that the pronunciation of vowels which had improved the most were /æ/ and /ʌ/. Apart from the general feeling of improvement in pronunciation, learners reported learning new vocabulary, and becoming more fluent, as well as feeling more self-confidence and motivation. Last, group comparisons yielded significant differences in the perception of difficulty, suggesting that the CG perceived the tasks as significantly more difficult and requiring more mental effort than simple tasks, but both groups equally enjoyed the tasks. However, the reasons for task difficulty differed between groups. Finally, whereas the CG perceived all target vowels as easier to pronounce than the SG, their overall feeling of pronunciation learning was similar in both groups, especially regarding the English vowel /ʌ/. Learners' perceptions of overall pronunciation and vocabulary learning were greater in the CG than the SG but the SG indicated having improved in other aspects such as motivation and self-confidence.

05 DISCUSSION

CHAPTER 5. DISCUSSION

Following a task-based approach for pronunciation teaching in an EFL context, the current study sought to examine whether focusing on phonetic form while interacting would lead to L2 phonological development and whether task complexity and learner factors would mediate gains in L2 vowel learning. To this end, we conducted a three-testing-point longitudinal study where L1-Catalan/Spanish EFL learners were tested on L2 vowel perception, lexical encoding, vowel production and occurrence of P-LRE after a 10-hour task-based pronunciation-focused intervention. Learners did either simple (N = 31) or complex (N = 32) versions of twenty decision-making tasks cognitively manipulated through ±reasoning demands (Robinson, 2001b, 2007a, 2011b), which were preceded by a pre-task and followed by a post-task (Willis, 1996). Overall conclusions were derived. First, the current study showed that TBPT led to overall improvement in L2 vowel acquisition, besides generalization and retention effects. Another interesting finding was that performing tasks with increased complexity led to larger L2 vowel gains in the long run. Third, this study showed that experiential, and especially, cognitive differences explained a fair amount of variability in L2 vowel performance but very little in L2 vowel gains. Last but not least, TBPT raised learners' awareness about the importance of pronunciation instruction in class, generated positive feelings and led to perceived improvement not only in English pronunciation, but also in vocabulary and overall speaking skills.

The discussion in this chapter is addressed in four main sections corresponding to each research question, and following the same order presented in the results: the effectiveness of TBPT for L2 vowel learning (Section 5.1.); task complexity effects on L2 vowel learning and P-LRE (Section 5.2.); the role of ID (i.e., past and recent English experience, L2 proficiency, WM, ASA) on L2 vowel performance and learning (Section 5.3.), and learners' perceptions of the TBPT intervention (Section 5.4.).

5.1. Effectiveness of TBPT

In answer to RQ1 regarding the main effects of TBPT on L1-Catalan/Spanish acquisition of L2 vowels, the results of the current study show an overall beneficial effect of TBPT for the development of L2 segmental pronunciation (Gurzynski-Weiss et al., 2017a; Mora & Levkina, 2017), in accordance with previous TBPT studies (Mora-Plaza et al., 2018; Mora & Levkina, 2018; Solon et al., 2017). The present findings also appear to be consistent with previous research on the advantages of form-focused communicative instruction for promoting L2 pronunciation learning (Darcy et al., 2021; Park, 2000; Ruan & Saito, 2023; Saito, 2015). The overall effectiveness of this TBPT intervention will be interpreted in light of the potential challenges (Section 1.3.4.4.) and solutions (Section 1.3.4.5.) that were presented with respect to pronunciation instruction in FL classrooms (Darcy, 2018; Levis, 2022).

The first challenge in the teaching of L2 pronunciation concerns the lack of time EFL teachers dedicate to it. As mentioned in the introduction of this dissertation, in Catalan secondary school education, English oral abilities and pronunciation are not usually taught (Tragant, 2009; Tragant et al., 2010) because teachers prioritize the teaching of English reading comprehension (and to a lesser extent, listening comprehension) and writing skills —rather than oral abilities— mainly because these skills are assessed in university entrance exams. The current study has shown that, by integrating pronunciation into listening and speaking tasks that are meaningful for the learners (Darcy, 2018; Levis & Echelberger, 2022), not only does TBPT enhance L2 pronunciation learning, but it also helps to develop other linguistic competences, such as listening, speaking, and the acquisition of vocabulary, as indicated by learners' perceptions of improvement and their feedback on the project in the qualitative analyses (Section 4.4.1.). In addition, 91% of the taught words, which were reported as unknown by experimental learners through the word familiarity questionnaire prior to the intervention, were learned after the treatment, as shown in the target word assessment results (Section 3.3.3.5)[73].

The second pronunciation teaching challenge concerns teachers' lack of methods and resources to teach pronunciation in class (Kirkova-Naskova et al., 2021; Levis & Sonsaat, 2016). Prior studies have noted the importance of implementing a dual focus on form and meaning for L2 pronunciation learning so that learners are able to notice cross-language differences between L1 and L2 phonologies while engaging in meaningful activities (Darcy et al., 2019; Guion & Pederson, 2007; Park, 2000; Pederson & Guion-Anderson, 2007; Saito, 2015). Either through reactive (e.g., corrective feedback; Saito, 2015; Park, 2000; Ruan & Saito, 2023) or proactive (e.g., Abe 2011; Sicola, 2008; Darcy et al., 2019, 2021) form-focused techniques, pronunciation learning requires opportunities for contextualized repetitive practice and elaboration in meaningful contexts (Darcy et al., 2019; Kissling, 2013; Trofimovich & Gatbonton, 2006; Sardegna 2022). Findings from this TBPT intervention —designed around Willis' (1996) Task-Based Learning framework — seem to echo the benefits for L2 pronunciation obtained when the Communicative Framework (Celce-Murcia et al., 2010) and ACCESS (Gatbonton & Segalowitz, 2005) frameworks are combined (Darcy et al., 2019, 2021), or

[73] One limitation of this study is that we did not use the same receptive or productive vocabulary test before and after the intervention as there was no initial purpose of evaluating learners' vocabulary knowledge. The word familiarity test was used for the purpose of discarding unknown words in the lexical decision task, and the vocabulary assessment test was used for the purpose of discarding words that learners had not learned during the intervention.

a strategy-based approach is followed (Sardegna, 2022). There are several hypothetical explanations for the long-term pronunciation gains (at perceptual, lexical encoding and production levels) obtained through this study's form-focused TBPT intervention: [74]

- The pre-task phase offered learners opportunities for being exposed to the target words through a variety of voices, contexts and speech rates, illustrating one of the main characteristics of HVPT that has been found to lead to generalization of phonetic learning (Carlet & Cebrian, 2019; Sakai & Moorman, 2018; Thomson, 2018) and aid vocabulary learning (Sinkeviciute et al., 2019).
- The target items in the task were *task essential* (i.e., task completion depended on learners' ability to distinguish between L2 contrastive vowels). Therefore, these pronunciation-*focused* tasks created a need for noticing (Leow, 2015; Schmidt, 1990) and directing conscious attention to phonetic form while communicating (Mora-Plaza et al., 2018; Mora & Levkina, 2018; Sicola, 2008; Solon et al., 2017). Had the tasks been pronunciation-*unfocused,* such difficult L2 phonological forms might not have been noticed even by advanced learners (Ellis, 2017). In addition, dual form-meaning attention may lead to proceduralization of learners' declarative knowledge in long-term memory (DeKeyser, 1998).
- Following up on the previous point, the design of the tasks (*information-split, two-way, close*, *information-required*) created genuine communicative pressure to interact (Darcy et al., 2019). According to the interaction hypothesis (Long, 1985, 2015), while negotiating for meaning, learners might have noticed phonological inaccuracies in the input, and engaged in hypothesis-testing —while relying in their linguistic resources— and externalized their form-meaning hypotheses

[74] The concept of "long-term gains" in the present dissertation refers to retention effects observed 11 weeks after the intervention.

through P-LRE (Swain, 1985, 1995), provoking adjustments until the message was clearly understood (Gass, 1997; Long, 2015).

- The repetition of target words and the task procedure along the twenty tasks might have raised learners' awareness about the relevant phonetic information needed to distinguish minimal pairs (Kissling, 2013; see Jung et al., 2017 for lexical stress) and to automatize L2 phonological processing (Segalowitz & Hulstijn, 2005). In addition, the planning and report stages in the tasks might have provided an opportunity to recycle those phonological targets that were used spontaneously during the task.
- The post-task phase might have consolidated L2 phonological forms by orienting learners' attention to articulatory gestures through awareness-raising activities (Willis & Willis, 2007).

From a TBLT perspective, an analytic FonF may have pushed learners through the SLA learning processes of input noticing (i.e., detection of L2 confusable vowels), intake processing (i.e., hypothesis testing about the phonological cross-linguistic differences between L1-L2 vowels) and L2 knowledge processing (i.e., analysis of L2 phonological representations) (Gilabert et al., 2016; Leow, 2015), as well as through reflections of form and meaning through output production (Swain, 1985, 1995), as observed by the elevated number of P-LRE dyads engaged in. From a pedagogic perspective, the departure from common teacher-centred approaches and outdated pronunciation materials may have also increased their motivation, as some learners pointed out in the post-intervention questionnaire (Section 4.4.1.).

The third and last challenge has to do with the lack of teaching priorities in the pronunciation curriculum. Despite not conducting a diagnosis prior to the treatment through a dictation or a short pronunciation questionnaire (Couper, 2006; Isbell, 2020;

Levis & Echelberger, 2022), this investigation targeted L2 English vowels which have been documented to be perceived as very similar L1 sounds by L1-Catalan/Spanish learners (SLM: Flege, 1995; PAM-L2: Best & Tyler, 2007), and have a high functional load (Munro & Derwing, 2006; Suzukida & Saito, 2022) due to the multiple minimal-pairs which these L2 vowel contrasts can form. Evidence from the elevated number of P-LRE at T1 showed that the mispronunciation of these contrasts may have led to a decrease in comprehensibility (Sewell, 2017).

The following discussion sections are dedicated to interpret the quantitative results regarding the main effects of TBPT on learners' perceptual discrimination (RQ1.1), lexical encoding (RQ1.2) and production (RQ1.3) of L2 English vowels, as well as the relationship between perception, lexical encoding and production performance and gains (RQ1.4). In order to do so, the findings from Section 5.1. will be centred on the experimental group (simple + complex) because the control group did not show any T1-to-T2 accuracy improvement in either perception or lexical encoding, or produced vowels significantly more target-like or distinctively after the intervention.[75] The lack of significant T1-T2 differences for the control group, who prior to the intervention (T1) shared the same L2 English experience and L2 vowel perception and production abilities with the experimental group, rejects the hypothesis that L2 pronunciation gains from the experimental group are caused by test-retest effects or pronunciation learning outside the TBPT intervention.

[75] However, the control group obtained similar reaction time gains to the experimental group in some cases.

5.1.1. Perceptual discrimination

Within RQ1, this study sought to examine whether TBPT led to improvement in the discrimination of L2 contrastive vowels (/iː-ɪ/, /æ-ʌ/), as measured through the proportion of correct responses (accuracy) and speed of response (RT).

Results revealed that learners in the experimental group (but not those in the control group) became significantly more accurate and faster at discriminating L2 vowels from T1 to T2, gaining 6% in accuracy overall (5% in /iː-ɪ/; 7% in /æ-ʌ/). These findings echoed those observed in studies that conducted implicit high-variability perceptual training of segmentals (Saito et al., 2022c: 6.2% for /æ-ʌ/; Lim and Holt, 2011: 8.5% for /r-l/) or form-focused communicative interventions (Ruan & Saito, 2023: 4% for /iː-ɪ/). As expected, gains were lower than those obtained from explicit HVPT involving fewer hours (Carlet & Cebrian, 2019: 10-15% → 2.5h; Rato, 2004: 15-30% → 5h). In line with HVPT studies, improvement from pre- to post-test showed evidence of generalization as testing stimuli were produced by novel voices (Carlet, 2017; Carlet & Cebrian, 2022; Thomson, 2012, 2018) and generalization to untaught words (Carlet & Cebrian, 2019, 2022; Iverson et al., 2005; Mora & Mora-Plaza, 2019; Ortega et al., 2021) in terms of accuracy and response time. Obtaining generalization from segmental instruction is a reliable indication of robust pre-lexical learning (Iverson et al., 2005) and suggests that the TBPT intervention may have led to the formation of long-term memory representations (Flege, 1995). In addition, according to Leow's (2015) model of second language learning, FonF may have had a positive impact in the internalization, modification and consolidation of L2 knowledge, in this case, in L2 pronunciation.

In addition, analyses showed that learners obtained overall larger accuracy and RT gains in the discrimination of English /æ-ʌ/ than /iː-ɪ/ but T1-T2 improvement was significant for both contrasts. This finding is contrary to PAM-L2's (Best & Tyler, 2007) predictions that category-goodness assimilations (/æ-ʌ/) are expected to result in poorer discrimination than uncategorized-categorized assimilations (/iː-ɪ/)[76]; however, it is in agreement with Cebrian's (2019) findings which showed that the /iː-ɪ/ pair is not better discriminated than the single-category pair /æ-ʌ/ by L1-Catalan/Spanish learners of English. This discrepancy with the PAM-L2 model might be attributed to the different L2-L1 assimilation patterns reported for the /iː-ɪ/ pair (Cebrian, 2006, 2019; Cebrian et al., 2021; Rallo Fabra & Romero, 2012) or differences in cue weighting of different acoustic cues (Mora & Fullana, 2007).

Finally, the accuracy gains obtained at T2 were maintained after the intervention and learners became increasingly faster at distinguishing L2 contrastive vowels. In line with previous HVPT (Carlet & Cebrian, 2019; Rato, 2014; Rato & Rauber, 2015) and form-focused instruction (Abe, 2011) findings from EFL contexts, vowel discrimination learning was retained 11 weeks (T3) after the intervention ended. Overall, learners gained 7% of accuracy and 151ms of speed between T1 and T3, but gains were larger for the discrimination of /æ/ - /ʌ/ (8.5%) than /iː/ - /ɪ/ (5%). In total, at T3 learners reached about a 75% of accuracy in vowel discrimination. Overall perceptual gains echoed those found by Saito et al. (2022c) after a shorter implicit multimodal HVPT, or Mora and Levkina (2018) after a shorter nonword TBPT intervention with advanced EFL learners.

A possible explanation for the smaller longitudinal gains obtained with respect to other explicit and implicit HVPT studies may be due to the perceptual nature of the intervention (vs. production-driven in this dissertation) and the control conditions under

[76] However, if the English high-vowel contrast /iː-ɪ/ is considered to be a category-goodness type of assimilation (Carlet, 2017) in PAM-L2 terms —where English /iː/ is strongly assimilated to Spanish /i/ and English /ɪ/ is perceived as a poorer fit of Spanish /i/— vowels in both /æ-ʌ/ and /iː-ɪ/ contrasts would be equally difficult to distinguish for L1 Catalan/Spanish learners of English.

which HVPT is usually conducted (vs. classroom context in this dissertation). Still, these findings support the idea that engaging in meaningful interactions that induce learners to notice the differences between their interlocutor's productions and their own's may raise awareness of the vowel quality differences between L2 contrasts, as well as between L1 and L2 vowel quality, leading to improvement in learners' discrimination of L2 vowel contrasts (see Gómez Lacabex et al., 2008; Kartushina et al., 2015; Ruan & Saito, 2023 for transfer of production training gains to perception).

5.1.2. Lexical encoding

The second sub-question within RQ1 inquired about the extent to which learners would improve in the lexical encoding of L2 vowel contrasts (/iː-ɪ, æ-ʌ/) after the TBPT intervention, as assessed through FLeC and LD tasks. In view of the fact that experimental and control groups became significantly faster from T1 to T2 at identifying words (FLeC) or rejecting nonwords (LD), but no differences existed between them, and gains in RT only considered correct responses —which did not provide much information about how accurately lexical entries were accessed— the remainder of the discussion will be focused on accuracy of the lexical encoding of L2 contrasts, as it has been previously reported (Amengual 2016; Llompart, 2021a, 2021b; Llompart & Reinisch, 2019, 2021). In addition, findings from this section need to be interpreted with caution due to the fact that these lexical decision tasks did not include the same number of test and filler items (See Limitations Section 6.2). Still, learners obtained higher accuracy scores on filler than on test items in both FLeC (91% vs 64%) and LD tasks (91% vs 41%).

Concerning the findings from the **FLeC** task, learners' distinction of L2 phonological contrasts in a lexical context improved from T1 to T2 (from 60% to 72%

accuracy) independently of the target L2 vowel contrast (/iː-ɪ/, /æ-ʌ/). Interestingly, word identification asymmetries were found for words with different L2 vowels (Darcy et al., 2013; Llompart, 2021b; Llompart & Reinisch, 2019; Melnik & Peperkamp, 2021). Overall higher accuracy (and speed) in the identification of words containing /ɪ/ and /ʌ/ may be explained by the fact that /iː/ and /æ/ are longer, more salient, and more peripheral in the vowel space than /ɪ/ and /ʌ/. Therefore, /iː/ and /æ/ serve as better perceptual anchors (Polka & Bohn, 2003, 2011) and may aid detection of an unexpected */iː/ or */æ/ in words that are pronounced with /ɪ/ and /ʌ/, than the other way around. In other words, it may have been easier for learners to identify words like *fish* or *drum* when they were presented together with **f[iː]sh* or **dr[æ]m*, than words like *jeans* or *jam* when they were paired with **j[ɪ]ns* or **j[ʌ]m*. These asymmetries in terms of vowel acoustics seem to be in line with Llompart's (2021b) findings. According to the author, vowel acoustics may affect the judgement of the lexicality of words and similar sounding nonwords. For example, mispronunciations from substitutions involving the less peripheral vowel by the more peripheral vowel (e.g., *sun* → *s*[æ]n*) have been found easier to detect than the opposite scenario. In addition, this study revealed retention effects at T3 by getting significantly more sensitive at the lexical encoding of L2 contrasts in real words. On average, results indicated that learners gained around an 18% of accuracy in the identification of L2 vowels in a lexical context from the pre-test to delayed post-test.

With regard to the **LD** task, the initial low scores obtained at T1 echo those obtained by intermediate-level EFL learners in Llompart (2021a). Pre- / post-test findings also showed an overall improvement from T1 to T2 (38% → 58% accuracy) in nonword rejection for both L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/). Results showed that learners became significantly more accurate at lexically encoding both target vowel contrasts (especially /æ-ʌ/) and nonword rejection gains were retained at T3. The slight superiority in the lexical encoding of /æ-ʌ/ may have to do with the transparent orthography of this contrast (relative to /iː-ɪ/), which has been found to have an impact on L2 lexical representations (Hayes-Harb et al., 2018; Charoy & Samuel, 2020). Slightly larger nonword rejection accuracy gains were obtained for the more centralized vowels /ɪ/ and /ʌ/ compared to /iː/ and /æ/, as in the FLeC task. Given that /iː/ is more peripheral than /ɪ/ and /æ/ is more peripheral than /ʌ/, mispronunciations involving the substitution of the less peripheral vowel (e.g., **f[iː]sh*) may have been more salient (i.e., have extreme articulatory-acoustic properties) and easier to reject than the opposite, according to the NRV framework (Polka & Bohn, 2003, 2011). Recall that the NRV claims that vowels that fall closer to the periphery (e.g., /i/, /a/, /u/) in the vowel space act as natural referent vowels, from where the other categories are formed and their salience and stability is due to formant frequency focalization (Polka & Bohn, 2003, 2011). Apart from periphery and natural reference, English /iː/ and /æ/ are perceptually a better match to Catalan/Spanish /i/ and /a/ and act as the "dominant" category for /iː-ɪ/ and /æ-ʌ/ pairs, respectively. In an eye-tracking experiment, Cebrian and Mora (2017) showed that during L2-word competition, confusable English minimal pairs (/iː-ɪ/, /æ-ʌ/) would trigger an asymmetry in perception, which depended on the acoustic-phonetic proximity to the nearest L1 category (Cutler et al., 2006).

Echoing the findings from the FLeC task, learners became increasingly more accurate from T2 to T3 (by 11%), demonstrating that the TBPT intervention had been effective at improving the perception of L2 sounds when they were embedded in lexical contexts. On average, learners' improvement in the lexical encoding of sounds was of approximately 30% between T1 and T3. However, individual variability in nonword rejection rate in a by-item analysis revealed that accuracy in L2 phonolexical encoding varied a lot (see Darcy & Thomas, 2019; Llompart & Reinisch, 2021). Hypothetically, this variability might be explained as a function of lexical frequency (Llompart, 2021b), word age (Darcy & Holliday, 2019; Llompart & Reinisch, 2021), orthography (Hayes-Harb et al., 2018; Charoy & Samuel, 2020), neighbourhood density (Llompart, 2021b), among other factors.

To summarize, these findings indicate that TBPT was not only beneficial at improving learners' perception at the pre-lexical level but it also had a positive impact on learners' encoding of L2 sounds in the L2 lexicon. In contrast to earlier findings which reported relatively small improvements in the sensitivity of L2 English contrasts after 2.5-hour HVPT interventions (Mora & Mora-Plaza, 2019; Mora-Plaza et al., 2022a), this study provides additional evidence that orienting attention to phonetic form either via form-focused techniques such as HVPT (Adrian & Mora, 2022; Melnik-Leroy & Peperkamp, 2021) or phonological specificity (i.e., phonologically-focused) training (Llompart & Reinisch, 2021) enhances the lexical encoding of challenging L2 vowel contrasts embedded in minimal pairs. Our results closely match those obtained in Melnik-Leroy and Peperkamp (2021), where L2 learners gained almost a 30% from T1 to T3 (4 months after) in the LD task. Accuracy score differences at delayed post-test (Melnik-Leroy & Peperkamp: 90%; this study: 70%) may be explained by the fact that this study was conducted with lower-level L2 learners and the learning context was much less controlled than Melnik-Leroy and Peperkamp's (2021) lab-setting.

In conclusion, attention to the phonological units making up the words in meaningful contexts appears to be helpful for the encoding of difficult L2 contrasts into lexical representations (in line with Llompart & Reinisch, 2020, 2021)[77]. The lack of

[77] Yet, we should be cautious with this statement given that improvement in word identification (FLeC) or nonword rejection (LD) may not directly represent improvement in lexical encoding because it is hard to determine which score would actually represent that lexical encoding has taken place, especially, if accuracy in the FLeC and LD tasks is low to medium.

improvement of the control group, who did not participate in any focus-on-phonetic form intervention, provides further support to the claim that unless attention is directed to the critical L2 sounds during word learning (e.g., by forcing learners to choose between members of a minimal pair), it is very difficult to establish a phonological contrast between L2 words (Llompart & Reinisch, 2020). Yet, these findings also demonstrate that meaning does not need to be sacrificed for the sake of form. In addition, this study partly corroborates the findings by Darcy and Holliday (2019) in that the phonological representation of well-known words (or "old words" in Darcy & Holliday's terms) can be updated through form-focused interventions giving support to the fact that a delay in phonological learning is not necessarily an obstacle for lexical improvement (Llompart & Reinisch, 2021). Nevertheless, it is important to bear in mind the possible bias in learners' responses (Section 6.2.) due to minimal exposure to filler items in both lexical decision tasks, and that a production-based task (i.e., picture naming) would have been more adequate to test whether learners were able to update their phonological representations (Hayes-Harb & Masuda, 2008).

Finally, overall moderate correlations for T1/T2/T3 performance between the FLeC and LD tasks showed that the two tasks employed were useful at measuring learners' discrimination of L2 vowel contrasts in lexical contexts. Overall, learners performed more accurately in the three testing times in the FLeC (T1: 60%, T2: 72%, T3: 78%) than the LD (T1: 38%, T2: 58%, T3: 69%) task, as previously corroborated (Kojima, 2019) who showed that the FLeC task reduced the cognitive load of the learners by presenting the target English word together with its nonword counterpart. However, overall learning gains were larger in the LD than FLeC task. Despite the FLeC task showing better results for beginner learners, the LD task seems to be an appropriate task for intermediate- and more advanced-level learners. Yet, further research is needed to validate the LD task appropriateness for low and intermediate-level learners.

5.1.3. Production

The third sub-question within RQ1 assessed whether learners' production of L2 vowels (/iː, ɪ, æ, ʌ/) embedded in words elicited in isolation (via DWR) and in sentence contexts (via DSR) became more distinct and target-like after the TBPT intervention. Apart from vowel *distinctiveness* and *nativelikeness*, potential changes in vowel duration were also evaluated.

Firstly, as far as the findings from the **DWR** task are concerned, learners who took part in the TBPT intervention progressively produced more centralized and lower /ɪ/, and slightly more front (less centralized) /iː/, both in the direction of native-speaker productions, from T1 to T3. As for the low vowels, learners' productions of /æ/ became more front and slightly lower, and /ʌ/ was produced with more tongue retraction and lower F1. Despite both vowels approaching the native speaker productions in terms of F2 at T2/T3, their productions were still higher in F1 than native speakers' productions. Some authors have speculated that given the strong influence of the Catalan vowel system, it seems EFL learners tend to give more weight to the F2 dimension than to the F1 dimension (height), hence, tend to modify their productions along the frontness parameter in the distinction of English vowel contrasts /iː-ɪ/ and /æ-ʌ/ (Rallo Fabra & Romero, 2012).

The aforementioned changes in F1 and F2 illustrated greater distinctiveness of the contrasting vowels as well as reduced distance with respect to native vowel productions from T1 to T2. Although improvement in vowel *distinctiveness* happened in a similar way for both L2 vowel contrasts, gains were larger for /æ-ʌ/ than /iː-ɪ/, in line with the findings for perceptual discrimination. In addition, learners produced more target-like productions of /iː/, /æ/ and /ʌ/ but learners' production of /ɪ/ did not seem to change significantly after the TBPT intervention, which might be explained by the fact that their productions of English /ɪ/ were quite accurate even at T1, as illustrated in the vowel plot (Figure 4.8.). Interestingly, learners' separation of contrastive vowels and approximation to native speakers' models happened similarly when these appeared in taught and untaught words and their learning in vowel *distinctiveness* and *nativelikeness* was retained 11 weeks after the intervention. According to the L2 speech literature, this suggests that learners could have developed robust phonetic categories with different levels of L2 processing (Iverson et al., 2005). Interestingly, a negative association between gains in the measures of vowel distinctiveness and nativelikeness indicated that, generally speaking, learners who produced more distinct vowel productions of /iː/ - /ɪ/ and /æ/ - /ʌ/, also produced more target-like productions, as shown by a reduced distance with native speakers' productions. These findings seem to give empirical support to common assumptions from HVPT studies which suggested that helping learners produce more distinct realizations of confusable vowels (e.g., /æ/ - /ʌ/) would lead to more target-like productions (Melnik-Leroy et al., 2022; Mora, 2021).

Last but not least, the experimental group's vowel duration increased over time, which may be explained by the fact that learners may have been speaking more slowly and carefully at T2 and T3 than at T1, and thus, lengthening their vowel productions. Moreover, the vowel DR for the experimental group increased significantly for /iː-ɪ/ but not for /æ-ʌ/. In other words, the duration difference between the high tense and lax vowel increased after the intervention. L1-Catalan/Spanish learners' overreliance on temporal cues to distinguish these target sounds in perception and production has been well-documented (Cebrian, 2006; Cerviño & Mora, 2009; Mora & Fullana, 2007). The fact that vowel duration is not a distinctive feature in learners' L1 may indicate that they relied on duration, which is acoustically more salient than vowel frequency (Bohn, 1995); however, it is important to notice that an increase in vowel duration was not detrimental for the improvement in vowel quality after this TBPT intervention.

Overall, these findings demonstrate that form-focused task-based instruction was successful at drawing L2 learners' attention towards phonetic form, while maintaining their focus on meaning, and that less explicit pronunciation instruction -with a repetition component- also allows for noticing the relevant features of L2 speech (Kissling, 2013). In line with Saito's (2013, 2015) findings, embedding L2 pronunciation in meaningful contexts that offer opportunities for repetitive practice of the target language led to generalization of gains to novel lexical contexts as well as long-term effects (Abe, 2011). Also, in accordance with Sicola's (2008) findings, reflection on phonetic form during communicative tasks that make pronunciation targets task-essential and necessitate agreement on a single correct outcome (Ellis, 2003) lead to more distinct and accurate realizations of L2 confusable sounds. In a way, manipulating tasks (e.g., through task essentialness, task demands) may have enhanced the noticing of challenging L2 phonological forms in the input and lead to intake of those forms by making form-meaning connections (Leow, 2015). Then, during output production, learners may have engaged in hypothesis testing through negotiation of form and meaning (Swain, 1985), resulting in progressive automaticity of L2 phonological gestures.

Secondly, formant analyses from the **DSR** task revealed that, after the TBPT intervention, learners' production of /iː/ and /ɪ/ became more dissimilar because they produced higher and more fronted /iː/ and more centralized and lower /ɪ/, and they did so in a target-like direction. In the case of the low vowels, /æ/ was produced with slightly lower tongue position but especially more fronted, whereas the productions of /ʌ/ became higher in tongue position and especially more centralized. As illustrated in the B1-B2 vowel plot, learners' productions of /iː/ and /ɪ/ were significantly more distinct and more target-like at T3 than learners' productions of /æ/ and /ʌ/, whose distinctiveness was evident but they were still somehow far from the native speaker productions.

Findings from acoustic distance measures (vowel *distinctiveness* and vowel *nativelikeness*) mirrored the aforementioned changes in formant frequencies, hence, TBPT helped learners to separate vowels that were minimally distinct at T1 as well as produce more accurate (i.e. target-like) vowel productions in meaningful contexts. Contrary to the results obtained when words were elicited in isolation (DWR task), learners produced slightly larger gains in vowel distinctiveness with the high vowel contrast (/iː-ɪ/) than the low vowel contrast (/æ-ʌ/), but T1-T2 comparisons revealed that learners learned to distinguish L2 vowels in both contrasts in similar ways. This slight inconsistency may be attributable to the fact that, learners may have paid more attention to temporal than spectral cues of vowels in the distinction of /iː-ɪ/ when the test entailed repeating in isolation, whereas attention to vowel duration could have been minimized when the elicitation method was through meaningful sentences, and hence, triggered more vowel distinctiveness in terms of vowel quality. Another possible explanation may have to do with the differential sentence contexts in which the words containing /iː-ɪ/ or /æ-ʌ/ were embedded in. Nevertheless, in line with the findings from the DWR test, only learners' productions of /æ/, /ʌ/ and /iː/ significantly approximated the vowel quality of native speakers' vowel productions at T2, whilst /ɪ/ did not seem to improve in accuracy after the TBPT intervention. The large inter-individual variability in the production of /ɪ/ may account for the lack of significant improvement from pre- to post-test. Generalization of vowel *distinctiveness* and *nativelikeness* to untaught words happened for all target vowels, indicating phonological acquisition beyond the lexical contexts they were familiar with and had been recently exposed to (Carlet & Cebrian, 2019, 2022). Finally, retention of the TBPT effects on their ability to distinguish confusable L2 vowels and producing them in a nativelike manner indicates robustness of L2 vowel acquisition (Iverson et al., 2005). As a matter of fact, a similar association between vowel *distinctiveness* and *nativelikeness* arose, which indicated that larger non-native vowel distances between /iː/-/ɪ/ and /æ/-/ʌ/ corresponded to smaller distances between non-native and native vowel productions, and gains in one dimension transferred to the other (Mora, 2021). Hence, findings provide some support for the assumption that effectiveness of pronunciation instruction on the distinctiveness of confusable L2 vowels may be extrapolated to more accurate vowel productions, provided that learners receive sufficient amount of comprehensible input (Tyler, 2019) and engage in activities that enhance the noticing of confusable L2 vowels. However, methodological studies that compare both types of distances (*distinctiveness* and *nativelikeness*) should be conducted to provide further evidence of the reliability of these measures in the assessment of pronunciation instruction gains (Mora, 2021).

Last, contrary to the DWR findings, learners did not increase vowel duration a lot from T1 to T2 when vowels were produced in words in sentence contexts, which could have been partly influenced by learners' speech rate in the utterance of sentences (vs. words in isolation). However, in terms of DR, learners significantly increase the duration difference between English /iː/ - /ɪ/, but not between /æ/ - /ʌ/. These findings align with previous research that learners use duration cues to distinguish /iː/ from /ɪ/ (Cebrian, 2006; Cerviño & Mora, 2009; Mora & Fullana, 2007), yet the TBPT intervention was effective at raising learners' awareness of differences in terms of vowel quality.

In sum, these findings indicate that the beneficial effects of task-based pronunciation-focused instruction were evident not only when words were produced in isolation but also when they were embedded in sentences, as it has been previously found in an HVPT study that reported training gains for learners who had been trained on nonwords under adverse conditions (Mora et al., 2022). Although we cannot assume that L2 production gains from words repeated in sentences mirrored learners' production in real-world contexts, vowel productions in meaning-focused sentences are more likely to reflect vowel quality as produced in spontaneous speech than vowels elicited from words in isolation, which would foster more conscious attention to phonetic form (Bradlow et al., 1999). In addition, learning the target words repetitively under authentic communicative contexts could have generated form-to-meaning connections (Leow, 2015), allowed for more attention for monitoring (Bygate et al., 2001) and enhanced more accurate phonological representations of words in the L2 mental lexicon, resulting in more accurate L2 pronunciation (Darcy & Holliday, 2019) in meaningful contexts. Overall, these results align with Sicola (2008) and Solon et al.'s (2017) findings that designing tasks, whose accurate (or distinct) production of L2 difficult sounds is necessary for task completion, may have raised learners' phonological awareness about L2 targets, and thus, modified learners' production of L2 segments in a target-like direction.

Finally, results from the DWR and DSR task were compared. Findings from the current study echo those in Mora et al. (2022) who found superior accuracy in L2 vowel production in words in isolation than words embedded in sentences, probably due to the fact that producing words in isolation allowed them to pay more attention the acoustic properties of the target words for accurate production (Trofimovich, 2008). Nevertheless, gains in vowel *distinctiveness* and *nativelikeness* (i.e., learners' ability to produce more distinct and accurate L2 vowel productions) were larger when the target vowels were elicited via meaningful sentences. These observations are in agreement with Hirata's (2004) findings which showed that, after a 10-session computer assisted pronunciation training of Japanese pitch and durational contrasts in words, phrases and sentences, the trained group made more robust improvements in the word-in-sentence contexts than in word-in-isolation contexts. However, it differs from Thomson and Derwing (2016), whose training consisted of identifying 10 English monophthongs embedded in either lexical or non-lexical tokens. The fact that none of Thomson and Derwing's groups significantly improved in the production of those monophthongs in meaningful sentences (elicited via a picture-based sentence production task), might mean that gains in sentence contexts may especially occur when L2 sounds are learned contrastively in meaningful contexts during phonetic training or pronunciation teaching intervention.

In terms of pronunciation assessment for TBPT interventions, this study suggests that delayed sentence repetition tasks may be appropriate for assessing gains in L2 pronunciation in intermediate-to-advanced FL learners especially when pronunciation instruction is meaning-oriented. In this way, learners' L2 pronunciation may reflect the current state of phonolexical representations which may then be activated and retrieved for L2 use during spontaneous conversation. Further studies should also incorporate spontaneous tasks that prompt students to use the target vowels embedded in minimal pairs while freely communicating (e.g., recording a voice message to a friend indicating the way to a specific place; or reporting on the objects that contained a suitcase which they lost in the airport) to assess transferability of L2 pronunciation gains to less controlled speech (Darcy & Rocca, 2023).

5.1.4. Discrimination - lexical encoding - production comparisons

When comparing learners' overall performance in the perception and production tasks, first of all, the two lexical encoding (FLeC and LD) and production (DWR and DSR) tasks appeared to be moderately related. This might be due to the fact that each pair of tasks was assessing the same construct (i.e., lexical encoding and controlled production, respectively) and the items were the same, hence, the target sounds were always presented in the same phonetic contexts by the same speakers. However, the lack of a stronger correlation could indicate that the LD and DSR may have been more cognitively demanding than the FLeC (Kojima, 2019) and DWR (Mora et al., 2022) tasks, respectively, involving distinct processing levels.

Concerning the comparison between discrimination and lexical decision, first of all, learners were overall more accurate in the task that involved the phonetic categorization of L2 sounds than in the two tasks which involved the recognition of words containing difficult L2 contrasts, in line with previous research (Amengual, 2016; Darcy, et al., 2013; Díaz et al., 2012; Llompart, 2021a; Llompart & Reinisch, 2019). These results offer evidence that learners were better able to perceptually discriminate the contrasts at the phonetic level than at the lexical processing level, hence, it could be stated that accurate perception of L2 sound contrasts does not necessarily reflect successful phonolexical encoding (Llompart, 2021a).

Second, learners' performance in vowel discrimination appeared to be weakly (to moderately) related to accuracy and speed scores in the lexical decision tasks, suggesting that learners who performed better at a pre-lexical perceptual level also performed better at a lexical perceptual level. This finding provides additional evidence to research which found positive correlations between accuracy in perception and lexical representations (Darcy & Holliday, 2019; Llompart, 2021a for intermediate learners). Nevertheless, this study is unable to demonstrate symmetries between accuracy gains obtained for pre-lexical discrimination and lexical encoding of L2 sounds, contrary to Melnik-Leroy and Peperkamp (2021), who found that the more effective the phonetic training was on pre-lexical perception, the larger the transfer effects onto the perception of L2 sounds in lexical contexts. This discrepancy could be attributed to the fact that scores in the lexical decision task were much higher in Melnik-Leroy and Peperkamp than in this study, whose learners had lower proficiencies. In addition, the lack of an association between gains in the ABX and LD tasks may indicate that individual improvement in lexical encoding does not always entail immediate changes in L2 sound discrimination, and vice versa (Darcy et al., 2013). However, whether the establishment/updating of an L2 lexical contrast preceded (Cutler et al., 2006; Darcy et al., 2012b, 2013) or followed the perceptual discrimination of the contrast at a pre-lexical level cannot be demonstrated with our data.

Last, gains in the production of words in isolation and sentence contexts were unrelated to gains in perception (Rallo Fabra & Romero, 2012) or lexical encoding (Amengual, 2016; Simonchyk & Darcy, 2021). This lack of relationship echoes the findings from Simonchyk and Darcy (2021) and Amengual (2006) who argued that pronunciation improvement may be independent of improvement in lexical encoding because it can be affected by factors such as orthography, and perception and production involve different processes such as audition/comprehension vs. oral-motor/articulation abilities. Rallo Fabra and Romero (2012) found a weak relationship between perception and production vowel performance only when learners were grouped by L2 proficiency. While perception may enhance production abilities in the early stages of L2 learning (Flege, 1995), as learners gain experience with the target language, perceptual abilities may be more resistant to change. Alternatively, production may lag behind perception and catch up with it at a later stage (Casillas, 2020; Nagle, 2021).

Nevertheless, comparable methodologies and more longitudinal research are needed to shed further light on the relationship between perception and production across time, and how improvement in L2 sound production may transfer to improvement in L2 sound perception (and the other way around). Finally, the use of non-behavioural measures such as such as event-related potentials or functional magnetic resonance imaging may give us more objective insights into the present relationship (Thomson, 2022).

5.2. Task complexity effects

The second main research question of the present study inquired about the effects of task complexity on L1-Catalan/Spanish learners' acquisition of English vowels and the occurrence of P-LRE. Overall, the results of this study suggest that increasing task complexity seems to have a positive effect on learners' pronunciation (in line with Robinson's (2011b) Cognition Hypothesis), as it has been previously shown in other pronunciation-focused task-based interventions (Gordon, 2021; Mora-Plaza et al., 2018; Solon et al., 2017), but this beneficial effect can be most clearly observed in the long-term.

Long-term retention effects as a result of increasing the task demands along resource-directing dimensions (while keeping resource-dispersing dimensions low) accords well with Robinson's (2001b, 2007b, 2011b) Cognition Hypothesis' predictions that heightened attention to input promoted by task complexity leads to greater depth of processing and retention as more attentional resources are directed towards linguistic form. Given that task demands can determine what is noticed (Gilabert et al., 2016; Schmidt, 1990, 2001), increasing the functional and conceptual demands of tasks may have geared learners' attention towards the overlap or divergence between L1/L2 form-meanings (Robinson, 2003b), hence, how certain tokens are phonologically encoded in the L2 relative to the L1 (Talmy, 2000). In fact, the progressive increase of P-LRE across times may have reflected learners' conscious attention to cross-linguistic phonological differences in the input, leading to subsequent negotiation of form and meaning (Long, 1985, 1996, 2015) and by hypothesis, uptake in L2 production (Robinson, 2011b) and enhanced pronunciation accuracy. In fact, according to Leow's (2015) model of L2 learning, higher task demands may have pushed learners through the mental processes (input > input processing > intake > intake processing > L2 knowledge > L2 knowledge processing > output) in much faster and efficient ways than the lower task demands generated by the simple task.

Last, instructing learners to face cognitively complex tasks, where they had to adapt their language under the pressure to meet the demands of the task, may have helped them attain sufficient control of the target pronunciation features to be able to use them in meaningful contexts, such as the production of L2 sentences, and successful interaction in real-world tasks. In accordance with Celce-Murcia et al. (2010), increasing the cognitive load of tasks may have forced learners to use L2 pronunciation under real life situations that require more attentional demands.

In the following subsections, we interpret the findings on the differential effects of task complexity on perceptual discrimination (RQ2.1.), lexical encoding (RQ2.2.) and production (RQ2.3.) of L2 English vowels. Furthermore, the last subsection is dedicated to explain the occurrence of P-LRE (RQ2.4.) as a function of group (simple/complex) and time (T1/T2/T3).

5.2.1. Perceptual discrimination

This study examined whether implementing tasks differing in cognitive complexity would lead to differential effects in the discrimination of L2 contrastive vowels (/iː-ɪ/, /æ-ʌ/), in terms of accuracy and response speed. To our knowledge, this is the first study reporting a long-term advantage of the group who performed complex tasks (CG) compared to the group who performed simple tasks (SG).

On the one hand, our findings revealed that whereas both SG and CG improved significantly in the discrimination of all L2 vowels after the TBPT intervention, the CG kept improving 11 weeks after the intervention, reaching a 75% of accuracy (and 8% of overall gain), whereas the SG reached a plateau (72% of accuracy; 6.5% of overall gain). In fact, the CG's gains mainly reflected learners' increased accuracy in the perception of /æ-ʌ/ at a pre-lexical level. On the other hand, learners' discrimination of L2 vowel contrasts became significantly faster from T1 to T2 irrespective of group, but the CG kept reducing their speed of discrimination at T3, hence, obtaining overall larger gains in speed of discrimination, relative to the SG.

All in all, these findings suggest that increasing the demands of a task along resource-directing variables (+reasoning demands) while keeping resource-dispersing variables low (+planning time, +familiarization, +repetition) in pronunciation-focused tasks seems to results in more accurate discrimination of difficult L2 vowel contrasts in the long-term, presumably due to enhanced attention to phonetic form. According to Robinson's Cognition Hypothesis (2001b, 2007b, 2011b), increased attention (alertness, focus and orientation) should enhance in-depth processing of new forms to a higher degree, resulting in more accurate L2 performance, and L2 speech perception in this case.

The beneficial effects of task complexity on L2 perception have only been previously evidenced by Mora and Levkina (2018) who, despite not manipulating task complexity, showed that a nonword-based TBTP intervention sequenced from simple-to-complex tasks led to robust gains in discrimination accuracy and speed of the English /iː-ɪ/ contrast. Hypothetically, increasing task demands in focused tasks may have allocated sufficient attentional and memory resources to notice the perceptual differences between confusable L2 vowels present in the input. However, further research should test this hypothesis by assessing to what extent manipulating task design may impact L2 speech perception through various behavioural tasks (e.g., identification, discrimination, categorization) and stimulated recalls (Gass & Mackey, 2017).

5.2.2. Lexical encoding

Apart from learners' vowel discrimination at the pre-lexical level (Section 5.2.1.), we were interested in the effects of task complexity on the perceptual sensitivity of L2 vowel contrasts (/iː-ɪ/, /æ-ʌ/) in lexical contexts via FLeC and LD tasks. Given the fact that both SG and CG obtained similar RT at T2 and T3, gain differences coming from initial differences at T1, and the lack of research assessing the interpretability of lexical encoding gains through a RT measure (Amengual, 2016; Llompart & Reinisch, 2019, 2021), this section will be centred on accuracy of the lexical encoding of L2 vowel contrasts. Recall that these results should be interpreted with caution due to the unbalanced number of test and filler items.

Concerning the **FLeC** task, findings revealed an overall advantage in the identification of English words containing the challenging L2 vowel contrasts for the CG, relative to the SG (75% vs 65% of identification accuracy). Both groups improved from T1 to T2. However, the CG improved significantly in the lexical encoding of both L2 vowel contrasts (whereas the SG only /æ-ʌ/), and correct identification of lexical items containing all vowels (whereas the SG only /ʌ/), and kept being significantly more accurate at T3, reaching 89% of accuracy. Instead, the SG did not show any further improvement 11 weeks after the intervention. These results echo the pre-lexical discrimination results previously discussed on the advantage of engaging in complex-decision making tasks for the development of robust L2 contrasts. In addition, the overall gains were significantly larger for the CG (28.5%) than the SG (7.5%), and the CG made particular improvement in the identification of English words containing /ʌ/.

In terms of nonword rejection accuracy from the **LD** task, both SG and CG got significantly more accurate immediate after the TBPT intervention but only the CG improved the lexical encoding of both L2 vowel contrasts. Instead, the SG only improved the lexical encoding of /æ-ʌ/ from T1 to T2. In addition, overall accuracy was higher for the CG (56%) than SG (45%), mirroring the results obtained from the FLeC task. In terms of robust learning, the CG showed significant improvement 11 weeks after the treatment (76%), whereas the SG maintained the learning but did not show long-term beneficial effects. Interestingly, the SG's improvement in nonword rejection concerned the correct rejection of nonwords that were formed by substituting /ʌ/ by */æ/, whereas the complex group improvement in nonword rejection was present for all L2 vowels at T2 and T3. Overall, the CG obtained twice the gains of the SG, considering the three testing times, and /ʌ/ was the vowel that both groups obtained more gains in nonword rejection accuracy, showing asymmetries in vowel acoustics (Darcy et al., 2013; Llompart, 2021b).

This is the first TBPT study to show beneficial effects of task complexity on the lexical encoding of L2 vowels, hence, we call for further replications of this study, focusing on the effects of task design during communicative form-focused instruction on the lexical encoding of L2 contrasts. However, it is interesting to note that interactive tasks which are difficult to solve, hence, require much more negotiation of phonetic form and meaning between interlocutors to reach a solution (using task essential language), may be useful to increase learners' perception of the L2 contrasts at the lexical level, and potentially lead to more accurate word recognition. This long-term effect of task complexity may be explained by the fact that increased awareness on the relevant phonological differences while making form-meaning connections (Leow, 2015; Long, 1996) results in robust learning (Llompart & Reinisch, 2021). Other learner factors under our control (e.g., motivation) that may have been enhanced by the nature of complex decision-making tasks (Robinson, 2001a, 2011b), may have also played a role in the updating of L2 phonological representations.

5.2.3. Production

In this subsection, task complexity effects on learners' production of L2 target vowels (/iː, ɪ, æ, ʌ/) is examined in relation to vowel quality descriptions, distances of *distinctiveness* and vowel *nativelikeness*, and vowel duration. We first discuss the findings regarding words elicited in isolation through a DWR task and then words elicited in sentence contexts through a DSR task.

When words were produced *in isolation* **(DWR),** both SG and CG's confusable L2 vowels became more distinct and nativelike from T1 to T3. From T1 to T3, both groups produced more fronted and higher /iː/ as well as more centralized and lower /ɪ/, in the direction of target-like productions, and CG's productions were especially close to the native-speaker productions. In terms of the low vowels (/æ, ʌ/), the SG's production of /æ/ was more fronted and /ʌ/ was produced with more centralization from T1 to T2, but not much distinction was made at T3. Learners in SG approached native speakers' vowel productions but still they were relatively far from the native model. From T1 to T3, the CG not only produced more fronted /æ/ but also modified their F1 towards the lower end. At the same time, /ʌ/ became more centralized and also higher, simulating the acoustic distance that exists between English /æ/ and /ʌ/.

In terms of Mahalanobis distances, the *distinctiveness* of confusable vowels across times by groups seems to align with findings for perceptual discrimination. Whereas the SG only improved the distinctiveness of vowels significantly from T1 to T2, the distances were maintained at T3 but did not keep increasing. Instead, the complex group not only increased the distinction of the confusable vowels from T1 to T2, but also from T2 to T3, and for both contrasts, especially between /æ/ - /ʌ/. Overall, the CG obtained larger gains in vowel distinctiveness than the SG. As for the Mahalanobis distances between native and non-native vowels (i.e., vowel *nativelikeness*), the SG's vowel productions became more target-like across times but T1-T2 differences did not reach significance; however, the CG's vowels were produced more accurately from T1 to T2, and especially at T3, with all vowels, but the biggest movement towards native-speaker quality values happened with /æ/ and /ʌ/ productions. At T3, overall, CG productions were closer to the native speakers' productions (i.e., more accurately produced) than SG productions. Finally, both SG and CG used duration over time to distinguish the L2 vowel contrasts and they had similar DR when producing words in isolation after the TBPT treatment. However, as previously found in the L2 speech literature, changes in duration ratio were only present for the high vowel contrast /iː- ɪ/ (but not /æ-ʌ/) because Catalan learners of

English tend to use duration cues to distinguish this L2 English vowel contrast (Cebrian, 2006; Mora & Fullana, 2007).

When words were produced *in sentence contexts* **(DSR)**, a similar picture emerged. In general, both SG and CG's vowels became more distinct and accurate from T1 to T2, so the TBPT intervention appeared to enhance more accurate L2 pronunciation. Immediately after the intervention, all learners produced more fronted and higher /iː/ (i.e., more peripheral productions) and more centralized and lower /ɪ/, but the CG seemed to make less overlap between vowel categories than the SG at T3. Both groups produced more accurate vowels at T2, but the SG's production of /iː/ became slightly less target-like at T3. As for the low vowels (/æ, ʌ/), whereas the CG's productions of /æ/ became considerably lower and fronted, the SG's productions of /æ/ did not change much after the TBPT intervention (T1-T3). Contrastively, the SG and CG centralization of vowel /ʌ/ created less of an overlap with /æ/ and was produced more accurately from T1 to T3, especially in the case of the CG.

As far as Mahalanobis distances are concerned, the CG made less of an overlap between confusable vowels than the SG. Despite both groups improving in vowel *distinctiveness* from T1 to T2, only the CG kept producing more distinct vowels at T3 for both L2 contrasts. In fact, the SG lost part of their distinctiveness in vowel production from T2 to T3, but still obtained gains with respect to T1. Overall, the CG obtained larger gains in the production of distinct vowels than the SG. In terms of vowel *nativelikeness*, the SG and CG produced more accurate vowels from T1 to T2, but this positive difference only reached significance for the CG, who kept reducing the distance with native-speaker productions significantly at T3 (vs. the SG which seemed to reach a plateau). Overall, CG's vowels became significantly more accurate from T1 to T3 and overall gains for the CG were larger than SG. Finally, none of the groups seemed to rely a lot on vowel duration across times. However, the DR increased significantly for both the SG and CG across the three times, indicating that, apart from vowel quality, learners used vowel duration to distinguish confusable tense and lax vowels, especially /iː-ɪ/. These findings echo what was previously found when learners produced words in isolation (DWR task), and are in line with previous research demonstrating L2 learners' overreliance of temporal cues in the production of English /iː-ɪ/ (Cebrian, 2006; Cerviño & Mora, 2009; Mora & Fullana, 2007).

Overall, these findings align with previous literature documenting the positive effects of task complexity on L2 accuracy. Psycholinguistic theories of task complexity and interactionist research claim that greater task complexity encourages more attention to form (Robinson, 2011b) as learners engage in form-meaning negotiations, reformulate their hypothesis and produce changes in the output (Long, 1996; Swain, 1995, 2005). While most research examining the potential links between task complexity and L2 accuracy has focused on grammar (Kim & Tracey-Ventura, 2011; Révész, 2009), lexis (Kim et al., 2018; Nuevo et al., 2011) or pragmatics (Gilabert & Barón, 2013; Márquez & Barón, 2021), this experiment has extended this line of work by showing that increasing the task demands raised learners' awareness of phonetic form and resulted in overall larger gains in L2 vowel production. This study complements Solon et al.'s (2017) cross-sectional investigation, which revealed that L2 Spanish vowels (especially /e/ productions) were produced more target-like by learners engaging in more complex tasks. While the Cognition Hypothesis' predictions only applied for a subset of L2 Spanish vowels in Solon et al. (2017), our longitudinal study showed that a 7-week complex-task intervention generated more distinct and accurate vowels immediately after and long after the intervention.

Interestingly, this long-term benefit of task complexity echoes findings from Kim and Taguchi (2015) who found increases in reasoning demands to relate to L2 English learners' development of request expressions on delayed post-tests. To sum up, this study provides additional evidence that the predictions of the Cognition Hypothesis (Robinson, 2001b, 2007b, 2011b) on L2 accuracy can extend to pronunciation (Gordon, 2021; Solon et al., 2017) as long as communicative tasks are designed in a way that L2 pronunciation is enhanced through proactive focus-on-form techniques such as insertion of task essential minimal pairs (Sicola, 2008), visual or auditory enhancement of L2 sounds (Ruan & Saito, 2023), auditory priming, etc. or reactive focus-on-form techniques such as the provision of feedback (Saito, 2015). Given that task demands are tightly linked to learners' allocation of attention to monitoring L2 speech (Kormos, 1999), increased complexity could be detrimental to L2 pronunciation (Crowther et al., 2018; Kuiken & Vedder, 2011) unless learners' attention is directed to phonetic form, as found by studies reporting general unawareness of phonological errors (i.e., operationalized as instances of self-repairs in SLA studies), with respect to grammatical or lexical errors, in cognitively demanding tasks (e.g., Kormos, 2000).

5.2.4. P-LRE

The last subsection focuses on the effects of task complexity on the frequency and duration of P-LRE which took place during the oral interactive task prior to and after the TBPT intervention. LRE were expected to manifest learners' attention to phonetic form during meaningful task-based interaction in dyads. In fact, this was the case. Globally, the large amount of P-LRE that experimental learners engaged in suggests that careful task design can clearly draw L2 learners' attention to form, raise their awareness, and become a spontaneous but standardized practice among learners.

Initially, it was hypothesized that increasing task complexity would generate more conscious reflections of form reflected through a higher frequency of LRE, as a wealth of research had found in L2 studies focusing on grammar (Baralt, 2014; Kim, 2009, 2012; Révész, 2011), lexis (Gilabert et al., 2009) and pragmatics (Kim & Taguchi, 2015). However, the findings of the current study do not support previous research. In general, dyads engaged in more P-LRE when performed complex than simple decision-making tasks, and the duration of those P-LRE was greater for the CG. Nevertheless, when the ratio of number of P-LRE per minute was calculated, significant group differences disappeared. These results seem to be consistent with Solon et al's (2017) study who did not find statistical differences in the occurrence of interactional features in simple versus complex tasks. The lack of differences in our study may be explained in part by the individual characteristics of interactional partners, which have been shown to directly impact interactive moves (Kim & McDonough, 2008); in particular, the lack of control of differences in L2 proficiency by dyad in both intervention and testing phases. Studies have shown that L1-unmatched or L2 proficiency-unmatched pairs tend to generate more LRE than L1-matched or pairs who share the same level of proficiency because familiarity with an L2 accent facilitates comprehension (Bueno-Alastuey, 2013; Loewen & Isbell, 2017). Also, the fact that the interactions in the intervention were conducted by student-selected pairs may have generated more off-talk behaviour (Mozaffari, 2017), due to their pre-existing friendship, mitigating the effects of task complexity on negotiation of form and meaning. Taking these factors into consideration in future studies may maximize the effect of task complexity on the occurrence of P-LRE.

Interestingly, the frequency of P-LRE increased across the three times and, albeit non-significantly, the CG engaged in more P-LRE than the SG in each one of the times. The observed increase in P-LRE across times may be attributed to an increased metalinguistic awareness after the TBPT intervention. In other words, having completed the 20 pronunciation-focused tasks, learners may have become more aware of the physical modification of articulators to produce the target sounds and may have acquired tools to verbally reflect on phonetic form while communicating (see DeKeyser's 1998 idea of automatization of declarative knowledge). The study of self-repairs, which is currently outside the scope of this study, could provide further evidence that, over time, learners' attentional resources were allocated to form during monitoring (Kormos, 2000).

Last, despite the duration of P-LRE being higher at T1 for the complex group, there was a lot of variability in the duration of P-LRE prior to the treatment probably because some dyads did not notice any phonological differences in the target minimal pairs (Kormos, 2000), hence, engaged in relatively short stretches of P-LRE and did not complete the interactive task successfully, whereas others realised about phonological differences in the interlocutors' speech and externalized their hypotheses (Swain, 1985) and engaged in longer metalinguistic discussions. Nevertheless, we must be extremely cautious with this assertion, as it is based on the researcher' qualitative perceptions. Further analyses on the vowel accuracy of words produced during P-LRE at T1 are needed to provide solid evidence for this interpretation. Overall, SG and CG neither differed in the amount of time they spent engaging in P-LRE, nor the duration ratio changed across times.

5.3. The role of ID

The research agenda of TBPT after the special SSLA issue on tasks and L2 pronunciation (Gurzynski-Weiss et al., 2017a; Mora & Levkina, 2017) outlined the need for consideration of individual factors that could mediate the effectiveness of task manipulations in driving pronunciation learning in FL contexts. This dissertation examined several experiential (i.e., past and recent English experience inside and outside the classroom, L2 proficiency) and cognitive (WM and ASA) factors that could potentially impact L2 speech performance and gains after the TBPT intervention, hence, could help explain inter-individual variation in L2 perception and production. Overall, the most relevant findings suggested that, whereas past and recent L2 experience did not explain much variance in L2 pronunciation performance/gains, L2 proficiency, WM and attention explained almost 40% of variance in performance and only less than 10% in gains. The discussion of the findings regarding RQ3 will be organized in three main blocks corresponding to the comparison of experiential and cognitive factors across groups and their relationship (RQ3.1.), the associations between experiential/ cognitive ID and L2 speech performance as well as the specific contribution of each individual factor to L2 speech performance (RQ3.2.) and the aforementioned associations and contributions with respect to L2 speech gains.

Firstly, the results revealed no significant differences in terms of experiential and cognitive factors across three groups (i.e., SG, CG, CTG), although ASA was slightly higher for the CG than the SG and CTG. These findings show that the three groups were similar in terms of their English experience in the past and present and, in general, in terms of their cognitive abilities, which makes the interpretation of the pre- to post-test results due to the presence or absence of the intervention clearer. In addition, the relatively weak to moderate correlations between experiential and cognitive factors justifies using these variables as separate predictors in the assessment of L2 speech performance and gains. When inspecting the relationships between experiential and cognitive factors of the experimental group, as expected, learners' recent English experience inside (e.g., hours per week of English classes) and outside (e.g., hours with native/non-native speakers) the classroom was weakly but significantly associated with learners' proficiency in English as well as, to a lesser extent, their PSTM. The fact that past experience did not seem to be associated with L2 proficiency or any other experiential or cognitive factor may be explained by the lack of variability in the years of instruction of these students, which had followed a similar education path. In line with previous studies (Kim et al., 2016), L2 proficiency was moderately associated with PSTM and ASA but was unrelated to complex WM suggesting that the EIT we used to measure proficiency, which is conducive to the retrieval of implicit knowledge with minimal attention capacity (Granena, 2016), was appropriate to test proficiency-dependent abilities without the overreliance of complex WM. Lastly, ASA, which requires the allocation of conscious attention to L2 speech was found to be moderately related to complex WM.

In terms of L2 speech performance, L2 learning experience did not seem to be related to L2 vowel accuracy at three different points in time. Whereas these factors may explain inter-individual variation in L2 perception and production when learners are in immersive settings, where the quality and quantity of input determines L2 acquisition (Flege, 1995), in FL settings where quantity of L2 input is rather scarce (Muñoz, 2014) and L2 input tends to be accented (Tyler, 2019), differences in self-reported L2 experience may not account for much variation in L2 speech performance. Still, learners' experience outside the classroom was moderately associated with better scores in the identification of words containing difficult L2 vowels (FLeC task). However, learners who had a higher L2 proficiency and PSTM obtained higher scores in perceptual discrimination, word identification (FLeC task) and higher accuracy of L2 vowel production in words in isolation (DWR task) and in sentences (DSR task). The mediating effect of L2 proficiency on learners' phonological accuracy was also shown by Saito et al. (2021a) after Japanese EFL learners completed an interaction-based intervention. Interestingly, our measure of complex WM was significantly and moderately related to learners' accuracy in the perception of L2 sounds at the pre-lexical and lexical levels (ABX, FLeC and LD) as well as learners' accuracy in the production of L2 sounds elicited through words and sentences (DWR and DSR). The present findings are consistent with Darcy et al.'s (2015) study who found that complex WM, assessed through a backward digit span task, was moderately related to individual L2-English phonological scores (a composite of segmental categorization, lexical stress and phonotactics). Moreover, it also supports previous research that WM contributes significantly to L2 speech imitation capacity. For example, Christiner and Reiterer (2016), who conducted a study with musicians, found that WM (also measured through a forward and backward-digit span test) was associated with L2 English accent imitation and English text reading. Although our L2 production tasks did not involve direct imitation from the auditory stimuli, these results add further evidence that WM plays a role in the production of difficult L2 sounds. Another important finding was that ASA was significantly and moderately associated with learners' discrimination of L2 confusable vowels (ABX task) and learners' perception of L2 vowels at the lexical level (FLEC and LD task). These results match those observed in earlier studies (Darcy et al., 2014; Mora & Mora-Plaza, 2019; Mora-Plaza et al., 2022a) that L2 learners that have high attention control may tune into the phonologically relevant acoustic information in the L2, and this may have an impact on learners' ability to discern between confusable vowels in perception.

When assessing the independent contributions of ID to L2 vowel performance, L2 proficiency, WM and attention explained almost 40% of the variance in vowel discrimination, and complex WM was the factor, together with ASA, that predicted differences in vowel perception. This suggests that learners with higher WM skills may be better able to rely on the spectral differences underlying the English vowel contrast, hence, PSTM may enhance more target-like L2 cue weighting (Safronova, 2016). Similarly, the aforementioned factors explained 40% of the variance in the identification of English words (FLeC task) and 25% in the rejection of nonwords (LD task) containing the target vowels. The factor which explained greater variance was ASA, hence, having higher selective attention seems to enhance the lexical encoding of L2 sounds. Last, complex WM seemed to be the greatest contributor to explain inter-individual variation in the accuracy of vowel production in words elicited in isolation and in sentences.

Lastly, the relationship between experiential/cognitive ID and L2 perception and production gains was overall weak. Contrary to our predictions, ASA was unrelated to L2 vowel discrimination gains immediately after the treatment. Although learners' ability to focus attention to specific speech dimensions was expected to be related to learners L2 phonological acquisition (Mora & Mora-Plaza, 2019), these findings further corroborate those obtained by Mora-Plaza et al. (2022a) that ASA may explain performance in L2 vowel discrimination but may not be necessarily related to perceptual discrimination gains. No relationship arose between L2 vowel gains and L2 proficiency which could be explained in part because learners' L2 pronunciation gains were independent of their L2 proficiency. One could speculate that focused tasks gave all students opportunities to practice pronunciation, even those with a limited linguistic repertoire. Only complex WM explained considerable variation in gains obtained for L2 vowel discrimination, speed in lexical encoding and accuracy in the production of words in isolation and in context. The advantages of higher WM in learners' ability to attend to relevant spectral cues have been previously attested. For example, in the domain of HVPT, Aliaga-Garcia et al. (2011) found L1-Spanish/Catalan trainees with higher PSTM capacity to obtain higher accuracy scores and larger perceptual accuracy gains than the lower PSTM capacity trainees in the discrimination of L2 English vowels, suggesting that PSTM may contribute significantly to the development of L2 speech perception. In sum, only 7% of the variance in L2 vowel gains was explained by experiential and cognitive factors. In a way, this result may suggest that instruction is facilitative of L2 pronunciation development regardless of learners' ID in L2 experience and L2 proficiency, WM and selective attention (Suzukida & Saito, 2023). These results contradict Lee's (2021) findings that attention control and WM predicted L2 grammatical gains supported by proactive form-focused instruction. Attention control has been identified as a major predictor for the acquisition of L2 morphosyntax (Trofimovich et al., 2007). However, given that L2 pronunciation learning is an acoustic-motor phenomenon (different from L2 grammar instruction which requires the processing of abstract rules), it may be the case that articulatory and auditory-based instruction may be beneficial for all learners, even if they vary in experiential/cognitive factors.

Finally, findings regarding the relationship between ID and L2 vowel gains (T1-T2) separately by group (SG, CG) revealed that, in general, gains from the CG correlated with more individual factors (and associations were overall stronger) than the SG's gains. In accordance with Robinson (2007b), cognitive differences may differentiate performance when task complexity is enhanced. Given the fact that the tests used to assess L2 vowel learning did not differ in complexity, the role of WM mediated by complexity may be difficult to assess in this study, with respect to other studies who assessed the effects of task complexity and WM in their predictions of oral performance while producing a spontaneous oral task (e.g., Kim et al., 2015). Still, the fact that correlations between cognitive factors and L2 pronunciation gains were stronger for learners who had completed complex tasks may suggest that gains enhanced by performing tasks with high demands may be partly explained by certain advantages at the cognitive level.

In the case of the SG, learners' past English experience was related to larger speed gains in vowel accuracy and larger gains in vowel *nativelikeness* in the production of words in isolation. In addition, learners with higher complex WM were better at vowel discrimination and faster at identifying L2 sounds in English words (FLeC). However, ASA was unrelated to any gains. As for the CG, learners' recent experience with English inside the classroom was related to gains in vowel discrimination and speed of lexical encoding, and learners who had had more English experience outside the classroom were overall faster in vowel discrimination. Interestingly, learners with higher proficiency and stronger PSTM were more accurate in the production of L2 vowels embedded in words in isolation, and the complex WM measure was associated with greater vowel *distinctiveness* and *nativelikeness* in the production of words and sentences. Finally, learners with greater ASA obtained larger gains in the distinctiveness of vowels elicited in words.

All in all, ID seem to be more strongly associated with L2 vowel perception and production performance rather than gains (as in Mora & Mora-Plaza, 2019). The relatively weak correlations between ID and L2 vowel gains, which were not consistent across groups, should be treated with caution. Further research should take into consideration other experiential (e.g., parents' educational background), cognitive / auditory (e.g., aptitude, general auditory processing) and affective (e.g., motivation, anxiety) factors that may help understand inter-individual differences in L2 vowel performance and gains. In addition, instead of assessing the role of selective attention to domain-specific information, as in the present study, it would be interesting to investigate how domain-general acoustic parameters (e.g., pitch, formants, duration), which explain individual variation, may relate to L2 phonology.

5.4. Learners' perception of the intervention

The last section of the discussion aims at examining learners' perceptions (beliefs, likeability, learning) of the TBPT intervention in general (RQ4.1.) and comparing learners' evaluation of the project and self-perception of improvement between the SG and CG (RQ4.2.). Apart from commenting on the results of the closed questions, some relevant comments were selected in order to illustrate the different views on the TBPT project.

First of all, learners' **beliefs** about the importance of teaching pronunciation changed substantially after the TBPT intervention, where over 80% of the learners considered pronunciation to be either important or extremely important, and the great majority claimed that it should be taught around 3-4 hours per month. Whilst the importance of pronunciation instruction was initially underrated, it can be speculated that the communicative nature of the intervention may have raised learners' self-awareness about the importance of receiving pronunciation instruction to improve their pronunciation as well as listening and speaking skills (Henrichsen & Stephens, 2015; Nguyen et al., 2021).

Concerning the **likeability** of the TBPT project, overall, pre-tasks were regarded as very useful to learn the meaning of the target words, and especially, their pronunciation. Thus, learners' opinions reflect the beneficial effect of pre-tasks to direct attention to both form and meaning. Although more than half of the learners found the pre-tasks interesting and enjoyable, one third of the learners were unsure about them or did not enjoy the pre-tasks, perhaps due to the low demands needed for pre-task completion or repetitiveness. The fact that over 70% of learners reported the listening comprehension activity to be easy may be due to the repetitive exposure to the target words (5 times over 20 tasks) and the increased familiarization with the stimuli voices, which may have reduced the cognitive load of the pre-task.

In terms of the task phase, almost 90% of the learners enjoyed doing the tasks. The source of enjoyment mainly came from the visual materials provided to solve the tasks as well as the topic of travelling. The likeability of the images in the flashcards was mainly linked to the positive reaction they generated (*They where funny and motivate you to make the task* [S68]), as well as their effectiveness in helping learners recall the meaning and pronunciation of the target words (*I remember the sounds with the images* [S32]) and achieving task success. Only a small percentage provided negative feedback considering them to be childish (*they were ok but some for children* [S33]). Interestingly though, the tasks learners enjoyed the most were those whose flashcards contained images that had to be manipulated or organized on the task board to create an album, a website, or dress different characters. In the case of a communication breakdown, the majority of learners asked the teacher or peers for support. An important 28% of the responses indicated learners' engagement in interaction moves by asking for clarification, paraphrasing or repeating the same information. Other strategies involved using gestures, switching to their L1 or placing special emphasis on pronunciation (*I exagerated the pronunciation of the word so she could hear the difference* [S12]). The use of emphasis to aid comprehensibility was also shown in Sicola (2008) whose EFL learners used pausing or rising intonation, to make their pronunciation of the target L2 consonants more salient to their partner during an interactive task where L2 pronunciation was task-essential.

Post-tasks were not perceived to be as useful to revise the meaning of the words as the pre-task probably due to learners' familiarization with the target vocabulary after performing the task. Nevertheless, the vast majority of learners believed post-tasks helped consolidate their pronunciation, which was one of the reasons for implementing Willis' (1996) framework, and half of them believed they were interesting and enjoyable. Finally, the overall impression of the TBPT project was very positive, with a great percentage of responses indicating the enjoyability of the project and the tasks and their impact on learning (*group tasks were really entertaining and I learned many many words and pronunciation* [S36]). Specifically, learners enjoyed the procedure (*What I enjoyed the most were the activities that we have to record* [S16]), working in small groups (*I loved it that we changed pairs and I could learn from other people* [S17]) and the different stages of the framework (report, planning and post-tasks). Responses about the likeability of the project also highlighted the wide diversity and originality of tasks (*they were fun and if I am honest they were a good excuse to stop doing activities from the coursebook* [S38]). The fact that pronunciation is hardly ever integrated in speaking tasks and pronunciation content is boring, often non-systematic and presented as add-ons in the margins of lessons (Derwing et al., 2012), may have enlightened learners about the potential benefits of learning pronunciation through communicative tasks. Four responses related the likeability of the tasks to how challenging tasks were to solve. Previous findings have also indicated that performing complex tasks may enhance learners' motivation (Robinson, 2001a, 2007b). In terms of learning, learners showed self-awareness of pronunciation learning while communicating (*I loved learning English pronunciation through different games* [S12]; *I loved practicing each one of the words'* *pronunciation when i try to solve a task* [S33]). Apart from learning pronunciation and vocabulary, learners commented on the benefits for speaking, fluency development and overall language (*my language was better after speaking with my classmates* [S62]). Given that focused tasks opened up opportunities for the use of unpredicted language (not only the preselected and predetermined language forms chosen by the teacher), learners' perceptions of improvement showed learning happened beyond L2 pronunciation. Although we did not empirically test how TBPT affected L2 speech fluency or lexico-grammatical complexity and accuracy, when learners are given the chance to attend the language in a meaningful way, this has the potential to drive acquisition (Erlam & Tolosa, 2022). Last, the most recurrent negative comments about the TBPT intervention were related to the testing phase of the experiment —as learners were not used to performing lengthy perception and production tests—, the repetitive nature of the tasks which recycled the same words several times, and the pre-task listening comprehension, which learners may have found too easy. Some others referred to their fear of public speaking during the report stage, task difficulty (i.e., too easy or too difficult) and inappropriate classmates' attitudes.

In terms of perception of pronunciation **performance** after the TBPT intervention, almost all learners felt more confident in the pronunciation of the target vowels, with only a very small percentage admitting that it was still (extremely) difficult. Over 90% of learners considered English /æ/ the easiest to pronounce and English /ɪ/ to be the most difficult, whilst English /iː/ was easier to pronounce than /ʌ/. Learners' perception of vowels /iː/ and /æ/ as easier to produce than /ɪ/ and /ʌ/ accords with the NRV framework (Polka & Bohn, 2003, 2011) that vowels that are more peripheral in the vowel space tend to function as perceptual anchors from which other categories are formed, hence, they are more stable, more salient, hence, easier to perceive and produce. Another possible explanation is that L2 English /iː/ and /æ/ have a higher degree of perceptual assimilation to L1 Catalan/Spanish /i/ and /a/ than /ɪ/ and /ʌ/ to the same L1 vowels (Cebrian, 2021), which may have influenced their self-perception of difficulty. Therefore, vowels /iː/ and /æ/ act as "dominant" and may be perceived as easier to pronounce because of their proximity to their closest L1 counterpart (Cebrian & Mora, 2017).

In general, almost 90% of learners considered that their English pronunciation had improved after the TBPT intervention, which provides additional support to the quantitative results that showed learning pronunciation communicatively through tasks was effective at improving learners' pronunciation. Specifically, around 80% of learners claimed that their pronunciation of words containing /iː/ and /ɪ/ had improved but less than 15% indicated considerable improvement. In contrast, between 25% and 40% of learners stated that they had improved *a lot* in the pronunciation of English /æ/ and /ʌ/, and approximately 90% of learners believed the pronunciation of English /æ/ and /ʌ/ ranged from somewhat to a lot. It is interesting that learners' perception of greater improvement in /æ/ and /ʌ/ aligns with the generalized superiority in performance and gains for /æ/ and /ʌ/ (relative to /iː/-/ɪ/) found in quantitative results. Except for vowel distinctiveness in the DSR task, learners' perception, lexical encoding and production accuracy (and distinctiveness) gains measured objectively appeared to be larger for /æ-ʌ/ than /iː-ɪ/. Responses from the open-ended question on perception of L2 learning after the intervention revealed that, apart from improving the pronunciation of words (*My english pronunciation, my skills to distinguish different words that sound similar* [S70]), learners perceived an increase in their vocabulary (*Learning new words with their correct pronunciation* [S33]) and a general improvement in their listening, speaking skills and fluency (*Saying many words without stopping and good pronunciation* [S37]). Although L2 fluency in learners' speech was not objectively measured, learners' perception of fluency development agrees with previous studies demonstrating that, regardless of the proficiency level, learners tend to show gains in oral fluency as a result of repeating a task, in terms of either procedure or content (Lambert et al., 2017), and this may lead to improved comprehensibility. One could speculate that integrating pronunciation in communicative classes may have fostered L2 oral production beyond L2 segmental improvement, as it was found in Darcy et al. (2021). Additionally, learners expressed their self-perceived improvement in other skills such as problem-solving and communication. In fact, phonological instruction may enhance listening comprehension and communication skills, as stated by Elliott (1997). Last, performing communicative tasks that focused on phonetic form also enhanced their motivation (*I am more motivated to speak English now* [S15]) and self-confidence (*The fact of communicating in English. I felt very confident and comfortable in the last classes* [S27]) and reduced their shyness when speaking English. These perceptions corroborate Willis' (1996) ideas that tasks promote spontaneous, exploratory talk and confidence-building as learners interact in small groups. Also, supportive and positive attitudes on behalf of the teacher during the task and post-task phases may have boosted learners' motivation towards L2 pronunciation learning and L2 speaking development.

All in all, learners' beliefs about the usefulness of pronunciation instruction seem to go beyond pronunciation learning. For example, Henrichsen and Stephens (2015) suggested that formal instruction in pronunciation, particularly communicative, motivating and learner-centred, is extremely valuable and beneficial from students' perspectives and it may result in increased awareness, improved listening skills, motivation, and learning of strategies for continuing improvement. Contrary to teacher-centred pronunciation instruction, in TBPT lessons, learners became "pronunciation detectives", who tried to discover the articulatory gestures to produce the sounds by themselves and improved L2 perception and production through trial and error.

Having discussed how all learners perceived the TBPT intervention, the second part of Section 4.4. compares evaluations of the project and perceptions of self-performance, and self-improvement between the SG and CG. First, in terms of mental effort and difficulty, the CG perceived tasks as more difficult and requiring more mental effort than SG, as previously reported in Table 3.11. (Section 3.3.2.2.), but both groups found tasks equally enjoyable. These qualitative findings agree with Robinson's (2001a; 2007b) predictions of task perception by learners. Robinson showed that increases in task complexity manipulated along resource-dispersing and resource-directing variables were accompanied by significantly higher learner ratings of task difficulty and stress, but non-significant differences in task interest. This finding seems to contradict Révész (2011) who found that learners perceived the more complex task to be more useful for L2 learning as well as more interesting. However, Révész's study was cross-sectional and learners' perceptions of task difficulty were only based on one task performance, rather than their performance after engaging in several tasks, as it is the case in this dissertation. Whereas almost 70% of the CG considered task conditions as the main reason for difficulty, the majority of the SG indicated that dealing with challenging pronunciation was the most difficult aspect. In terms of the sources of task enjoyability, no significant differences arose between groups due to the great variability in responses but, whereas the CG indicated difficulty made tasks enjoyable, together with the topic, the appealing nature of the images was the most popular source of likeability reported by the SG. Concerning the strategies used during communicative breakdowns, interaction moves seemed to be more present in the CG, probably due to the greater discussion of form and meaning, whereas the SG reported using more gestures or emphasis to make themselves understood.

Finally, the CG considered all vowels to be overall easier to pronounce than the SG, especially vowels /ɪ/ and /ʌ/, probably due to self-perceived sense of achievement. However, no significant group differences were found in terms of overall perception of improvement, suggesting that both groups felt their pronunciation of English pronunciation had improved after the TBPT intervention. Except for vowel /ʌ/, the CG reported greater improvement in the high and low vowels than the SG. Interestingly, English /ʌ/ was the vowel both groups reported they had improved the most, followed by /æ/, /ɪ/ and /iː/. Learners' subjective impressions of L2 vowel improvement seem to reflect some of the results previously reported throughout this section, and they provide additional qualitative data supporting the beneficial effects of the TBPT intervention on all L2 target vowels, especially the low vowel contrast, and specifically, the English vowel /ʌ/, which initially was not easy to perceive or produce at all.

Last but not least, responses from open-ended questions varied widely between groups. The CG reported having gained more in pronunciation, vocabulary and English skills in general, but differences were minor with respect to the SG. Whilst the responses from the problem-solving strategies came from the CG, comments regarding affective factors (motivation, self-confidence, social skills) were greater in the SG, who performed tasks which had lower cognitive demands.

06 CONCLUSIONS

CHAPTER 6. CONCLUSIONS

The final chapter of this dissertation summarizes the main findings of the study (Section 6.1.); acknowledges the study's limitations (Section 6.2.); paves the way for further research in the area of TBLT and pronunciation instruction (Section 6.3.); proposes several pedagogical implications for pronunciation instruction in foreign language contexts (Section 6.4.); and ends with some concluding remarks (Section 6.5.) regarding the contributions of this doctoral dissertation to the field of pronunciation instruction and learning and a call for further teacher-researcher collaborations.

6.1. Summary of findings

The present study set out to investigate to what extent task design and manipulation would enhance a focus-on-phonetic form during authentic learner interactions, resulting in L2 phonological learning. With this objective, 63 Catalan/Spanish EFL learners received twenty TBPT lessons during 7 weeks consisting of a pre-task, an interactive task, and a post-task. Thirty-one learners carried out simple and thirty-two complex decision-making tasks, which were manipulated in terms of ±reasoning demands (Robinson, 2007a, 2011b). A class of twenty-nine learners belonged to a control group that did not receive task-based pronunciation instruction. Learners were tested at three different times (T1, T2, T3) on perceptual discrimination, lexical encoding and production of four target vowels, as well as frequency and duration of P-LRE. Finally, the study explored how L2 learners' speech development could be mediated by experiential and cognitive factors, and learners' perceptions of the TBPT intervention. The main findings and contributions of the study are presented under each research question.

RQ1. Is TBPT effective at improving L1-Catalan/Spanish learners' perception, lexical encoding and production of L2 English vowels?

Overall, the findings revealed that adolescent L1-Catalan/Spanish learners improved their pronunciation of English vowels after taking part in task-based lessons which were designed to focus on L2 pronunciation while communicating (Gurzynski-Weiss et al., 2017b). From a qualitative perspective, this intervention showed that integrating pronunciation in task-based lessons is beneficial for the development of L2 pronunciation and oral production, and can be easily incorporated in L2 speaking-oriented lessons. Also, implementing a dual focus on form and meaning through proactive techniques such as *task essentialness* and repetition increases the opportunities for learning L2 pronunciation in context, leading to automatization of the phonological targets (Darcy et al., 2019; Trofimovich & Gatbonton, 2006) in the long run.

The quantitative results confirmed our hypothesis that TBPT helped to improve learners' discrimination of L2 contrastive vowels in accuracy and speed, which generalized to new voices and lexical contexts that learners were unfamiliar with and was retained 11 weeks after the classroom intervention. As for the lexical encoding of L2 vowels, learners' identification of words (FLeC task) and rejection of nonwords (LD task) containing the target L2 vowels also significantly improved immediately after the TBPT intervention and this beneficial effect was maintained at the delayed post-test. Interestingly, whilst learners obtained higher scores in the FLeC than the LD task in each one of the testing times, probably due to the lower demands of the FLeC task (Kojima, 2019), the LD task (i.e., nonword rejection) obtained the largest gains. In relation to vowel production, pronunciation-focused tasks helped learners produce significantly more distinct as well as more accurate L2 vowels in words elicited in isolation and sentences, and such improvement generalized to untaught voices and words. These gains were still present at delayed post-test, and for vowel production in isolated words, learners' vowel *distinctiveness* and *nativelikeness* kept increasing significantly (See Table 6.1. for a summary of results). Interestingly, learners who produced L2 contrastive vowels more distinctively after the TBPT intervention, also produced L2 vowels in a more target-like manner. When comparing the vowel quality of words produced in isolation (DWR) and in sentences (DSR), learners produced overall greater distances between non-native vowels and closer distances to native speakers' vowel qualities in the DWR task, but overall, gains in vowel distinctiveness and nativelikeness were greater in the DSR task, indicating that TBPT was effective at improving learners' vowel productions when they also had to process meaning.

Table 6.1. Summary of main findings in relation to RQ1 for the experimental group.

	PERCEPTION					
	Discrimination		**Word Id. (FLeC)**		**Nonword Rej. (LD)**	
	Accuracy	*Speed*	*Accuracy*	*Speed*	*Accuracy*	*Speed*
Improvement (T1-T2)	✓	✓	✓	✓	✓	✓
Vowel improvement	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/æ-ʌ/
Generalization	✓	✓	*N/A*	*N/A*	*N/A*	*N/A*
Retention (T2-T3)	✓	✓	✓	=	✓	✓
Gains (T1-T3)	7%	151ms	18%	51ms	29.50%	58ms
	PRODUCTION					
	Words in isolation (DWR)		**Words in sentences (DSR)**			
	Dist.	*Nat.*	*Dist.*	*Nat.*		
Improvement (T1-T2)	✓	✓	✓	✓		
Vowel changes	/iː-ɪ/, /æ-ʌ/	/æ/, /ʌ/	/iː-ɪ/, /æ-ʌ/	/æ/, /ʌ/, /iː/		
Generalization	✓	✓	✓	✓		
Retention (T2-T3)	✓	✓	✓	✓		
Gains (T1-T3)	3.75SD	4.75SD	5.20SD	5.60SD		

Note. Shading indicates statistical significance. Id: identification, Rej: Rejection, Dist: distinctiveness, Nat: nativelikeness, ms: milliseconds, SD: standard deviations, N/A: non-applicable

As a matter of fact, post-intervention changes in vowel duration were mainly present in the production of words in isolation. Vowel duration in words produced in sentences minimally increased for the tense vowels, and did not increase for the lax vowels over time. Yet, the DR of /iː-ɪ/ (but not of /æ-ʌ/) increased significantly from pre- to post- test, indicating that, apart from vowel quality, learners used vowel duration to produce vowels /iː/ and /ɪ/ distinctively in words in isolation and sentences. Finally, while accuracy and response speed in the pre-lexical (ABX discrimination) and lexical (FLeC/LD) perception of L2 sounds seemed to be moderately related, this study could neither provide empirical evidence of the direct relationship between discrimination and lexical encoding gains nor support a causal relationship between perception and production gains.

RQ2. Does task complexity play a role in L1-Catalan/Spanish learners' perception, lexical encoding and production of L2 English vowels, as well as the occurrence of P-LRE?

In general, the results of the current investigation showed that learning L2 pronunciation through tasks that are cognitively demanding seem to bring about more benefits in the discrimination, lexical encoding and production of L2 vowels than performing cognitively simple tasks, in line with the predictions of the Cognition Hypothesis (Robinson, 2001b, 2007b, 2011b). This result concurs with previous studies showing that task complexity may enhance the development of L2 phonology in form-focused communicative interventions (Gordon, 2021; Solon et al., 2017, Mora & Levkina, 2018; Mora-Plaza et al., 2018). One of the most significant findings to emerge from this longitudinal study is that the potential advantage of task complexity was particularly noticeable 11 weeks after the intervention, suggesting that, directing attentional and memory resources to L2 phonological forms by increasing the demands of a task, may result in higher chances of incorporating these newly acquired/updated L2 phonological forms in long-term memory (Robinson, 2011b), hence, more robust L2 vowel learning.

Table 6.2. Summary of main findings in relation to RQ2 for the simple/complex groups.

	PERCEPTION					
	Discrimination		**Word Id. (FLeC)**		**Nonword Rej. (LD)**	
Simple	*Accuracy*	*Speed*	*Accuracy*	*Speed*	*Accuracy*	*Speed*
Improvement (T1-T2)	✓	✓	✓	✓	✓	✓
Vowel improvement	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/æ-ʌ/	/iː-ɪ/	/æ-ʌ/	/iː-ɪ/, /æ-ʌ/
Retention (T2-T3)	×	✓	×	=	✓	✓
Gains (T1-T3)	6.5%	117ms	7.5%	85.8ms	19.75%	36.9ms
Complex						
Improvement (T1-T2)	✓	✓	✓	✓	✓	✓
Vowel improvement	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/	/iː-ɪ/, /æ-ʌ/
Retention (T2-T3)	✓	✓	✓	=	✓	✓
Gains (T1-T3)	8%	188ms	28.5%	72.9ms	40%	92.5ms
	PRODUCTION					
	Words in isolation (DWR)		**Words in sentences (DSR)**			
Simple	*Dist.*	*Nat.*	*Dist.*	*Nat.*		
Improvement (T1-T2)	✓	✓	✓	✓		
Vowel change	/æ-ʌ/	/iː/, /æ/, /ʌ/	/iː-ɪ/, /æ-ʌ/	/ʌ/		
Retention (T2-T3)	✓	✓	×	×		
Gains (T1-T3)	3.0SD	2.2SD	2.5SD	2.1SD		
Complex						
Improvement (T1-T2)	✓	✓	✓	✓		
Vowel change	/iː-ɪ/, /æ-ʌ/	/iː/, /ɪ/, /æ/, /ʌ/	/iː-ɪ/, /æ-ʌ/	/iː/, /ɪ/, /æ/		
Retention (T2-T3)	✓	✓	✓	✓		
Gains (T1-T3)	4.5SD	7.4SD	7.8SD	9.0SD		

Note. Shading indicates statistical significance. Id: identification, Rej: Rejection, Dist: distinctiveness, Nat: nativelikeness, ms: milliseconds, SD: standard deviations

As for L2 vowel perception, SG and CG learners improved significantly after the intervention but the CG obtained higher accuracy, and discriminated L2 confusable vowels more accurately and faster at T3 than the SG, who obtained overall less gains. When it comes to lexical encoding, results from word identification (FLeC) and nonword rejection (LD) echoed similar results. Although both groups improved significantly at the level of lexical perception of the L2 vowels, the CG obtained higher proportion of correct responses than the SG and improvement showed for all vowels, not only a subset, as in the case of the SG. In addition, only the CG kept making significant improvement in the lexical encoding of L2 vowel contrasts at T3. As for L2 vowel production (*distinctiveness* and *nativelikeness*), between-group findings were analogous for words produced in isolation and in sentences. Generally speaking, all learners learned to distinguish L2 confusable vowels to a larger extent after the intervention but only all CG's vowel productions became significantly more accurate at T2, and only the CG kept producing more distinct and target-like vowels at T3 (See Table 6.2. for a summary of results). Whilst gains in vowel *distinctiveness* and *nativelikeness* were overall larger in the CG than the SG, both groups used vowel duration similarly to distinguish L2 vowel contrasts, especially, /iː-ɪ/. Finally, contrary to previous research on LRE and L2 oral performance (Kim, 2009, 2012; Révész, 2011), task complexity did not have a significant impact on the frequency (Solon et al., 2017) and duration of LREs. However, the frequency of P-LRE increased across the three times, perhaps reflecting progressive awareness of L2 phonologically distinct forms. To close, these findings advance existing research (Gordon, 2021; Solon et al., 2017) assessing the longitudinal benefits of manipulating task complexity, following the Triadic Componential Framework (Robinson, 2007a, 2011b), on L2 phonological development.

RQ3. Do individual differences in L2 experiential and cognitive factors explain L1-Catalan/Spanish learners' performance and gains in English vowels?

Having assessed potential learner factors mediating the effectiveness of L2 pronunciation instruction, the results speak to the complex relationships between experiential/cognitive factors and L2 vowel performance and gains. Overall, self-reported L2 learning experience did not seem to explain much variance in either L2 vowel performance or gains, indicating that L2 proficiency, WM, and ASA contributed significantly more than L2 English learning experience to explain different phonological outcomes, which seems reasonable considering that this study was conducted in a FL context where L2 input is relatively scarce (Muñoz, 2014), thus, inter-individual variation in terms of L2 exposure and production may be subtle.

Concerning L2 vowel performance across the three testing times, the most obvious finding to emerge from this study is that complex WM and ASA were the cognitive factors which seemed to explain larger inter-individual variation in L2 vowel perception, lexical encoding and production (Darcy et al., 2015; Mora & Mora-Plaza, 2019). L2 proficiency and PSTM were positively related to perceptual discrimination, word identification and accuracy in the production of words and sentences. Last, learners with greater complex WM obtained higher accuracy in all perception and production tasks. In line with Mora-Plaza et al. (2022a), ASA was moderately associated with perceptual discrimination and lexical encoding.

As for L2 vowel gains, the correlations with experiential and cognitive factors were relatively weak; only complex WM seemed to be related to vowel discrimination, speed of lexical encoding and accuracy in vowel production. However, only 7% of L2 vowel gains were explained by past and recent L2 learning experience (Suzukida & Saito, 2023). Interestingly, ID in L2 experience and cognition were more strongly correlated with L2 vowel gains from the CG than SG. Still these results should be treated with caution due to the weak strength of the correlations.

RQ4. Which were L1-Catalan/Spanish learners' perceptions of the TBPT intervention?

When it comes to learners' perceptions after participating in the TBPT intervention, a recurrent theme in the post-intervention questionnaire was a general sense of enjoyment after performing the communicative tasks and a feeling of improvement in L2 pronunciation and oral skills as well as acquisition of new vocabulary.

First, self-awareness of the importance of learning pronunciation in class changed dramatically after learners practiced L2 pronunciation communicatively under a task-based approach. Second, in terms of learners' evaluation of the project, in general, responses reflected the usefulness of pre-tasks and post-tasks to learn the meaning and pronunciation of the target words, respectively, and there was general consensus about the likeability of the interactive tasks. Whilst a minority mentioned that the repetitive nature of the tasks and the testing phase (amongst others) made the project less enjoyable, the great majority provided a very positive feedback about the intervention regarding the interactive tasks and the overall procedure, as well as highlighted the beneficial effects on L2 learning. A common view amongst learners was that the intervention helped them improve L2 pronunciation and learn new vocabulary; however, they also expressed benefits for speaking, fluency development and overall language skills.

Last, there was a general sense of overall improvement in L2 pronunciation after the TBPT intervention. Learners confidently rated the target vowels as relatively easy to pronounce, especially English vowels /iː/ and /æ/, and stated greater self-perceived improvement in /æ/ and /ʌ/, which accords with the generalized superiority in performance and gains found in quantitative analyses. Additionally, responses from open-ended questions revealed that, apart from learning pronunciation and new vocabulary, learners' listening, speaking and problem-solving skills had improved and, on the affective level, performing engaging communicative tasks had enhanced their motivation to learn English and increased their self-confidence when communicating in English.

Finally, learners who performed complex tasks (CG) perceived tasks to be cognitively more difficult and requiring more mental effort than the SG (Robinson, 2001a, 2007b), but both groups considered tasks equally interesting. Both groups expressed an overall perception of improvement but the CG reported significantly greater improvement for all vowels than the SG, except for /ʌ/, which was the English vowel both groups reported having improved the most, which seems to echo the findings obtained from quantitative analyses.

6.2. Limitations of the study

There are several limitations that need to be acknowledged so as to consider them for future research and for replication purposes.

First, an important limitation lies in the fact that this study did not measure L2 vowel production during spontaneous speech. A less structured task such as a monologic picture narrative (Thomson & Derwing, 2015) or an information-gap task prompting the target L2 words would have provided useful insights into learners' vowel production in real-world communication (see Darcy & Rocca, 2023). Still, the advantages of assessing L2 pronunciation through a delayed sentence repetition task ensured that the learners produced all segments targeted in the study and allowed precise analyses of learners' vowel production for pre-, post- and delayed post-test, as well as a fair between-learner group comparison. Following up on pronunciation assessment methods, this study could not evaluate the state of learners' phonolexical representations. Instead of using two perceptual lexical decision tasks differing in difficulty, further studies could employ a picture-naming task (Hayes-Harb & Masuda, 2008; Simonchyk & Darcy, 2021) to assess changes in L2 phonolexical representations at the production level. In a similar vein, the results of the perceptual lexical decision should be treated with extreme caution due to the unbalanced number of test and filler trials, which may have drawn learners' attention to the phonological form of words exclusively, rather than both form and meaning. In addition, stimulated recalls could not be conducted after the oral interactive tasks due to time constraints but would have been useful at determining whether intended task complexity did indeed create different levels of cognitive processes in the tasks and provide insights into learners' attentional allocation during communication (Gass & Mackey, 2017).

Second, the current study examined L2 phonological accuracy in production through acoustic measures (i.e., vowel quality in Mahalanobis distances and vowel duration), but was not specifically designed to evaluate pronunciation with global scoring methods such as comprehensibility or accentedness (e.g., Darcy et al., 2019; Gordon, 2021). Although the effects of instruction tend to be less robust when L2 pronunciation is measured through listener-based ratings rather than objective measurements (Lee et al., 2015), employing additional subjective assessments of spontaneous speech may have provided valuable information of how L2 vowel learning influenced global L2 pronunciation proficiency (Saito & Plonsky, 2019). Furthermore, it was beyond the scope of this study to assess the impact of TBPT on L2 oral production dimensions (i.e., CAF) but further research should investigate whether an increase in pronunciation accuracy resulting from task complexity induces a decrease in speaking fluency or whether pronunciation accuracy and fluency compete with lexical and grammatical accuracy and complexity. Last, the reader should bear in mind that the present study was focused on learners' improvement in L2 vowel perception, lexical encoding and production, hence, findings regarding the occurrence of P-LRE remain rather exploratory. For example, we did not provide a thorough analysis on the type of P-LRE (e.g., clarification request, confirmation check) dyads engaged in, or took into consideration other kinds of LRE beyond pronunciation (e.g., lexis, grammar, pragmatics), as Solon et al. (2017) did.

Third, the relationships between L2 past and recent experience and L2 vowel performance and learning need to be interpreted cautiously as the data from learners' L2 experience was self-reported. More accurate measurements of quantity and quality of L2 experience such as the use of technological devises to track learners' L2 interactions would provide more precise information about learners' L2 experiences inside and outside the FL classroom. Furthermore, the present dissertation only focused on a number of experiential and cognitive factors that have been found to mediate L2 speech development (Darcy et al., 2015; Mora et al., 2022; Suzukida & Saito, 2023); however, we did not control for learners' vocabulary size (Daidone & Darcy, 2021; Llompart, 2021a); musical expertise (Christiner & Reiterer, 2013, 2016); other socio-affective aspects such as motivation (Saito et al., 2018) and pronunciation-specific anxiety (Baran-Łucarz, 2016, 2022); or the socioeconomic and cultural status of the families (CSd'A, 2020). Further studies should include as many potentially related variables in a single study as possible to be able to assess the joint and unique contribution of each variable (Mora, 2022; Suzukida, 2021), especially in longitudinal studies.

Fourth, provided that this study was conducted in a FL classroom context, learners were unavoidably exposed to L1-accented input from the peers during the interactive task. In the case of low-proficiency leaners, foreign accented conversations might have reduced the opportunities to develop awareness of cross-language differences between L1 and L2 segmental phonology (Tyler, 2019). Given the interactive nature of the intervention, individualized immediate corrective feedback could not be provided during spontaneous speech performance, which would have certainly helped weaker learners. In addition, due to practical constraints, we could not control for interlocutor effects and the possibility of speech accommodation (Bueno-Alastuey, 2011, 2013; Mozaffari, 2017). Conversational partners may have been self-selected on the basis of friendships (i.e., someone who they felt comfortable speaking in English), or proficiency (i.e., someone with a higher proficiency in English) or, by default, may have interacted with someone who they did not know very well. Although conversational partners changed after performing five tasks, future studies should take into consideration how this may affect the occurrence of P-LRE and L2 oral production.

Fifth, although the intervention stimuli were fairly balanced in terms of lexical frequency and word familiarity and fairly consistent in terms of orthography (phoneme-grapheme correspondence) and phonetic context by presenting the target vowels in L2 minimal pairs, future designs should try to reduce the variability of the speech materials in terms of word familiarity, lexical frequency, cognate occurrence, and neighbourhood density. These variables have been found to potentially affect the learning of L2 vowels in lexical contexts (e.g., Cook & Gor, 2015; Llompart, 2021b).

Last but not least, the project used a convenience sample (i.e., three intact EFL classes), and only focused on adolescent learners who shared the same L1. Further research should examine how the results of this study can be extrapolated to other L2 learner populations from different settings (e.g., immersive contexts), different ages (e.g., young EFL learners) and different L1s. In addition, more TBPT research needs to be conducted with learners that are not from WEIRD (Western, Educated, Industrialized, Rich and Democratic) contexts to be able to generalize these findings to wider populations.

6.3. Directions for further research

It is hoped that the current study paves the way for further research in the field of TBLT and pronunciation instruction by throwing up numerous questions in need of further investigation.

To start with, in order to reduce learners' exposure to L1-accented input, a possible avenue of future research would be to explore task designs that contemplate the possibility of including native input and corrective feedback (Lyster et al., 2013; Saito & Lyster, 2012) while performing communicative tasks. For example, by using a computer-based collaborative map task in the L2 speech laboratory, Mora and Levkina's (2018) EFL learners could hear the name of the streets containing problematic L2 vowels produced by English native speakers while engaging in a conversation. Whether this can be applicable to the context of real classrooms is a question that deserves further investigation. Given the affordances that technology can offer nowadays (see Section 1.3.3.), video-based interactive tasks would be useful to assess the effect of the interlocutor L1 background and L2 proficiency, for example, in the occurrence of P-LRE and, subsequently, L2 pronunciation learning. In addition, virtual reality could help create interactive virtual scenarios where learners would need to navigate through L2 phonetic forms to complete virtual tasks.

Second, it would be interesting to compare the effectiveness of form-focused communicative interventions —based on Celce-Murcia et al.'s (2010) Communicative Framework and/or Trofimovich and Gatbonton's (2006) ACCESS framework — with task-based approaches for pronunciation instruction —grounded on Willis' (1996) framework of Task-Based Learning— on L2 segmental and suprasegmental learning. While these different approaches may be complementary and applicable in different FL contexts, further research may help determine which approach leads to larger gains in L2 pronunciation instruction. A further study could also assess the advantages of providing explicit instruction on L2 phonology prior to task-supported pronunciation instruction (Gordon, 2021; Ruan & Saito, 2023), or how CAPT resources (e.g., *English Accent Coach*, Thomson, 2012) may help consolidate L2 segmental learning after a TBPT intervention, especially, for those learners who have lower L2 proficiency, aptitude or suffer from foreign classroom anxiety (Baran-Łucarz, 2016; Saito et al., 2018). In all cases, longitudinal research (e.g., 1-year intervention) is urgently needed to assess the potential beneficial effects of task-based/task-supported pronunciation teaching on L2 pronunciation development.

Third, further investigation and experimentation into task design and manipulation would be desirable to determine which methodological aspects of task design may enhance L2 pronunciation development. For example, it would be worth exploring how many L2 segmental and suprasegmental targets should be included in the design of a task and determine how often learners should be exposed to them. More broadly, exploring the interaction effects between variables presented in the Triadic Componential Framework (Robinson, 2007a, 2011b) and validating the theoretical constructs in theories of task complexity by investigating the cognitive processes in which learners engage during task-based performance (e.g., Révész et al., 2022 for behavioural and neurocognitive approaches) would theoretically and methodologically advance the field of TBPT. It is recommended that further research be undertaken in the following TBLT manipulation techniques: task types (i.e., focused, unfocused, closed, open, one-way, two-way); task sequencing (i.e., the effects of sequencing task complexity and phonetic complexity, for example, consider simple-to-complex syllable structures, short vs. long words, single vs. multiple-talker exposure); task repetition (i.e., the role of repeating L2 input for developing efficiency in the processing of L2 speech); task modality (i.e., face-to-face vs. computer-mediated synchronous communication); and task complexity.

Following on the previous paragraph, we would like to call for further research on the role of task complexity for L2 pronunciation learning. On the one hand, future studies could consider triangulation of methods (e.g., expert ratings, learner ratings, interviews, stimulated recalls) to ensure an appropriate operationalization of task complexity. While the current dissertation showed that task complexity resulted in long-term benefits in L2 pronunciation learning when the phonological targets were made task essential, future investigations could potentially assess task complexity effects in pronunciation-unfocused tasks, where attention is not directed to L2 pronunciation. On the other hand, more research is needed to better understand what effects this type of TBPT intervention may have on suprasegmental aspects of L2 pronunciation (e.g., lexical stress, rhythm), L2 oral production and specific linguistic dimensions (e.g., grammar and vocabulary) or L2 oral comprehension.

Last, more work is needed to assess L2 pronunciation development in terms of acoustic measurements and global assessments of accentedness and comprehensibility to be able to adequately characterize the impact of task design manipulations on L2 pronunciation development at fine-grained, and also, perceptible levels (Saito & Plonsky, 2019). The use of validated automated comprehensibility assessment (Saito et al., 2022a) may speed up the time-consuming (and often costly) task of eliciting listeners' L2 speech assessment. In addition, categorical ABX discrimination tasks could be employed along with categorization and goodness ratings of the stimuli tokens so as to be able to determine learners' assimilation patters for non-native contrasts precisely (Tyler et al., 2014), and thus, be able to carefully interpret pronunciation instruction outcomes in view of L2 speech acquisition models (i.e., PAM-L2, Best & Tyler, 2007).

6.4. Pedagogical implications

Taken together, the findings of this study suggest a number of important implications for future pedagogical practices.

First of all, this thesis has shown that implementing form-focused communicative instruction promotes robust learning of challenging L2 phonological forms in authentic meaningful speech. Instead of teaching pronunciation explicitly, often in a decontextualized manner, L2 phonological learning can happen incidentally as long as learners engage in form-focused problem-solving tasks which have clear communicative goals. In terms of task design, the current dissertation showed that it is important to illustrate a connection between reality and teaching by selecting tasks that are purely meaningful for learners. In this case, organizing a trip is an activity which was part of learners' reality and they could easily relate to. In addition, adopting a learner-centred, task-based approach to the teaching of L2 pronunciation with interaction as the central component generated a fair balance between form and meaning as tasks were completed successfully while there was room for negotiation of form, as instantiated through P-LRE.

Furthermore, by integrating pronunciation in the teaching of L2 vocabulary at the earliest possible time and revisiting already known words while focusing on phonetic form (Darcy & Holliday, 2019), learners may be able to develop accurate L2 phonolexical representations, and thus, more accurate L2 pronunciation. At the same time, it is possible to integrate pronunciation in the course curriculum so as to bring new challenges and experiences that will reinforce the different skills involved in L2 learning. Despite not being tested in this dissertation, TBPT may progressively lead to an increase in overall intelligibility and comprehensibility (Gordon, 2021).

One of the great advantages of these pronunciation-based tasks is that they are adaptable to different proficiency levels. By creating focus-on-form opportunities during interaction such as making L2 problematic phonological forms *task essential* and exposing learners to the same words in different contexts (if possible through multiple different voices), less advance learners can develop pronunciation awareness, without the need to produce immediate modified output. Also, this dissertation showed that it is important to first introduce segmental and suprasegmental features that may have a higher impact on intelligibility (Alnafisah et al., 2022; Munro & Derwing, 2006) and, as learners' pronunciation proficiency increases, shift the main goal towards comprehensibility while the focus is on meaning (Darcy et al., 2012). An alternative is to conduct a pronunciation diagnosis (Levis & Echelberger, 2022), assist learners in determining their goals and targets for L2 pronunciation and develop realistic tasks and assessment based on these goals.

This dissertation showed that increasing the cognitive demands of a task seems to foster L2 pronunciation development and lead to longer-term effects than performing cognitively simple tasks. Therefore, another important pedagogical implication from these findings is that task designers should consider adding complexity to tasks with the aim of directing learners' attentional and memory resources to L2 phonological forms. Although our study did not assess the effect of task sequencing, these findings point to the need to experiment with sequences and combinations (e.g., simple-to-complex, complex-to-simple, simple-to-simple to complex-to-complex, simple-to-complex to simple-to-complex) in order to find a task sequence that will maximize learners' motivation and learning. For example, if sufficient time is allotted to practicing L2 oral skills, teachers could introduce simpler tasks to familiarize learners with the meaning and phonological form of the words and increase task complexity progressively. All in all, task sequencing is a very promising avenue for further research with clear pedagogical implications for the design of task syllabi and TBLT programs.

Furthermore, TBPT has been found to be generally beneficial regardless of learners' differences in L2 experience and cognitive factors. In order to boost the quality of L2 pronunciation development, teachers should enhance learners' motivation and induce positive emotional states through careful design of communicative tasks, avoiding "drill-like" pronunciation activities appearing at the end of lessons. In light of learners' perceptions of increased motivation and self-confidence reported after the TBPT intervention and based on my personal experience in the EFL classroom, I recommend teachers to create a friendly and pleasant classroom environment in which the target language can be used frequently and freely, provide positive feedback throughout all stages of the task cycle, place attention on those learners who struggle to communicate in the second language and create low-risk situations such as completing the task in pairs or small groups to mitigate the effects of public speaking anxiety (Baran-Łucarz, 2022).

Last, some of the materials that are part of this TBPT implementation have been publicly shared in the Task Bank (Gurzynski-Weiss & IATBLT, n.d.) and L2 Speech Tools (Mora-Plaza et al., 2022b) repositories with the aim of making these free, ready-to-use, downloadable materials accessible to the teaching community. However, we also encourage teachers to integrate L2 pronunciation in their TBLT classes by using proactive form-focused techniques such as task essentialness, input flooding or enhancement that can lead to L2 pronunciation learning during meaningful communication.

6.5. Concluding remarks

The present investigation has been one of the first attempts to examine the benefits of task-based language teaching for L2 pronunciation development in a real FL classroom context. On the one hand, the current findings add to a growing body of literature on TBPT (Gurzynski-Weiss et al., 2017a) demonstrating the potential of task design and manipulation to induce a focus on phonetic form during meaningful interaction. On the other hand, this study confirms previous findings (Darcy et al., 2021) and contributes additional evidence that L2 pronunciation can be learned through communicative approaches. Methodologically speaking, this study has also advanced in the assessment of L2 pronunciation by employing tasks previously validated for phonetic training and pronunciation instruction studies in a real classroom setting.

Yet, many fundamental questions in TBPT still need to be explored empirically to gain a better understanding of the role of tasks in developing L2 pronunciation and which assessment methods can best capture improvement in L2 pronunciation. We hope that the insights gained from this study may be of assistance to teachers, task designers and material developers to increase the use of tasks to teach pronunciation in the classroom. We also call for future dialogue between researchers and teachers to understand the realities of actual classroom practice and how TBPT may be maximally relevant to the instructional contexts where FL teachers operate.

REFERENCES

Abe, A. (2011). Effects of form-focused instruction on the acquisition of English weak forms by Japanese EFL learners. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 184-187), Hong Kong.

Adrian, M. & Mora, J. C. (2022, May 3-4). *Is phonetic training effective at updating phonolexical representations? Interactions between training materials and stages in phonological development* [Paper presentation]. Bilingual Lexical Representations Workshop, Marseille, France.

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

Aliaga-Garcia, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics, 47*(1), 1–14.

Alnafisah, M., Goodale, E., Rehman, I., Levis, J., & Kochem, T. (2022). The impact of functional load and cumulative errors on listeners' judgments of comprehensibility and accentedness. *System, 110*, 102906.

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

Anderson, N. & McCutcheon, N. (2019). *Activities for Task-Based Learning: Integrating a fluency first approach into the ELT classroom*. DELTA Publishing.

Andringa, S., & Dąbrowska, E. (2019). Individual differences in first and second language ultimate attainment and their causes. *Language Learning, 69*, 5-12.

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics, 40*(4), 582-594.

Astheimer, L. B., Berkes, M., & Bialystok, E. (2016). Differential allocation of attention during speech perception in monolingual and bilingual listeners. *Language, Cognition and Neuroscience, 31*(2), 196-205.

Awwad, A., & Tavakoli, P. (2022). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching*, *60*(2), 169-196.

Bachman, L. & A. Palmer (1996). *Language testing in practice*. Oxford: Oxford University Press.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*(1), 136-163.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*(4), 689-725.

Baralt, M. (2014). Task complexity and task sequencing in traditional versus online language classes. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 95–122). London, UK: Bloomsbury.

Baralt, M., Gilabert, R. & Robinson, P. (2014). *Task sequencing and instructed second language learning*. London, UK: Bloomsbury.

Baran-Łucarz, M. (2016). Conceptualizing and measuring the construct of pronunciation anxiety. In M. Pawlak (Ed.), *Classroom-Oriented Research, Second Language Learning and Teaching* (pp. 39–56). Cham: Springer International Publishing.

Baran-Łucarz, M. (2022). Language Anxiety. In Derwing, T. M., Munro, M. J., & Thomson, R. I. (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 83-96). New York (USA): Routledge.

Barón, J., Celaya, M. L. & Levkina, M. (2020). Learning pragmatics through tasks. When interaction plays a role. *Applied Pragmatics, 21*, 1-25.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

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

Bitchener, J. (2004). The relationship between the negotiation of meaning and language learning: A longitudinal study. *Language Awareness*, *13*(2), 81-95.

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.09). http://www.praat.org/.

Bohn, O. S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 379- 410). Timonium, MD: York Press.

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.

Bowles, M. A., Adams, R. J., & Toth, P. D. (2014). A comparison of L2–L2 and L2–heritage learner interactions in Spanish language classrooms. *The Modern Language Journal*, *98*(2), 497-517.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *State of-the-art issues in second language phonology* (pp. 287-308). Amsterdam, The Netherlands: John Benjamins.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61*(2), 206-219.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. London, UK: Sage.

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology*, *18*(3), 328-352.

Breen, M. P. (1987). Contemporary paradigms in syllabus design. Part I. *Language teaching*, *20*(2), 81-92.

Brown, A. (1991). Functional load and the teaching of pronunciation. In A. Brown (Ed.), *Teaching English Pronunciation: A Book of Readings* (pp. 221–224). Routledge, London.

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language teaching research*, *20*(4), 436-458.

Brunfaut, T., Kormos, J., Michel, M., & Ratajczak, M. (2021). Testing young foreign language learners' reading comprehension: Exploring the effects of working memory, grade level, and reading task. *Language testing*, *38*(3), 356-377.

Bueno-Alastuey, M. C. (2011). Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. *Computer Assisted Language Learning*, *24*, 419–432.

Bueno-Alastuey, M. C. (2013). Interactional feedback in synchronous voice-based computer mediated communication: Effect of dyad. *System*, *41*, 543–559.

Burri, M. (2021). "Teaching pronunciation is always on my mind": A 5-year longitudinal study on a Japanese English teacher's developing practices and cognition about pronunciation. *JALT Journal*, *43*(2), 143–166.

Bygate, M. Skehan, P. and Swain, M. (Eds.). (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. Harlow: Longman.

Byrne, D. (1986). *Teaching Oral English.* New Edition. Harlow, UK: Longman.

Candlin, C. (1987). Toward task-based learning. In C. Candlin & D. Murphy (Eds.), *Language learning tasks* (pp. 5–22). Englewood Cliffs, NJ: Prentice-Hall.

Carlet, A. (2017). *L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study*. (Unpublished doctoral dissertation). Barcelona: Universitat Autònoma de Barcelona.

Carlet, A., & Cebrian, J. (2019). Assessing the Effect of Perceptual Training on L2 Vowel Identification, Generalization and Long-term Effects. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters—In honor of Ocke-Schwen Bohn* (pp. 91-119). Department of English, School of Communication & Culture, Aarhus University.

Carlet, A., & Cebrian, J. (2022). The roles of task, segment type, and attention in L2 perceptual training. *Applied Psycholinguistics, 43*(2), 271-299.

Carlet, A., & De Souza, H. K. (2018). Improving L2 pronunciation inside and outside the classroom: Perception, production and autonomous learning of L2 vowels. *Ilha do Desterro, 71*(3), 99–123.

Carlet, A., & Rato, A. (2015). Non-native perception of English voiceless stops. In E. Babatsouli, & D. Ingram (Eds.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015* (pp. 57−67). Chania, Greece: Institute of Monolingual and Bilingual Speech.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87-136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge, UK: Cambridge University Press.

Casillas, J. V. (2020). The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning, 70*(3), 768-806.

Cauldwell, R. (2018). *A syllabus for listening: Decoding*. Speech in Action.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*, 372-387.

Cebrian, J. (2007). Old sounds in new contrasts: L2 production of the English tense-tax vowel distinction. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1637-1640). Saarbrucken: University of Saarland.

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.

Cebrian, J. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *Journal of the Acoustical Society of America*, *149*(4), 2671–2685.

Cebrian, J., & Mora, J. C. (2017, September 7-9). *Crosslinguistic perceptual similarity and asymmetric lexical competition in L2 spoken-word recognition* [Paper presentation]. The 23th International Conference on Architectures and Mechanisms of Language Processing, Lancaster, UK.

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*, 660917.

Cebrian, J., Mora, J. C. & Aliaga-García, C. (2011). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In Wrembel, M., Kul, M., Dziubalska-Kołaczyk, K. (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 41-52). Frankfurt am Main: Peter Lang.

Celce-Murcia, M. (1983). Teaching Pronunciation Communicatively. *Mextesol Journal*, *7*(1), 10-25.

Celce-Murcia, M, Brinton, D, & Goodwin, J. (2010). *Teaching Pronunciation: A Reference Book for Teachers (2nd edition)*. New York, US: Cambridge University Press.

Cerviño, E., & Mora, J. C. (2009). Duration as a phonetic cue in the categorization of /iː/-/ɪ/ and /s/-/z/ by Spanish/Catalan learners of English. *The Journal of the Acoustical Society of America*, *125*(4), 2764.

Charoy, J., & Samuel, A. G. (2020). The effect of orthography on the recognition of pronunciation variants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1121–1145.

Christiner, M., & Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in psychology*, *4,* 874.

Christiner, M., & Reiterer, S. M. .(2016). Music, song and speech: A closer look at the interfaces between musicality, singing and individual differences in phonetic language aptitude. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 131–156). Amsterdam, The Netherlands: John Benjamins.

Cook, S. V., & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *The mental lexicon*, *10*(2), 247-270.

Cooke, M., & García Lecumberri, M. L. (2018). Effects of exposure to noise during perceptual training of non-native language sounds. *The Journal of the Acoustical Society of America*, *143*(5), 2602-2610.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*, 46–66.

Couper, G. (2016). Teacher cognition of pronunciation teaching amongst English language teachers in Uruguay. *Journal of Second Language Pronunciation*, *2*(1), 29-55.

Couper, G. (2022). Effective Feedback for Pronunciation Teaching. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 174–193). Hoboken, NJ: John Wiley & Sons, Inc.

Crookes, G. (1986). Towards a validated analysis of scientific text structure. *Applied linguistics*, *7*(1), 57-70.

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, *40*(2), 443-457.

Cruttenden, A. (2014). *Gimson's Pronunciation of English* (8th ed). London: Routledge.

Crystal, T. H., & House, A. S. (1988). The duration of American-English vowels: An overview. *Journal of Phonetics*, *16*(3), 263-284.

CSd'A (2020). *L'avaluació de l'expressió oral a quart d'ESO: Curs 2018-2019*. Barcelona, Departament d'Ensenyament, Consell d'Avaluació Superior del Sistema Educatiu (Informes d'Avaluació 23).

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269-284.

Daidone, D. (2020). *How Learners Remember Words in their Second Language: The Impact of Individual Differences in Perception, Cognitive Abilities, and Vocabulary Size*. Dissertation, Bloomington (IN): Indiana University.

Daidone, D., & Darcy, I. (2021). Vocabulary Size Is a Key Factor in Predicting Second Language Lexical Encoding Accuracy. *Frontiers in psychology*, *12*, 688356.

Dao, D. (2018). Learners' perspectives on English pronunciation teaching and learning: A preliminary study in the Vietnamese context. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 86-99). Ames, IA: Iowa State University.

Darcy, I. (2018). Powerful and Effective Pronunciation Instruction: How Can We Achieve It? *Catesol Journal*, *30*(1), 13-45.

Darcy, I., & Holliday, J. (2019). Teaching an old word new tricks: Phonological updates in the L2 lexicon. In J. Levis, C. Nagle & E. Todey (Eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference* (pp. 10–26). Ames, IA: Iowa State University.

Darcy, I., & Mora, J. C. (2016). Executive control and phonological processing in language acquisition: The role of early bilingual experience in learning an additional language. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive*

Individual Differences in Second Language Processing and Acquisition (pp. 249-277). Amsterdam, The Netherlands: John Benjamins.

Darcy, I. & Rocca, B. (2023). Comprehensibility improvements in integrated pronunciation instruction: A comparison of instructional methods and task effects. *Journal of Second Language Pronunciation*, *8*(3), 328–362.

Darcy, I. & Thomas, T. (2019). When blue is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism: Language and Cognition*, *22*(5), 1141–1159.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The mental lexicon*, *8*(3), 372-420.

Darcy, I., Ewert, D, & Lidster, R. (2012a). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, September 2011 (pp. 93-108). Ames, IA: Iowa State University.

Darcy, I., Mora, J.C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Concordia Working Papers in Applied Linguistics*, *5*, 115–129.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*(4), 741-773.

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63-72.

Darcy, I., Rocca, B., & Hancock, Z. (2021). A window into the classroom: How teachers integrate pronunciation instruction. *RELC Journal, 52*(1), 110-127.

Darcy, I., Hancock, Z., Lee, J., & Rocca, B. (2019, June 6-7). *Improving pronunciation in spontaneous speech? A comparison of two instructional methods* [Paper presentation]. L2 pronunciation research workshop: Bridging the gap between research and practice, Barcelona, Spain.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012b). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research, 28,* 5-40.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics, 13*, 1-24.

De Ridder, I., Vangehuchten, L., & Seseña Gomez, M. (2007). Enhancing automaticity through task-based language learning. *Applied Linguistics, 28,* 309–315.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 42–63). Cambridge, UK: Cambridge University Press.

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2013). Production training in Second Language Acquisition: A comparison between objective measures and subjective judgments. *Proc. Interspeech, 14,* 2375–2379.

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

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research* (Vol. 42). Amsterdam, The Netherlands: John Benjamins.

Derwing, T. M., & Munro, M. J. (2022). Pronunciation learning and teaching. In T. M. Derwing, M. J. Munro, R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 147–159). New York: Routledge.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied language learning*, *13*(1), 1-17.

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, *30*(1), 22-44.

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

Deterding, D. (1990). *Speaker Normalization for Automatic Speech Recognition*. Doctoral dissertation. Cambridge, UK: Cambridge University.

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association, 27*(1-2), 47-55.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning an.d Individual Differences, 22*, 680-689.

Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., ... & Gutierrez-Osuna, R. (2019). Golden speaker builder–An interactive tool for pronunciation training. *Speech Communication, 115*, 51-66.

Dörnyei, Z. (2009). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Bristol, UK: Multilingual Matters.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in second language acquisition, 20*(3), 349-385.

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in Second Language Research: Construction, Administration, and Processing* (2nd Ed.). Routledge.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (ed). *Cognition and Second language Instruction* (pp. 206–257). New York: Cambridge University Press.

Doughty, C., & Pica, T. (1986). "Information gap" tasks: Do they facilitate second language acquisition?. *TESOL quarterly, 20*(2), 305-325.

Doughty, C. J. & Williams, J. (1998). *Focus on form in classroom second language acquisition.* Cambridge, UK: Cambridge University Press.

Draine, S. (1999). Inquisit Lab (Version 5.0.14.0) [Computer software]. Seattle, WA: Millisecond Software. Retrieved from http://www.millisecond.com/

Duff, P. (1986). Another look at interlanguage talk: taking task to task. In R. R. Day, (Ed.). *Talking to learn: conversation in second language acquisition*, (pp. 147-181). Rowley, MA: Newbury House.

East, M. (2012). *Task-based language teaching from the teachers' perspective: Insights from New Zealand*. Amsterdam, The Netherlands: John Benjamins.

Elliott, R. A. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal, 79,* 530–542.

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

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246.

Ellis, R. (2017). Position paper: Moving task-based language teaching forward. *Language Teaching, 50*(4), 507–526.

Ellis, R. (2018). Taking the critics to task: The case for task-based teaching. In I. Walker, D. Chan, M. Nagami & C. Bourguignon (Eds.) *New Perspectives on the Development of Communicative and Related Competence in Foreign Language Education* (pp. 23–40). Berlin: De Gruyter.

Ellis, R. (2021). Task-Based Language Teaching: Early Days, Now and into the Future. In Sudharshana, N. P., Mukhopadhyay, L. (Eds.) *Task-Based Language Teaching and Assessment* (pp. 39–61). Singapore: Springer.

Ellis, R., Baskturkmen, H., & Loewen, S. (2001). Learning uptake in communicative ESL lessons. *Language Learning*, *51*, 281–318.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge, UK: Cambridge University Press.

Erlam, R., & Tolosa, C. (2022). *Pedagogical realities of implementing task-based language teaching in the classroom*. Amsterdam, Netherlands/Philadelphia, PA: John Benjamins.

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

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.

Flanders, B. N., & Nuthall, G. (1972). The classroom behavior of teachers: An overview. *International Review of Education*, 427-429.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (2016, June 10-12). *The role of phonetic category formation in second language speech acquisition* [Paper presentation]. The 8th International Conference on Second Language Speech, Aarhus, Denmark.

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, *10*(9781108886901.002).

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

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*(1), 1–26.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic systems. *Speech Communication*, *40*, 467–491.

Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, *3*(1), 34-56.

Foote, J. A., & Trofimovich, P. (2018). Second Language Pronunciation Learning: An overview of theoretical perspectives. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (pp. 75-90). London and New York: Routledge.

Forster, K. I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, 35*(1), 116-124.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition*, *18*(3), 299-323.

Fouz-González, J. (2015). Trends and directions in computer assisted pronunciation training. In J. A. Mompean, & J. Fouz-González (Eds.), *Investigating English Pronunciation: Trends and Directions* (pp. 314–342). Basingstoke and New York: Palgrave Macmillan.

Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology, 24*(1), 62–85.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349–366.

Fraser, H. (2011). Teaching teachers to teach /ɹ/ and /l/ to Japanese learners of English: An integrated approach. In M. Ashby (Ed.), *Proceedings of the Phonetics Teaching and Learning Conference* (pp. 11-15).

Fullana, N. (2006) The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (Ed.) *Age and the Rate of Foreign Language Learning* (pp. 41–64). Clevedon, UK: Multilingual Matters.

Fullana, N., & MacKay, I. R. A. (2003). Production of English sounds by EFL learners: The case of /i/ and /i /. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1525–1528). Barcelona: Causal Productions.

García Lecumberri, M. L., & Gallardo del Puerto, F. (2003). English foreign language sounds in school learners of different ages. In M. P. García Mayo & M. L. García Lecumberri (Eds.*), Age and the acquisition of English as a foreign language* (pp. 115–135). Clevedon, Avon: Multilingual Matters.

García Mayo, M. D. P., & García Lecumberri, M. L. (2003). *Age and the acquisition of English as a foreign language*. Clevedon, Avon: Multilingual Matters.

Gass, S. (1997). *Input, Interaction, and the Second Language Learner*, Mahwah, NJ: Lawrence Erlbaum.

Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research* (2nd Ed.). New York, NY: Routledge.

Gass, S. M., & Varonis, E. M. (1985). Variation in native speaker speech modification to non-native speakers. *Studies in second language acquisition*, *7*(1), 37-57.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL quarterly*, *22*(3), 473-492.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*(3), 325-353.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259.

Ghaffarvand Mokari, P., & Werner, S. (2019). On the role of cognitive abilities in second language vowel learning. *Language and Speech*, *62*(2), 260-280.

Gilabert, R. (2005). *Task Complexity and L2 Narrative Oral Production*. Dissertation, Barcelona: Universitat de Barcelona.

Gilabert, R. (2007a). The simultaneous manipulation of task complexity along (+/- planning time and here-and-now): Effects on L2 oral production. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44–68). Clevedon, UK: Multilingual Matters.

Gilabert, R. (2007b). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching, 45*, 215-240.

Gilabert, R. & Barón, J. (2013). The impact of increasing task complexity on L2 pragmatic moves. In A. Mackey & K. McDonough (Eds.), *Second language interaction in diverse educational settings* (pp. 45-69). Amsterdam, The Netherlands: John Benjamins.

Gilabert, R., & Malicka, A. (2021). From Needs Analysis to Task Selection, Design, and Sequencing. In M. Ahmadian & M. Long (Eds.), *The Cambridge Handbook of Task-Based Language Teaching* (pp. 226-249). Cambridge, UK: Cambridge University Press.

Gilabert, R., & Muñoz, C. (2010). Differences in attainment and performance in a foreign language: The role of working memory capacity. *International Journal of English Studies*, *10*(1), 19-42.

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, *47*, 367–395.

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in Theoretical and Empirical TBLT Research: Advancing Research Agendas. *Annual Review of Applied Linguistics, 36*, 117-135.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and language*, *109*(2-3), 55-67.

Gómez Lacabex, E., García Lecumberri, M. L., & Martin Cooke, J. (2008). Identification of the contrast full vowel-schwa: Training effects and generalization to a new perceptual context. *Ilha do Desterro, 55*, 173-196.

Gordon, J. (2021). Pronunciation and task-based instruction: Effects of a classroom intervention. *RELC Journal*, *52*(1), 94-109.

Gordon, J. (2022). Making the Teaching of Segmentals Purposeful. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 61–84). Hoboken, NJ: John Wiley & Sons, Inc.

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, *2*(1), 56-92.

Gordon, J. & Darcy, I. (2019). Teaching segmentals vs. suprasegmentals: different effects of explicit instruction on comprehensibility. In J. Levis, C. Nagle & E. Todey

(Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp 116–126). Ames, IA: Iowa State University.

Gordon, J., & Darcy, I. (2022). Teaching segmentals and suprasegmentals: Effects of explicit pronunciation instruction on comprehensibility, fluency, and accentedness. *Journal of Second Language Pronunciation*, *8*(2), 168-195.

Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 194–206). Ames: Iowa State University.

Granena, G. (2016). Elicited imitation as a measure of implicit L2 knowledge: The role of working memory and short-term memory capacity. In G. Granena, D. Jackson & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 185-204). Amsterdam, The Netherlands: John Benjamins.

Grant, L. (2014). Prologue to the myths: What teachers need to know. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 1-33). Ann Arbor: University of Michigan Press.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. J. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 57-77). Amsterdam, The Netherlands: John Benjamins.

Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition*, *36*, 1–37.

Gurzynski-Weiss, L., & IATBLT (n.d.). The TBLT Language Learning Task Bank. https://tblt.indiana.edu/

Gurzynski-Weiss, L., Long, A., & Solon, M. (Eds.). (2017a). TBLT and L2 Pronunciation: Do the Benefits of Tasks Extend Beyond Grammar and Lexis? [Special issue]. *Studies in Second Language Acquisition*, *39*(2).

Gurzynski-Weiss, L., Long, A., & Solon, M. (2017b). TBLT and L2 Pronunciation: Do the Benefits of Tasks Extend Beyond Grammar and Lexis? Studies in Second Language Acquisition, *39*(2), 213-224.

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. *Proceedings of 159th Meeting Acoustical Society of America*, Baltimore, MD, 1–10.

Hancock, M. (2017). *PronPack 3: Pronunciation Pairworks*. Chester: Hancock McDonald ELT.

Hanson, S. (2022, August 29-31). *Technology-mediated TBLT and task complexity: Acquisition of second language Spanish stops /p t k/ and rhotics /r ɾ/* [Paper presentation]. The 9th International Conference on Task-Based Language Teaching, Innsbruck, Austria.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5-33.

Hayes-Harb, R., Brown, K., & Smith, B. L. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech*, *61*(4), 547–564.

Henderson, A., Curnick, L., Frost, D., Kautzsch, A., Kirkova-Naskova, A., Levey, D., Tergujeff, E., & Waniek-Klimczak, E. (2015). The English pronunciation teaching in Europe survey: Factors inside and outside the classroom. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Current trends and directions* (pp. 260–292). Basingstoke, UK/New York, NY: Palgrave Macmillan.

Henrichsen, L., & Stephens, C. (2015). Advanced adult ESL students' perspectives on the benefits of pronunciation instruction. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (pp. 197- 205). Ames, IA: Iowa State University.

Higgins, J. (2017). Minimal pairs for English RP: lists by John Higgins. Retrieved on 15 February 2019 from http://l2phon.lt.cityu.edu.hk/marlodge/min-pairs/minimal.html/

Hillenbrand, J. M., Clark M. J., & Houde R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America, 108*, 3013-3022.

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning, 17*(3-4), 357-376.

Horgues, C., & Scheuer, S. (2014). "I understood you, but there was this pronunciation thing…": L2 pronunciation feedback in English/French tandem interactions. *Research in Language, 12*(2), 145-161.

Housen, A., Kuiken, F. & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). Philadelphia: John Benjamins.

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and language, 127*(3), 366-376.

Huensch, A. (2016). Perceptual phonetic training improves production in larger discourse contexts. *Journal of Second Language Pronunciation, 2*(2), 183-207.

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America, 120*(5), 2926-2937.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475-505.

Isbell, D. R. (2020). Diagnostic language assessment for L2 pronunciation: A worked example. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the*

11th Pronunciation in Second Language Learning and Teaching conference (pp. 127–140). Ames, IA: Iowa State University.

Ishikawa, T. (2008). The effect of task demands of intentional reasoning on L2 speech performance. *The Journal of Asia TEFL, 5,* 29–63.

Iverson, P., and 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*, 2842–2854.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, *63*(2), 330-367.

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, *20*(3), 123–140.

Jenkins, J. (2000). *The Phonology of English as an International Language: New Models, New Norms, New Goals*. Oxford: Oxford University Press.

Jones, T. (Ed.). (2016). *Pronunciation in the classroom: The overlooked essential*. Alexandria, VA: TESOL Press.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language teaching*, *44*(2), 137-166.

Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-stress paterns through auditory priming tasks. *Studies in Second Language Acquisition*, *39*(2), 319–346.

Kabakoff, H., Go, G., & Levi, S. V. (2020). Training a non-native vowel contrast with a distributional learning paradigm results in improved perception and production. *Journal of phonetics*, *78*, 100940.

Kahneman, D. (1973). *Attention and Effort*. Upper Saddle River, NJ: Prentice Hall.

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.

Kartushina, N., Soto, D., & Martin, C. (2022). Metacognition in Second Language Speech Perception and Production. *Language Learning*.

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

Keck, C. M., Iberri-Shea G., Tracy-Ventura N., Wa-Mbaleka S. (2006). Investigating the empirical link between interaction and acquisition: A quantitative meta-analysis. In Ortega L., Norris J. (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). Amsterdam, The Netherlands: John Benjamins.

Kelly, L. G. (1969). *25 centuries of language teaching*. Newbury House.

Kennedy, S., Guénette, D., Murphy, J., & Allard, S. (2015). Le rôle de la prononciation dans l'intercompréhension entre locuteurs de français lingua franca. *Canadian Modern Language Review*, *71*, 1–25.

Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System*, *37*, 254–268.

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition*, *34*, 627–658.

Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. In K. Dziubalska-Kolaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech* (pp. 251-256). Poznań: School of English, Adam Mickiewicz University.

Kim, Y., & McDonough, K. (2008). The effect of interlocutor proficiency on the collaborative dialogue between Korean as a second language learners. *Language teaching research*, *12*(2), 211-234.

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *Modern Language Journal*, 99, 656–677.

Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 287-306). Amsterdam, The Netherlands: John Benjamins.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition*, *37*(3), 549-581.

Kim, Y., Tracy-Ventura, N., & Jung, Y. (2016). A measure of proficiency or short-term memory? Validation of an Elicited Imitation Test for SLA research. *The Modern Language Journal*, *100*(3), 655-673.

Kim, Y., Crossley, S., Jung, Y., Kyle, K., & Kang, S. (2018). The effects of task repetition and task complexity on L2 lexicon use. In M. Bygate (Ed.), *Language learning through task repetition* (pp. 75-96). Amsterdam, The Netherlands: John Benjamins.

Kirkova-Naskova, A., Henderson, A., & Fouz-González, J. (2021). Advancing towards research-informed pronunciation pedagogy. In Kirkova-Naskova, A., Henderson, A. & Fouz, J. (Eds.) *English Pronunciation Instruction: Research-based insights* (pp. 1-13). Amsterdam, The Netherlands: John Benjamins.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The modern language journal*, *97*(3), 720-744.

Kojima, H. (2019). *Lexical encoding of length contrasts in learners of Japanese as a second language*. Dissertation, Bloomington (IN): Indiana University.

Kormos, J. (1999). Monitoring and self-repair in L2. *Language learning*, *49*(2), 303-342.

Kormos, J. (2000). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition*, *22*(2), 145-167.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, *11*(2), 261-271.

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: researching the cognition hypothesis of language learning and performance* (pp. 267-289). Amsterdam, The Netherlands: John Benjamins.

Kostromitina, M., & Kang, O. (2021). The effects of ESL immersion and proficiency on learners' pronunciation development. *Frontiers in Communication*, *6*, 636122.

Krahnke, K. (1987). *Approaches to syllabus design for foreign language teaching*. Englewood Cliffs, NJ: Prentice-Hall.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.

Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 91–104). Amsterdam, The Netherlands: John Benjamins.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39*(1), 167-196.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*. New York: Routledge.

Lee, J. F. (2000). *Tasks and communicating in language classrooms*. Boston, MA: McGraw-Hill

Lee, A. H. (2021). The effects of proactive form-focused instruction and individual differences on second language acquisition. *Language Teaching Research*.

Lee, D. Y., & Baese-Berk, M. M. (2021). Non-native English listeners' adaptation to native English speakers. *JASA Express Letters*, *1*(10), 105201.

Lee, J., Jang, J., & Plonsky, L. (2015). The Effectiveness of Second Language Pronunciation Instruction: A Meta-analysis. *Applied Linguistics, 36*, 345-366.

Lennon, P. (1984). Retelling a story in English as a second language. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 50-68). Tübingen: Günter Narr.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in second language acquisition*, *22*(4), 557-584.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach.* New York, NY: Routledge.

Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, *47*, 36-46.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

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

Levis, J. M. (2018). Technology and second language pronunciation. *Journal of Second Language Pronunciation*, *4*(2), 173-181.

Levis, J. M. (2022). Teaching pronunciation: Truths and lies. In C. Bardel, C. Hedman, K. Rejman, & E. Zetterholm (Eds.), *Exploring Language Education: Global and local perspectives*, (pp. 39–72). Stockholm: Stockholm University Press.

Levis, J. M., & Echelberger, A. (2022). Integrating Pronunciation into Language Instruction. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 19-41). Hoboken, NJ: John Wiley & Sons, Inc.

Levis, J. M., & Sonsaat, S. (2016). Pronunciation materials. In M. Azarnoosh, M. Zeraatpishe, A. Faravani, & et al. (Eds.), *Issues in Materials Development* (pp. 109-119). Rotterdam, The Netherlands: Sense Publishers.

Levis, J. M., & Sonsaat, S. (2017). Pronunciation teaching in the early CLT era. In O. Kang, R. Thomson & J. Murphy (Eds.), *The Routledge handbook of English pronunciation*, (pp. 267-283). London: Routledge.

Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In Housen, A., Kuiken, F., & Vedder, I. (Eds.), *Dimensions of L2* *performance and proficiency: investigating complexity, accuracy, and fluency in SLA* (pp. 171–198). Amsterdam, The Netherlands: John Benjamins.

Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in System. *System, 84,* 93-109.

Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an Alien World: Videogame training improves non-native speech categorization. *Cognitive science*, *35*(7), 1390-1405.

Llompart, M. (2021a). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition*, *24,* 481–496.

Llompart, M. (2021b). Lexical and Phonetic Influences on the Phonolexical Encoding of Difficult Second-Language Contrasts: Insights from Nonword Rejection. *Frontiers in. Psychology*, *12*:659852.

Llompart, M., & Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*, *22*(5), 1085-1100.

Llompart, M., & Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(8), 1590–1610.

Llompart, M., & Reinisch, E. (2021). Lexical representations can rapidly be updated in the early stages of second-language word learning. *Journal of Phonetics, 88*:101080.

Loewen, S., & Isbell, D. R. (2017). Pronunciation in face-to-face and audio-only synchronous computer-mediated learner interactions. *Studies in Second Language Acquisition*, *39*(2), 225-256.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language training. In K. Hyltenstam and M. Pienemann (Eds.), *Modelling and* *assessing second language acquisition* (pp. 77-99). Clevedon: Multilingual Matters.

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

Long, M. H. (1998). Focus on form in task-based language teaching. *University of Hawai'i Working Papers in ESL, 16*(2), 35-49.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. West Sussex, UK: Wiley Blackwell.

Long, M. H. (2016). In defence of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics, 36*, 5–33.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge, UK: Cambridge University Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes and S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, UK: Multilingual Matters.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach* (Vol. 18). Amsterdam, The Netherlands: John Benjamins.

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

MacKay, I., & Fullana, N. (2009). Starting age and exposure effects on EFL learners' sound production in a formal learning context. In M. Watkins, A. Rauber, & B. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 43-61). Newcastle upon Tyne: Cambridge Scholars Publishing.

Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning.* Oxford: Oxford University Press.

Mackey, A. & Gass, S. (2016). *Second Language Research: Methodology and Design*. 2nd Ed. New York: Routledge.

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language learning*, *60*(3), 501-533.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, *2*(1), 49– 55.

Márquez, D., & Barón, J. (2021). Effects of task complexity on L2 suggestions: An exploratory study on trade-offs between accuracy and complexity. *TASK*, *1*(2), 227-265.

Martin, I. A., & Sippel, L. (2021). Is giving better than receiving?: The effects of peer and teacher feedback on L2 pronunciation skills. *Journal of Second Language Pronunciation*, *7*(1), 62-88.

Martin, I. A., & Inceoglu, S. (2022). The laboratory, the classroom, and online: What works in each context. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 254–272). Hoboken, NJ: John Wiley & Sons, Inc.

Matsumoto, Y. (2011). Successful ELF communications and implications for ELT: Sequential analysis of ELF pronunciation negotiation strategies. *Modern Language Journal*, *95*(1), 97–114.

McAndrews, M. M., & Thomson, R. I. (2017). Establishing an empirical basis for priorities in pronunciation teaching. *Journal of Second Language Pronunciation*, *3*(2), 267-287.

McDougall, K., & Nolan, F. (2007). Discrimination of speakers using the formant dynamics of/uː/in British English. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1825-1828).

McManus, K., & Liu, Y. (2022). Using elicited imitation to measure global oral proficiency in SLA research: A close replication study. *Language Teaching*, *55*(1), 116-135.

Melnik-Leroy, G.A., & Peperkamp. S. (2021). High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Bilingualism: Language and Cognition, 24*(3), 497–506.

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). 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.

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 141–173). Amsterdam, The Netherlands: John Benjamins.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*, 241–259.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*(1), 8-14.

Montero Pérez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749-773.

Mora, J. C. (2021). Assessing L2 vowel production gains after high-variability phonetic training: acoustic measurements vs. perceptual judgements. In the *Proceedings of the 3rd International Symposium on Applied Phonetics (ISAPh 2021)* (pp. 9-18).

Mora, J. C. (2022). Aptitude and Individual Differences. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *Routledge Handbook of Second Language Acquisition and Speaking* (pp. 68-82). London: Routledge.

Mora, J. C. & Darcy, I. (2016). The relationship between cognitive control and pronunciation in a second language. In T. Isaacs & P. Trofimovich (Ed.), *Second Language Pronunciation Assessment: Interdisciplinary Perspectives* (pp. 95-120). Bristol, UK: Multilingual Matters.

Mora, J. C., & Fullana, N. (2007, August). Production and perception of English /i:/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. In *Proceedings of the 16th international congress of phonetic sciences* (Vol. 3, pp. 1613-1616). Saarbrücken: Universität des Saarlandes.

Mora, J. C., & Levkina, M. (2017). Task-Based Pronunciation Teaching and Research: Key Issues and Future Directions. *Studies in Second Language Acquisition, 39*(2), 381-399.

Mora, J. C. & Levkina, M. (2018). Training vowel perception through map tasks: The role of linguistic and cognitive complexity. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 151- 162). Ames, IA: Iowa State University.

Mora, J. C. & Mora-Plaza, I. (2019) Contributions of cognitive attention control to L2 speech learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters—In honor of Ocke-Schwen Bohn* (pp. 477-499). Department of English, School of Communication & Culture, Aarhus University.

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Improving the pronunciation of L2 vowels under different training conditions: the use of non-lexical materials and masking noise. *Phonetica, 79*(1), 1-43.

Mora-Plaza, I., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis, C. Nagle & E. Today (Eds.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp.174-184). Ames, IA: Iowa State University.

Mora-Plaza, I., Ortega, M., & Mora, J. C. (2022a). High-Variability Phonetic Training under Different Conditions: Individual Differences in Auditory Attention Control. In V. G. Sardegna, & A. Jarosz (Eds.) *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*: *Studies in Honour of Ewa Waniek-Klimczak* (pp. 241-260). Springer, Cham.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J-M., & Tierney, A. (2022b). *Tools for second language speech research and teaching*. http://sla-speech-tools.com.

Mott, B. (2011) *English Phonetics and Phonology for Spanish Speakers*. Barcelona, Publicacions i Edicions Universitat de Barcelona, 2nd Ed.

Moyer, A. (2014). The social nature of second language pronunciation. In J. M. Levis & A. Moyer (Eds.), *Social dynamics in second language acquisition* (pp. 11– 29). Boston: De Gruyter.

Mozaffari, S. H. (2017). Comparing student-selected and teacher-assigned pairs on collaborative writing. *Language Teaching Research*, *21*(4), 496-516.

Munro, M. J., & Derwing, T. M. (1995). Foreign Accent, Comprehensibility and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *49*(1), 285-310.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520-531.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language learning*, *58*(3), 479-502.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*, 463–482.

Murphy, J. (2014). Teacher training programs provide adequate preparation in how to teach pronunciation. In L. Grant (Ed.), *Pronunciation myths: applying second language research to classroom teaching* (pp. 188-224). University of Michigan Press.

Murphy, J., & Baker, A. (2015). History of ESL pronunciation teaching. In J. Levis and M. Reed (Eds), *Wiley-Blackwell handbook of English pronunciation* (pp. 36–65). Chichester: Blackwell.

Nagle, C. L. (2021). Revisiting perception–production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Language Learning*, *71*(1), 243–279.

Nagle, C. L. (2022). Rethinking pronunciation posttesting. *Journal of Second Language Pronunciation*, *8*(2), 161-167.

Newton, J. (2018). Pronunciation and speaking. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (pp. 337-351). New York/London: Routledge.

Newton, J., & Kennedy, G. (1996). Effects of communication tasks on the grammatical relations marked by second language learners. *System*, *24*(3), 309-322.

Nguyen, L. T., Hung, B. P., Duong, U. T. T., & Le, T. T. (2021). Teachers' and learners' beliefs about pronunciation instruction in tertiary English as a foreign language education. *Frontiers in Psychology, 12*.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, *50*(3), 417-528.

Nuevo, A.-M., Adams, R., & Ross-Feldman, L. (2011). Task complexity, modified output, and L2 development in learner-learner interaction. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 176–201). Amsterdam, The Netherlands: John Benjamins.

Nunan, D. (1989). *Designing Tasks for the Communicative Classroom*. Cambridge, UK: Cambridge University Press.

Nunan, D. (2004). *Task-based language teaching*. Cambridge, UK: Cambridge University Press.

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

Ortega, L., Iwashita, N., Norris, J. & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*, Toronto.

Ortega, M., Mora-Plaza, I., & Mora, J. C. (2021). Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In Kirkova-Naskova, A., Henderson, A., & Fouz-González, J. (Eds.), *English pronunciation instruction: Research-based insights*. Amsterdam, The Netherlands: John Benjamins.

Ota, M., Hartsuiker, R.J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, *111*, 263-269.

Ou, J., Law, S. P., & Fung, R. (2015). Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review, 22*(6), 1725-1732.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based vs. abstract lexical entries. *Psychological Science, 12*(6), 445-449.

Parlak, Ö., & Ziegler, N. (2017). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition*, *39*(2), 257-285.

Park, J. K. (2000). *The effects of forms and meaning-focused instruction on ESL learners' phonological acquisition* (Unpublished doctoral dissertation). Philadelphia: University of Pennsylvania.

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, *127*(2), EL54-EL59.

Pennington, M.C. & Rogerson-Revell, P. (2019). *English Pronunciation Teaching and Research: Contemporary Perspectives.* London: Palgrave Macmillan.

Perdue, C. (Ed.). (1993). *Adult language acquisition: Cross-linguistic perspectives: Vol. 2. The results*. Cambridge, UK: Cambridge University Press.

Pica, T. (1984). An Accent on Communication. *A Journal for the Teacher of English Outside the United States*, *22*, 2.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language learning*, *44*(3), 493-527.

Pica, T., Kanagy, R. & Falodun, J. (1993). Choosing and Using Communicative Tasks for Second Language Instruction. In G. Crookes, & S. Gass (Eds.), *Tasks and Language Learning: Integrating Theory and Practice* (pp. 9-34). Clevedon, UK: Multilingual Matters.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics, 29*(2), 191-215.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 221–231.

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. *Journal of Phonetics, 39,* 467–478.

Prabhu, N. S. (1987). *Second language pedagogy* (Vol. 20). Oxford: Oxford university press.

Prieto, P. (2004). *Fonètica i fonologia: els sons del català*. Editorial UOC.

Rallo Fabra, L. (2005). Predicting ease of acquisition of L2 speech sounds. A perceived dissimilarity test. *Vigo International Journal of Applied Linguistics,* (2), 75-92.

Rallo Fabra, L. (2022). Factors Affecting Pronunciation Accuracy in English as a Foreign Language: The Case of Spanish-Catalan Intermediate Learners. *Atlantis, 44*(2), 45-64.

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics, 40,* 491–508.

Ranta, L., & Lyster, R. (2018). Form-focused instruction. In P. Garrett & J. M. Vots (Eds.), *The Routledge handbook of language awareness* (pp. 40– 56). New York: Routledge.

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Concordia Working Papers in Applied Linguistic*, *5,* 529-546.

Rato, A., & Rauber, A. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In *Proceedings of ICPhS 2015*. Glasgow, UK: University of Glasgow.

Recasens, D. (1993). *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition, 31,* 437–470.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*, 162-181.

Révész, A., & Gurzynski-Weiss, L. (2016). Teachers' Perspectives on Second Language Task Difficulty: Insights From Think-Alouds and Eye Tracking. *Annual Review of Applied Linguistics*, *36*, 182-204.

Révész, A., & Han, Z. (2006). Task content familiarity, task type and efficacy of recasts. *Language Awareness*, *15*, 160–179

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A Validation Study. *Studies in Second Language Acquisition, 38*(4), 703-737.

Révész, A., Sachs, R., & Mackey, A. (2011). Task complexity, uptake of recasts, and L2 development. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 203–235). Amsterdam, The Netherlands: John Benjamins.

Révész, A., Jeong, H., Suzuki, S., Cui, H., Matsuura, S., Saito K., & Sugiura, M. (2022, August 29-31). *Exploring the neurocognitive correlates of task complexity during speech production* [Paper presentation]. The 9th International Conference on Task-Based Language Teaching, Innsbruck, Austria.

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

Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57.

Robinson, P. (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge, UK: Cambridge University Press.

Robinson, P. (2003a). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, *21*(2), 45-105.

Robinson, P. (2003b). Attention and memory. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631-678). Oxford, UK: Blackwell.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*, 1-32.

Robinson, P. (2007a). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Clevedon, UK: Multilingual Matters.

Robinson, P. (2007b). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, *45*, 193-213.

Robinson, P. (2010). Situation and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz, & L. Sicola, (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243-268). Amsterdam, The Netherlands: John Benjamins.

Robinson, P. (2011a). Task-based language learning: A review of issues. *Language learning*, *61*, 1-36.

Robinson, P. (2011b). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3-37). Amsterdam, The Netherlands: John Benjamins.

Robinson, P., & Gilabert, R. (Eds.). (2007). Task complexity, the Cognition Hypothesis and second language instruction [Special issue]. *International Review of Applied Linguistics*, *45*(3).

Róg, T. (2021). The impact of task complexity and task repetition on L2 lexical complexity. *Konin Language Studies*, *9*(4), 409-433.

Rogers, J., & Cheung, A. (2021). Does it matter when you review?: Input spacing, ecological validity, and the learning of L2 vocabulary. *Studies in Second Language Acquisition*, *43*(5), 1138-1156.

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, *52*(1), 189-205.

Ruan, Y., & Saito, K. (2023). Communicative focus on phonetic form revisited: Less precise auditory processing limits instructed L2 speech learning. *System*, 103020.

Safronova, E. (2016). *The role of cognitive ability in the acquisition of second language perceptual phonological competence*. (Unpublished doctoral dissertation). Barcelona: Universitat de Barcelona.

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

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

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in second language acquisition, 35*(1), 1-29.

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English/ɹ/without explicit instruction. *Applied Psycholinguistics*, *36*(2), 377-409.

Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English /ɹ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics, 40*(1), 183-204.

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing Auditory Precision Hypothesis-L2. *Language Teaching,* 1-17.

Saito, K. & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics, 37,* 813–840.

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

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

Saito, K., & Tierney, A. (2022). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 1-25.

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*(1), 201-225.

Saito, K., Dewaele, J. M., Abe, M. & In'nami. Y. (2018). Motivation, Emotion, Learning Experience, and Second Language Comprehensibility Development in Classroom Settings: A Cross-Sectional and Longitudinal Study. *Language Learning, 68*, 709–743.

Saito, K., Suzuki, S., Oyama, T., & Akiyama, Y. (2021a). How does longitudinal interaction promote second language speech learning? Roles of learner experience and proficiency levels. *Second Language Research, 37*(4), 547-571.

Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021b). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. *Language Learning*, *71*(3), 669-715.

Saito, K., Macmillan, K., Kachlicka, M., Kunihara, T., & Minematsu, N. (2022a). Automated assessment of second language comprehensibility: Review, training, validation, and generalization studies. *Studies in Second Language Acquisition*, *45*(1), 234-263.

Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (2022b). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(12), 1410.

Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022c). Incidental and multimodal high variability phonetic training: Potential, limits, and future directions. *Language Learning*, *72*(4), 1049-1091.

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.

Samuda, V. & Bygate, M. (2008). *Tasks in second language learning*. Basingstoke: Palgrave MacMillan.

Sardegna, V. G. (2022). Evidence in favor of a strategy-based model for English pronunciation instruction. *Language Teaching*, *55*(3), 363-378.

Sardegna, V. G., & Jarosz, A. (2022). Exploring how YouGlish supports learning English word stress: A perception study. In V. G. Sardegna, & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training: Studies in honour of Ewa Waniek-Klimczak* (pp. 165-184). Cham: Springer International Publishing.

Sardegna, V. G., & McGregor, A. (2022). Classroom research for pronunciation. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 107–128). Hoboken, New Jersey: John Wiley & Sons.

Scheuer, S., & Horgues, C. (2021). Corrective feedback and unintelligibility. In Kirkova-Naskova, A., Henderson, A. & Fouz, J. (Eds.) *English Pronunciation Instruction: Research-based insights* (pp. 223-252). Amsterdam, The Netherlands: John Benjamins.

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

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & cognition*, *33*(4), 644-653.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). Oxford: Oxford University Press.

Sewell, A. (2017). Functional load revisited: Reinterpreting the findings of “lingua franca” intelligibility studies. *Journal of Second Language Pronunciation*, *3*(1), 57-79.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied psycholinguistics*, *3*(3), 243-261.

Sicola, L. (2008). "*No, they won't 'just sound like each other'": NNS-NNS negotiated interaction and attention to phonological form on targeted L2 pronunciation tasks* (Unpublished doctoral dissertation). Philadelphia, PA: University of Pennsylvania.

Simonchyk, A., & Darcy, I. (2021). Development of production skills in the absence of precise phonolexical representations. *Second Language Research.*

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*, *41*(4), 795-820.

Skehan, P. (1988). Language testing. Part 1. *Language teaching*, *21*(4), 211-221.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. (2007). Language Instruction Through Tasks. In J. Cummins, J. & C. Davison (Eds.) *International Handbook of English Language Teaching. Springer International Handbooks of Education* (pp. 289–301). Springer, Boston, MA.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge, UK: Cambridge University Press.

Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 199-220). Philadelphia: John Benjamins.

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, *39*(2), 347-380.

Solon, M., Park, H. I., Henderson, C., & Dehghan-Chaleshtori, M. (2019). Revisiting the Spanish elicited imitation task: A tool for assessing advanced language learners? *Studies in Second Language Acquisition*, *41*(5), 1027-1053.

Sonsaat-Hegelheimer, S., & McCrocklin, S. (2022). Informed Materials for Pronunciation Teaching. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 293-313). Hoboken, New Jersey: John Wiley & Sons.

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

Sudharshana, N. P. (2021). From Cognitive Grammar to Pedagogic Grammar: Macrostrategies for Designing Form-Focused Tasks. In N. P. Sudharshana, & L. Mukhopadhyay (Eds.), *Task-Based Language Teaching and Assessment: Contemporary Reflections from Across the World* (pp. 163-181). Singapore: Springer.

Sudharshana, N. P., & Mukhopadhyay, L. (2021). *Task-Based Language Teaching and Assessment: Contemporary Reflections from Across the World*. Singapore: Springer.

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical implications. *RELC Journal, 52*, 48–61.

Suzukida, Y., & Saito, K. (2022). What is second language pronunciation proficiency? An empirical study. *System*, *106*, 102754.

Suzukida, Y., & Saito, K. (2023). Detangling experiential, cognitive, and sociopsychological individual differences in second language speech learning: Cross-sectional and longitudinal investigations. *Bilingualism: Language and Cognition*, 1-14.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input and Second Language Acquisition* (pp. 235–53). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.) *Principle and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson* (pp. 125-44). Oxford: Oxford University Press.

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

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99–118). London, UK: Longman.

Swales, J. (1990). *English in academic and research settings*. Cambridge, UK: Cambridge University Press.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America, 79*(4), 1086-1100.

Talmy, L. (2000). *Toward a Cognitive Semantics.* Cambridge, MA: MIT Press.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowels: Perception improves pronunciation, *CALICO Journal, 28*(3), 744–765.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4),1231–1258.

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.

Thomson, R. I. (2022). The relationship between L2 speech perception and production. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.) *The Routledge Handbook of Second Language Acquisition and Speaking* (p.14). London: Routledge.

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

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds.). *Proceedings of the 7th Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 88–97). Ames, IA: Iowa State University.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in second language acquisition*, *16*(2), 183-203.

Tragant, E. (2009). Trilingualism in Catalan schools and universities? Not yet. *APAC Quarterly Magazine, 65*, 33-38.

Tragant, E., Serrrano R., Miralpeix I., Navés T., Pahissa I., Serra N. & Gilabert R. (2010). *Estudi de sis instituts amb resultats destacables en la prova d'anglès de les PAU*. Universitat de Barcelona.

Traunmüller, H. (1997). Perception of speaker sex, age, and vocal effort. *Phonum*, *4*, 183-186.

Trofimovich, P. (2008). What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research, 37*, 309-329.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal*, *90*(4), 519-535.

Trofimovich, P., Ammar, A., & Gatbonton. E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171 – 195). Oxford: Oxford University Press.

Trofimovich, P., McDonough, K., & Foote, J. A. (2014). Interactive alignment of multisyllabic stress patterns in a second language classroom. *TESOL Quarterly*, *48*(4), 815-832.

Turner, J. (2022). Analysing the relationship between L2 production and different stages of L2 processing: Eye-tracking and acoustic evidence for a novel contrast. *Journal of Phonetics, 91*, 101134.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters—In honor of Ocke-Schwen Bohn* (pp. 607–630). Department of English, School of Communication & Culture, Aarhus University.

Tyler, M. D. (2021). Phonetic and phonological influences on the discrimination of non-native phones. In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 157-174). Cambridge, UK: Cambridge University Press.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4-21.

Uchida, Y., & Sugimoto, J. (2018). A survey of pronunciation instruction by Japanese teachers of English: Phonetic knowledge and teaching practice. *Journal of the Tokyo University of Marine Science and Technology*, *14*, 65–75.

Van de Ven, M., Segers, E., & Verhoeven, L. (2019). Enhanced second language vocabulary learning through phonological specificity training in adolescents. *Language Learning*, *69*(1), 222-250.

Van den Branden, K. (2006). Task-based language education: From theory to practice. *Language*, *53*(82), T35.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176-1190.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of memory and language*, *50*(1), 1-25.

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics*, *39*(4), 703-728.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35–44.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13-40.

Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Human skills* (pp. 71–105). John Wiley & Sons.

Willis, J. (1996). *A Framework for Task-Based Learning*. Harlow: Longman.

Willis, J. (2021). An evolution of a framework for TBLT: What trainers and teachers need to know to help learners succeed in task-based learning. In N. P. Sudharshana, & L. Mukhopadhyay (Eds.), *Task-based language teaching and assessment: Contemporary reflections from across the world* (pp. 63-92). Singapore: Springer.

Willis, D. & Willis, J. (2007). *Doing Task-based Teaching*. Oxford University Press.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, *42*(3), 599-624.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology, 33*, 101–111.

Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign language annals, 46*(4), 680-704.

Wu, S. L., Tio, Y. P., & Ortega, L. (2022). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, *44*(1), 271-300.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*(4), 497-528.

Yeldham, M., & Choy, V. (2022). The effectiveness of direct articulatory–abdominal pronunciation instruction for English learners in Hong Kong. *Language, Culture and Curriculum*, *35*(2), 184-199.

Zielinski, B., & Yates, L. (2014). Pronunciation instruction is not appropriate for beginning-level learners. In L. Grant (Ed.), *Pronunciation myths: applying second language research to classroom teaching* (pp. 56-79). Ann Arbor, MI: University of Michigan Press.

APPENDIX

Appendix A. Information sheets and consent forms

A.1. Teachers

ACTIVITATS ORALS PER AL DESENVOLUPAMENT DE LA PRONUNCIACIÓ EN ANGLÈS[1]

La idea d'aquest projecte sorgeix per la necessitat d'augmentar el nombre de pràctiques d'expressió oral a les classes de secundària de Catalunya. En el marc del projecte RecerCaixa, aquest treball pretén promoure la millora de la pronunciació de l'anglès mitjançant tasques pedagògiques comunicatives que segueixen la línia de les activitats orals que es fan a les escoles i que han estat testades científicament i presentades a congressos internacionals.

El projecte es durà a terme amb la investigadora Íngrid Mora i, de forma puntual, les investigadores Miren Adrián, Daniela Avello i Valeria Galimberti ajudaran a l'Íngrid Mora durant les proves d'avaluació.

Els alumnes que facin les activitats orals durant un dels tres trimestres faran uns tests d'avaluació abans i després del període de pràctica oral per tal d'avaluar l'eficàcia de les esmentades activitats. Una vegada finalitzat l'estudi, els professors rebran els materials que s'han utilitzat durant les classes d'oralitat, i que prèviament hauran treballat en el curs de formació del Juliol 2019, per tal de poder fer-ne ús a classe.

Abans d'establir la planificació de les tasques i dades, convindria parlar sobre la disponibilitat d'horaris que podria oferir el centre per la posada en pràctica de les activitats orals. Una possibilitat és dedicar una hora i mitja d'anglès a la setmana a aquestes activitats durant 1 mes i mig. Tot seguit mostro un exemple de planificació de les tasques en el període de setembre de 2019 a febrer de 2020 pels grups experimentals. El grup control faria el mateix sense fer les tasques indicades en taronja al calendari.

ESTUDI EXPERIMENTAL: 3 grups de Batxillerat -*curs 2019-2020*- (90-100 alumnes)

TASQUES	*DURADA*	*PERIODE*	*ESPAIS i DISTRIBUCIÓ*
PRESENTACIÓ -Presentació del projecte als grups de 1r de BTX -Consentiment i recollida de dades personals	30 min	*1BA & 1BB*: Setmana del 16/09-20/09 *1BC*: Setmana del 16/09-20/09	A classe
QUESTIONARIS ONLINE -Qüestionari demogràfic/lingüístic -Qüestionari sobre coneixement de vocabulari (VKS)	15-20 min 10 min 10 min	*1BA & 1BB*: Setmana del 16/09-20/09 *1BC*: Setmana del 16/09-20/09	A casa

[1] Aquestes activitats s'emmarquen dins d'un projecte de tesi que ha estat aprovat per un comitè internacional.

PRE-TEST (PROVA 1) -Tasca interactiva -Prova de percepció ABX -Prova de lèxic FLEC i LD -Prova de producció de paraules DWR -Prova de producció de frases DSR -Prova de proficiència oral	60 min x grups de 8 5-7 min 10 min 8-10 min 5-7 min 10 min 7 min	*1BA*: Setmana del 23/09-04/10 *1BB*: Setmana del 23/09-04/10 *1BC:* Setmana del 23/09-04/10 o 11/10	Aula on pugui col·locar 8 ordinadors portàtils.
TASQUES ORALS -Pre-tasca -Tasca simple/complexa -Post-tasca -Qüestionari de d'aprenentatge paraules i valoració tasques	60 min x classe 10 min 20 min 20 min 10 min	*3 dies x setmana (30 min)* *1BA*: Setmana del 30/09 o 07/10 – 05/12 *1BB*: Setmana del 30/09 o 07/10 – 05/12 *1BC*: Setmana del 07/01 – 07/02	A classe. En parelles A casa
POST-TEST (PROVA 2) -Tasca interactiva -Prova de percepció ABX -Prova de lèxic FLEC i LD -Prova de producció de paraules DWR -Prova de producció de frases DSR -Prova d'atenció	60 min x grups de 8 5-7 min 10 min 8-10 min 5-7 min 10 min 7-9 min	*1BA*: Setmana del 02/12-20/12 *1BB*: Setmana del 02/12-20/12 *1BC*: Setmana del 02/12-20/12	Aula on pugui col·locar 8 ordinadors portàtils.
DELAYED POST-TEST (PROVA 3) -Tasca interactiva -Prova de percepció ABX -Prova de lèxic FLEC i LD -Prova de producció de paraules DWR -Prova de producció de frases DSR -Prova de memòria operativa	60 min x grups de 8 5-7 min 10 min 8-10 min 5-7 min 10 min 10 min	*1BA*: Setmana del 10/02-28/02 *1BB*: Setmana del 10/02-28/02	Aula on pugui col·locar 8 ordinadors portàtils.

LLEGENDA:

- Presentació
- Prova 1
- Tasques orals
- Prova 2
- Prova 3

CONFIDENCIALITAT

Tota la informació que recollim dels estudiants es tractarà amb plena confidencialitat. Les dades seran totalment anònimes i es guardaran en un lloc segur. La identitat dels estudiants mai es revelarà en els llocs on es publiquin les dades de l'estudi. Únicament la investigadora principal d'aquest estudi podrà tenir accés als enregistraments de veu. S'assignarà un codi a cada estudiant perquè el seu nom no es pugui identificar amb cap enregistrament ni resultats dels tests.

COMPENSACIÓ PER LA PARTICIPACIÓ

Els estudiants rebran un 15% de la nota global de l'assignatura d'anglès en el primer trimestre si completen totes les proves d'avaluació (pre-test i post-test) i formen part de la intervenció pedagògica. En el segon trimestre, rebran un 5% de la nota global de l'assignatura d'anglès si participen en la prova d'avaluació final al febrer de 2020 (delayed post-test).

CONTACTE EN CAS DE PREGUNTES O PROBLEMES

Si necessiteu contactar per parlar sobre l'estudi o teniu qualsevol dubte o pregunta, poseu-vos en contacte amb l'Íngrid Mora a imoraplaza@ub.edu

NATURALESA VOLUNTÀRIA DE L'ESTUDI

La participació d'aquest estudi és totalment voluntària. Qualsevol estudiant pot decidir no formar part de l'estudi en qualsevol moment. Tanmateix, els pares/tutors poden no autoritzar la participació dels seus fills/es. Això no pot afectar de cap manera la relació amb la investigadora o els professors.

Nom i cognoms del director en representació dels mestres d'anglès:	
Estic d'acord que l'escola participi en aquest projecte d'investigació	
	Signatura

A.2. Parents

Barcelona, 16 de setembre de 2019

Benvolgudes famílies,

Em dic Ingrid Mora i soc estudiant de doctorat del programa *Ciència Cognitiva i Llenguatge* de la Universitat de Barcelona. Estem duent a terme el projecte RecerCaixa "Habilitats orals per a les professions del futur: un programa d'intervenció en el currículum de secundària i batxillerat a Catalunya". Aquest té com a objectiu general sensibilitzar la comunitat educativa de la importància de potenciar les habilitats orals i alhora dotar-la d'un programa d'intervenció pedagògica que integri de forma transversal els coneixements científics més recents sobre tècniques d'ensenyament i millora de les habilitats orals. El curs conté una sèrie de tallers dissenyats per un equip de professionals per a treballar de forma transversal les habilitats essencials de la comunicació oral a l'aula o de forma individual.

Degut a que aquestes habilitats en llengua estrangera no es treballen suficientment, la meva recerca se centra en l'elaboració de tasques comunicatives per millorar l'expressió oral i, en concret, la pronunciació de l'anglès a 1r de Batxillerat. Amb aquesta finalitat, els/les vostres fills/es duran a terme unes proves individuals d'avaluació i unes tasques comunicatives per treballar la pronunciació de l'anglès a classe durant el primer trimestre (30 minuts, 3 vegades a la setmana). El/la vostre/a fill/a podrà decidir no contestar a les preguntes o aturar la seva participació en qualsevol moment. Finalment, signaran un consentiment a classe conforme estan d'acord que la seva veu sigui enregistrada i les seves dades recollides de manera anònima i privada. Si desitgen que el seu fill/a no participi a l'estudi, si us plau, prego poseu-vos en contacte amb mi a través del correu imoraplaza@ub.edu.

Moltes gràcies.

Quedo a la vostra disposició.

Ingrid Mora Plaza
Departament de Llengües i Literatures Modernes i d'Estudis Anglesos
Facultat de Filologia i Comunicació, Universitat de Barcelona
Gran Via de les Corts Catalanes, 585 - 08007 Barcelona

A.3. Students

Si us plau llegeix les següent frases i fes un tic  si hi estàs d'acord

	He escoltat i entès l'explicació del projecte. Sé que puc deixar de participar-hi en qualsevol moment.	
	Estic d'acord a participar en les sessions d'expressió oral i a ser avaluat al final.	
	Estic d'acord que la meva veu sigui enregistrada.	
	Estic d'acord a ser fotografiat.	
	Estic d'acord que la informació que doni sigui compartida només amb l'equip del projecte	

____

Nom i cognoms

Signatura

____

Data

Appendix B. Learner demographic and linguistic information

Table B.1. Participant demographic information (experimental [simple+complex], control). *M= Mean, SD= Standard deviation, RH= Right-handed, LH= Left-handed, AD= Ambidextrous, Dys= Dyslexic, HI= Hearing impairment.*

***N*=92** *(47 male, 45 female)*	**Experimental** (*N*=63)			**Control** (*N*=29)		
Sex (N)	Male=33 (52.4%)	Female=30 (47.6%)		Male=14 (48,3%)	Female=15 (51,7%)	
Age at testing (M) [16-17]	*M*=16.05	*SD*=.215	*Range [16-17]*	*M*=16.07	SD=.258	*Range [16-17]*
Age at testing (N)	16 → 60 (95.2%)	17 → 3 (4.8%)		16 → 27 (93,1%)	17→ 2 (6,9%)	
Hand-dominance (N)	RH=53 (84.1%)	LH=10 (15.9%)	AD=0 (0%)	RH=23 (79,3%)	LH=5 (17.2%)	AD=1 (3.4%)
Pathologies	NO=58 (92.1%)	YES=5 (7.9%)		NO=29 (100%)		
Pathology type	Dys=2 (40%)	HI=3 (60%)		-	-	

Table B.2. Participant linguistic information (experimental [simple+complex], control). *M= Mean, SD= Standard deviation, AoL= Age of learning, Cat= Catalan, Sp= Spanish, Ar= Arabian, Ge= German, Fr= French, It= Italian, Jap= Japanese, Ru= Russian, FCE= English First Certificate, RP= Received Pronunciation, GA= General American.*

***N*=92**	**Experimental** (*N*=63)			**Control** (*N*=29)		
Mother tongue (N)	Cat/Sp=27 (42.9%)	Cat=25 (39.7%)	Sp=11 (17.5%)	Cat/Sp=7 (24,1%)	Cat=11 (37.9%)	Sp= 9 (31.0%) Sp/Ar=2 (6,9%)
L1-dominance (N)	Cat=45 (71.4%)	Sp=18 (28.6%)		Cat=20 (69.0%)	Sp=9 (31.0%)	
L1 daily use (%)	Cat. *M*=62.94/ *SD*=21.32	Sp= *M*=31.22/ *SD*=20.25	Cat range [1-100] Sp range [0-94]	Cat. *M*=61.59/ *SD*=29.52	Sp. *M*=37.86/ *SD*=27.36	Cat range [5-100] Sp range [2-100]
L2 (N)	En=63 (100%)			En=29 (100%)		
L2 AoL (M)	*M*=5.63	*SD*=1.91	Range [3-9]	*M*=6.14	*SD*=1.62	Range [3-10]
L3 (N)	Ge=23 (71.9%)	Fr=8 (25.0%)	It=1 (3.1%)	Ge=13 (68.4%)	Fr=5 (26.3%)	Ar=1 (5.3%)
L3 AoL (M)	*M*=10.88	*SD*=2.47	Range [4-14]	*M*=11.79	*SD*=1.84	Range [5-14]
L4 (N)	It=1 (33.3%)	Jap=1 (33.3%)	Ru=1 (33.3%)	Jap=1 (100%)		
L4 AoL (M)	*M*=14.00	*SD*=1.73	Range [12-15]	*M*=12.00	*SD*= 0	Range [12]
L2 Instr. (years)	*M*=10.37	*SD*=1.91	Range [7-13]	*M*=9.86	*SD*=1.62	Range [6-13]
L2 School exposure (h/week)	*M*=3.46	*SD*=1.06	Range [2-8]	*M*=3.31	*SD*=.541	Range [2-4]
L2 Academy exposure (h/week)	*M*=1.39	*SD*=1.25	Range [0-4]	*M*=1.14	*SD*=1.19	Range [0-3]
L2 certificates (N)	NO=52 (82.5%)	FCE=11 (17.5%)		NO=28 (96.6%)	FCE=1 (3.4%)	
L2 use NNS (h/week)	*M*=3.46	*SD*=3.12	Range [0-14]	*M*=2.69	*SD*=1.53	Range [0-6]
L2 use NS	*M*=1.37	*SD*=1.75	Range [0-7]	*M*=1.10	*SD*=1.37	Range [0-5]

(h/week)						
British exposure	*M*=51.75%	*SD*=26.73%	Range [0-100]	*M*=54.48%	*SD*=25.15%	Range [0-100]
American exposure	*M*=48.25%	*SD*=26.73%	Range [0-100]	*M*=45.52%	*SD*=25.15%	Range [0-100]
Accent production	RP=33 (52.4%)	GA=30 (47.6%)		RP=19 (65.5%)	GA=10 (34.5%)	
Spoken L2 INPUT	*M*=3.57	*SD*=1.04	Range [1-5]	*M*=3.50	*SD*=1.03	Range [1-5]
Written L2 INPUT	*M*=1.96	*SD*=0.96	Range [1-5]	*M*=1.74	*SD*=0.73	Range [1-5]
Spoken L2 OUTPUT	*M*=2.79	*SD*=1.02	Range [1-5]	*M*=2.46	*SD*=1.03	Range [1-5]
Written L2 OUTPUT	*M*=2.05	*SD*=1.09	Range [1-5]	*M*=1.86	*SD*=1.02	Range [1-5]
Reading L2 Prof	*M*=6.46	*SD*=1.71	Range [2-9]	*M*=5.76	*SD*=1.84	Range [1-9]
Listening L2 Prof	*M*=6.06	*SD*=1.94	Range [1-9]	*M*=5.83	*SD*=1.89	Range [1-9]
Speaking L2 Prof	*M*=5.90	*SD*=1.89	Range [1-9]	*M*=5.55	*SD*=1.92	Range [2-9]
Writing L2 Prof	*M*=5.84	*SD*=1.77	Range [2-9]	*M*=5.66	*SD*=1.54	Range [2-8]
Pron. L2 Prof	*M*=5.81	*SD*=2.03	Range [1-9]	*M*=5.62	*SD*=2.21	Range [1-9]
L2 overall evaluation	*M*=6.01	*SD*=1.60	Range [1-9]	*M*=5.68	*SD*=1.68	Range [1-9]
Elicited imitation	*M*=71.19	*SD*=20.06	Range [32-113]	*M*=71.68	*SD*=21.18	Range [35-116]
Stays abroad (+3w)	NO=51 (81.0%)	YES=12 (19.0%)		NO=27 (93.1%)	YES=2 (6.9%)	
SA duration	*M*=12.33	*SD*=14.79	Range [3-48]	*M*=7.50	*SD*=6.36	Range [3-12]

Table B.3. Specific sources of L2 spoken and written input and output in counts and percentages (%) by group (experimental [simple+complex], control)

		Never	**Few times x year**	**Monthly**	**Weekly**	**Daily**
Experimental	TV series	17 (27.0%)	16 (25.4%)	9 (14.3%)	11 (17.5%)	10 (15.9%)
	Newspapers/magazines	25 (39.7%)	23 (36.5%)	7 (11.1%)	3 (4.8%)	5 (7.9%)
	Books	28 (44.4%)	22 (34.9%)	7 (11.1%)	4 (6.3%)	2 (3.2%)
	Songs	2 (3.2%)	6 (9.5%)	5 (7.9%)	11 (17.5%)	39 (61.9%)
	Films/videos	3 (4.8%)	8 (12.7%)	12 (19.0%)	17 (27.0%)	23 (36.5%)
	Speaking NS	12 (19.0%)	19 (30.2%)	11 (17.5%)	19 (30.2%)	2 (3.2%)
	Speaking NNS	14 (22.2%)	5 (7.9%)	20 (31.7%)	21 (33.3%)	3 (4.8%)
	Writing	24 (38.1%)	21 (33.3%)	12 (19.0%)	3 (4.8%)	3 (4.8%)
Control	TV series	8 (27.6%)	9 (31.0%)	2 (6.9%)	7 (24.1%)	3 (10.3%)
	Newspapers/magazines	13 (44.8%)	9 (31.0%)	5 (17.2%)	2 (6.9%)	0 (0.0%)
	Books	16 (55.2%)	9 (31.0%)	3 (10.3%)	1 (3.4%)	0 (0.0%)
	Songs	0 (0.0%)	2 (6.9%)	3 (10.3%)	6 (20.7%)	18 (62.1%)
	Films/videos	3 (10.3%)	4 (13.8%)	3 (10.3%)	12 (41.4%)	7 (24.1%)
	Speaking NS	8 (27.6%)	8 (27.6%)	5 (17.2%)	6 (20.7%)	2 (6.9%)
	Speaking NNS	9 (31.0%)	7 (24.1%)	6 (20.7%)	6 (20.7%)	1 (3.4%)

Writing	13 (44.8%)	11 (37.9%)	1 (3.4%)	4 (13.8%)	0 (0.0%)

Table B.4. Participant demographic information (simple, complex). *M= Mean, SD= Standard deviation, RH= Right-handed, LH= Left-handed, Dys= Dyslexic, HI= Hearing impairment.*

***N*=63** *(33 male, 30 female)*	**Simple** (*N*=31)		**Complex** (*N*=32)	
Sex (N)	Male=17 (54.8%)	Female=14 (45.2%)	Male=16 (50.0%)	Female=16 (50.0%)
Age at testing (M) [16-17]	*M*=16.03	*SD*=.180	*M*=16.06	SD=.246
Age at testing (N)	16 → 30 (96.8%)	17 → 1 (3.2%)	16 → 30 (93.8%)	17→ 2 (6.3%)
Hand-dominance (N)	RH=27 (87.1%)	LH=4 (12.9%)	RH=26 (81.3%)	LH=6 (18.8%)
Pathologies	NO=28 (90.3%)	YES=3 (9.7%)	NO=30 (93.8%)	YES=2 (6.3%)
Pathology type	Dys=2 (66.7%)	HI=1 (33.3%)		HI=2 (100%)

Table B.5. Participant linguistic information (simple, complex). *M= Mean, SD= Standard deviation, AoL= Age of learning, Cat= Catalan, Sp= Spanish, Ar= Arabian, Ge= German, Fr= French, It= Italian, Jap= Japanese, Ru= Russian, FCE= English First Certificate, RP= Received Pronunciation, GA= General American.*

***N*=63**	**Simple** (*N*=31)			**Complex** (*N*=32)		
Mother tongue (N)	Cat/Sp=15 (48.4%)	Cat=9 (29.0%)	Sp=7 (22.6%)	Cat/Sp=12 (37.5%)	Cat=16 (50.0%)	Sp= 4 (12.5%)
L1-dominance (N)	Cat=19 (61.3%)	Sp=12 (38.7%)		Cat=26 (81.3%)	Sp=6 (18.8%)	
L1 daily use (%)	Cat. *M*=62.29/ *SD*=19.68	Sp. *M*=31.26/ *SD*=19.19	Cat range [25-97] Sp range [0-00]	Cat. *M*=63.56/ *SD*=23.10	Sp. *M*=31.19/ *SD*=21.54	Cat range [1-100] Sp range [2-94]
L2 (N)	En=31 (100%)			En=32 (100%)		
L2 AoL (M)	*M*=5.74	*SD*=2.08	Range [3-9]	*M*=5.53	*SD*=1.75	Range [3-9]
L3 (N)	Ge=8 (57.1%)	Fr=5 (35.7%)	It=1 (7.1%)	Ge=15 (83.3%)	Fr=3 (16.7%)	
L3 AoL (M)	*M*=10.29	*SD*=2.70	Range [4-12]	*M*=11.33	*SD*=2.24	Range [4-14]
L4 (N)	It=1 (100%)			Jap=1 (50%)	Rus=1 (50%)	
L4 AoL (M)	*M*=12.00	*SD*=0		*M*=15.00	*SD*= .000	
L2 Instr. (years)	*M*=10.26	*SD*=2.08	Range [7-13]	*M*=10.47	*SD*=1.75	Range [7-13]
L2 School exposure (h/week)	*M*=3.77	*SD*=1.28	Range [2-8]	*M*=3.16	*SD*=.667	Range [2-5]
L2 Academy exposure (h/week)	*M*=1.46	*SD*=1.25	Range [0-3]	*M*=1.31	*SD*=1.26	Range [0-4]
L2 certificates (N)	NO=25 (80.6%)	FCE=6 (19.4%)		NO=27 (84.4%)	FCE=5 (15.6%)	
L2 use NNS (h/week)	*M*=4.13	*SD*=3.69	Range [0-14]	*M*=2.81	*SD*=2.33	Range [0-10]
L2 use NS (h/week)	*M*=1.56	*SD*=1.80	Range [0-7]	*M*=1.19	*SD*=1.71	Range [0-6]
British exposure	*M*=51.61%	*SD*=27.70%	Range [0-100]	*M*=51.87%	*SD*=26.20%	Range [0-90]
American exposure	*M*=48.39%	*SD*=27.70%	Range [0-100]	*M*=48.13%	*SD*=26.20%	Range [10-100]

Accent production	RP=15 (48.4%)	GA= 16 (51.6%)		RP=18 (56.3%)	GA=14 (43.8%)	
Spoken L2 INPUT	*M*=3.63	*SD*=.936	Range [1.6-5]	*M*=3.52	*SD*=1.15	Range [1-5]
Written L2 INPUT	*M*=2.01	*SD*=1.03	Range [1-4.5]	*M*=1.92	*SD*=.89	Range [1-5]
Spoken L2 OUTPUT	*M*=2.80	*SD*=1.10	Range [1-5]	*M*=2.78	*SD*=0.96	Range [1-4.5]
Written L2 OUTPUT	*M*=1.74	*SD*=1.09	Range [1-5]	*M*=2.34	*SD*=1.03	Range [1-5]
Reading L2 Prof	*M*=6.39	*SD*=1.72	Range [2-8]	*M*=6.53	*SD*=1.72	Range [2-9]
Listening L2 Prof	*M*=6.16	*SD*=2.03	Range [1-9]	*M*=5.97	*SD*=1.87	Range [1-8]
Speaking L2 Prof	*M*=5.77	*SD*=2.07	Range [1-9]	*M*=6.03	*SD*=1.71	Range [2-9]
Writing L2 Prof	*M*=5.97	*SD*=1.92	Range [2-9]	*M*=5.72	*SD*=1.63	Range [2-9]
Pron. L2 Prof	*M*=5.94	*SD*=1.87	Range [1-9]	*M*=5.69	*SD*=2.20	Range [1-9]
L2 overall evaluation	*M*=6.04	*SD*=1.71	Range [1.8-8.2]	*M*=5.98	*SD*=1.53	Range [1.8-8]
Elicited imitation	*M*=72.28	*SD*=23.94	Range [32-113]	*M*=70.13	*SD*=15.77	Range [40-101]
Stays abroad (+3w)	NO=28 (90.3%)	YES=3 (9.7%)		NO=23 (71.9%)	YES=9 (28.1%)	
SA duration	*M*=18.33	*SD*=16.62	Range [3-36]	*M*=10.33	*SD*=14.62	Range [3-48]

Table B.6. Specific sources of L2 spoken and written input and output in counts and percentages (%) by group (simple/complex).

		Never	**Few times x year**	**Monthly**	**Weekly**	**Daily**
Simple	TV series	6 (19.4%)	9 (29.0%)	7 (22.6%)	4 (12.9%)	5 (16.1%)
	Newspapers/magazines	14 (45.2%)	10 (32.3%)	3 (9.7%)	1 (3.2%)	3 (9.7%)
	Books	12 (38.7%)	12 (38.7%)	2 (6.5%)	4 (12.9%)	1 (3.2%)
	Songs	1 (3.2%)	2 (6.5%)	1 (3.2%)	6 (19.4%)	21 (67.7%)
	Films/videos	1 (3.2%)	5 (16.1%)	7 (22.6%)	7 (22.6%)	11 (35.5%)
	Speaking NS	9 (29.0%)	7 (22.6%)	5 (16.1%)	9 (29.0%)	1 (3.2%)
	Speaking NNS	6 (19.4%)	2 (6.5%)	9 (29.0%)	12 (38.7%)	2 (6.5%)
	Writing	17 (54.8%)	9 (29.0%)	3 (9.7%)	0 (0.0%)	2 (6.5%)
Complex	TV series	11 (34.4%)	7 (21.9%)	2 (6.3%)	7 (21.9%)	5 (15.6%)
	Newspapers/magazines	11 (34.4%)	13 (40.6%)	4 (12.5%)	2 (6.3%)	2 (6.3%)
	Books	16 (50.0%)	10 (31.3%)	5 (15.6%)	0 (0.0%)	1 (3.1%)
	Songs	1 (3.1%)	4 (12.5%)	4 (12.5%)	5 (15.6%)	18 (56.3%)
	Films/videos	2 (6.3%)	3 (9.4%)	5 (15.6%)	10 (31.3%)	12 (37.5%)
	Speaking NS	3 (9.4%)	12 (37.5%)	6 (18.8%)	10 (31.3%)	1 (3.1%)
	Speaking NNS	8 (25.0%)	3 (9.4%)	11 (34.4%)	9 (28.1%)	1 (3.1%)
	Writing	7 (21.9%)	12 (37.5%)	9 (28.1%)	3 (9.4%)	1 (3.1%)

Appendix C. Intervention stimuli

Table C.1. Intervention stimuli organized by L2 Vowel (/iː/,/ɪ/,/æ/,/ʌ/), Type (minimal pairs/ extra words also containing the target L2 vowels) and Syllable (1/2).

	/iː/	**/ɪ/**	**/æ/**	**/ʌ/**
Minimal pairs				
1 syllable	bean	bin	bag	bug
	cheek	chick	bat	butt
	feast	fist	cap	cup
	peel	pill	cat	cut
	sheep	ship	mag	mug
	teen	tin	ram	rum
2 syllables	heating	hitting	amber	umber
	keeper	kipper	ankle	uncle
	lever	liver	babble	bubble
	sleeper	slipper	batter	butter
	sneakers	Snickers	carry	curry
	weeping	whipping	natty	nutty
Extra words				
1 syllable	leave	kill	act	run
	weed	fish	hat	drum
	tea	chips	ham	bun
	jeans	pin	jam	gun
2 syllables	illegal	bitter	jacket	public
	kiwi	whiskey	baggy	nugget
	Peter	Jimmy	Patrick	Luster
	Sheila	Lily	Cathy	Sunset

Appendix D. Testing stimuli

Table D.1. ABX stimuli organized by L2 vowel (/iː/,/ɪ/,/æ/,/ʌ/), Word Type (taught/untaught minimal pairs [MP]), Item Type (test/practice/control) and Syllable (1/2).

	/iː/	/ɪ/	/æ/	/ʌ/
		Test		
Taught MP				
1 syllable	bean	bin	bag	bug
	cheek	chick	bat	butt
	feast	fist	cap	cup
	peel	pill	cat	cut
	sheep	ship	mag	mug
	teen	tin	ram	rum
2 syllables	heating	hitting	amber	umber
	keeper	kipper	ankle	uncle
	lever	liver	babble	bubble
	sleeper	slipper	batter	butter
	sneakers	Snickers	carry	curry
	weeping	whipping	natty	nutty
Untaught MP				
1 syllable	beef	biff	crash	crush
	feel	fill	lag	lug
	seal	sill	stab	stub
2 syllables	greeting	gritting	attar	utter
	litre	litter	bagger	bugger
	weaner	winner	clatter	clutter
		Practice		
	gene	gin	mat	mutt
	feet	hill	rat	sun
		Control		
	bleak	pin	black	pun
	cheating	fizzy	chatting	fuzzy

Table D.2. FLeC and LD stimuli organized by L2 vowel (/iː/,/ɪ/,/æ/,/ʌ/), Token Type (word/nonword), Item Type (test/practice/filler) and Syllable (1/2).

	/iː/	/ɪ/	/æ/	/ʌ/
		Test		
Words				
1 syllable	keep	kill	act	run
	weed	fish	hat	drum
	tea	chips	ham	bun
	jeans	pin	jam	gun

2 syllables	illegal kiwi Peter Sheila	bitter whiskey Jimmy Lily	jacket baggy Patrick Cathy	public nugget Luster Sunset
Nonwords				
1 syllable	ki:ll fi:sh chi:ps pi:n	kɪp wɪd tɪ jɪns	ræn dræm bæn gæn	ʌct hʌt hʌm jʌm
2 syllables	bi:tter whi:skey Ji:mmy Li:ly	illɪgal Pɪter Shɪla kɪwi	pæblic nægget Læster Sænset	jʌcket bʌggy Pʌtrick Cʌthy
		Practice		
Words				
	pass, bird neck, work			
Nonwords				
	*kass, *tird *meck, *yerk			
		Fillers		
Words				
	pen, merry horse, morning			
Nonwords				
	*poon, *murry *hars, *marning			

Table D.3. DWR stimuli organized by L2 vowel (/i:/,/ɪ/,/æ/,/ʌ/), Type (minimal pairs [MP] /extra words also containing the target L2 vowels) Word Type (taught/untaught), Item Type (test/practice) and Syllable (1/2).

	/i:/	**/ɪ/**	**/æ/**	**/ʌ/**
		Test		
Taught MP				
1 syllable	bean cheek feast peel sheep teen	bin chick fist pill ship tin	bag bat cap cat mag ram	bug butt cup cut mug rum
2 syllables	heating keeper lever sleeper sneakers weeping	hitting kipper liver slipper Snickers whipping	amber ankle babble batter carry natty	umber uncle bubble butter curry nutty

Untaught MP				
1 syllable	beef	biff	crash	crush
	feel	fill	lag	lug
	seal	sill	stab	stub
2 syllables	greeting	gritting	attar	utter
	litre	litter	bagger	bugger
	weaner	winner	clatter	clutter
Extra Words				
1 syllable	keep	kill	act	run
	weed	fish	hat	drum
	tea	chips	ham	bun
	jeans	pin	jam	gun
2 syllables	illegal	bitter	jacket	public
	kiwi	whiskey	baggy	nugget
	Peter	Jimmy	Patrick	Luster
	Sheila	Lily	Cathy	Sunset
Practice				
	gene	gin	mat	mutt
	feet	hill	rat	sun

Table D.4. List of sentences containing the DWR stimuli (in bold) organized by L2 vowel (/iː/,/ɪ/,/æ/,/ʌ/).

/iː/	/ɪ/
The **bean** is green	The **bin** is empty
The **cheek** is swollen	The **chick** has wings
The **feast** starts soon	The **fist** is protected
The **peel** is orange	The **pill** tastes awful
The **sheep** give milk	The **ship** attacks coasts
The **teen** organizes parties	The **tin** contains peas
The **heating** is on	The **hitting** happens today
The **keeper** avoids goals	The **kipper** is fresh
The **lever** is static	The **liver** cleans blood
The **sleeper** looks beautiful	The **slipper** has stripes
The **sneakers** are dirty	The **Snickers** are delicious
The **weeping** is annoying	The **whipping** seems terrible
The **beef** is raw	The **biff** was horrible
They **feel** true love	They **fill** a teacup
The **seal** looks ill	The **sill** is broken
The **greeting** is friendly	The **gritting** is scary
The **litre** is useless	The **litter** is full
The **weaner** eats carrots	The **winner** smiled widely
They **leave** the house	They **kill** annoying flies

The **weed** is dangerous	The **fish** swim slowly
The **tea** tastes delicious	The **chips** are salty
The **jeans** have cuts	The **pin** is green
The **illegal** are caught	The **bitter** is tasty
The **kiwi** has seeds	The **whiskey** is Irish
Peter loves sweet potatoes	**Jimmy** loves English grammar
Sheila enjoys horror films	**Lily** hates selfish people
The **gene** is unknown	The **gin** has grapes
The **feet** smell bad	The **hill** seems high

/æ/	**/ʌ/**
The **bag** is full	The **bug** eats leaves
The **bat** sleeps a lot	The **butt** is visible
The **cap** looks cool	The **cup** has drawings
The **cat** is hidden	The **cut** is huge
The **mag** is interesting	The **mug** is washed
The **ram** eats greens	The **rum** smells pleasant
The **amber** is fashionable	The **umber** has spots
The **ankle** has bruises	The **uncle** lives near
The **babble** is funny	The **bubble** goes up
The **batter** is sweet	The **butter** adds flavour
They **carry** many eggs	They **curry** the chicken
The **natty** is fabulous	The **nutty** has almonds
The **crash** surprises me	The **crush** is blond
The **lag** takes hours	The **lug** carries fruit
They **stab** the criminal	They **stub** out cigarettes
The **attar** smells well	They **utter** a word
The **bagger** helps customers	The **bugger** is stupid
The **clatter** scares me	The **clutter** needs cleaning
They **act** with respect	They **run** 20 kilometres
The **hat** shows high-class	The **drum** is cool
The **ham** tastes great	The **bun** isn't crunchy
The **jam** is red	The **gun** is heavy
The **jacket** has buttons	The public are cheap
The **baggy** are comfortable	The **nugget** seems overcooked
Patrick cooks every day	**Luster** organizes a party
Cathy cleans on Saturdays	**Sunset** draws a penguin
The **mat** seems dirty	The **mutt** is silly
The **rat** eats cheese	The **sun** is shiny

Appendix E. Pre-tasks, Tasks and Post-tasks

E.1. Pre-tasks

Task Nº	Task title	L2 vowel contrast	Target minimal pairs	Extra words	Instructions for the listening comprehension
1	ORGANIZING YOUR TRIP	i:-ɪ	Feast-fist Sheep-ship Teen-tin Heating-hitting Sneakers-Snickers Weeping-whipping	Weed, Tea, Illegal, Fish, Chips, Bitter, Peter, Lily	Circle the activities that the speakers want to do
2	DECIDING ON ACCOMMODATION	æ-ʌ	Bag-bug Cap-cup Cat-cut Ankle-uncle Carry-curry Natty-nutty	Hat, Ham, Jacket, Patrick, Sunset, Run, Bun, Public	Tick the accommodation where the speakers decide to go
3	CREATING A SPOTIFY LIST	i:-ɪ	Teen-tin Feast-fist Sheep-ship Lever-liver Heating-hitting Weeping-whipping	Jeans, Tea, Sheila, Kiwi, Kill, Chips, Jimmy, Bitter	Underline the songs which the speakers will listen to
4	PACKING YOUR SUITCASE	æ-ʌ	Bag-bug Cap-cup Mag-mug Amber-umber Carry-curry Natty-nutty	Hat, Jam, Jacket, Cathy, Luster, Run, Gun, Nugget	Fill the gaps of the following statements
5	GOING OUT FOR LUNCH	i:-ɪ	Cheek-chick Peel-pill Bean-bin Keeper-kipper Lever-liver Sneakers-Snickers	Tea, Leave, Peter, Kiwi, Fish, Chips, Lily, Whiskey	Underline the chosen starter, main course and dessert
6	TRACING PATHS IN A CITY MAP	æ-ʌ	Bat-butt Cat-cut	Hat, Ham, Jacket, Patrick, Sunset, Run, Public, Drum	Identify which speaker expresses each idea

			Ram-rum Amber-umber Batter-butter Natty-nutty		
7	VISITING AN AUCTION HOUSE	i:-ɪ	Feast-fist Sheep-ship Teen-tin Keeper-kipper Weeping-whipping Lever-liver	Leave, Weed, Kiwi, Sheila, Pin, Kill, Jimmy, Whiskey	Cross out the pictures the speakers do not choose
8	DOING A "ROOM ESCAPE" ADVENTURE"	æ-ʌ	Bag-bug Bat-butt Mag-mug Ankle-uncle Babble-bubble Amber-umber	Act, Hat, Baggy, Cathy, Luster, Drum, Gun, Nugget *(+ Pennsylvania, NY, Country, Moon, London, Boston)*	Choose the correct path
9	GOING TO A SHOPPING CENTRE	i:-ɪ	Bean-bin Feast-fist Peel-pill Keeper-kipper Sleeper-slipper Sneakers-Snickers	Leave, Jeans, Illegal, Peter, Pin, Chips, Bitter, Lily	Make an arrow from the objects to the basket
10	VISITING THE CITY'S HISTORY MUSEUM	æ-ʌ	Bag-bug Bat-butt Mag-mug Amber-umber Ankle-uncle Babble-bubble	Act, Ham, Baggy, Patrick, Sunset, Drum, Gun, Public	Draw the pictures the speakers choose to see
11	DISCOVERING LONDON'S ZOO	i:-ɪ	Bean-bin Peel-pill Sheep-ship Heating-hitting Keeper-kipper Weeping-whipping	Weed, Jeans, (Il)legal Sheila, Kill, Fish, Bitter, Jimmy	Paint the bubbles which refer to the decisions the speakers make
12	COOKING A TYPICAL ENGLISH RECIPE	æ-ʌ	Cap-cup Mag-mug Ram-rum Batter-butter Carry-curry Natty-nutty	Ham, Jam, Cathy, Baggy, Luster, Bun, Run, Nugget	Order the steps which speakers follow to make the recipe

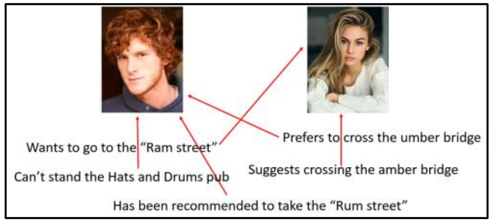
13	CHOOSING CLOTHES FOR A PARTY	i:-ɪ	Bean-bin Cheek-chick Teen-tin Heating-hitting Lever-liver Sleeper-slipper	Jeans, Leave, Peter, Kiwi, Fish, Pin, Lily, Whiskey	Choose the correct answer (A, B or C) about the speakers' friends' outfits
14	PARTICIPATING IN A "ROLE PLAY" PARTY	æ-ʌ	Bat-butt Cap-cup Cat-cut Batter-butter Ankle-uncle Babble-bubble	Act, Jam, Baggy, Patrick, Sunset, Bun, Gun, Nugget *(+dull & mad)*	Relate the objects to each of the speaker
15	SHOPPING IN A TRADITIONAL MARKET	i:-ɪ	Cheek-chick Peel-pill Bean-bin Lever-liver Sleeper-slipper Sneakers-Snickers	Weed, Tea, (il)legal, Sheila, chips, Kill, Jimmy, Bitter	Write down the chosen products in the receipt
16	GOING OUT FOR DINNER	æ-ʌ	Cap-cup Cat-cut Ram-rum Batter-butter Carry-curry Natty-nutty	Ham, Jam, Cathy, Jacket, Luster, Run, Bun, Nugget	Complete the dialogue happening while the speakers are having dinner
17	BUYING LONDON SOUVENIRS	i:-ɪ	Peel-pill Cheek-chick Sheep-ship Keeper-kipper Sleeper-slipper Sneakers-Snickers	Leave, Tea, Peter, Kiwi, Fish, Pin, Lily, Whiskey (+*children & live*)	Cross out the souvenirs speakers will not buy
18	DECIDING ON FILMS TO WATCH TOGETHER	æ-ʌ	Cat-cut Mag-mug Ram-rum Babble-bubble Carry-curry Batter-butter	Act, Jam, Patrick, Baggy, Sunset, Drum, Public, Bun	Order the top 3 films which the speakers are going to watch
19	UPLOADING PHOTOS IN THE SCHOOL WEBSITE	i:-ɪ	Cheek-chick Feast-fist Teen-tin Heating-hitting	Jeans, weed, illegal, Sheila, Kill, Pin, Jimmy, Whiskey	Say if the statements are true or false

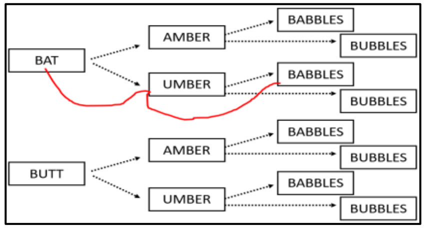
			Sleeper-slipper Weeping-whipping		
20	DESIGNING AN ALBUM AFTER YOUR HOLIDAY	æ-ʌ	Bag-bug Bat-butt Ram-rum Amber-umber Ankle-uncle Babble-bubble	Act, Jacket, Cathy, Hat, Luster, Drum, Gun, Public *(+running)*	Decide if the sentences are true or false

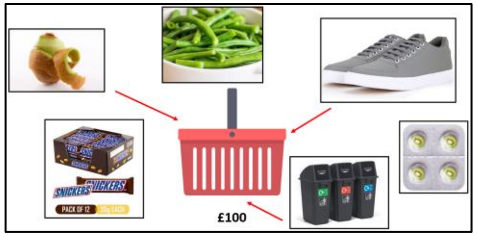
Pre-task listening comprehension KEY

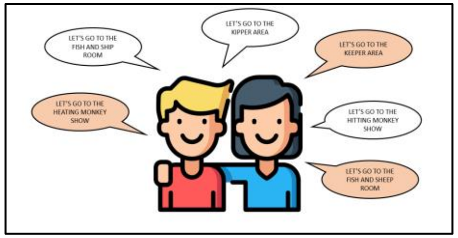
1. Luster has a warm amber jacket.
2. Cathy prefers the umber jacket because she doesn't like the amber colour.
3. Cathy agrees on bringing bags but not cups because they take up too much space.
4. Luster suggests packing a plastic gun to play.
5. Cathy can't carry the jam for her host family because liquids are not allowed.
6. At the end, they will just get the amber jacket and the bags.

Starter	CHIPS ON LEVER	CHIPS ON LIVER
Main	EGG KEEPER	EGG KIPPER
Dessert	SNEAKER CHEESECAKE	SNICKER CHEESECAKE







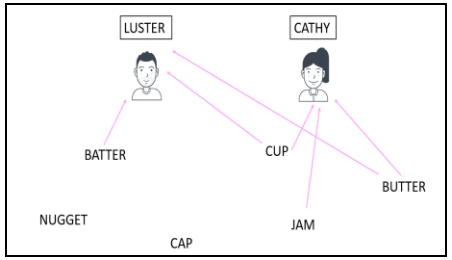


Add some ham 3
Put in the oven 4
~~Carry it on the table~~
Take the ram 1
~~Add nuggets~~
~~Add the rum~~
Curry it on the table 2

1. **Which t-shirt is really childish for Jimmy?**
   a. Bean t-shirt
   b. Bin t-shirt
   c. Pin t-shirt

2. **What does Lily suggest Sheila should wear?**
   a. Bin t-shirt and tins jeans
   b. Kiwi top and teens jeans
   c. Kiwi top and tins jeans

3. **What do they agree on Sheila's outfit?**
   a. Whiskey bottle and teens jeans
   b. Cool pins and teens jeans
   c. Cool pins and bin t-shirt



HANNA'S SHOP

Camden market, London

C_HICK	£10
S_LEEPERS	£25
S_LIPPERS	£15
TOTAL	**£50**

-Hi Luster! Shall we go out for dinner?

-Hey, sure! I want to tell you news about my life and work.

-Great. I thought we could go to Rosie's restaurant.

-I love it. Let's go!

-OK, shall we choose some dishes to share?

-Good idea! I love RAM and vegetables. Do you fancy it?

-I'm sorry but I am vegetarian and I'd rather have FISH WITH RUM.

-Rum? Alright, I'd love a little taste of alcohol in the sauce. What about the dish CURRY THE CHICKEN?

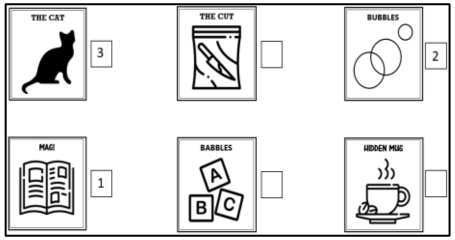
-CARRY OR CURRY?

-CURRY THE CHICKEN

-As I told you before, I can't eat meat, but we could have a NUTTY CAKE.

-NATTY OR NUTTY?

-Both seem delicious. Let's try out both!



-Jimmy prefers the Picture showing their fist FALSE

-The picture of the heating area is funny TRUE

-Sheila would take the whipping horses picture FALSE

-Cathy finds the picture of the group with zoo bugs funny F

-Luster finds the picture of the group with zoo bags hilarious F

-Luster says that the picture of the baby making babbles is his favourite T

-Cathy prefers the picture of the baby making bubbles T

-Cathy wants to include the picture of Patrick's gun T

-Luster has no problem with Patrick's gun and Luster's drum pictures F

E.2. Tasks

Task Nº	Instructions	Conditions SG (simple)	Conditions CG (complex)	Outcome
1	You are Peter and Lily. Your class has voted London to be your next trip. You need to organize several afternoon activities from Monday to Sunday. Your preferences may be the same or different. Please talk to your partner and decide on 1 activity to do each day of the week (total=**7 activities**) → **Sheet:** Instructions + A timetable from Monday to Sunday with empty squares for the afternoon → **Student A:** Drinking tea, buying sneakers, seeing a fist fight, visiting *The Tins bar*, watching *The Hitting film*, going to Jimmy's whipping show, taking illegal weed → **Student B:** Eating fish and chips, buying Snickers, seeing a feast fight, visiting *The Teens* bar, watching *The Heating* film, going to Jimmy's weeping show, taking bitter pills	**Student A** 1. You want to eat healthily **Student B** 1. You'd rather avoid violence	**Student A** 1. You want to eat healthily 2. You'd be happy to skip crowded places 3. You prefer not to try out unknown pills **Student B** 1. You'd rather avoid violence 2. You suffer a lot when hearing about drugs 3. You may not like new experiences	LIST & CLASSIFY
2	You are Patrick and Sunset. You will share information about several hotels & Bed and Breakfast in London. You have done some research and noted down your favourite. You need to talk about all the options but only decide on the **TOP 6**. The whole class will vote for the best accommodation. → **Sheet:** Instructions + a table with six squares with pictures of their 6 hotels → **Student A**: Bugs 'n cups**, The witches' cat***, The fat uncle**, Curry's**, The natty bun*****, The public run***** → **Student B**: Bags 'n caps*, The witches' cut***, The fat ankle*, Carry's****, The nutty bun****, Jackets 'n hats**	**Student A:** 1. You don't like 1-star hotels (*) **Student B:** 1. You don't want to pay for a 5-star hotel (*****)	**Student A:** 1. You don't like 1-star hotels (*) 2. You would prefer a "breakfast included" option 3. You love hotels with spa **Student B:** 1. You don't want to pay for a 5-star hotel (*****) 2. You'd prefer the check out to be after 11:00 a.m. 3. You would like your dog to accompany you.	LIST & ORDER
3	You are Jimmy and Sheila. You have finished your school lessons and you are meeting a friend. You decide to create a shared Spotify playlist for your trip to London. Although you may have different opinions, you need to make a list of your **TOP 8 songs**. → **Sheet:** Instructions + a table with several rows to put their TOP 8 songs	**Student A:** 1. You dislike the song "Amber chips" **Student B:** 1. You can't stand the songs "Red lever"	**Student A:** 1. You dislike the song "Amber chips" 2. The song "show your feast" is not your favourite 3. "The teens" song reminds you of your ex-partner	LIST & WRITE

	→ **Student A**: Kill me, Free sheep, The tins, Red lever, The hitting's on, Whipping horses, Amber jeans, Show your fist → **Student B:** Bitter tea, Free ship, The teens, Red liver, The heating's on, Weeping horses, Amber chips, Show your feast		**Student B:** 1. You can't stand the songs "Red lever" 2. You dislike songs about blood and violence 3. You would stop the song "Free sheep" quickly.	
4	You are Luster and Cathy. You are getting ready for the trip. You just want to bring one suitcase for the two of you. Each of you finds several objects essential so you need to discuss and decide on which objects you will bring in your shared suitcase. **The maximum is 8**. Then stick them in your suitcase so they all fit inside. → **Sheet:** Instructions + 8 squares to put the objects on. → **Student A:** Nuggets, Amber jacket, Bags, Cups, Gun Mag, Carry a chicken, Nutty cake, run sneakers → **Student B**: Ugly hat, Umber jacket, Bugs, Caps, Gun mug, Curry a chicken, Natty cake, jam	**Student A:** 1. You prefer other colours than umber **Student B:** 1. You would like to avoid amber colours	**Student A:** 1. You prefer other colours than umber 2. You can't bring dangerous objects 3. Spices & bugs are not allowed **Student B:** 1. You would like to avoid amber colours 2. You are allergic to nuts 3. Bags and cups take up too much space.	LIST & STICK
5	You are Peter and Lily. You are already in London. You are really hungry and you find a very modern restaurant in the city centre. You will create one full menu (starter, main, dessert) but you want to share everything. Therefore, you need to share your preferences and decide on 6 dishes you want to try. Write them down to create your personalized menu. Then, tell the waiter your choices. The waiter will be the classmate sitting next to you → **Sheet:** Instructions + a table with 2 squares for starters, 2 for mains and 2 for desserts. → **Student A:** ST => Orange pill, chips on lever MA=> Whiskey cheek, The egg kipper, fish & beans DE=>Tea, Sneakers cheesecake → **Student B:** ST=> Orange peel, chips on liver	**Student A:** 1. You are a vegetarian (you can't eat meat) **Student B:** 1. You hate tea and the Sneakers cheesecake	**Student A:** 1. You are a vegetarian (you can't eat meat) 2. You prefer fish & beans 3. You don't like fruit **Student B:** 1. You hate tea and the Sneakers cheesecake 2. You really like meat 3. You'd rather eat healthy dishes.	MAKE A MENU

	MA=> Whiskey chick, the egg keeper, fish & bins DE=>Kiwi, Snickers cheesecake			
6	You are Patrick and Sunset. You have been visiting many museums in the city. Your next destination is the public library. Each of you have different instructions on how to get there. Please trace in the map the shortest way from where you are at the moment until the public library (**max. 7 streets**) → **Sheet**: Instructions + a map with different paths to trace and include illustrations of famous buildings & monuments. → **Student A**: Start in Bat street, go straight to Cat street, turn right to Ram street, cross the umber bridge, continue to Butter buns street, turn right to Public ham square, run to the Nutty jacket shop to get discounts. The public library is in front of you. → **Student B**: Start in Butt street, turn right to Cut street, go straight to Rum street, cross the amber bridge, continue to Batter Buns street, turn left to Hats and Drums pub, run to the Natty jacket shop to get discounts. The public library is in front of you.	**Student A:** 1. You don't want to walk near pubs **Student B:** 1. You are against streets with animal names	**Student A:** 1. You don't want to walk near pubs 2. You can't stand offensive street names 3. You feel nostalgic in "batter buns" street **Student B:** 1. You are against streets with animal names 2. The umber colour makes you feel sad 3. You hate everything having to do with nuts	MAKE A ROUTE
7	You are Sheila and Jimmy. Today you are visiting Sotheby's, which is a place where people buy and sell works of art. There is an auction at the time you are there. Your task is to take a look at different paintings and **decide on 7** you would like to buy. The most voted one in the class will be the one that we will give a price for. → **Sheet**: Instructions+ 7 big boxes to put the paintings on. → **Student A:** Lily's feast, Leaving the ship, green liver, Whiskey keeper, Tins, Weeping, Weed banning → **Student B:** Lily's fist, Leaving the sheep, umber lever, Whiskey kipper, Teens, Whipping and killing, The royal pin	**Student A:** 1. You dislike animals appearing in pictures **Student B:** 1. You are happy with objects present in pictures	**Student A:** 1. You dislike animals appearing in pictures 2. You love pop art 3. You prefer to avoid violent pictures **Student B:** 1. You are happy with objects present in pictures 2. You love animals appearing in pictures 3. You can't stand much green.	LIST & VOTE
8	You are Luster & Cathy. It's the afternoon and you are doing a "room escape" adventure called PRISON. In order to find the three keys that will get you out of the prison, you need to get through different challenges **(=12 actions)**. You and your pair will need to share the information you know to answer the questions that you find on your way.	**Student A:** 1. You prefer not to take big heavy objects **Student B:**	**Student A:** 1. You prefer not to take big heavy objects 2. You think the object next to your ankle is relevant	SOLVE & ROUTE

	→ **Sheet:** Instructions + decoration of the sheet + a path with arrows to follow to find the three keys! → **Student A:** Act & make babbles, take the bugs, eat some nuggets, look at your ankle, run after the bat, find a mug, find an amber drum, catch the baggy hat. → **Student B:** Act & make bubbles, take the bags, eat some nuggets, look at your uncle, run after the butt, find a mag, find an umber drum, catch the gun.	2. You believe that amber colours are a false clue	3. You need an object to keep water in **Student B:** 1. You'd like not to eat or drink anything. 2. You believe that amber colours are a false clue 3. You hate chasing animals	
9	You are Peter and Lily. You want to spend the morning in a shopping centre. You have £200 for the two of you and you may have different shopping preferences. Take a look at your objects and share the information with your partner until you reach a final list of **8 objects.** → **Sheet:** Instructions + the drawing of a basket or trolley in the middle to put the objects on. → **Student A:** A pair of sleepers (£20), Snickers (£5), dry pills (£25), zoo keeper t-shirt (£40), Illegal jeans (£60), 'feast' jacket (£40), bin bags (£5), pins (£5) → **Student B:** A pair of slippers (£25), sneakers (£55), dry peels (£25), zoo kipper t-shirt (£20), Legal jeans (£30), 'fist' jacket (£20), bean bags (£15), "leaving or living" top (£10)	**Student A:** 1. You'd rather avoid main meals in your clothes **Student B:** 1. You don't want to buy anything that costs <£10	**Student A:** 1. You'd rather avoid main meals in your clothes 2. You want to bring your sister's baby a present 3. You fell in love with the "feast" jacket **Student B:** 1. You don't want to buy anything that costs <£10 2. You need new sneakers 3. You don't want to spend more than £40 in a jacket or jeans.	LIST & STICK
10	You are Patrick & Sunset. Today you are visiting London's history museum. After you have walked through the museum, your teacher gives you a sheet with several pictures of paintings and objects you have just seen. In pairs, select those objects you would decorate your class with. Try to reach an agreement on 7 pictures. → **Sheet:** Instructions + 7 squares to put the pictures on → **Student A:** Public act, amber drum, amber bugs, Cathy's bat, 60's mag, Luster's uncle, the baby making babbles → **Student B:** Guns & ham, umber drum, baggy umber bags, Cathy's butt, 60's mug, Luster's ankle, the baby making bubbles	**Student A:** 1. You do not welcome umber colours **Student B:** 1. You would avoid animals or too many people	**Student A:** 1. You do not welcome umber colours 2. You don't like naked bodies in pictures 3. You would love animals and people appearing together **Student B:** 1. You would avoid animals or too many people 2. You love showing naked parts of the body 3. The 60's mug is more interesting than the mag	LIST & VOTE

11	You are Jimmy and Sheila. You are going to London's zoo with your class. Once you get there, your teacher asks you to visit the different places/rooms in pairs. On the one hand, it is exciting to walk around with your friend. On the other hand, you need to agree on 7 places to go. You may have different preferences. Discuss and reach an agreement. → **Sheet**: Instructions + 7 squares to put the pictures on → **Student A**: The sheep room, Legal weed shop, hitting room, zoo keepers point, the pills area, whipping camels, Bins shop → **Student B:** The ship room, Bitter fish restaurant, heating room, zoo kippers point, the peels area, weeping camels, Beans shop.	**Student A:** 1. You can't stand high temperatures **Student B:** 1. You'd rather not see shocking scenes	**Student A:** 1. You can't stand high temperatures 2. You love buying objects in the zoo 3. You prefer not to spend money on food **Student B:** 1. You'd rather not see shocking scenes 2. You enjoy eating in the zoo 3. You would never buy alcohol or drugs	LIST & ROUTE
12	You are Luster and Cathy. Your host mother has explained to both of you the typical English recipe "sweet ram with curry". However, you have understood different ingredients. Share the information with your partner until you create the final version of the recipe. There are only **8 possible steps** which make sense: → **Sheet:** Instructions + An opened recipe book with different squares for students to put the pictures on and create the recipe. → **Student A**: Run to the kitchen, put your cup on the table, take the mag, add the ram, cover it in batter, curry it on the table, leave it natty, accompany with buns & jam → **Student B:** Walk to the kitchen, put your cap on the table, take the mug, add the rum, cover it in butter, carry it on the table, leave it nutty, accompany with a baggy jacket.	**Student A:** 1. You want to avoid slow and useless steps **Student B:** 1. You are allergic to food containing eggs	**Student A:** 1. You want to avoid slow and useless steps 2. You are allergic to nuts 3. You prefer batter than butter **Student B:** 1. You are allergic to food containing eggs 2. You know the recipe by heart 3. You don't like sweet ingredients	LIST & ORDER
13	You are Peter and Lily. There is a party tonight. You have already chosen your own clothes but your best friends have asked you to help them. Both of you need to share your preferences and come up with the perfect outfits for the party. Select a maximum of **8 pieces**. → **Sheet:** Instructions + sheet with two people you will need to wear. → **Student A:** Bean t-shirt, Slippers, Liver jumper, kiwi top, teen jeans, heating shirt, Pig's chicks top, keep your fish watch. → **Student B:** Bin t-shirt, Sleepers, Lever jumper, whiskey top, tin jeans, hitting shirt, Pig's cheeks top, leave your pin watch.	**Student A:** 1. You don't like childish clothes or violence **Student B:** 1. You prefer to wear other colours than green	**Student A:** 1. You don't like childish clothes or violence 2. You prefer not to wear clothes with drawing referring to drinks 3. You love being warm **Student B:** 1. You prefer to wear other colours than green	LIST & WEAR

			2. You want to leave your pin watch at home 3. You already have old slippers	
14	You are Patrick and Sunset. You are organizing a “role play” party tonight and you need to decide what Luster and Cathy will bring and will make. Take a look at your preferences and decide together on **8 objects/ actions** for each friend. 1) Agree on the objects and actions. 2) Write down the invitation to the party and the instructions 3) Swap papers with a pair and change what you don’t agree. → **Sheet**: Instruction+ invitation for Luster and Cathy → **Student A:** Bring (a) ... cap, baggy cat trousers, fish in butter, your uncle’s tattoo, make babbles, bun, jam, show your butt and act as mad. → **Student B:** Bring (a) ... cup, baggy cut trousers, fish in batter, your ankle’s tattoo, make bubbles, gun, nugget, show your bat and act as dull.	**Student A:** 1. You prefer Cathy not to show signs of violence **Student B:** 1. You’d rather Luster not to act as crazy	**Student A:** 1. You prefer Cathy not to show signs of violence 2. You don’t like Luster to make bubbles 3. You know Luster doesn’t like animals at all **Student B:** 1. You’d rather Luster not to act as crazy 2. You prefer Cathy to bring fish in batter 3.You think Luster can have an ankle’s tattoo	LIST & WRITE
15	You are Jimmy and Sheila. Your host father and your host mother have given you a different list of products you need to buy in the market. You only have the money to **get 8 products** between the two lists. 1) Discuss with your partner which the best choices are 2) Place the objects on the basket 3) Make a guess about the money that your family gave you according to your opinion. → **Sheet:** Instruction + trolley to put the products in → **Student A:** Cheeks, orange peels, beans, a liver, a couple of slippers, Snickers, legal weed, bitter tea. → **Student B:** Chicks, orange pills, bins, a lever, a couple of sleepers, sneakers, chips, kill(ing) knife	**Student A:** 1. You think buying food is more urgent than objects **Student B:** 1. You are not sure about buying dangerous health-products.	**Student A:** 1. You think buying food is more urgent than objects 2. You love pig cheeks with orange peels 3. You urgently need a pair of slippers **Student B:** 1. You are not sure about buying dangerous health-products. 2. You really like chicks, pills & chips 3. You are not sure about bitter tea	LIST & GUESS
16	You are Luster and Cathy.	**Student A:** 1. You want to avoid hot/spicy plates	**Student A:** 1. You want to avoid hot/spicy plates	MENU & STICK

	You have a date. Because it is a very special occasion, you decide to go to one of the most traditional pubs in London. Instead of choosing a starter, main and dessert, you decide to try different small plates to share. 1) Agree on **7 of them** to share. 2) Stick a GREEN STICKER on the selected plates 3) When you have the name of the plates, get together with another pair and choose who the waiter and clients will be. Practice it! → **Sheet:** Instructions + the drawing of a big table to put the different plates on. → **Student A:** Ham in butter, The nutty bun, Carry the ram experience, The black cut, The cup, The jam nuggets, A Jacket cake → **Student B:** Ham in batter, The natty bun, Curry the rum experience, The black cat, The cap, The hot run, A Jacket shot	**Student B:** 2. You are allergic to nuts	2. You don't want to drink any alcohol 3. You love ham in butter **Student B:** 1. You would like to avoid sweet plates 2. You are allergic to nuts 3. You dislike meat plates	
17	You are Peter and Lily. This is your last day in London. Your sister and you want to bring your family several souvenirs. Since your parents have only given you £100 to share, you will need to discuss which are the best presents for the family and **agree on 8.** → **Sheet:** Instructions + The design of a shop with objects to cross out → **Student A:** Blue kipper (£5), delicious orange peels (£5), Union jack flag slippers (£20), Amber Snickers (£5), The children's chick t-shirt (£25), a sheep (£10), Whiskey (£15), fish (£10), tea (£5). → **Student B:** Blue keeper (£10), delicious orange pills (£5), Union jack flag sleepers (£10), Umber sneakers (£30), The children's cheek t-shirt (£15), a ship (£10), a pin (£5), kiwis (£5), Leaving or living t-shirt (£10).	**Student A:** 1. You'd rather avoid very patriotic souvenirs **Student B:** 1. You would prefer not to buy any animal-related object	**Student A:** 1. You'd rather avoid very patriotic souvenirs 2. You don't want to spend = or >£30 3. You love the children's chick t-shirt and the sheep **Student B:** 1. You would prefer not to buy any animal-related object 2. You don't want to buy any liquids or bitter food 3. You really like the umber colour	LIST & CALCULATE
18	You are Patrick and Sunset. This afternoon you arrange to watch a film with your friend. You have different preferences but you only have the time to watch one film. Therefore, you need to make a list of **8 films** you both like and come up with a final one. → **Sheet:** Instructions + the design of a set of shelves with numerous films. → **Student A:** The public act, The cat, My sweet ram, Bubbles, Hot mug!, Curry it, Butter & buns, Jam 'n drum	**Student A:** 1. You don't like violent films **Student B:** 1. You dislike animal films	**Student A:** 1. You don't like violent films 2. You hate films which start with "ba" 3. You enjoy films that deal with food **Student B:** 1. You dislike animal films	LIST & NARROW

	→ **Student B:** The cut, My sweet rum, Babbles, Hot mag!, Carry it, Batter & buns, Baggy t-shirts, the sunset		2. You like the film which is called like you 3. You prefer not to watch "the public act"	
19	You are Jimmy and Sheila. You are back to Vilassar. Your teacher has announced that you will need to choose several photos to upload in the school website. Take a look at your preferences and decide on **9 pictures together.** → **Sheet:** Instructions + the setting of the school's website. → **Student A:** Jimmy's chicks, illegal weed, Sheila's pin, Peter's feast, teens, the house heating, Lily's slippers, Weeping on the horse, Amber jeans. → **Student B:** Jimmy's cheeks, whiskey, Sheila's killing knife, Peter's fist, tins, the house hitting, Lily's sleepers, Whipping on the horse, Umber jeans	**Student A:** 1. You can't stand violence **Student B:** 1. You want to upload pictures with umber colours	**Student A:** 1. You can't stand violence 2. You'd rather select pictures with amber colours 3. You love Lily's slippers **Student B:** 1. You want to upload pictures with umber colours 2. You like the pop art tins but not Sheila's pin 3. You are not allowed to upload anything illegal	SELECT & POST
20	One week has passed by and you meet in Cathy's house to make an album. You will need to select **8 photos** you want to include in the album but, be careful, you may have very different opinions. Then, stick your photos in the album. Add the stickers you like to decorate it. → **Sheet**: Instructions + an opened album to put the pictures on → **Student A**: Group with zoo bugs, Students' bats, Cathy with ram, Amber jackets, Patrick's uncle, baby making babbles, Luster's drum, the group acting in the street → **Student B**: Group with zoo bags, Students' butts, Cathy with rum, Umber jackets, Patrick's ankle, baby making bubbles, Luster's gun, the group running in the street	**Student A:** 1. You prefer amber colours in the pictures **Student B:** 1. You dislike animals in the pictures	**Student A:** 1. You prefer amber colours in the pictures 2. You are against any type of violence or offensive pictures 3. You really like the baby making babbles **Student B:** 1. You dislike animals in the pictures 2. You love umber colours 3. You find the picture of Patrick showing his ankle really funny	SELECT & DESIGN

E.3. Post-tasks: Practice stage

Task Nº	Post-task title	Target words	Instructions
1	WHISPER CHALLENGE	feast, fist, sheep, ship, teen, tin, heating, hitting, sneakers, snickers, weeping, whipping	In a group of 5-6 students, the first person in the circle whispers a word into the ear of the person sitting to their right. Players whisper the word to their neighbours until it reaches the last player. The last player says the word out loud so everyone in the group can hear how much it has changed from the first whisper at the beginning of the circle.
2	HOW GOOD IS YOUR MEMORY?	Bug, carry, cat, uncle, public, natty, umber, amber, cap, Sunset, mag, mug	Individually, memorize the words which contain the sound in “cup” and write them down as a list. Then, describe a type of accommodation with the target words
3	RAP	feast, fist, sheep, ship, teen, tin, heating, hitting, lever, liver, weeping, whipping	In pairs or groups of 3, create a short rap with 8 of the target words. If there is time, present it in front of another group.
4	GUESS & DRAW	amber watch, umber wallet, mag, mug, bag, bug, natty cake, nutty salad	Individually, draw the objects I dictate and later compare them with your partner. Have you understood the same?
5	LUNCH TIME!	bin cake, bean, pig cheek, chicks, orange peels, orange pills, egg keeper, kipper, chips on lever, chips on liver, snickers, sneaker cheesecake, whiskey, kiwi, tea.	In pairs, think about one friend you share and hypothesize whether they would ask for the following meals (see target words). Give reasons about your choices.
6	MAP DESIGNERS	butt, bat, cat, cut, amber, umber, ram, rum, batter, butter, natty, nutty	In pairs, you need to design a map of London and include the following streets: STUDENT A: butt, bat, cat, cut, amber, umber STUDENT B: ram, rum, batter, butter, natty, nutty Once all streets are included, you need to ask the other person to follow a route you have come up with.
7	READY FOR SOME SHOPPING?	Feasts, fists, free sheep, free ship, teens, tins, old egg keeper, egg kipper, lever, liver, weeping, whipping	In pairs, try to convince your partner to buy some of your favourite pictures. Your partner may agree or disagree and will also share his/her preferences.
8	ESCAPING THE ROOM	Bag, bug, bat, butt, mag, mug, amber, umber, ankle, uncle, babble, bubble	You are trapped in a room. In pairs or groups of 3, you need to escape but first, you need to guess what the following objects (see target words) may be used for. Then, choose 8 and ask the teacher if they are the ones that will get them out of the room. The group with a greater number of correct answers is the winner.
9	BINGO!	Bean, bin, feast, fist, peel, pill, keeper, kipper, sleepers, slippers, sneakers, snickers, jeans, illegal, Peter, chips, pin, bitter, Lily, leave	The whole group participates in this activity. Each student has a Bingo card which contains some of the target words. Students need to cross out the words they hear. The winner(s) are the ones who finish crossing out all words first. Answers are checked orally in front of the other classmates.

10	IMAGINE	Bag, bug, bat, butt, mag, mug, amber, umber, ankle, uncle, babble, bubble, ham, gun, Patrick, Sunset	In pairs, one imagines a picture which contains 8 of the target words and tries to describe it to his/her classmate. The other person draws the imaginary picture while the description is taking place.
11	A CHAT IN THE ZOO	Bean, bin, sheep, ship, peel, pill, heating, hitting, keeper, kipper, weeping, whipping	You decide to spend the day in the zoo. In pairs or groups of 3, create a short dialogue in which 8 of the following target words appear. Which places are you going to visit?
12	YUM, YUM!	Ram, rum, carry, curry, batter, butter, natty, nutty	In groups of 4, write down a recipe which must include the following actions/ingredients. It can be an invented recipe but it needs to make sense and needs to be edible.
13	PICTIONARY	Bean, bin, cheek, chick, teen, tin, heating, hitting, lever, liver, sleepers, slippers	Students look at the target words on the blackboard. In pairs, one chooses one of the words and draws it in a maximum of 1 min. The other partner needs to guess it before the time is up.
14	MIME	Bat, butt, cap, cup, cat, cut, ankle, uncle, babble, bubble, batter, butter, jam, gun, act, bun	It is time to express yourselves with mime. In groups of 3 or 4, one needs to mime one of the target words and the others need to guess which object the performance represents. The first to know the answer wins!
15	THE POEM	bean, bin, cheek, chick, peel, pill, lever, liver, sleepers, slippers, sneakers, snickers, weed, chips	Are you a good poet? Individually, write a poem which includes a maximum of 8 words. Share it with the person sitting behind you. How do they feel about it?
16	SHARE YOUR MEAL	cat, cut, ram, rum, batter, butter, natty, nutty	Your classmate and you do not have enough money for two full meals so you decide to share one. Talk about the following meals (containing the target words) and prepare a menu for the two. Then, compare your menu with other classmates.
17	"IMPROVERSEM"	Cheek, chick, peel, pill, sheep, ship, keeper, kipper, sneakers, snickers, sleepers, slippers	Individually, choose a song you like and know its lyrics. Change the original lyrics to fit the target words. If there is some time left, record it and send it to your best friend. What is his/her reaction?
18	FILM REVIEW	Cats, Cuts, The ram, The rum, Carry it, curry it, Babbles, Bubbles	In pairs, choose four of these eight films. Each of you will write a good and a bad review about two of them. Read them orally in front of the class. Who can you convince?
19	POSTING!	Cheek, chick, feast, fist, teen, tin, heating, hitting, sleepers, slippers, weeping, whipping	You have created your own blog about the last trip you made. Individually, write down your post sharing your feelings and experiences. Include 6 of the target words.
20	THE STORY OF MY TRIP	Bag, bug, bat, butt, ram, rum, amber, umber, ankle, uncle, babble, bubble, hat, drum, act, gun	In pairs, make the story of your trip by using 8 target words. Make it memorable and share it with your mates.

Appendix F. Pre-task listening scripts

TASK 1: ORGANIZING YOUR VISITS DURING THE TRIP
-Good morning Lily!
-Hi Peter! What would you like to do on Monday afternoon?
-I'd love to buy a pair of SNEAKERS.
-SNEAKERS OR SNICKERS? Because I'd rather get some Snickers.
-I'm sorry. I am cutting down on chocolate and I don't want to spend money on food.
-OK! Let's get the SNEAKERS, then. Oh! And can we visit the TEENS BAR on Tuesday afternoon?
-Sure, although I'm also interested in going to the TINS BAR. I love seeing all kinds of TINS!
-My friends will be in THE TEENS BAR. We can get a cup of tea
-Perhaps we can have dinner there. I'll ask for FISH AND CHIPS
-Really? I can't stand the smell of fish. I'll just have the TEA.
-No problem.
-We'll have a great time, I'm sure.

TASK 2: DECIDING ON ACCOMODATION
-Patrick, after doing some research on London's accommodation, I think we could go to the BUGS BED & BREAKFAST
-Well Sunset, this is only a 2-star hotel and why would you choose a hotel with 'insects' in its name? So, what about THE BAGS hotel?
-Isn't it the same? No, I said BAGS! It's a 4-star hotel near Buckingham palace.
-That sounds really expensive. I also liked UNCLE T. HOTEL. They allow animals in this hotel.
-That's fine but I'd love to go to the spa in the afternoons. The ANKLE T. HOTEL offers a 12-hour-spa with massages and it is very near our language school.
-Yes Sunset, but animals can't go inside. We should look for another one.
-I see... We could always try the CAT'S HOUSE
-The CUT'S HOUSE? Where you can check out after 11:00 a.m.?
-No, the CAT'S HOUSE, you know CAT as in the animal, which has free WI-FI, a small spa and really good food.
-That sounds cool! Will the other classmates agree?
-I think so! Our second possibility could be the CUT'S HOUSE, though.
-Definitely!

TASK 3: CREATING A SPOTIFY PLAYLIST
-Hello Sheila! Are you ready for our trip to London?
-Almost! Why don't we create a shared Spotify list for our trip?
-Yes, why not? We have very similar tastes.
-Do you like the song FEAST NIGHT, which talks about celebrating the beginning of summer?
-You mean the song FIST NIGHT, the one about boxing?
-No no, FEAST NIGHT, by the famous Spanish DJ.
-I remember. I don't like it at all. I prefer to listen to the LIVER. Their group name is a part of their body, isn't it strange?
-I like the LIVERS and the LEVERS, too.
-Isn't it the same?
-No, they are two different groups.
-Alright, what about the WEEPING HORSES?
-The WHIPPING HORSES sounds excellent.
-No, the WEEPING HORSES, a 90s song about horses crying
-We can choose the WHIPPING HORSES and I'll let you choose the KIWI song, which I know you love.
-Great, it's a deal.

TASK 4: PACKING UP YOUR SUITCASE
-Hi Luster! Shall we pack up together so we only bring one bag?
-Yes! I have a very warm AMBER JACKET which will be really useful in London.
-I prefer the UMBER JACKET you have. I hate the AMBER COLOUR.
-OK! That's fine. Could you bring CUPS AND BAGS?
-I agree that small BAGS may be really practical but I'd rather not bring CUPS as they take up too much space and we can buy them there.
-Shall we bring BAGS AND CAPS for the sun, then?
-Sorry, did you say CAPS? I meant CUPS!
-Oh sorry! I didn't understand you. Finally, would you add a PLASTIC GUN to play with?
-I don't think that's appropriate. We may have problems at the airport. We could bring JAM for our host family.
-I'd love to, but jam contains liquid so I'm afraid we may get stuck at the security check.

TASK 5: GOING OUT FOR LUNCH
-Good afternoon Lily!
-Hello! I'm really happy we are having lunch together! I have plenty of news to tell you.
-Great! Shall we look at the menu first?
-Sure, we could share it! I'd love to eat CHIPS ON LIVER as a starter. What do you think?
-Yuk! I hate meat. Why don't we go for the CHIPS ON LEVER? It's a beautiful dish where the chips are on a metallic bar.
-Sounds good. Then, we can have the EGG KIPPER.
-The EGG KIPPER or EGG KEEPER? I'd prefer the EGG KEEPER because I dislike fish.
-I would like to try the EGG KIPPER, though. I'll let you choose the dessert.
-Ok... Should we finish off with the SNEAKER CHEESECAKE?
-Did you say the SNICKER CHEESECAKE?
-KIWI AND SNEAKER CHEESECAKE. I hate cakes with too much chocolate.
-Great. We've made up our minds! What was the news you wanted to share with me?

TASK 6: TRACING PATHS IN A CITY MAP
-Hello Patrick. Did you have fun last night?
-Hi Sunset! Yes, I met lots of new people at the party. Anyway, I'd love to go to George's church today. Do you want to join me?
-Sure! I was thinking about going there, too.
-Alright. My host family gave me some instructions. Did yours too?
-Yes! She said we should take RAM STREET.
-RUM, as in the drink, or RAM STREET, as in the animal?
-RAM STREET
-My host mother recommended RUM STREET but let's try this one.
-Yes, it's better, because I don't like streets with lots of pubs. What about crossing the AMBER BRIDGE?
-I'd prefer to cross the UMBER BRIDGE. We will reach the PUBLIC SQUARE faster.
-Ok, but first, let's stop at the HATS AND DRUMS PUB!
-I'm sorry but you know I can't stand PUBS...

TASK 7: VISITING AN AUCTION HOUSE
-Good afternoon Sheila!
-Hi Jimmy! What's up? I'm really excited about visiting the auction house.
-Me too. One of my favourite pictures is the Free SHEEP.
-The FREE SHIP. Mine too!
-Yes, I love animals in pictures.
-Oh, you said the FREE SHEEP? I don't like animals in pictures, I'm sorry. Can we choose the FREE SHIP, please?
- Alright. What about the WEEPING HORSE?
-I prefer the 19th century WHIPPING HORSE. I really like the history behind that picture.
-I really dislike seeing someone hitting a horse. If you agree, we could go for the WEEPING HORSE even if it's a sad scene. Is that fine by you?
-Okay... but then, we see the ROYAL PIN instead of the weed banning picture. I'd rather avoid contemporary art

TASK 8: DOING A ROOM ESCAPE ADVENTURE
-Okay then!
-Are you ready Cathy? We have 60 min to find the key that will get us out of this prison.
-Alright, the first question is: what is the name of the animal that sleeps upside down?
-BUTT
-Did you say BAT?
-No, BUTT.
-Sorry, but it's BAT, BUTT is just a part of the body.
-I see. Then we need to take a card and solve the logarithms.
-Done! Next question?
-Which colour is most similar to brown?
-AMBER!
-AMBER or UMBER? Amber cards are said to give false clues.
-UMBER, sorry. This brings us to the next question: what does this sound represent, a baby making BABBLES OR BUBBLES?
-Clearly BABBLES! The baby is trying to speak!
-Are you sure? They may be BUBBLES.
-Ok, let's SAY BUBBLES and see if the door opens.
-It doesn't. It's a baby making BABBLES.

TASK 9: GOING TO A SHOPPING CENTRE
-Good morning Lily! I am going to the shopping centre next to our town. Are you coming?
-Sure, Peter. How much money do we have?
-We've got £100. I need something to help me go to the toilet! I've got a bit of stomach ache!
-Should we get these brown PEELS? They cost £5!
-PILLS are not natural!
-Yes, I am talking about PEELS. Should we get these KIWI PEELS? I am sure they'll be good for your body.
-Perfect! What do you need from the shopping centre?
-I saw very cheap SNEAKERS on that shop website.
-Yes, I saw them too, but they are not really healthy.
-Healthy? Are you talking about SNICKERS OR SNEAKERS?
-Oh sorry, what a misunderstanding! You mean SNEAKERS.
-Before we leave, should we get BINS for the barbecue?
-I agree, let's take some ... BEANS?
-No, BINS, they are £40 and we can afford them.
-Great then, let's do that!
-Yes, I think that'll be fine.

TASK 10: VISITING THE CITY'S HISTORY MUSEUM
-Hello Patrick. How are you doing?
-I'm great, thanks. Should we go into the history museum? I guess the teacher is already there.
-Sure. Let me go to the toilet, first.
-Luster, the teacher has given us a sheet with the pictures that we are going to see. We need to go around and select the ones that we could decorate our class with.
-Look, I love this painting representing a 60's MAG. The Beatles were just starting!
-I love it too but I'd rather choose the 60'S MUG because this MUG has a beautiful design.
-What about the huge sculpture of the ANKLE?
-Well it's really cool but I think the UNCLE's sculpture is even more impressive.
-Alright, then let me choose the 90's GUN AND DRUM painting. I love dark colours in pictures!
-I'm sorry I hate this one. Can I choose the 60's MAG instead?
-OK but I'll keep the GUN AND DRUM painting in my bedroom!

TASK 11: DISCOVERING LONDON'S ZOO
-Good morning Sheila!
-Hello! I'm very sorry that I got here so late! We just have 30 min to visit the zoo. Shall we select 3 places we MUST go?
-Of course. I'd love to see the HEATING MONKEY show. They react to high temperatures in a very strange way.
-I'm afraid I can't see this. I hate violence.
-Violence? I said HEATING, NOT HITTING!
-Then it's great. I also like the FISH AND SHIP room.
-SHIP OR SHEEP?
-I am sorry but I'd rather see the SHEEP ROOM. It would to be more fun!
-OK but then, let me go to the KEEPER AREA, please.
-Why is it so interesting?
-They keep the eggs of the KIPPERS.
-Impressive. I think the route is ready. Should we start then?
-Yes, let's go!

TASK 12: COOKING A TYPICAL ENGLISH RECIPE
-Hi Luster!
-How are you doing Cathy?
-I'm great, thanks. Would you like to help me prepare a RAM DISH?
-Sure! I have been given the recipe but I'm not sure if I understood it well.
-We should get THE RAM first.
-THE RAM or THE RUM?
-THE RAM. I hate alcohol, sorry!
-Take THE RUM for me then. Should we carry it to the table?
-No! What we should do is CURRY it!
-CURRY? What do you mean?
-Yes, to cover it in CURRY.
-I hope it is not too spicy.
-Should we also cook some NUGGETS?
-Perhaps it will be too much. Let's add some HAM though, and put it in the oven.

TASK 13: CHOOSING CLOTHES FOR THE PARTY

-Good morning Lily! Did you know Jimmy asked me to help him choose clothes for tomorrow's party?
-Really? Sheila asked me too!
-Let's create the perfect outfit for them!
-Ok! Why don't we try this BEAN t-shirt on Jimmy?
-Did you say BEAN, like the vegetable?
-Yes, BEAN, not BIN, where you put the rubbish.
-I'm really sorry but I find this t-shirt really childish.
-What do you prefer then?
-I really like the BIN t-shirt or the KIWI Top.
-I suggest Sheila should wear the KIWI top and the TINS jeans.
-The TINS jeans? I don't like drinks on her clothes.
-Alright. Any other option?
-Yes, the KIWI top and the TEENS jeans.
-And she'll wear cool pins.
-Yes, bring PINS instead of the WHISKEY bottle.

TASK 14: PARTICIPATING IN A "ROLE PLAY" PARTY

-Hi Patrick! Today we have a very difficult task. We need to organize the end-of-the-course party and we need to assign different actions and objects to people. Shall we start with Luster and Cathy?
-Sure! Luster could bring a CUP, a tea cup, and Cathy a CAP, you know, a hat.
-That would be great but Cathy hates things on her head. I think Luster and Cathy could bring different CUPS.
-That's a good idea! I know both love cooking so perhaps they could bring BUTTER.
-Well Cathy may bring BUTTER to make a cake and Luster could bring BATTER to bake fish. What do you think?
-Alright, it's a great decision.
-Finally, Cathy could bring some JAM to put on the cake and Luster could bring some NUGGETS.
-NUGGETS? Oh no, he hates fast food. Perhaps he can bring some butter, too.
-Sounds like a plan!
-OK, let's continue with Jamie and Anne.

TASK 15: SHOPPING IN A TRADITIONAL MARKET

-Good morning Sheila!
-What's up Jimmy?
-Our host mother has given us a list of products to buy in the market. Do you have money to get them?
-I just have £50 so we may only have enough money for 3 or 4 products.
-I think that we need to get a CHICK for lunch.
-A CHICK OR A CHEEK?
-A CHICK, otherwise it'll be too expensive.
-Alright, let's get the CHICK and not the pig CHEEK. Are you happy with it?
-Yes, but we could also get a LIVER!
-Livers are disgusting! What about the LEVER for our host father?
-Yes, but our host mother doesn't want to have more tools.
-Well, in this case, let's choose another thing.
-We also need some SLEEPERS for the baby.
-True, SLIPPERS for the baby's feet!
-SLEEPERS or SLIPPERS?
-I guess both will be useful. Let's take them and see if we have enough with £50.

TASK 16: GOING OUT FOR DINNER

-Hi Luster! Shall we go out for dinner?
-Hey, sure! I want to tell you news about my life and work.
-Great. I thought we could go to Rosie's restaurant.
-I love it. Let's go!
-OK, shall we choose some dishes to share?
-Good idea! I love RAM and vegetables. Do you fancy it?
-I'm sorry but I am vegetarian and I'd rather have FISH WITH RUM.
-Rum? Alright, I love a little taste of alcohol in the sauce. What about the dish CURRY THE CHICKEN?
-CARRY OR CURRY?
-CURRY THE CHICKEN, which is a little spicy.
-As I told you before, I can't eat meat, but we could have a NUTTY CAKE.
-NATTY OR NUTTY?
-Both seem delicious. Let's try out both!

TASK 17: BUYING LONDON SOUVENIRS
-Good morning Lily! I'd love to get some London souvenirs for my family!
-So would I, but we can't spend more than £50
-No worries. What do you think of this SHIP?
-SHEEP or SHIP?
-That famous old SHIP!
-Well, I'd rather take the SHEEP because it is more representative of the English countryside
-Alright. I think the English SLEEPERS are really good.
-Yes, everyone needs a pair.
-A pair of SLEEPERS? You mean SLIPPERS, don't you?
-Yes, sorry, I'd love to get these SLIPPERS but I am happy with the SLEEPERS too!
-Fine. Finally, should we take this London WHISKEY or the PINS?
-I'm sorry but I can't take liquids back home and THE PINS are too expensive.
-Let's leave it this way and see if the total is less than £50!

TASK 18: DECIDING ON FILMS TO WATCH TOGETHER
-Hello Patrick! Do you fancy coming home to watch a film?
-Sure, I do! Do you like THE CUT, which is all about blood?
-I'm sorry. I hate horror films. What about THE CAT? It's a comedy!
-Mmmmm... OK, but films about animals are not my favourite.
-I also like the film whose cover has a baby MAKING BUBBLES.
-A baby MAKING BABBLES, like speaking, OR BUBBLES?
-The one MAKING BUBBLES is great for me. Dramas aren't my favourite either but I guess you prefer it to THE CAT.
-If not, there's an action film called the HIDDEN MAG, which happens in the 90's and stars Tom Cruise. They need to find a mag which contains secret information.
-The HIDDEN MUG?
-No, MAG! I've read great reviews about it. Shall we choose this one then?
-Yes, and if we don't like it, we go for the drama one!
-Do you want me to bring some popcorn?
-That would be really cool, thanks!
-No problem!

TASK 19: UPLOADING PHOTOS IN THE SCHOOL WEBSITE
-Good afternoon Sheila!
-Hi Jimmy! How are you? I am very sad that the trip is over!
-Me too. Our teacher asked us to upload some photos on the school website
-Yes, do you want to start looking at them now?
-Sure!
-Do you fancy this picture of the group showing their FIST?
-FIST OR FEAST?
-FIST.
-Oh no, I prefer the picture showing their FEAST, which shows a big celebration!
-What about the HITTING SCENE?
-HITTING is violence again, but the picture of the HEATING AREA in the zoo is very funny! Do you remember how hot it was?
-Of course, I do!
-Finally, we have two interesting pictures here: the WEEPING MONKEYS and the WHIPPING HORSES
-I would take the WEEPING one as it is really curious to see a crying monkey.
-Great. Let's upload these three for now.

TASK 20: DESIGNING AN ALBUM AFTER YOUR HOLIDAY
-Hello Luster! Welcome to my house.
-Hi Cathy. I'm really excited about creating our trip's album with you.
-Me too. Look at this funny picture of the group with the ZOO BAGS!
-You mean the one with THE ZOO BUGS?
-No, no, ZOO BAGS! Do you like it?
-To be honest, I find the picture of the group with ZOO BUGS hilarious!
-Perhaps, we can choose both.
-What about the baby MAKING BABBLES?
-I prefer the baby MAKING BUBBLES
-I'm sorry but the baby MAKING BABBLES is my favourite. He is trying to utter his first words!
-Alright! Then, you let me include the picture of PATRICK'S GUN.
-But this is really violent. I can't accept this. If we choose the picture of PATRICK'S DRUM, you are happy because Patrick appears but no violence is shown.
-OK but now I'll decide on the following ones.

Appendix G. Target word assessment

SUBJECT CODE: CLASS:
DATE:

"OUR TRIP TO LONDON": VOCABULARY TEST

WORDS	TRANSLATION/EXPLANATION	WORDS	TRANSLATION/EXPLANATION
bean		bin	
cheek		chick	
feast		fist	
peel		pill	
sheep		ship	
teen		tin	
heating		hitting	
keeper		kipper	
lever		liver	
sleeper		slipper	
sneakers		Snickers	
weeping		whipping	
leave		kill	
keep		fish	
tea		chips	
jeans		pin	
illegal		bitter	
kiwi		whiskey	
bag		bug	
bat		butt	
cap		cup	
cat		cut	
mag		mug	
ram		rum	
amber		umber	
ankle		uncle	
babble		bubble	
batter		butter	
carry		curry	
natty		nutty	
act		run	
hat		drum	
ham		bun	
jam		gun	
jacket		public	
baggy		nugget	

Appendix H. Learners' language background questionnaire

Personal Data (Dades personals)

Surnames (Cognoms) *

Name (Nom) *

Sex (Sexe) *

- Male(masculí)
- Female(femení)

Class (Classe de BTX) *

- 1r Batx. A
- 1r Batx. B
- 1r Batx. C

Date of Birth (Data de naixement) *
mm/dd/yyyy

You are (ets): *

- right-handed (dretà)
- left-handed (esquerrà)
- both (tots dos)

Do you have any speech or hearing problems? Any other pathology? (Tens algun problema de la parla o oïda? Alguna altra patologia [ex. dislèxia...?]) *

- Yes
- No

If YES, please, specify (Si és així, si us plau, especifica): *

____
____
____
____
____

Language Profile (Perfil lingüístic)

1. Indicate which language(s) you normally speak on daily basis (indica quines llengües parles diàriament): *

You can tick more than one box!

- [ ] Catalan (Català)
- [ ] Spanish (Castellà)
- [ ] English (Anglès)

Others: [ ] ______________________________

2. Indicate your Mother tongue [First Language or L1] (Indica la teva llengua mare): *

Language(s) you learnt to speak from birth. You can tick more than one box!

- [ ] Catalan (Català)
- [ ] Spanish (Castellà)
- [ ] English (Anglès)

Others: [ ] ______________________________

3. Indicate language(s) spoken at home most of the time (Indica les llengües que parles a casa la majoria del temps): *

You can tick more than one box

- ☐ Catalan (Català)
- ☐ Spanish (Castellà)
- ☐ English (Anglès)

Others: ☐ ______________________

4.1. Please list all the languages you know in order of dominance and age at which you started to learn them (Fes una llista de les llengües que saps de major a menor coneixement i posa-hi l'edat quan les vas aprendre): *

For example, Catalan/0

____

4.2. *

For example, Spanish/2

____

4.3. *

For example, English/8

____

4.4.

For example, German/18

____

4.5.

Russian/20

____

Language Use (Ús lingüístic)

5.1. Estimate the % of DAILY use of Catalan (Escriu el % d'ús diari del català): *

Type in the number from 0 to 100. Please, make sure that the percentages in the next 3 questions add up to 100. For example: Catalan 70, Spanish 20 and English 10 = 100 (Comprova que la suma dels tres percentatges sigui 100).

5.2. Estimate the % of DAILY use of Spanish (Escriu el % d'ús diari del castellà): *

5.3. Estimate the % of DAILY use of English (Escriu el % d'ús diari de l'anglès): *

English Learning Experience (Experiència en l'aprenentatge de l'anglès)

6. Age at which you started learning English (A quina edat vas començar a aprendre l'anglès?) *

7. How many hours a week do you study English in high school (Quantes hores setmanals fas d'anglès al Batxillerat)? *

For example: 2 hours

8. How many hours a week do you study English outside the school? Do you go to an English academy? (Quantes hores semanals estudies anglès fora de l'escola? Vas a una acadèmia d'anglès?)

9. Do you have any Certificate of English level (Tens un certificat de nivell d'anglès)? *

If YES, please, specify whether it is First Certificate/Advanced/Proficieny. If NO, type in NO. (Si el tens, especifica si és First/Advanced/Proficiency. Si no, escriu NO).

10. Estimate the number of hours spent WEEKLY speaking English with Non-Native Speakers of English (Escriu el número d'hores setmanals parlant anglès amb no nadius): *

For example, 14 hours

11. Estimate the number of hours spent WEEKLY speaking English with Native Speakers of English (Escriu el número d'hores setmanals parlant anglès amb nadius): *

For example, 14 hours

12. Estimate the % of exposure to British and American English (escriu el % d'exposició a l'anglès britànic i americà): *

For example: British 50% - American 50%

- 0%/100%
- 10%/90%
- 20%/80%
- 30%/70%
- 40%/60%
- 50%/50%
- 60%/40%
- 70%/30%
- 80%/20%
- 90%/10%
- 100%/0%

13. Do you think your pronunciation is more British- or American-like (creus que la teva pronunciació és més britànica o americana)? *

- British (britànica)
- American (americana)

14. For each of the items below, choose the response that corresponds to the amount of time you spend on average doing each activity in English (per cada un dels ítems, selecciona la resposta que correspon al temps que passes fent cada activitat en anglès). *

Please, use this scale 1 - never (mai); 2 - a few times a year (uns quants cops a l'any); 3 - monthly (cada mes); 4 - weekly (cada setmana); 5 - daily (cada dia)

	1	2	3	4	5
Watching TV series in English (Mirar series de televisió en anglès)					
Reading newspapers/magazines in English (Llegir diaris/revistes en anglès)					
Reading books in English (Llegir llibres en anglès)					
Listening to songs in English (Escoltar cançons en anglès)					
Watching movies or videos in English (Veure pel·lícules o vídeos en anglès)					
Speaking English with native or fluent speakers (Parlar anglès amb parlants nadius o quasi nadius)					
Speaking English with non-native speakers (Parlar anglès amb no-nadius)					
Writing e-mails/ letters in English (Escriure e-mails/ cartes en anglès)					

15. FAMILY. Please, estimate how many hours per WEEK you use English with your parents/relatives (si us plau, anota quantes hores setmanals utilitzes anglès amb els teus pares/familiars): *

16. Rate your command of English on the scale: 1=VERY POOR and 9=NEAR-NATIVE (evalua el teu nivell d'anglès en l'escala: 1=MOLT POBRE i 9=NADIU): *

	1	2	3	4	5	6	7	8
Reading (Comprensió escrita):	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Listening (Comprensió oral):	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Speaking (Parlar):	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Writing (Escriure):	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Pronunciation (Pronúncia):	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

17.1. Have you ever lived in an English-speaking country more than 2 weeks (has viscut en un país de parla anglesa durant més de 2 setmanes)? *

- ⬭ Yes
- ⬭ No

17.2. If YES, please, specify how many times, where, when, for how long and the purpose of the stay (Si és així, si us plau, especifica quantes vegades, on, quan, durant quant temps i l'objectiu de l'estada): *

Appendix I. Native speakers' language background questionnaire

*Obligatory

Personal information

1. Name *

2. Surname *

3. Sex *

- Male
- Female

4. Date of birth (XX/XX/XXXX) *

5. Place of birth (e.g. London, UK) *

6. Place of residence (e.g. Barcelona, Spain) *

7. Occupation *

8. Education (e.g. BA, MA, PhD, CELTA, other courses...) *

Linguistic Background

9. Mother tongue (L1) (E.g. English) *

10. Language(s) you speak on a DAILY basis (%). [E.g. -> Catalan (70%), English (30%)] -the sum must be 100%- *

11. Language(s) you speak with your family (%). [E.g. -> English (100%)] -the sum must be 100%- *

12. Foreign Languages (L2, L3, L4...) from more to less dominant (E.g. Italian (B2), French (A2)) *

13. Age of Learning of these Foreign Languages (E.g. Italian/18, French/12) *

14. Foreign Language use [hours/week]. E.g. Italian (10h), French (5h) *

15. Do you have any certificate in Foreign Languages (E.g. Italian -> CELI 3)? *

16. Have you taken any courses about Applied Linguistics/Second Language Acquisition? *

- Yes
- No

17. Have you taken any courses about English Phonetics & Phonology? *

- Yes
- No

18. Do you have any hearing/speech disorders or vision problems? If so, which one(s)? *

Appendix J. Sample of the vocabulary knowledge scale questionnaire

Dear students,

This is NOT A TEST so you will not receive a mark out of these results.

1. READ THE WORD & click on one of the options.
2. BE HONEST! Tell me how familiar you are with the word.

Thank you.
*Compulsory to answer

1. Name *
2. Surnames *
3. Class *

CHEEK *

- I know what this word MEANS and I can USE it in a sentence.
- I know what this word MEANS, but I'm NOT SURE how to USE it.
- I've SEEN this word before, but I DON'T KNOW what it MEANS.
- I've NEVER SEEN this word before.

FEAST *

- I know what this word MEANS and I can USE it in a sentence.
- I know what this word MEANS, but I'm NOT SURE how to USE it.
- I've SEEN this word before, but I DON'T KNOW what it MEANS.
- I've NEVER SEEN this word before.

PEEL *

- I know what this word MEANS and I can USE it in a sentence.
- I know what this word MEANS, but I'm NOT SURE how to USE it.
- I've SEEN this word before, but I DON'T KNOW what it MEANS.
- I've NEVER SEEN this word before.

Appendix K. Post-intervention perceptions' questionnaire

Dear students,

Thanks a million for participating in my project about English pronunciation. I hope you have enjoyed it as much as I have :)

Once you answer this questionnaire and do the tests, we will give you a mark for your participation in the project.

It's been a great pleasure!

Ingrid

*Compulsory answers

1. Please write your NAME *

____

2. Please write your SURNAMES *

____

3. What CLASS are you in? (A quina classe estàs?)

- ◯ Ir Batx A
- ◯ Ir Batx B
- ◯ Ir Batx C

4. BEFORE doing the tasks, HOW IMPORTANT did you think it was to work on English pronunciation in class? (Abans de fer les tasques, quina importància li donaves a treballar la pronunciació a classe?) *

	1	2	3	4	5	6	7	
Not important at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely important

5. AFTER doing the tasks, HOW IMPORTANT do you think it is to work on English pronunciation in class? (Després de fer les tasques, quina importància li dones a treballar la pronunciació a classe?) *

	1	2	3	4	5	6	7	
Not important at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely important

6. AFTER this project, if you could choose, how many HOURS PER MONTH would you like to work on English pronunciation in class? (Si poguessis triar, quantes hores al mes treballaries la pronunciació a classe?) *

- ⬭ 0-1h
- ⬭ 1-2h
- ⬭ 2-3h
- ⬭ 3-4h
- ⬭ 4-5h
- ⬭ 5-10h
- ⬭ >10h

7. The PRE-TASKS -the listening + repeating words you did before the tasks in pairs- helped you learn new words. (Les pre-tasques et van ajudar a aprendre noves paraules.) *

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

8. The PRE-TASKS -the listening + repeating words you did before the tasks in pairs- helped you improve your pronunciation. (Les pre-tasques et van ajudar a millorar la pronunciació) *

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

9. How difficult did you find the LISTENING of the PRE-TASK to understand? (Com de difícils consideres que eren d'entendre els *listening* de les pre-tasques?) *

	1	2	3	4	5	6	7	
Extremely easy	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

10. I think the PRE-TASKS were INTERESTING and ENJOYABLE. (Considero que les pre-tasques eren interessants i entretingudes.) *

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

11. How much MENTAL EFFORT did you have to put to perform the TASKS? (Quant esforç mental vas haver de posar per resoldre les tasques?) *

	1	2	3	4	5	6	7	
Very low mental effort	◯	◯	◯	◯	◯	◯	◯	Very high mental effort

12. How DIFFICULT did you find the TASKS? (Com de difícils consideres que eren les tasques?) *

	1	2	3	4	5	6	7	
Extremely easy	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

13. I had to make an EFFORT to solve the TASKS due to... (Vaig haver de fer un esforç per resoldre les tasques degut a...) *

- ☐ TASK CONDITIONS: Les condicions de la tasca (les que apareixien en negreta)
- ☐ THE PRONUNCIATION: La pronunciació de les vocals
- ☐ BOTH THINGS: Ambdues coses per igual

14. I think the TASKS were INTERESTING and ENJOYABLE (Considero que les tasques eren interessants i entretingudes.) *

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

15. Which TASKS did you enjoy the most and the least? (Quines tasques vas gaudir més i quines menys?) *

*You can choose more than one

	I didn't do it (No la vaig fer)	I really disliked it	I disliked it	I liked it	I enjoyed it	I really enjoyed it
Task 1: Activities in London	☐	☐	☐	☐	☐	☐
Task 2: Accommodation in London	☐	☐	☐	☐	☐	☐
Task 3:Spotify playlist	☐	☐	☐	☐	☐	☐
Task 4: Suitcase	☐	☐	☐	☐	☐	☐
Task 5: Lunch in a restaurant	☐	☐	☐	☐	☐	☐
Task 6: City map	☐	☐	☐	☐	☐	☐
Task 7: Auction house	☐	☐	☐	☐	☐	☐
Task 8: Escape room	☐	☐	☐	☐	☐	☐
Task 9: Shopping centre	☐	☐	☐	☐	☐	☐
Task 10: History museum	☐	☐	☐	☐	☐	☐
Task 11: London's zoo	☐	☐	☐	☐	☐	☐
Task 12: Recipe	☐	☐	☐	☐	☐	☐
Task 13: Clothes	☐	☐	☐	☐	☐	☐
Task 14: Roleplay party	☐	☐	☐	☐	☐	☐
Task 15: Market	☐	☐	☐	☐	☐	☐

Task 16: Dinner in a pub	☐	☐	☐	☐	☐	☐
Task 17: Souvenirs	☐	☐	☐	☐	☐	☐
Task 18: Films	☐	☐	☐	☐	☐	☐
Task 19: Photos in the website	☐	☐	☐	☐	☐	☐
Task 20: Photos in the album	☐	☐	☐	☐	☐	☐

16. What makes a TASK more ENJOYABLE? (Què fa que una tasca sigui divertida?)

*You can choose more than one

- ☐ The topic (el tema)
- ☐ The images/drawings (les il·lustracions)
- ☐ The target words (les paraules que teníeu)
- ☐ The difficulty (la dificultat per resoldre la tasca)

17. I think the TASKS were REALISTIC and could have taken place in London. (Considero que les tasques eren realistes i podien haver tingut lloc a Londres.)

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

18. What did you do if you had PROBLEMS communicating with your classmates? (Què feies si tenies problemes per comunicar-te amb els teus companys?) *

If you didn't have any, please write "X"

19. What do you think of the IMAGES/DRAWINGS in the flashcards? (Què penses de les imatges als cartellets?) *

20. The POST-TASKS - the short “games” after the main task - helped you revise the meaning of the target words you had learned during the task. (Les post-tasques -els jocs curtets després de la tasca principal- et van ajudar a repassar el significat de les paraules que havies après durant la tasca.) *

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

21. The POST-TASKS - the short “games” after the main task - helped you revise the pronunciation of the target words you had learned during the task. (Les post-tasques -els jocs curtets després de la tasca principal- et van ajudar a repassar la pronunciació de les paraules que havies après durant la tasca.) *

	1	2	3	4	5	6	7	
Strongly disagree	◯	◯	◯	◯	◯	◯	◯	Strongly agree

22. I think the POST-TASKS -the short “games” after the main task- were INTERESTING and ENJOYABLE (Considero que les post-tasques -els jocs curtets després de la tasca principal- eren interessants i entretingudes.) *

	1	2	3	4	5	6	7	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly agree

23. Now that the project is over, how difficult do you think are the following sounds to PRONOUNCE? (Ara que el projecte ha acabat, com de difícil creus que són els següents sons de pronunciar): *

1: the easiest / 4: the most difficult (Give only one number for each word)

/i:/ (e.g. TEEN)

	1	2	3	4	5	6	7	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

/ɪ/ (e.g. TIN)

	1	2	3	4	5	6	7	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

/æ/ (e.g. CAP)

	1	2	3	4	5	6	7	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

/ʌ/ (e.g. CUP)

	1	2	3	4	5	6	7	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

24. After doing this project, I think I have improved my PRONUNCIATION in English. (Després de fer aquest projecte, crec que he millorat la meva pronunciació en anglès) *

	1	2	3	4	5	6	7	
Strongly disagree								Strongly agree

25. Indicate how much you think you have improved the pronunciation of /iː/ in "TEEN" (Indica quant creus que has millorat la pronunciació de /iː/ a "TEEN") *

1: no improvement at all (no he millorat gens la meva pronunciació)/ 7: a lot of improvement (he millorat molt la meva pronunciació)

	1	2	3	4	5	6	7	
Not at all								A lot

26. Indicate how much you think you have improved the pronunciation of /ɪ/ in "TIN" (Indica quant creus que has millorat la pronunciació de /ɪ/ a "TIN") *

1: no improvement at all (no he millorat gens la meva pronunciació)/ 7: a lot of improvement (he millorat molt la meva pronunciació)

	1	2	3	4	5	6	7	
Not at all								A lot

27. Indicate how much you think you have improved the pronunciation of /æ/ in "CAP" (Indica quant creus que has millorat la pronunciació de /æ/ a "CAP") *

1: no improvement at all (no he millorat gens la meva pronunciació)/ 7: a lot of improvement (he millorat molt la meva pronunciació)

	1	2	3	4	5	6	7	
Not at all								A lot

28. Indicate how much you think you have improved the pronunciation of /ʌ/ in "CUP" (Indica quant creus que has millorat la pronunciació de /ʌ / a "CUP") *

1: no improvement at all (no he millorat gens la meva pronunciació)/ 7: a lot of improvement (he millorat molt la meva pronunciació)

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	A lot

29. In general, what did you most LIKE/DISLIKE about doing the project? (En general, què et va agradar més i menys de fer aquest projecte?) *

30. Finally, what do you think you have IMPROVED after doing the tasks? (Finalment, què creus que has millorat després de fer les tasques?) *

THANK YOU FOR PARTICIPATING!

Appendix L. Outputs of statistical tests

**Note.* Reference categories indicated in brackets

Table L.1. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on accuracy scores in the **ABX task.**

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	.558	.112	4.960	<.001	.338	.779
Group	.530	.113	4.706	<.001	.309	.751
Time	.120	.059	2.036	.042	.004	.236
Group x Time	-.423	.072	-5.859	<.001	-.565	-.282

Table L.2. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on reaction times (RT) of correct responses in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	6.926	.027	255.805	<.001	6.873	6.979
Group	-.017	.032	-.524	.600	-.079	.046
Time	.117	.009	13.515	<.001	.100	.134
Group x Time	-.029	.010	-2.808	.005	-.049	-.009

Table L.3. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on accuracy scores in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.411	.117	12.111	<.001	1.183	1.640
Time	-.395	.063	-6.309	<.001	-.517	-.272
Contrast	-.610	.144	-4.230	<.001	-.893	-.327
Time x Contrast	.159	.084	1.893	.058	-.006	.324

Table L.4. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on reaction times (RT) of correct responses in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	6.890	.020	340.870	<.001	6.851	6.930
Time	.082	.008	10.984	<.001	.068	.097
Contrast	.039	.013	2.918	.004	.013	.065
Time x Contrast	.012	.011	1.079	.281	-.010	.033

Table L.5. Parameter estimates of the GLMM on the effects of Time (T1/T2), Word Type (taught/untaught) and their interaction on accuracy scores in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.055	.121	8.710	<.001	.818	1.292
Time	-.212	.059	-3.605	<.001	-.328	-.097
Word Type	.093	.153	.603	.547	-.209	.394

Time x Word Type	-.191	.083	-2.278	.023	-.355	-.027

Table L.6. Parameter estimates of the GLMM on the effects of Time (T1/T2), Word Type (taught/untaught) and their interaction on reaction times (RT) of correct responses in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	6.921	.020	335.962	<.001	6.880	6.961
Time	.082	.008	10.382	<.001	.067	.098
Word Type	-.022	.014	-1.517	.129	-.050	.006
Time x Word Type	.011	.011	1.002	.316	-.011	.033

Table L.7. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on accuracy scores in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.199	.101	11.897	<.001	1.001	1.396
Time [T1]	-.388	.044	-8.894	<.001	-.474	-.302
Time [T2]	-.080	.044	-1.801	.072	-.167	.007

Table L.8. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on reaction times (RT) of correct responses in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	6.848	.020	347.018	<.001	6.810	6.887
Time [T1]	.153	.006	26.607	<.001	.142	.164
Time [T2]	.063	.006	11.338	<.001	.052	.074

Table L.9. Parameter estimates of the GLMM on the effects of Item Type (test/filler) on accuracy scores in the **FLeC task**.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	2.422	0.2811	8.617	0.000	1.871	2.973
Item Type	-1.890	0.2544	-7.428	0.000	-2.389	-1.391

Table L.10. Parameter estimates of the GLMM on the effects of Item Type (test/filler) on reaction times (RT) of correct responses in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	1506.222	37.9880	39.650	0.000	1431.705	1580.740
ItemType	225.536	26.7549	8.430	0.000	173.054	278.019

Table L.11. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on accuracy scores in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	0.451	0.1608	2.805	0.005	0.136	0.766
Group	0.593	0.1227	4.832	0.000	0.352	0.833
Time	-0.084	0.1135	-0.737	0.461	-0.306	0.139
Group x Time	-0.497	0.1393	-3.567	0.000	-0.770	-0.224

Table L.12. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on reaction times (RT) of correct responses in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	7.447	0.0227	327.649	0.000	7.403	7.492
Group	-0.017	0.0214	-0.814	0.416	-0.059	0.025
Time	0.040	0.0117	3.438	0.001	0.017	0.063
Group x Time	-0.004	0.0140	-0.322	0.748	-0.032	0.023

Table L.13. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on accuracy scores in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	1.045	0.1893	5.520	0.000	0.674	1.416
Time	-0.696	0.1134	-6.134	0.000	-0.918	-0.473
Contrast	-0.033	0.2613	-0.124	0.901	-0.545	0.480
Time x Contrast	0.247	0.1607	1.535	0.125	-0.068	0.562

Table L.14. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on reaction times (RT) of correct responses in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	7.432	0.0236	315.346	0.000	7.386	7.478
Time	0.026	0.0113	2.325	0.020	0.004	0.049
Contrast	-0.004	0.0293	-0.151	0.880	-0.062	0.053
Time x Contrast	0.019	0.0157	1.183	0.237	-0.012	0.049

Table L.15. Parameter estimates of the GLMM on the effects of Time (T1/T2), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on accuracy scores in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	1.026	0.2610	3.932	0.000	0.514	1.538
Time	-0.883	0.1576	-5.606	0.000	-1.192	-0.574
Vowel [iː]	-0.297	0.3628	-0.818	0.413	-1.008	0.414
Vowel [ɪ]	0.286	0.3677	0.777	0.437	-0.435	1.007
Vowel [æ]	0.030	0.3658	0.082	0.935	-0.687	0.747

Time x Vowel [iː]	0.514	0.2209	2.327	0.020	0.081	0.947
Time x Vowel [ɪ]	0.338	0.2309	1.466	0.143	-0.114	0.791
Time x Vowel [æ]	0.394	0.2272	1.736	0.083	-0.051	0.840

Table L.16. Parameter estimates of the GLMM on the effects of Time (T1/T2), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on reaction times (RT) of correct responses in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	7.440	0.0321	231.917	0.000	7.377	7.503
Time	0.036	0.0161	2.262	0.024	0.005	0.068
Vowel [iː]	0.017	0.0426	0.394	0.693	-0.067	0.100
Vowel [ɪ]	-0.038	0.0424	-0.899	0.369	-0.121	0.045
Vowel [æ]	-0.015	0.0425	-0.348	0.728	-0.098	0.069
Time x Vowel [iː]	-0.025	0.0226	-1.115	0.265	-0.069	0.019
Time x Vowel [ɪ]	0.037	0.0217	1.710	0.087	-0.005	0.080
Time x Vowel [æ]	-0.019	0.0222	-0.863	0.388	-0.063	0.024

Table L.17. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on accuracy scores in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	1.418	0.1482	9.566	0.000	1.127	1.709
Time [T1]	-0.948	0.0859	-11.029	0.000	-1.116	-0.779
Time [T2]	-0.356	0.0887	-4.017	0.000	-0.530	-0.182

Table L.18. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on reaction times (RT) of correct responses in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	7.432	0.0182	409.311	0.000	7.396	7.467
Time [T1]	0.036	0.0078	4.638	0.000	0.021	0.052
Time [T2]	-0.001	0.0074	-0.070	0.944	-0.015	0.014

Table L.19. Parameter estimates of the GLMM on the effects of Item Type (test/filler) on accuracy scores in the **LD task**.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	2.107	0.2409	8.746	0.000	1.635	2.580
ItemType	-2.157	0.2401	-8.982	0.000	-2.627	-1.686

Table L.20. Parameter estimates of the GLMM on the effects of Item Type (test/filler) on reaction times (RT) of correct responses in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1447.141	28.4266	50.908	0.000	1391.341	1502.942
ItemType	65.031	26.7819	2.428	0.015	12.459	117.604

Table L.21. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on accuracy scores in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	-0.536	0.1616	-3.315	0.001	-0.852	-0.219
Group	0.760	0.1974	3.851	0.000	0.373	1.147
Time	-0.207	0.1953	-1.060	0.289	-0.590	0.176
Group x Time	-0.678	0.2382	-2.847	0.004	-1.145	-0.211

Table L.22. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on reaction times (RT) of correct responses in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	7.291	0.0163	448.324	0.000	7.259	7.322
Group	0.004	0.0197	0.220	0.826	-0.034	0.043
Time	0.019	0.0158	1.228	0.219	-0.012	0.050
Group x Time	0.008	0.0191	0.443	0.657	-0.029	0.046

Table L.23. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on accuracy scores in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	0.518	0.1553	3.336	0.001	0.214	0.823
Time	-1.095	0.1937	-5.655	0.000	-1.475	-0.716
Contrast	-0.486	0.1185	-4.100	0.000	-0.718	-0.253
Time x Contrast	0.300	0.1633	1.839	0.066	-0.020	0.621

Table L.24. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on reaction times (RT) of correct responses in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	7.285	0.0117	625.180	0.000	7.262	7.308
Time	0.037	0.0119	3.128	0.002	0.014	0.061
Contrast	0.021	0.0122	1.714	0.087	-0.003	0.045
Time x Contrast	-0.018	0.0178	-1.004	0.316	-0.053	0.017

Table L.25. Parameter estimates of the GLMM on the effects of Time (T1/T2), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on accuracy scores in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	0.226	0.1776	1.271	0.204	-0.123	0.574

Time	-1.175	0.2298	-5.111	0.000	-1.625	-0.724
Vowel [iː]	-0.263	0.1682	-1.564	0.118	-0.593	0.067
Vowel [ɪ]	-0.122	0.1675	-0.726	0.468	-0.450	0.207
Vowel [æ]	0.598	0.1700	3.520	0.000	0.265	0.932
Time x Vowel [iː]	0.411	0.2349	1.748	0.081	-0.050	0.871
Time x Vowel [ɪ]	0.343	0.2337	1.469	0.142	-0.115	0.802
Time x Vowel [æ]	0.118	0.2335	0.504	0.614	-0.340	0.576

Table L.26. Parameter estimates of the GLMM on the effects of Time (T1/T2), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on reaction times (RT) of correct responses in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	7.305	0.0150	485.843	0.000	7.275	7.334
Time	0.029	0.0188	1.537	0.125	-0.008	0.066
Vowel [iː]	-0.005	0.0181	-0.302	0.763	-0.041	0.030
Vowel [ɪ]	0.007	0.0178	0.412	0.681	-0.028	0.042
Vowel [æ]	-0.034	0.0162	-2.126	0.034	-0.066	-0.003
Time x Vowel [iː]	-0.002	0.0266	-0.071	0.944	-0.054	0.050
Time x Vowel [ɪ]	-0.017	0.0263	-0.631	0.528	-0.068	0.035
Time x Vowel [æ]	0.016	0.0244	0.637	0.524	-0.032	0.063

Table L.27. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on accuracy scores in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	1.075	0.2014	5.337	0.000	0.680	1.470
Time [T1]	-1.760	0.2187	-8.048	0.000	-2.189	-1.331
Time [T2]	-0.717	0.2211	-3.244	0.001	-1.151	-0.284

Table L.28. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on reaction times (RT) of correct responses in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	7.281	0.0108	672.514	0.000	7.260	7.302
Time [T1]	0.045	0.0090	5.027	0.000	0.028	0.063
Time [T2]	0.016	0.0086	1.819	0.069	-0.001	0.032

Table L.29. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the **DWR task.**

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	13.811	1.6853	8.195	0.000	10.507	17.115
Group	2.925	2.0457	1.430	0.153	-1.085	6.935
Time	-0.121	0.9095	-0.133	0.894	-1.904	1.662

Group x Time	-2.703	1.0681	-2.531	0.011	-4.797	-0.609

Table L.30. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	20.173	2.3991	8.409	0.000	15.470	24.876
Group	-2.987	2.9060	-1.028	0.304	-8.683	2.710
Time	-1.362	1.8916	-0.720	0.472	-5.070	2.346
Group x Time	3.704	2.2286	1.662	0.097	-0.665	8.072

Table L.31. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	17.981	1.2702	14.157	0.000	15.491	20.471
Time	-3.906	0.6832	-5.717	0.000	-5.245	-2.566
Contrast	-1.362	0.6884	-1.979	0.048	-2.712	-0.013
Time x Contrast	2.142	0.9368	2.287	0.022	0.306	3.978

Table L.32. Parameter estimates of the GLMM on the effects of Time (T1/T2), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	6.404	2.2530	2.842	0.004	1.987	10.820
Time	4.981	2.2659	2.757	0.049	-0.461	8.423
Vowel [iː]	7.709	2.2337	3.451	0.001	3.330	12.087
Vowel [ɪ]	26.828	2.2337	12.010	0.000	22.449	31.207
Vowel [æ]	8.433	2.2337	3.775	0.000	4.054	12.811
Time x Vowel [iː]	-1.981	3.1589	-0.627	0.531	-8.174	4.211
Time x Vowel [ɪ]	-6.525	3.1589	-2.066	0.039	-12.718	-0.333
Time x Vowel [æ]	4.033	3.1589	0.644	0.020	-4.160	8.225

Table L.33. Parameter estimates of the GLMM on the effects of Time (T1/T2), Word Type (taught/untaught) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI*	
					Lower	*Upper*
Intercept	17.013	1.1327	15.020	0.000	14.793	19.233
Time	-2.566	0.6774	-3.789	0.000	-3.894	-1.238
Word Type	-0.263	0.6581	-0.399	0.690	-1.553	1.028
Time x Word Type	-0.522	0.9274	-0.563	0.573	-2.340	1.296

Table L.34. Parameter estimates of the GLMM on the effects of Time (T1/T2), Word Type (taught/untaught) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	14.963	1.3234	11.306	0.000	12.368	17.557
Time	4.271	1.3771	3.101	0.002	1.571	6.970
Word Type	1.310	1.3542	0.967	0.333	-1.345	3.965
Time x Word Type	-4.009	1.8957	-2.115	0.035	-7.725	-0.292

Table L.35. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	18.458	1.1985	15.401	0.000	16.108	20.807
Time [T1]	-3.697	0.5159	-7.166	0.000	-4.709	-2.686
Time [T2]	-0.984	0.4972	-1.980	0.048	-1.959	-0.010

Table L.36. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	13.528	0.9324	14.508	0.000	11.700	15.356
Time [T1]	5.055	0.9771	5.173	0.000	3.139	6.970
Time [T2]	2.804	0.9530	2.943	0.003	0.936	4.673

Table L.37. Parameter estimates of the LMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on duration ratio of /iː-ɪ/ in the DWR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.179	0.025	178	46.487	0.000	1.129	1.229
Group	0.059	0.030	178	1.929	0.055	-0.001	0.119
Time	-0.019	0.036	178	-0.521	0.603	-0.089	0.052
Group x Time	-0.067	0.043	178	-1.552	0.122	-0.152	0.018

Table L.38. Parameter estimates of the LMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on duration ratio of /æ-ʌ/ in the DWR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.195	0.022	178	54.330	0.000	1.152	1.239
Group	0.045	0.026	178	1.707	0.090	-0.007	0.097
Time	-0.008	0.031	178	-0.257	0.798	-0.069	0.053
Group x Time	-0.029	0.037	178	-0.787	0.432	-0.103	0.044

Table L.39. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the **DSR task.**

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	10.773	1.7872	6.201	0.000	7.367	14.178
Group	8.269	2.5091	3.921	0.000	4.135	12.404
Time	1.274	0.9583	1.403	0.161	-0.506	3.054
Group x Time	-7.464	1.5671	-7.001	0.000	-9.554	-5.374

Table L.40. Parameter estimates of the GLMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	27.033	2.0501	13.316	0.000	23.054	31.012
Group	-5.592	2.6423	-2.290	0.022	-10.380	-0.805
Time	-2.033	1.9531	-1.097	0.273	-5.666	1.599
Group x Time	6.629	2.3567	3.032	0.002	2.343	10.916

Table L.41. Parameter estimates of the GLMM on the effects of Time (T1/T2), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	18.164	1.5550	14.027	0.000	15.626	20.703
Time	-5.660	0.4931	-8.286	0.000	-6.999	-4.321
Contrast	1.799	0.7055	2.613	0.009	0.449	3.149
Time x Contrast	-1.058	0.9965	-1.130	0.259	-2.895	0.778

Table L.42. Parameter estimates of the GLMM on the effects of Time (T1/T2), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	8.586	2.8966	3.072	0.002	3.106	14.066
Time	6.572	2.3504	2.869	0.004	2.082	11.062
Vowel [iː]	8.710	3.9854	2.983	0.003	2.987	14.433
Vowel [ɪ]	26.690	4.2511	11.599	0.000	22.179	31.201
Vowel [æ]	11.246	3.6669	4.381	0.000	6.214	16.278
Time x Vowel [iː]	-1.830	3.6980	-0.572	0.567	-8.099	4.439
Time x Vowel [ɪ]	-6.019	3.1851	-1.914	0.056	-12.184	0.147
Time x Vowel [æ]	0.003	3.1774	0.001	0.999	-6.204	6.210

Table L.43. Parameter estimates of the GLMM on the effects of Time (T1/T2), Word Type (taught/untaught) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	19.030	1.1921	16.362	0.000	16.750	21.310

Time	-5.863	0.7366	-8.637	0.000	-7.194	-4.532
Word Type	-0.162	0.7855	-0.245	0.806	-1.455	1.131
Time x Word Type	-0.651	0.9922	-0.701	0.484	-2.473	1.171

Table L.44. Parameter estimates of the GLMM on the effects of Time (T1/T2), Word Type (taught/untaught) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	18.217	1.4457	13.688	0.000	15.608	20.826
Time	6.648	1.5012	4.814	0.000	3.941	9.356
Word Type	1.360	1.4883	1.002	0.317	-1.302	4.023
Time x Word Type	-4.029	1.4589	-2.120	0.034	-7.756	-0.303

Table L.45. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	18.551	1.2522	15.000	0.000	16.127	20.975
Time [T1]	-5.051	0.6274	-9.754	0.000	-6.067	-4.036
Time [T2]	0.944	0.3322	1.893	0.058	-0.034	1.923

Table L.46. Parameter estimates of the GLMM on the effects of Time (T1/T2/T3) on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DSR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	17.881	1.1299	17.723	0.000	15.903	19.858
Time [T1]	5.746	1.0670	5.503	0.000	3.699	7.792
Time [T2]	1.211	1.0247	1.189	0.234	-0.784	3.206

Table L.47. Parameter estimates of the LMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on duration ratio of /iː-ɪ/ in the DSR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.205	0.030	178	40.720	0.000	1.146	1.263
Group	0.044	0.036	178	1.229	0.221	-0.026	0.114
Time	-0.024	0.042	178	-0.566	0.572	-0.106	0.059
Group x Time	-0.054	0.050	178	-1.070	0.286	-0.153	0.045

Table L.48. Parameter estimates of the LMM on the effects of Group (experimental/control), Time (T1/T2) and their interaction on duration of /æ-ʌ/ in the DSR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.217	0.025	178	48.594	0.000	1.168	1.267
Group	0.041	0.030	178	1.352	0.178	-0.019	0.100
Time	-0.010	0.035	178	-0.287	0.775	-0.080	0.060
Group x Time	-0.024	0.043	178	-0.572	0.568	-0.108	0.060

Table L.49. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on accuracy scores in the **ABX task.**

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.697	0.1484	11.438	0.000	1.406	1.988
Group	-0.230	0.1505	-1.528	0.127	-0.525	0.065
Time [T1]	-0.661	0.0958	-6.897	0.000	-0.849	-0.473
Time [T2]	-0.353	0.0974	-3.630	0.000	-0.544	-0.163
Contrast	-0.755	0.1723	-4.378	0.000	-1.092	-0.417
Group x Time [T1]	0.228	0.1319	1.728	0.084	-0.031	0.487
Group x Time [T2]	0.422	0.1364	3.092	0.002	0.154	0.689
Group [simple] x Time [T1] x Contrast	0.191	0.1270	1.506	0.132	-0.058	0.440
Group [simple] x Time [T2] x Contrast	-0.054	0.1311	-0.409	0.683	-0.310	0.203
Group [complex] x Time [T1] x Contrast	0.397	0.1269	3.131	0.002	0.149	0.646
Group [complex] x Time [T2] x Contrast	0.300	0.1283	2.338	0.019	0.049	0.552

Table L.50. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on reaction times (RT) of correct responses in the ABX task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	-0.081	12636.1975	0.000	1.000	-24768.991	24768.828
Group	6.931	12636.1975	0.001	1.000	-24761.979	24775.840
Time [T1]	0.182	0.0115	15.876	0.000	0.160	0.205
Time [T2]	0.092	0.0111	8.254	0.000	0.070	0.114
Contrast	0.049	0.0173	2.829	0.005	0.015	0.083
Group x Time [T1]	-0.064	0.0162	-3.933	0.000	-0.095	-0.032
Group x Time [T2]	-0.052	0.0157	-3.334	0.001	-0.083	-0.022
Group [simple] x Time [T1] x Contrast	0.008	0.0170	0.481	0.630	-0.025	0.042
Group [simple] x Time [T2] x Contrast	-0.015	0.0164	-0.934	0.350	-0.047	0.017
Group [complex] x Time [T1] x Contrast	-0.005	0.0167	-0.281	0.778	-0.037	0.028
Group [complex] x Time [T2] x Contrast	-0.004	0.0162	-0.235	0.814	-0.035	0.028

Table L.51. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on accuracy scores in the **FLeC task.**

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	2.392	0.3619	6.608	0.000	1.682	3.101
Group	-1.795	0.3346	-5.366	0.000	-2.451	-1.139
Time [T1]	-1.966	0.3132	-6.276	0.000	-2.580	-1.352
Time [T2]	-1.131	0.3250	-3.479	0.001	-1.768	-0.494
Vowel [iː]	-0.419	0.4705	-0.891	0.373	-1.342	0.503
Vowel [ɪ]	0.685	0.5259	1.302	0.193	-0.346	1.716
Vowel [æ]	0.053	0.4865	0.109	0.913	-0.901	1.007

Group [simple] x Time [T1]	1.234	0.3955	3.120	0.002	0.459	2.009
Group [simple] x Time [T2]	1.402	0.4100	3.419	0.001	0.598	2.206
Group [simple] x Time [T1] x Vowel [iː]	1.100	0.3977	2.767	0.006	0.321	1.880
Group [simple] x Time [T1] x Vowel [ɪ]	0.288	0.4641	0.622	0.534	-0.621	1.198
Group [simple] x Time [T1] x Vowel [æ]	0.608	0.4175	1.456	0.145	-0.211	1.427
Group [simple] x Time [T2] x Vowel [iː]	0.126	0.4026	0.312	0.755	-0.664	0.915
Group [simple] x Time [T2] x Vowel [ɪ]	-0.209	0.4743	-0.440	0.660	-1.139	0.721
Group [simple] x Time [T2] x Vowel [æ]	0.067	0.4255	0.158	0.874	-0.767	0.902
Group [simple] x Time [T3] x Vowel [iː]	0.456	0.4015	1.136	0.256	-0.331	1.243
Group [simple] x Time [T3] x Vowel [ɪ]	-0.138	0.4699	-0.295	0.768	-1.060	0.783
Group [simple] x Time [T3] x Vowel [æ]	0.209	0.4227	0.495	0.620	-0.619	1.038
Group [complex] x Time [T1] x Vowel [iː]	0.180	0.3954	0.455	0.649	-0.595	0.955
Group [complex] x Time [T1] x Vowel [ɪ]	-0.374	0.4624	-0.810	0.418	-1.281	0.532
Group [complex] x Time [T1] x Vowel [æ]	0.157	0.4172	0.375	0.707	-0.661	0.974
Group [complex] x Time [T2] x Vowel [iː]	0.067	0.4094	0.163	0.870	-0.736	0.870
Group [complex] x Time [T2] x Vowel [ɪ]	-0.584	0.4776	-1.222	0.222	-1.520	0.353
Group [complex] x Time [T2] x Vowel [æ]	-0.124	0.4314	-0.288	0.773	-0.970	0.721

Table L.52. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on reaction times (RT) of correct responses in the FLeC task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	7.435	0.0221	336.429	0.000	7.391	7.478
Group	-0.003	0.0320	-0.093	0.926	-0.066	0.060
Time [T1]	0.041	0.0231	1.765	0.078	-0.005	0.086
Time [T2]	0.006	0.0215	0.302	0.763	-0.036	0.049
Vowel [iː]	0.004	0.0225	0.171	0.864	-0.040	0.048
Vowel [ɪ]	-0.034	0.0212	-1.602	0.109	-0.076	0.008
Vowel [æ]	0.004	0.0220	0.176	0.861	-0.039	0.047
Group [simple] x Time [T1]	0.015	0.0346	0.440	0.660	-0.053	0.083
Group [simple] x Time [T2]	-0.001	0.0315	-0.024	0.981	-0.062	0.061
Group [simple] x Time [T1] x Vowel [iː]	0.001	0.0341	0.022	0.983	-0.066	0.068
Group [simple] x Time [T1] x Vowel [ɪ]	0.036	0.0331	1.082	0.279	-0.029	0.101

Group [simple] x Time [T1] x Vowel [æ]	-0.016	0.0340	-0.465	0.642	-0.083	0.051
Group [simple] x Time [T2] x Vowel [iː]	0.008	0.0321	0.237	0.813	-0.055	0.071
Group [simple] x Time [T2] x Vowel [ɪ]	-0.003	0.0304	-0.088	0.929	-0.062	0.057
Group [simple] x Time [T2] x Vowel [æ]	-0.008	0.0314	-0.246	0.806	-0.069	0.054
Group [simple] x Time [T3] x Vowel [iː]	0.025	0.0324	0.757	0.449	-0.039	0.088
Group [simple] x Time [T3] x Vowel [ɪ]	0.003	0.0310	0.099	0.921	-0.058	0.064
Group [simple] x Time [T3] x Vowel [æ]	0.005	0.0318	0.147	0.883	-0.058	0.067
Group [complex] x Time [T1] x Vowel [iː]	-0.050	0.0334	-1.510	0.131	-0.116	0.015
Group [complex] x Time [T1] x Vowel [ɪ]	0.011	0.0315	0.343	0.731	-0.051	0.073
Group [complex] x Time [T1] x Vowel [æ]	-0.063	0.0324	-1.956	0.051	-0.127	0.000
Group [complex] x Time [T2] x Vowel [iː]	0.013	0.0310	0.425	0.671	-0.048	0.074
Group [complex] x Time [T2] x Vowel [ɪ]	-0.012	0.0298	-0.401	0.689	-0.070	0.046
Group [complex] x Time [T2] x Vowel [æ]	-0.024	0.0306	-0.770	0.441	-0.083	0.036

Table L.53. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on accuracy scores in the **LD task.**

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.911	0.3607	5.298	0.000	1.204	2.618
Group	-1.664	0.4750	-3.503	0.000	-2.595	-0.733
Time [T1]	-2.812	0.4072	-6.905	0.000	-3.610	-2.013
Time [T2]	-1.289	0.4185	-3.079	0.002	-2.109	-0.468
Vowel [iː]	-0.294	0.3158	-0.930	0.353	-0.913	0.325
Vowel [ɪ]	-0.146	0.3160	-0.461	0.645	-0.765	0.474
Vowel [æ]	0.633	0.3263	1.941	0.052	-0.006	1.273
Group [simple] x Time [T1]	1.532	0.5453	2.809	0.005	0.463	2.601
Group [simple] x Time [T2]	1.039	0.5520	1.882	0.060	-0.043	2.122
Group [simple] x Time [T1] x Vowel [iː]	0.515	0.3919	1.313	0.189	-0.254	1.283
Group [simple] x Time [T1] x Vowel [ɪ]	0.360	0.3920	0.918	0.359	-0.409	1.128
Group [simple] x Time [T1] x Vowel [æ]	0.186	0.3978	0.468	0.640	-0.594	0.966
Group [simple] x Time [T2] x Vowel [iː]	-0.026	0.3908	-0.067	0.946	-0.793	0.740
Group [simple] x Time [T2] x Vowel [ɪ]	-0.013	0.3900	-0.032	0.974	-0.777	0.752
Group [simple] x Time [T2] x Vowel [æ]	-0.083	0.3985	-0.208	0.836	-0.864	0.699
Group [simple] x Time [T3] x Vowel [iː]	0.071	0.3990	0.178	0.858	-0.711	0.853

Group [simple] x Time [T3] x Vowel [ɪ]	0.042	0.3986	0.105	0.917	-0.740	0.823
Group [simple] x Time [T3] x Vowel [æ]	0.226	0.4085	0.553	0.581	-0.575	1.027
Group [complex] x Time [T1] x Vowel [iː]	0.370	0.3931	0.940	0.347	-0.401	1.140
Group [complex] x Time [T1] x Vowel [ɪ]	0.380	0.3918	0.971	0.332	-0.388	1.148
Group [complex] x Time [T1] x Vowel [æ]	-0.004	0.3983	-0.010	0.992	-0.785	0.777
Group [complex] x Time [T2] x Vowel [iː]	0.069	0.4116	0.169	0.866	-0.738	0.876
Group [complex] x Time [T2] x Vowel [ɪ]	0.059	0.4112	0.145	0.885	-0.747	0.866
Group [complex] x Time [T2] x Vowel [æ]	0.100	0.4224	0.237	0.813	-0.728	0.928

Table L.54. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on reaction times (RT) of correct responses in the LD task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	6.358	0.2476	25.675	0.000	5.872	6.843
Group	-0.018	0.0308	-0.600	0.548	-0.079	0.042
Time [T1]	0.064	0.0265	2.409	0.016	0.012	0.116
Time [T2]	0.013	0.0250	0.500	0.617	-0.037	0.062
Vowel [iː]	0.136	0.0464	2.938	0.003	0.045	0.227
Vowel [ɪ]	0.116	0.0376	3.071	0.002	0.042	0.189
Vowel [æ]	-0.035	0.0247	-1.434	0.152	-0.084	0.013
Group [simple] x Time [T1]	-0.021	0.0378	-0.543	0.587	-0.095	0.054
Group [simple] x Time [T2]	0.018	0.0351	0.521	0.603	-0.051	0.087
Group [simple] x Time [T1] x Vowel [iː]	-0.001	0.0374	-0.025	0.980	-0.074	0.072
Group [simple] x Time [T1] x Vowel [ɪ]	-0.016	0.0376	-0.414	0.679	-0.089	0.058
Group [simple] x Time [T1] x Vowel [æ]	0.047	0.0346	1.346	0.178	-0.021	0.115
Group [simple] x Time [T2] x Vowel [iː]	0.003	0.0356	0.090	0.928	-0.067	0.073
Group [simple] x Time [T2] x Vowel [ɪ]	0.006	0.0357	0.161	0.872	-0.064	0.076
Group [simple] x Time [T2] x Vowel [æ]	0.040	0.0321	1.249	0.212	-0.023	0.103
Group [simple] x Time [T3] x Vowel [iː]	0.030	0.0359	0.850	0.396	-0.040	0.101
Group [simple] x Time [T3] x Vowel [ɪ]	0.027	0.0358	0.746	0.456	-0.043	0.097
Group [simple] x Time [T3] x Vowel [æ]	0.064	0.0323	1.987	0.047	0.001	0.128
Group [complex] x Time [T1] x Vowel [iː]	-0.002	0.0380	-0.043	0.966	-0.076	0.073
Group [complex] x Time [T1] x Vowel [ɪ]	-0.037	0.0375	-0.991	0.322	-0.111	0.036
Group [complex] x Time [T1] x Vowel [æ]	0.042	0.0349	1.211	0.226	-0.026	0.111

Group [complex] x Time [T2] x Vowel [iː]	0.012	0.0365	0.317	0.751	-0.060	0.083
Group [complex] x Time [T2] x Vowel [ɪ]	-0.015	0.0358	-0.409	0.682	-0.085	0.055
Group [complex] x Time [T2] x Vowel [æ]	0.022	0.0328	0.683	0.495	-0.042	0.087

Table L.55. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the **DWR task**.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	13.380	0.9692	13.805	0.000	11.480	15.279
Group	-1.297	1.4043	-0.923	0.356	-4.049	1.456
Time [T1]	-4.463	1.0443	-4.274	0.000	-6.510	-2.416
Time [T2]	-1.138	1.0126	-1.124	0.261	-3.123	0.847
Contrast	3.165	0.9993	3.168	0.002	1.207	5.124
Group x Time [T1]	1.189	1.4630	0.813	0.416	-1.679	4.057
Group x Time [T2]	2.198	1.4406	1.525	0.127	-0.626	5.021
Group [simple] x Time [T1] x Contrast	-0.903	1.4246	-0.634	0.526	-3.695	1.890
Group [simple] x Time [T2] x Contrast	-4.263	1.4246	-2.992	0.003	-7.056	-1.470
Group [complex] x Time [T1] x Contrast	0.188	1.4133	0.133	0.894	-2.582	2.959
Group [complex] x Time [T2] x Contrast	-0.841	1.4133	-0.595	0.552	-3.612	1.929

Table L.56. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DWR task.

Predictor	*β*	*SE*	*t*	*p*	*95%CI* Lower	*95%CI* Upper
Intercept	6.813	2.6730	2.549	0.011	1.574	12.053
Group	0.246	3.8728	0.064	0.949	-7.345	7.838
Time [T1]	2.513	3.0017	0.837	0.402	-3.371	8.397
Time [T2]	-0.266	2.9483	-0.090	0.928	-6.045	5.513
Vowel [iː]	5.313	2.9265	1.816	0.069	-0.424	11.050
Vowel [ɪ]	6.358	2.9265	2.172	0.030	0.621	12.094
Vowel [æ]	4.917	2.9265	1.680	0.093	-0.820	10.654
Group [simple] x Time [T1]	2.357	4.2357	0.556	0.578	-5.946	10.660
Group [simple] x Time [T2]	-0.450	4.1980	-0.107	0.915	-8.679	7.779
Group [simple] x Time [T1] x Vowel [iː]	1.523	4.1719	0.365	0.715	-6.655	9.700
Group [simple] x Time [T1] x Vowel [ɪ]	7.976	4.1720	1.912	0.056	-0.202	16.154
Group [simple] x Time [T1] x Vowel [æ]	5.762	4.1719	1.381	0.167	-2.416	13.940
Group [simple] x Time [T2] x Vowel [iː]	5.154	4.1719	1.235	0.217	-3.024	13.331
Group [simple] x Time [T2] x Vowel [ɪ]	20.644	4.1719	4.948	0.000	12.466	28.822

Group [simple] x Time [T2] x Vowel [æ]	6.189	4.1719	1.484	0.138	-1.989	14.367
Group [simple] x Time [T3] x Vowel [iː]	3.505	4.1719	0.840	0.401	-4.673	11.683
Group [simple] x Time [T3] x Vowel [ɪ]	16.351	4.1719	3.919	0.000	8.174	24.529
Group [simple] x Time [T3] x Vowel [æ]	6.757	4.1719	1.620	0.105	-1.421	14.934
Group [complex] x Time [T1] x Vowel [iː]	-0.659	4.1387	-0.159	0.873	-8.772	7.453
Group [complex] x Time [T1] x Vowel [ɪ]	19.718	4.1387	4.764	0.000	11.606	27.831
Group [complex] x Time [T1] x Vowel [æ]	5.341	4.1387	1.291	0.197	-2.772	13.454
Group [complex] x Time [T2] x Vowel [iː]	-0.276	4.1387	-0.067	0.947	-8.389	7.837
Group [complex] x Time [T2] x Vowel [ɪ]	20.302	4.1387	4.906	0.000	12.190	28.415
Group [complex] x Time [T2] x Vowel [æ]	0.925	4.1387	0.224	0.823	-7.187	9.038

Table L.57. Parameter estimates of the LMM on the effects of Group (simple/complex), Time (T1/T2/T3) and their interaction on duration ratio of /iː-ɪ/ in the DWR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.289	0.026	183	50.068	0.000	1.238	1.340
Group	-0.041	0.037	183	-1.115	0.266	-0.113	0.031
Time [T1]	-0.162	0.036	183	-4.460	0.000	-0.234	-0.091
Time [T2]	-0.062	0.036	183	-1.711	0.089	-0.134	0.010
Group [simple] x Time [T1]	0.092	0.052	183	1.782	0.076	-0.010	0.195
Group [complex] x Time [T2]	0.063	0.052	183	1.214	0.226	-0.039	0.165

Table L.58. Parameter estimates of the LMM on the effects of Group (simple/complex), Time (T1/T2/T3) and their interaction on duration ratio of /æ-ʌ/ in the DWR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* Lower	Upper
Intercept	1.239	0.021	183	60.354	0.000	1.199	1.280
Group	0.031	0.029	183	1.043	0.298	-0.027	0.088
Time [T1]	-0.046	0.029	183	-1.594	0.113	-0.104	0.011
Time [T2]	0.002	0.029	183	0.066	0.947	-0.055	0.059
Group [simple] x Time [T1]	-0.010	0.041	183	-0.247	0.805	-0.092	0.071
Group [complex] x Time [T2]	-0.032	0.041	183	-0.777	0.438	-0.114	0.050

Table L.59. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Contrast (/iː-ɪ/, /æ-ʌ/) and their interaction on Mahalanobis distances between non-native contrastive vowels (i.e., distinctiveness) in the **DSR task**.

Predictor	β	SE	t	p	95%CI Lower	Upper
Intercept	13.254	0.7220	18.356	0.000	11.839	14.669
Group	-3.071	1.0293	-2.983	0.003	-5.089	-1.053
Time [T1]	-6.420	1.0211	-6.287	0.000	-8.421	-4.418
Time [T2]	-1.223	1.0211	-1.198	0.231	-3.225	0.779
Contrast	6.165	1.0211	6.038	0.000	4.164	8.167
Group x Time [T1]	3.636	1.4557	2.497	0.013	0.782	6.489
Group x Time [T2]	4.473	1.4557	3.073	0.002	1.619	7.326
Group [simple] x Time [T1] x Contrast	-3.902	1.4557	-2.681	0.007	-6.756	-1.049
Group [simple] x Time [T2] x Contrast	-4.263	1.4557	-2.929	0.003	-7.117	-1.410
Group [complex] x Time [T1] x Contrast	-2.801	1.4441	-1.940	0.052	-5.632	0.029
Group [complex] x Time [T2] x Contrast	-0.425	1.4441	-0.294	0.769	-3.255	2.406

Table L.60. Parameter estimates of the GLMM on the effects of Group (simple/complex), Time (T1/T2/T3), Vowel (/iː/, /ɪ/, /æ/, /ʌ/) and their interaction on Mahalanobis distances between native and non-native vowels (i.e., nativelikeness) in the DSR task.

Predictor	β	SE	t	p	95%CI Lower	Upper
Intercept	5.706	2.9178	1.955	0.051	-0.014	11.425
Group	7.118	4.2312	1.682	0.093	-1.176	15.412
Time [T1]	9.115	3.2477	2.806	0.005	2.748	15.481
Time [T2]	5.220	3.1892	1.637	0.102	-1.032	11.471
Vowel [iː]	5.261	3.1654	1.662	0.097	-0.944	11.466
Vowel [ɪ]	20.805	3.1654	6.573	0.000	14.600	27.010
Vowel [æ]	11.404	3.1654	3.603	0.000	5.199	17.609
Group [simple] x Time [T1]	-4.236	4.5823	-0.924	0.355	-13.219	4.746
Group [simple] x Time [T2]	-9.927	4.5411	-2.186	0.029	-18.829	-1.026
Group [simple] x Time [T1] x Vowel [iː]	1.575	4.5125	0.349	0.727	-7.271	10.420
Group [simple] x Time [T1] x Vowel [ɪ]	-6.463	4.5125	-1.432	0.152	-15.308	2.383
Group [simple] x Time [T1] x Vowel [æ]	-0.725	4.5125	-0.161	0.872	-9.570	8.121
Group [simple] x Time [T2] x Vowel [iː]	8.582	4.5125	1.902	0.057	-0.264	17.428
Group [simple] x Time [T2] x Vowel [ɪ]	9.196	4.5125	2.038	0.042	0.350	18.042
Group [simple] x Time [T2] x Vowel [æ]	2.729	4.5125	0.605	0.545	-6.117	11.575
Group [simple] x Time [T3] x Vowel [iː]	3.718	4.5125	0.824	0.410	-5.127	12.564
Group [simple] x Time [T3] x Vowel [ɪ]	1.904	4.5125	0.422	0.673	-6.942	10.749
Group [simple] x Time [T3] x Vowel [æ]	0.404	4.5125	0.090	0.929	-8.442	9.250
Group [complex] x Time [T1] x Vowel [iː]	-0.607	4.4766	-0.136	0.892	-9.382	8.168

Group [complex] x Time [T1] x Vowel [ɪ]	5.271	4.4766	1.177	0.239	-3.505	14.046
Group [complex] x Time [T1] x Vowel [æ]	-1.146	4.4766	-0.256	0.798	-9.921	7.629
Group [complex] x Time [T2] x Vowel [iː]	-4.094	4.4766	-0.915	0.360	-12.869	4.681
Group [complex] x Time [T2] x Vowel [ɪ]	1.854	4.4766	0.414	0.679	-6.921	10.629
Group [complex] x Time [T2] x Vowel [æ]	-4.702	4.4766	-1.050	0.294	-13.477	4.073

Table L.61. Parameter estimates of the LMM on the effects of Group (simple/complex), Time (T1/T2/T3) and their interaction on duration ratio of /iː-ɪ/ in the DSR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* *Lower*	*Upper*
Intercept	1.314	0.031	183	41.798	0.000	1.252	1.376
Group	-0.019	0.045	183	-0.423	0.673	-0.107	0.069
Time [T1]	-0.174	0.044	183	-3.915	0.000	-0.262	-0.086
Time [T2]	-0.077	0.044	183	-1.722	0.087	-0.164	0.011
Group [simple] x Time [T1]	0.082	0.063	183	1.288	0.199	-0.043	0.207
Group [complex] x Time [T2]	0.041	0.063	183	0.646	0.519	-0.084	0.166

Table L.62. Parameter estimates of the LMM on the effects of Group (simple/complex), Time (T1/T2/T3) and their interaction on duration ratio of /æ-ʌ/ in the DSR task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* *Lower*	*Upper*
Intercept	1.276	0.025	183	51.587	0.000	1.227	1.324
Group	0.064	0.035	183	1.819	0.071	-0.005	0.134
Time [T1]	-0.065	0.035	183	-1.858	0.065	-0.134	0.004
Time [T2]	-0.014	0.035	183	-0.414	0.679	-0.083	0.055
Group [simple] x Time [T1]	-0.037	0.050	183	-0.750	0.454	-0.136	0.061
Group [complex] x Time [T2]	-0.070	0.050	183	-1.402	0.163	-0.168	0.028

Table L.63. Parameter estimates of the LMM on the effects of Group (simple/complex) and Time (T1/T2/T3) on the number of **P-LRE** per minute in the oral interaction task.

Predictor	*β*	*SE*	*df*	*t*	*p*	*95%CI* *Lower*	*Upper*
Intercept	1.067	0.100	87	10.671	0.000	0.868	1.265
Group	-0.169	0.144	87	-1.175	0.243	-0.454	0.117
Time [T1]	-0.661	0.141	87	-4.677	0.000	-0.942	-0.380
Time [T2]	-0.154	0.141	87	-1.093	0.277	-0.435	0.126
Group x Time [T1]	0.044	0.203	87	0.215	0.830	-0.360	0.448
Group x Time [T2]	0.054	0.203	87	0.267	0.790	-0.350	0.458

Table L.64. Parameter estimates of the LMM on the effects of Group (simple/complex) and Time (T1/T2/T3) on the duration of P-LRE per minute in the oral interaction task.

Predictor	*β*	*SE*	df	*t*	*p*	*95%CI* *Lower*	*Upper*
Intercept	0.047	0.006	70	8.390	0.000	0.036	0.058
Group	-0.002	0.008	70	-0.212	0.833	-0.018	0.015

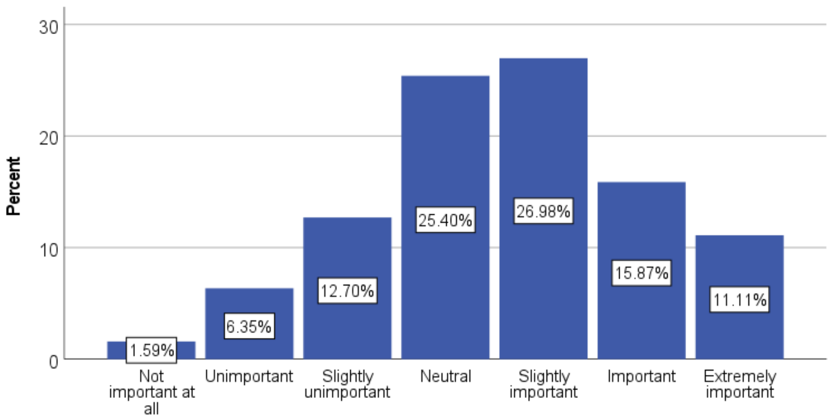
Time [T1]	0.014	0.009	70	1.493	0.140	-0.005	0.033
Time [T2]	0.004	0.008	70	0.543	0.589	-0.012	0.020
Group x Time [T1]	-0.016	0.014	70	-1.184	0.241	-0.043	0.011
Group x Time [T2]	-0.003	0.012	70	-0.244	0.808	-0.026	0.020

Table L.65. Standard multiple regression of individual differences (i.e., L2 proficiency, phonological short-term memory [PSTM] – Forward digit span, complex working memory [WM] – Backward digit span, auditory selective attention [ASA]) in ABX discrimination, FLeC, LD, DWR and DSR accuracy.

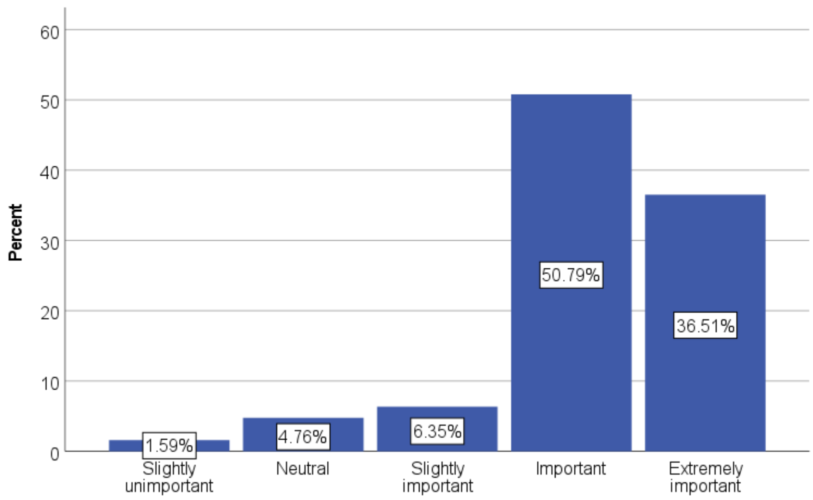
	R	*Adj. R²*	*F Change*	*Sig. F Change*	*B*	*SE B*	*β*	*p*
ABX	.65	.38	9.61	<.001				
Constant					.495	.056		
Proficiency					.001	.000	.210	.126
PSTM-Forward DS					.010	.011	.139	.346
Complex WM-Backward DS					.025	.008	.407	.004
ASA					.003	.001	.344	.009
FLeC	.64	.41	8.99	<.001				
Constant					.316	.077		
Proficiency					.001	.001	.157	.257
PSTM-Forward DS					.002	.015	.023	.877
Complex WM-Backward DS					.024	.012	.277	.047
ASA					.004	.001	.379	.005
LD	.55	.25	5.34	.001				
Constant					.279	.081		
Proficiency					.000	.001	.100	.507
PSTM-Forward DS					-.008	.015	-.082	.877
Complex WM-Backward DS					.019	.012	.237	.116
ASA					.004	.002	.400	.006
DWR	.74	.52	16.12	<.001				
Constant					24.29	4.73		
Proficiency					-.054	.041	-.158	.190
PSTM-Forward DS					-4.28	.892	-.620	<.001
Complex WM-Backward DS					-4.39	.709	-.733	<.001
ASA					-.009	.089	-.011	.922
DSR	.66	.39	9.96	<.001				
Constant					26.36	5.36		
Proficiency					-.071	.041	-.200	.134
PSTM-Forward DS					-3.27	1.10	-.471	.002
Complex WM-Backward DS					-3.90	.803	-.648	<.001
ASA					-.025	.101	-.031	.808

Appendix M. Results of post-intervention perceptions' questionnaire

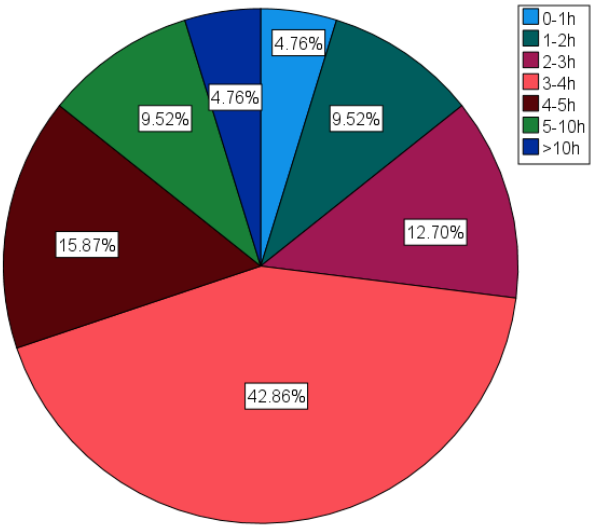
4. BEFORE doing the tasks, HOW IMPORTANT did you think it was to work on English pronunciation in class?



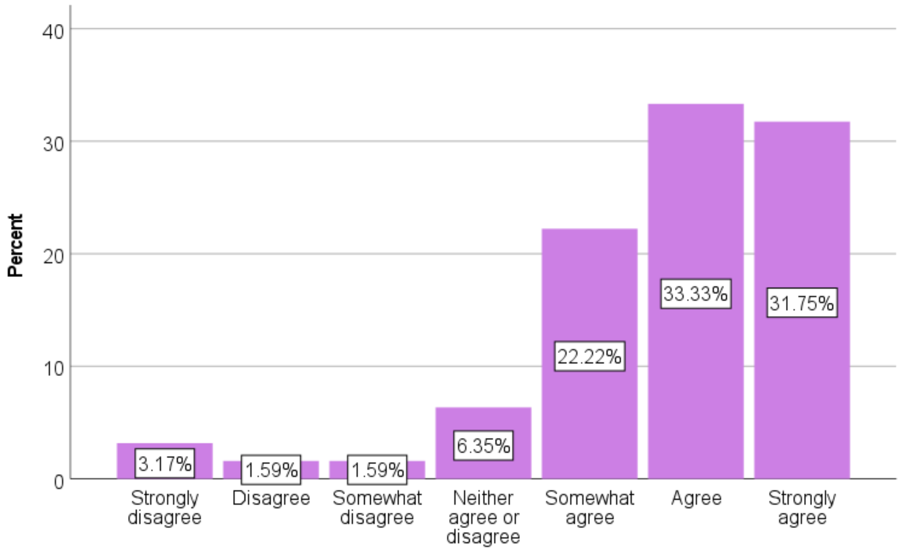
5. AFTER doing the tasks, HOW IMPORTANT do you think it is to work on English pronunciation in class?



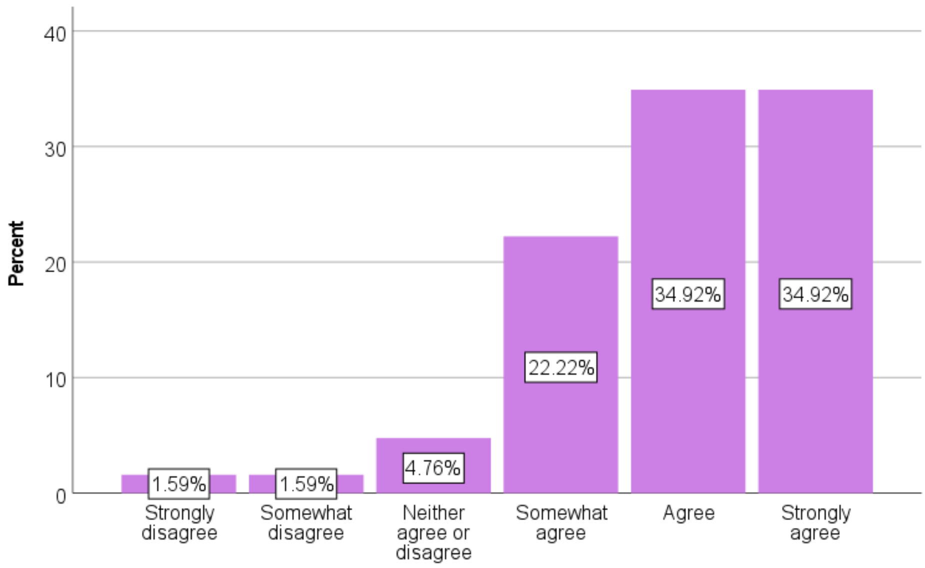
6. AFTER this project, if you could choose, how many HOURS PER MONTH would you like to work on English pronunciation in class?



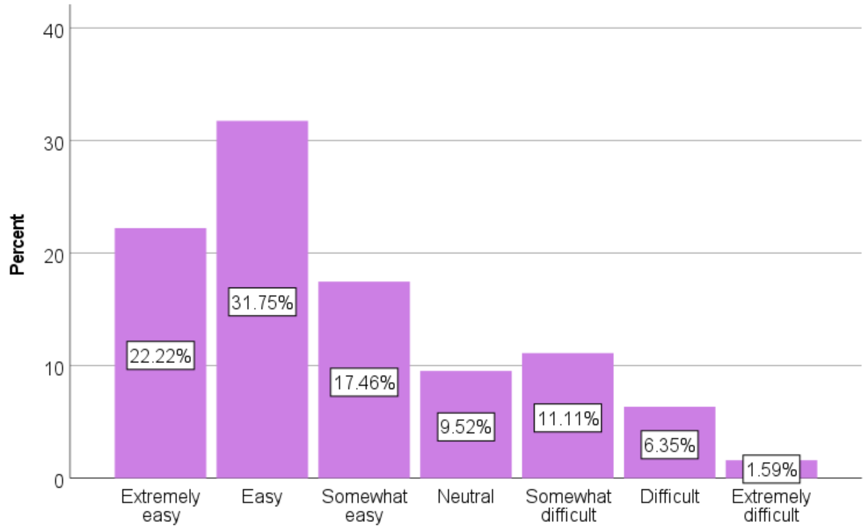
7. The PRE-TASKS -the listening + repeating words you did before the tasks in pairs- helped you learn new words.



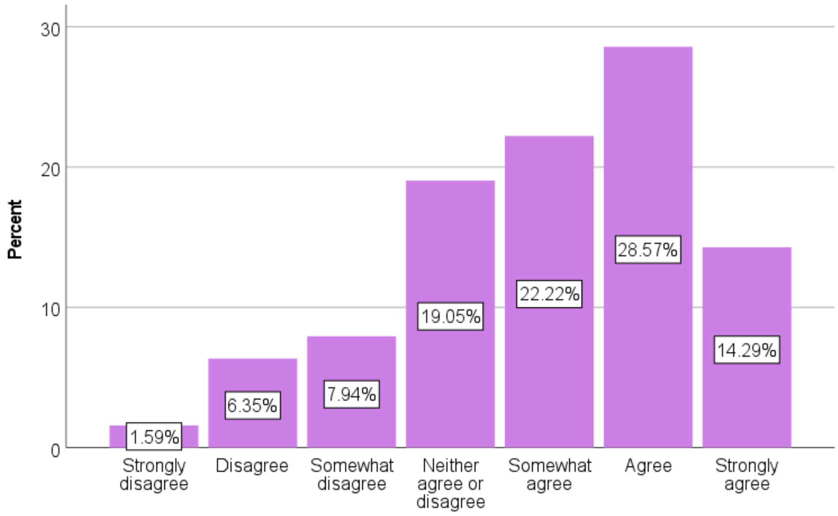
8. The PRE-TASKS -the listening + repeating words you did before the tasks in pairs- helped you improve your pronunciation.



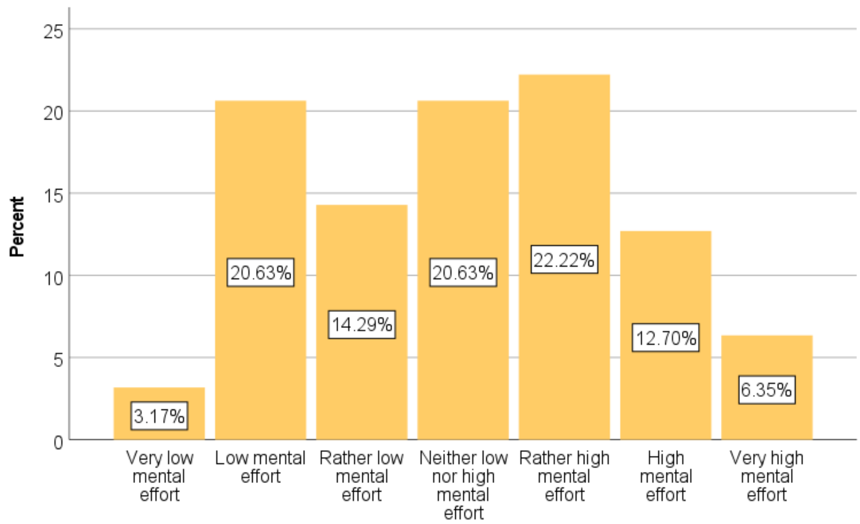
9. How difficult did you find the LISTENING of the PRE-TASK to understand?



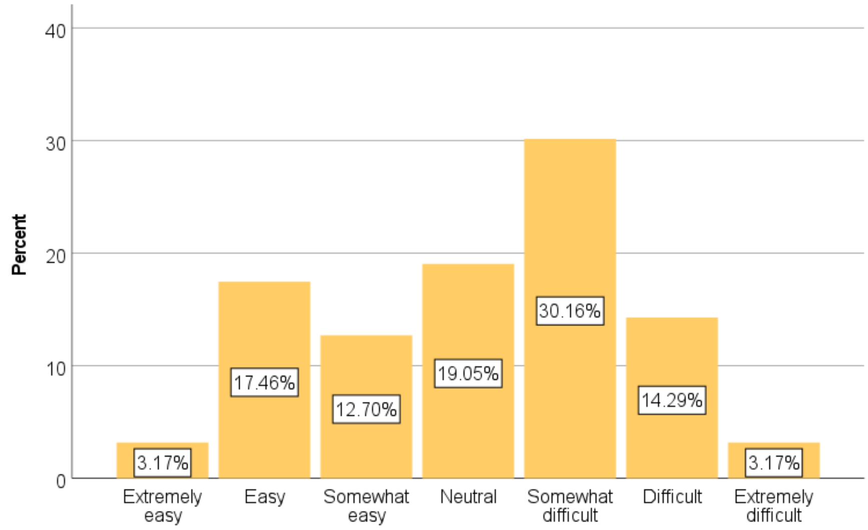
10. I think the PRE-TASKS were INTERESTING and ENJOYABLE.



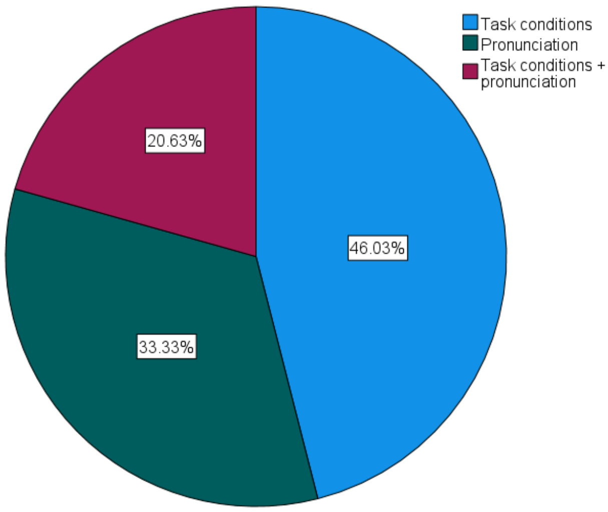
11. How much MENTAL EFFORT did you have to put to perform the TASKS?



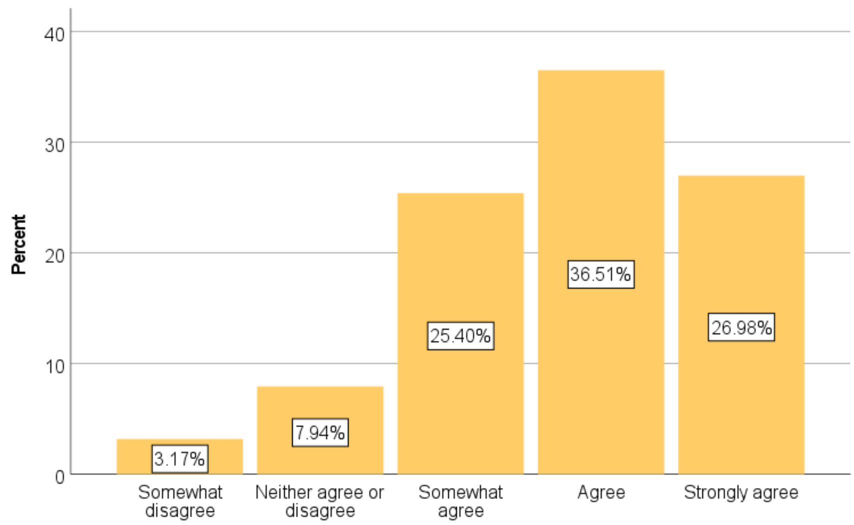
12. How DIFFICULT did you find the TASKS?



13. I had to make an EFFORT to solve the TASKS due to...



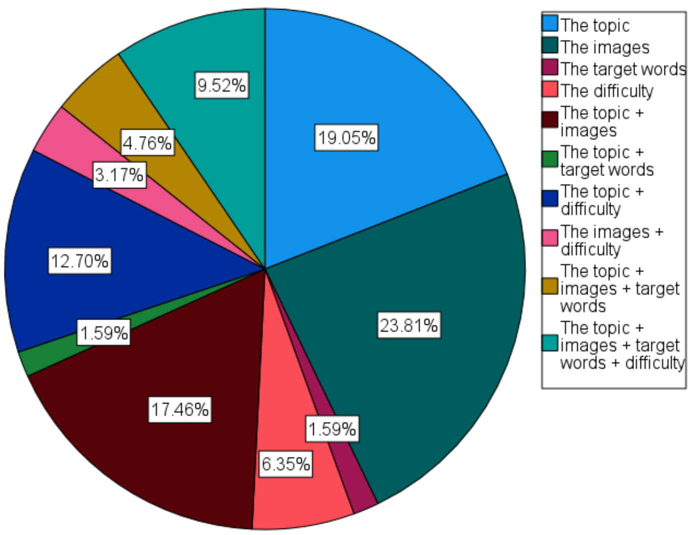
14. I think the TASKS were INTERESTING and ENJOYABLE



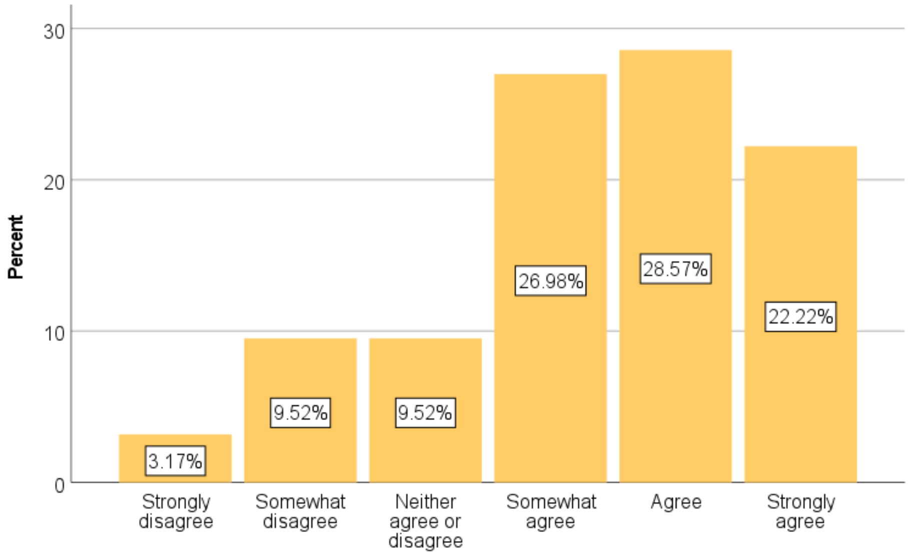
15. Which TASKS did you enjoy the most and the least? (%)

	I didn't do it	I really disliked it	I disliked it	I liked it	I enjoyed it	I really enjoyed it
Task 1: Activities in London	3.2	0.0	1.6	39.7	38.1	17.5
Task 2: Accommodation in London	0.0	0.0	1.6	46.0	34.9	17.5
Task 3: Spotify playlist	0.0	0.0	0.0	31.7	36.5	31.7
Task 4: Suitcase	0.0	0.0	1.6	41.3	36.5	20.6
Task 5: Lunch in a restaurant	0.0	0.0	3.2	41.3	36.5	19.0
Task 6: City map	0.0	0.0	4.8	33.3	38.1	23.8
Task 7: Auction house	3.2	0.0	1.6	39.7	39.7	15.9
Task 8: Escape room	1.6	0.0	3.2	28.6	41.3	25.4
Task 9: Shopping centre	0.0	1.6	6.3	30.2	42.9	19.0
Task 10: History museum	0.0	0.0	11.1	41.3	34.9	12.7
Task 11: London's zoo	1.6	0.0	4.8	39.7	34.9	19.0
Task 12: Recipe	0.0	0.0	6.3	33.3	36.5	23.8
Task 13: Clothes	0.0	0.0	4.8	30.2	46.0	19.0
Task 14: Roleplay party	1.6	1.6	1.6	34.9	42.9	17.5
Task 15: Market	1.6	1.6	3.2	36.5	38.1	19.0
Task 16: Dinner in a pub	0.0	0.0	4.8	34.9	31.7	28.6
Task 17: Souvenirs	0.0	0.0	4.8	30.2	44.4	20.6
Task 18: Films	1.6	0.0	1.6	30.2	38.1	28.6
Task 19: Photos in the website	0.0	0.0	0.0	23.8	46.0	30.2
Task 20: Photos in the album	0.0	0.0	0.0	20.6	22.2	57.1

16. What makes a TASK more ENJOYABLE?



17. I think the TASKS were REALISTIC and could have taken place in London.

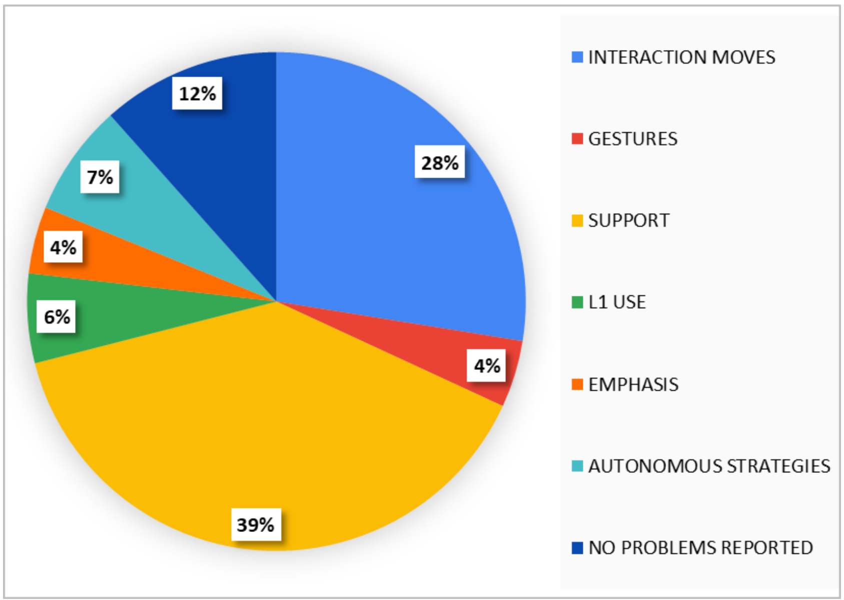


18. What did you do if you had PROBLEMS communicating with your classmates? (If you didn't have any, please write "X")

	Theme	*Code*	*Learners' response*
INTERACTION MOVES	REPETITION	S10	I **repeated** different times
		S14	say it loud and **repetitively** so she understands
		S44	**Repeat** the words or phrase a lot of times
		S54	I say, can you say again please, and i **repeat** too
		S06	I tried to find a solution by **repeating** or saying it in a different way
	CLARIFICATION	S61	**ask if I understood well** and if not he said it again
		S08	**Ask** again **more details**
		S69	**Define** the words and help my friend
		S72	**Ask** my partner to **repeat** it
		S54	I say, **can you say again please**, and i repeat too
		S52	say to my classmate if he can **say it again** or differently
		S05	Say **can you repeat** please?
		S48	Ask to **say again**
		S49	**ask to repeat** and find a solution together
	PARAPHRASING	S06	I tried to find a solution by repeating or **saying** it **in a different way**
		S09	I **defined** the words in a **similar way**
		S43	say in **a similar way** or use catalan if it was impossible to resolve
		S65	Look for **alternatives**
		S69	**Define** the words and help my friend
GESTURES		S04	**Use my hands** to help me say it
		S18	**Gestures** like party&co
		S41	**Made mimics** to help me communicate
SUPPORT	PEER	S27	**Preguntar** al **company** [Ask the classmate]
		S28	**Discuss** the pronuntiation with **classmate**
		S31	**Ask** to my **partner**
		S45	**Ask** my **partner**
		S47	**Ask** to the **partners**
		S62	**asked my friend** Tomeu to help me solve the task
		S02	I asked my mate to **say it again**
		S69	Define the words and **help my friend**

		S46	I did it wrong and then **my friend say it correctly** so helped me understand
		S66	**ask** my intelligent **friend** how to say the word
	TEACHER	S07	I would **ask Ingrid** to help me
		S11	**Ask ingrid**
		S13	I would **ask the teacher** for help
		S16	**Ask Ingrid**, but i hadn't many problems
		S17	**Ask the teacher**
		S20	**Ask** for help to my **teacher**
		S25	**Ask to Ingrid**
		S34	**Ask** to **Ingrid and Montse**
		S38	I **ask** the **teacher**
		S51	I **asked Ingrid.**
		S55	**Ask** the **teacher.**
		S59	ask to **Ingrid** or **Monse**
		S67	**Ask Ingrid**
		S68	**Ask Ingrid** to help me understand my classmate
		S70	**Ask Ingrid.**
	MATERIALS	S26	look at the **drawing on the board** to remember the pronunciation
		S58	**Watch the board** with the **example** of good pronunciation
L1 USE		S19	I say it in **Catalan**, sorry teacher
		S50	use **translator**
		S37	say in **spanish or catalan**
		S43	say in a similar way or use **catalan** if it was impossible to resolve
EMPHASIS		S12	I **exagerated** the pronunciation of the word so she could hear the difference
		S56	Try to **say the words slow** and with good pronunciation
		S14	**say it loud** and repetitively so she understands
AUTONOMOUS STRATEGIES		S30	I **improvised**
		S33	I stopped to **think**
		S36	Try to do it **working hard**
		S40	**take the flow** you know
		S53	me and my partner **try** to do it
NO PROBLEMS REPORTED		S15	X
		S23	X
		S29	X
		S32	X
		S39	X
		S60	X
		S64	X
		S71	X

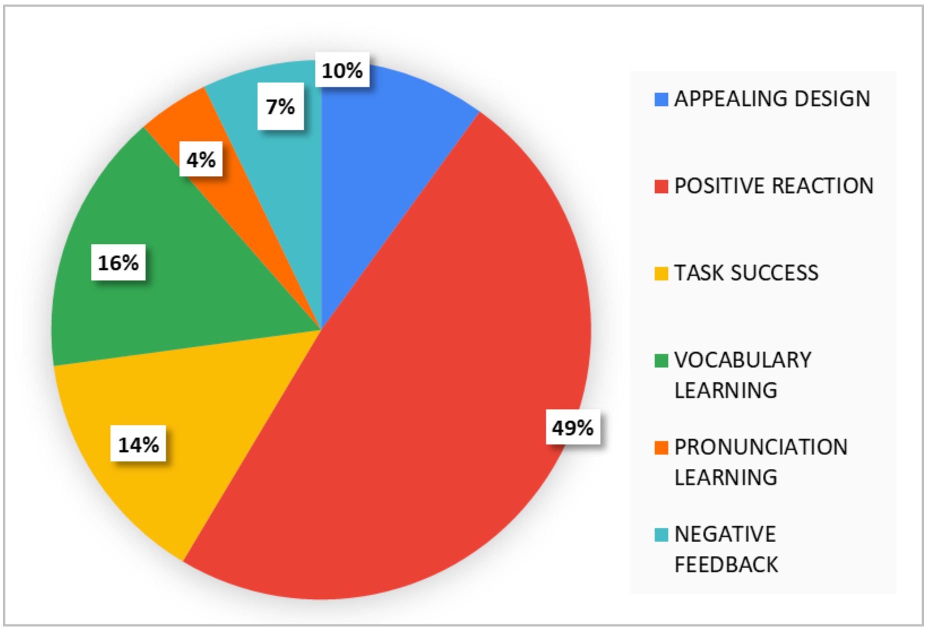
Note. Key words in bold. Translations from L1 Catalan to L2 English in brackets.



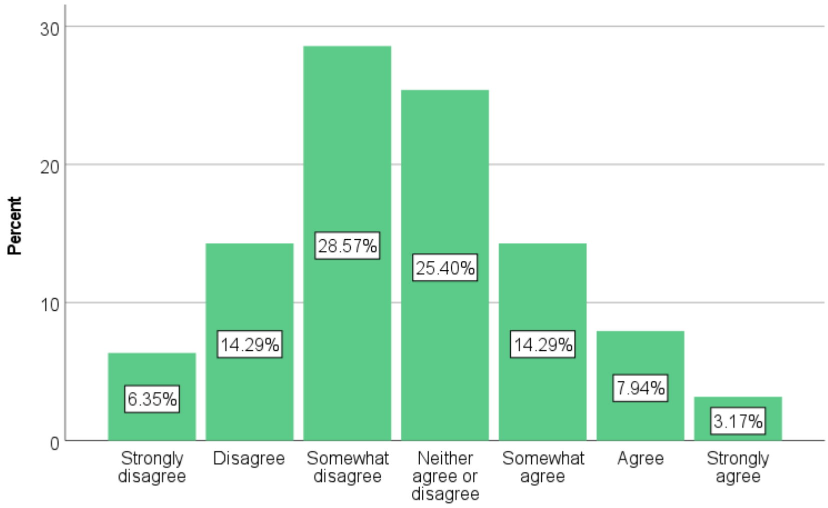
19. What do you think of the IMAGES/DRAWINGS in the flashcards?

Theme	*Code*	*Learners' response*
APPEALING DESIGN	S05	They are **cute**
	S08	Hahaha they were so **cool**, you should start a clothing line, for real.
	S15	They were very **beautiful**
	S44	Really **cool** design
	S56	Very **originals** if you did it with iPAD
	S61	Very **well done**
	S62	They were quite good, **pretty cute**.
POSITIVE REACTION	S72	Useful and **cool**
	S06	**Cool**
	S09	**Cool**
	S10	they are **good**
	S11	very **cool**
	S16	Almost all were **so cool** and useful
	S17	They were **right**
	S18	very **good**
	S27	**good**
	S36	**good**
	S41	they were **great**
	S45	**Good**
	S47	Are the **best**
	S48	They're **fine**
	S51	They were **cool**
	S53	**good**
	S54	were **cool**
	S67	They're **ok**
	S69	**Correct**
	S28	**Cool** and funny
	S23	motivating and **cool**
	S04	Exciting and **funny**
	S13	Really **funny**
	S14	**fun**
	S28	Cool and **funny**
	S37	**fun** and **useful**
	S40	very **funny!**
	S46	They were **funny**

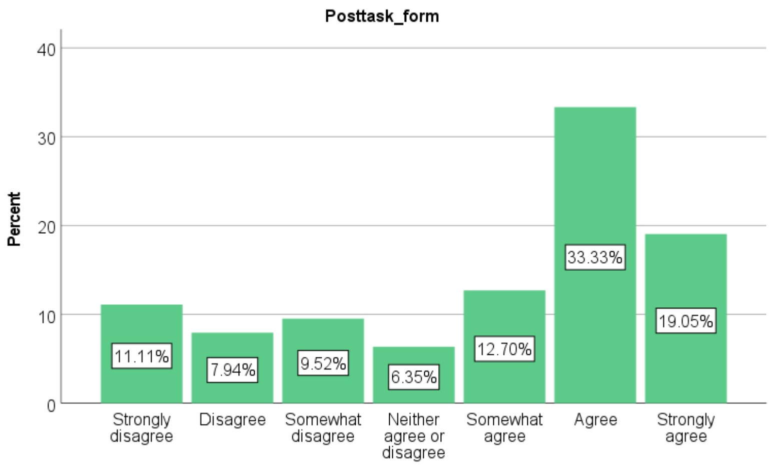
	S65	They're **funny.**
	S70	Very **funny!**
	S38	They where **funny** and motivate you to make the task
	S23	**motivating** and cool
	S55	They were **cool**, but sometimes innecessary.
	S38	They where funny and **motivate** you to make the task
TASK SUCCESS	S12	**help** me **solve** the task
	S30	They **helped me solve** the task
	S31	It **helps you do** the task
	S34	They are perfect for each activity to **help solve** the tasks
	S49	Very **usefully**
	S52	Necessary and **useful**
	S59	that **help you to solve** the tasks
	S71	Are **important** for **do** the task
	S72	**Useful** and cool
	S16	Almost all were so cool and **useful**
VOCABULARY LEARNING	S02	I think that images were a good idea because **helped me remember the meaning**
	S07	They **were related to the question** and they **helped to understand**
	S19	Very important to **remember the words** we learn
	S20	They were really **helpful** to know **what means**
	S25	I like it amd i think that are important to **know the meaning**
	S39	**Reinforce** your initial **ideas**
	S43	**self-explanatory**
	S58	The pictures helped you **remember the word** I learn
	S60	the pictures helped me **understand the words**
	S68	They are good for **help you remember the words**
	S26	Helped me **remember the meaning** and the pronunciation
PRONUNCIATION LEARNING	S26	Helped me remember the meaning and the **pronunciation**
	S29	Very **important**. Without them i thought the **pronunciation** was the same
	S32	I remember **the sounds** with the images
NEGATIVE FEEDBACK	S50	Adapted for **children**
	S66	Some of them were a little **childish**
	S33	they were ok but some **for children**
	S64	Some of them where a bit **strange**
	S55	They were cool, but sometimes **innecessary**.



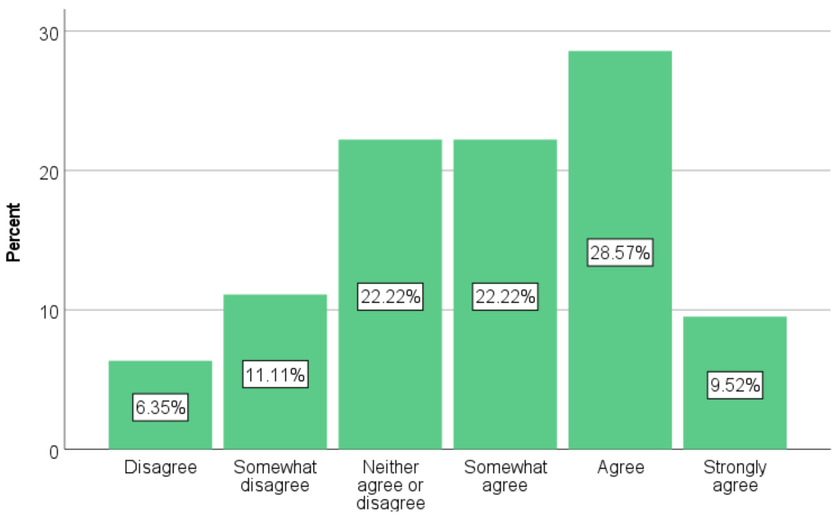
20. The POST-TASKS - the short “games” after the main task - helped you revise the meaning of the target words you had learned during the task.



21. The POST-TASKS - the short “games” after the main task - helped you revise the pronunciation of the target words you had learned during the task.



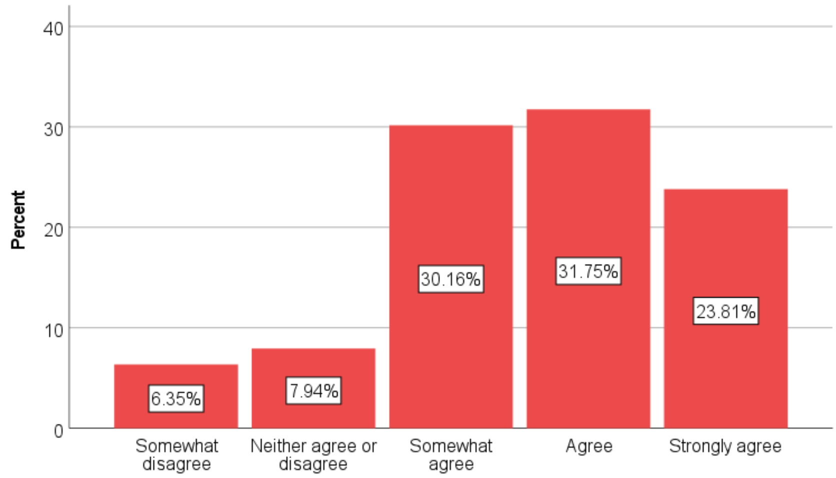
22. I think the POST-TASKS -the short “games” after the main task- were INTERESTING and ENJOYABLE.



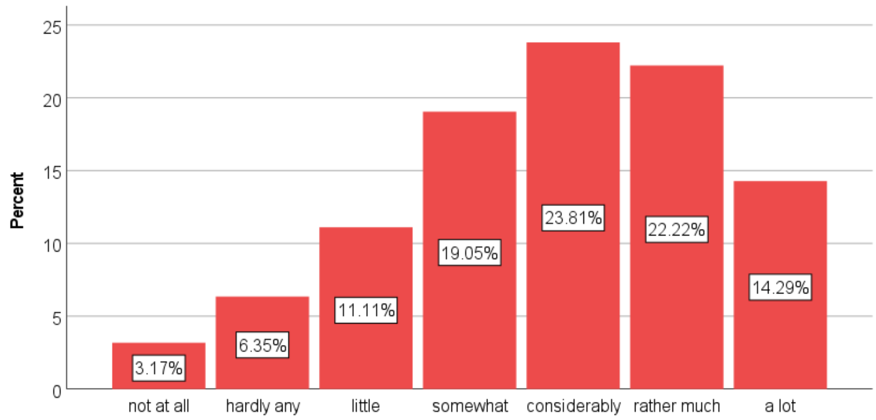
23. Now that the project is over, how difficult do you think are the following sounds to PRONOUNCE? (%)

	Extremely easy	Easy	Somewhat easy	Neutral	Somewhat difficult	Difficult	Extremely difficult
/i:/ (e.g. TEEN)	28.6	23.8	27.0	17.5	3.2	0.0	0.0
/ɪ/ (e.g. TIN)	11.1	19.0	30.2	20.6	14.3	4.8	0.0
/æ/ (e.g. CAP)	50.8	22.2	20.6	6.3	0.0	0.0	0.0
/ʌ/ (e.g. CUP)	30.2	15.9	23.8	23.8	4.8	1.6	0.0

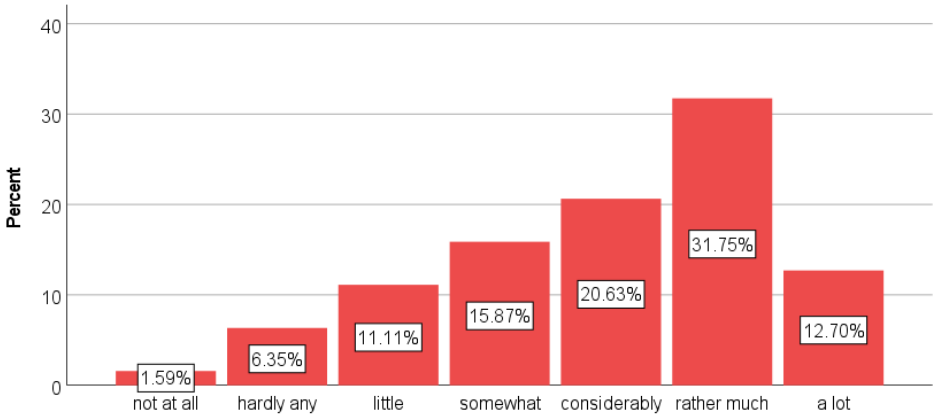
24. After doing this project, I think I have improved my PRONUNCIATION in English.



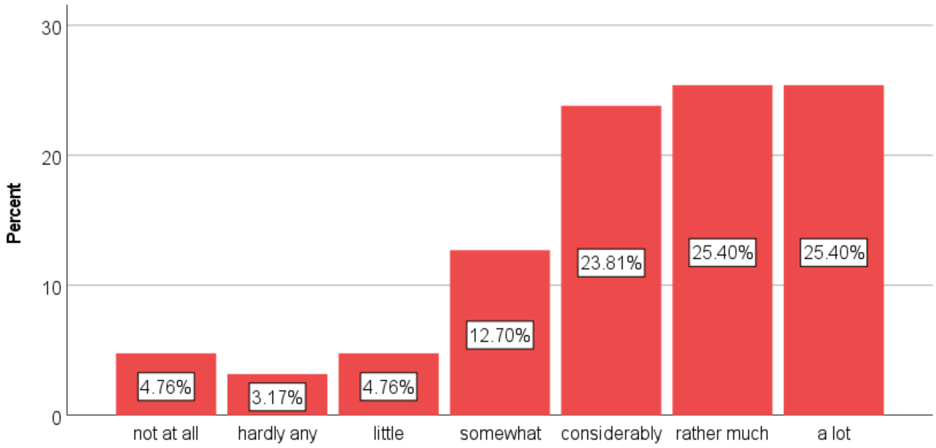
25. Indicate how much you think you have improved the pronunciation of /i:/ in "TEEN"



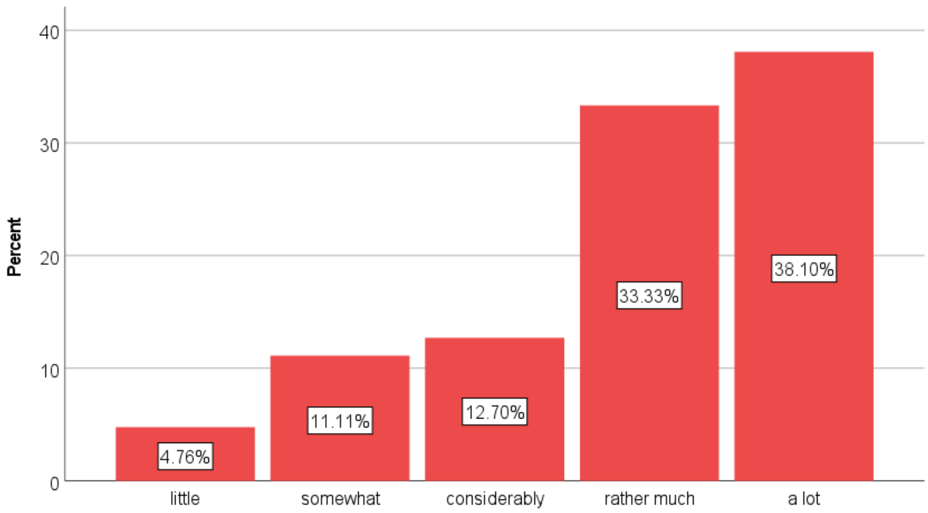
26. Indicate how much you think you have improved the pronunciation of /ɪ/ in “TIN”



27. Indicate how much you think you have improved the pronunciation of /æ/ in “CAP”



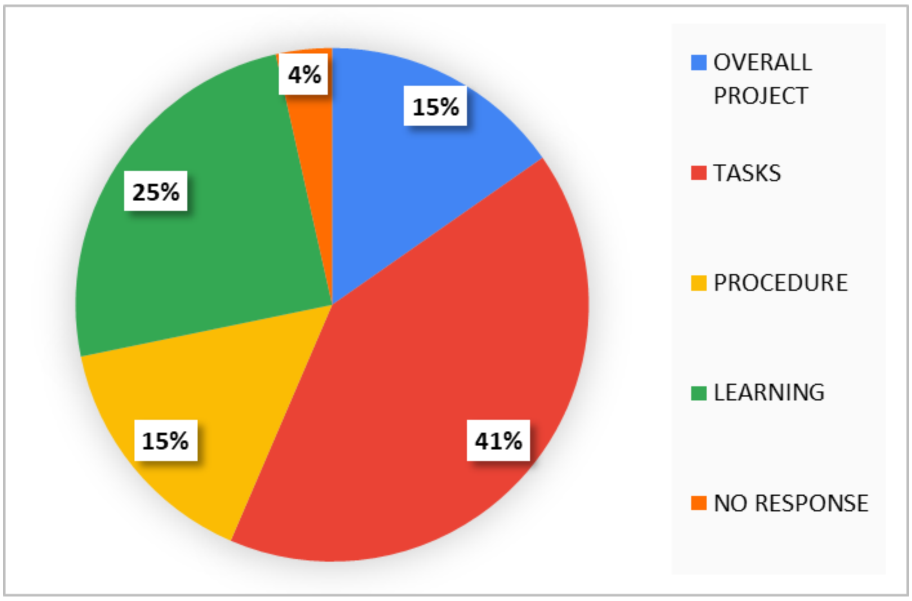
28. Indicate how much you think you have improved the pronunciation of /ʌ/ in “CUP”



29. In general, what did you most LIKE/DISLIKE about doing the project?

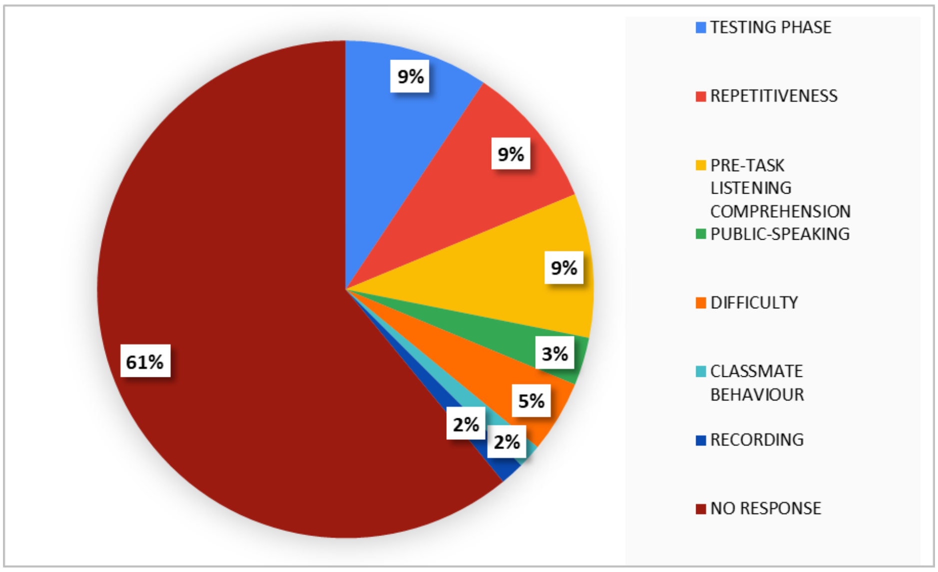
Theme		*Code*	*Learners' response (LIKEABILITY)*
OVERALL PROJECT		S02	I **liked** everything
		S18	I **loved** it
		S25	I **like it all**, and im sad that this finish
		S43	I **like** everything
		S41	It was **funny** and I enjoyed the new ideas so I would **totally repeat it**.
		S54	**I really enjoyed** the project cause I think the topic was great and that made the project different **and enjoyable**
		S07	We were learning many words during a lot of time and I have improved my pronunciation. I would **do** this project **again**.
		S13	I loved doing the recording in class. I **love** this project
		S23	the project was **very fun** and helped me learn pronunciation
		S70	it's a good idea to help students learn about pronunciation and i think it was very **fun and helpful**.
		S51	I really **enjoy the project**, i learned a lot
		S27	the class **dynamics**
		S67	I've loved the tasks in pairs. I'd **do** this project **again!**
TASK	ENJOYABILITY	S36	group tasks were really **entertaining** and I learned many many words and pronunciation
		S46	the images in the tasks make it **funny and entertaining**
		S49	I **loved the task** in pairs
		S67	I've **loved the tasks** in pairs. I'd do this project again!
		S72	I would love to participate again. I **love the tasks**
		S40	I **love** the speaking **tasks**, and speaking with different people
		S14	**do tasks** with my friends sharing the solutions with the class
		S59	i liked draw and sing a rap. Also the **tasks in pairs**
		S26	the **tasks** were really **funny** and games at the end too
		S31	The **tasks were interesting** anf it helps me to improve my pronunciation. I do this project again
		S15	**doing** the **tasks** and planning our speaking
		S68	i was **motivated** to go to english class **to do the tasks**
		S52	the tasks were **really useful** and we learning pronunciation in the same time
	DIFFICULTY	S39	I **liked** that I have to **make efforts** to solve the task and think how to do it
		S44	the **many** different **challenges** we found
		S58	**solving difficult** tasks
		S71	the **conditions to solve** the task
	PLANNING	S15	doing the tasks and **planning** our speaking
	REPORT	S14	do tasks with my friends **sharing** the solutions with the **class**
	POST-TASKS	S59	i liked **draw** and **sing a rap**. Also the tasks in pairs
		S29	I **loved** the **rap**
		S26	the tasks were really funny and **games at the end** too
	DIVERSITY	S04	The **different topics**
		S06	many many **different** super original tasks
		S10	The **wide variety** of activities
		S20	What I liked the most was **the different types** of activities.
		S44	the many **different** challenges we found
	ORIGINALITY	S19	like we **don't do activities** from the **book**
		S38	they were fun and if I am honest they were **a good excuse** to **stop** doing **activities from the coursebook**
		S53	I like practicing speaking because we don't do it. **Always study** from the **book**
		S66	**not doing** the **typical exercises** in class
		S06	many many different super **original tasks**
		S41	It was funny and I enjoyed **the new ideas** so I would totally repeat it.
		S54	I really enjoyed the project cause I think **the topic was great** and that **made the project different** and enjoyable

		S69	**creativity of the tasks** to learn english
PROCEDURE	RECORDING	S09	**Recording** the tasks with my classmates
		S13	I loved doing **the recording** in class. I love this project
		S16	What I enjoyed the most were the activities that we have to **record**.
		S32	i like to **record** my task with my colleague
		S37	I liked **doing the audios**
		S48	**record** my task with my friend
	TEAM WORK	S34	I really had an **amazing time** each day with **my partners**
		S55	What I liked the most was to **talk with my partner**. Can we do this again?
		S56	I loved **doing tasks with different people** in the class and **helping them**
		S17	I loved it that we **changed pairs** and I could **learn from other** people
		S30	my **friend help me** doing the tasks
		S40	I love the speaking tasks, and **speaking** with **different people**
		S11	**Speaking** with my **classmates**
LEARNING	PRONUNCIATION	S70	it's a good idea to **help** students **learn about pronunciation** and i think it was very fun and helpful.
		S61	What I like the most is **know the correct pronunciaton** of words
		S60	i like that i **improved a lot** my english **pronunciation** and of course i would do it again
		S23	the project was very fun and helped me **learn pronunciation**
		S31	The tasks were **interesting** anf it helps me to **improve my pronunciation**. I do this project again
		S45	learn new words and **new pronunciation**
		S36	group tasks were really entertaining and I learned many many words and **pronunciation**
		S28	practice **pronunciation** when speaking, not repeating the teacher
		S33	I loved **practicing** each one of the words' **pronunciation** when i try to solve a task
		S52	the tasks were really useful and we **learning pronunciation** in the same time
		S65	learn **pronunciation** with games
		S12	I loved **learning English pronunciation** through different games
		S07	We were learning many words during a lot of time and I have **improved my pronunciation**. I would do this project again.
	VOCABULARY	S45	learn **new words** and new pronunciation
		S07	We were learning **many words** during a lot of time and I have improved my pronunciation. I would do this project again.
		S36	group tasks were really entertaining and I learned many **many words** and pronunciation
	SPEAKING	S50	practicing **speaking in pairs**
		S53	I like **practicing speaking** because we don't do it. Always study from the book
	OVERALL	S51	I really enjoy the project, i **learned** a lot
		S62	my **language was better** after speaking with my classmates
		S64	at the ending the tasks became **more fluid and easy.**
NO RESPONSE		S05	
		S47	I **don't know**
		S08	



Theme	*Code*	*Learners' response (DISLIKEABILITY)*
TESTING PHASE	S29	I dislike **the tests** outside
	S36	the **tasks outside** the class
	S37	the **task** of **repeating words**
	S50	the **task** in sala d'actes were too long
	S61	listening and **exams**
	S27	**the tests** in sala d'actes
REPETITIVENESS	S06	It was great but **repetitive** at the end so I wasn't so motivated
	S53	**repetitive tasks**
	S48	I dislike that it was always **the same to speak**
	S05	I disliked how **repetitive** it became to say the **same words**
	S18	**repetitive, same words**
	S69	The **words** in the tasks were so **repetitive**
PRE-TASK LISTENING COMPREHENSION	S26	DISLIKE: The **listenings**
	S40	the **listenings** at the beginning
	S45	The **listenings** were too easy.
	S49	doing the **listening** before the task
	S61	**listening** and exams
	S16	I don't really dislike anything but if i have to choose one i will say **the listenings.**
PUBLIC-SPEAKING	S30	i don't like to **speak in front of the class**
	S46	**speaking in front of the class**
TASK DIFFICULTY	S09	sometimes **too easy**
	S19	I finish too fast because the task is **easy**
	S65	some tasks were **very difficult**
CLASSMATE BEHAVIOUR	S08	I disliked some of the other **classmates' attitude**.
RECORDING	S66	having to **record** it
NO RESPONSE	S02	
	S04	
	S07	
	S10	
	S11	
	S12	
	S13	
	S14	
	S15	
	S17	
	S20	

	S23	
	S25	
	S28	
	S31	
	S32	
	S33	
	S34	
	S38	
	S39	
	S41	
	S43	
	S44	
	S47	
	S51	
	S52	
	S54	
	S55	
	S56	
	S58	
	S59	
	S60	
	S62	
	S64	
	S67	
	S68	
	S70	
	S71	
	S72	



30. Finally, what do you think you have IMPROVED after doing the tasks?

Theme	*Code*	*Learners' response*
PRONUNCIATION	S02	The **pronunciation** of words in English
	S05	I think my **pronunciation** got a little bit better
	S11	the **pronunciation**
	S14	the **pronunciation**
	S18	the **pronunciation**
	S19	the **pronunciation**
	S23	the **pronunciation**

	S30	**pronunciation**
	S38	The **pronunciation**
	S39	The **pronunciation**
	S46	the **pronunciation**
	S48	the **pronunciation**
	S51	My **pronunciations**
	S54	**pronunciation** in English
	S59	the **pronunciation**
	S60	my **pronunciation**
	S65	**Pronunciation**
	S66	A little bit in **pronunciation**
	S68	My **pronunciation** and relationship with my classmates
	S71	The **pronunciation** of new words now i feel more confident talking in english.
	S09	**Pronunciation** and less "vergonya" [shyness]
	S26	The vocabulary and **pronunciation** of words
	S28	A lot of **pronuntiation** and vocabulary.
	S31	New vocabulary and the **pronuciation**
	S33	Learning new words with their correct **pronunciation**
	S36	**Pronunciation** and knowing more words
	S40	My **pronunciation** and my vocabulary.
	S41	My english and my **pronunciation**, also my vocabulary incremented a lot.
	S43	Especially vocabulary and **pronunciation**
	S50	**Pronunciation** and learning new words
	S52	**Pronunciation** and I have learned a lot of new words
	S55	Principally my **pronounciation** and I also learnt some words.
	S69	The **pronunciation** in some words and I also learned some vocabulary
	S72	I know how to **pronounce** better and I have learned many new words
	S44	**Pronunciation** and solving problems
	S64	I think that I have improved my **pronunciation** and my solving-things skills.
	S45	Comunication and **pronunciation**
	S07	The pronunciation **of "I" and "A" sounds** .
	S08	**General U** (f. Ex. Cup) **sound** pronunciation.
	S16	The different pronunciation **between "A" and "U"**. **Especially the "u"** in words like "cup"
	S29	I can **differenciate** words that **sound similar** but are two different words
	S49	The **pronuntiation** of very **similar** or equal **words**.
	S67	I've learned lots of new words and I've learned to **pronounce** words that are written different but **sound "the same"**.
	S70	My english **pronunciation**, my skills to **distinguish different words that sound similiar**
	S13	My **pronunciation** and **differentiation of similar words**
	S04	my **pronunciation** and speaking in general
	S37	Saying many words without stopping and **good pronunciation**
	S56	I speak faster and with better **pronunciation**
	S62	My response time while speaking and **pronunciation**.
VOCABULARY	S06	i learned a lot of **new words**
	S26	The **vocabulary** and pronunciation of words
	S28	A lot of pronuntiation and **vocabulary.**
	S31	New **vocabulary** and the pronuciation
	S33	Learning **new words** with their correct pronunciation
	S36	Pronunciation and knowing **more words**
	S40	My pronunciation and my **vocabulary.**
	S41	My english and my pronunciation, also my **vocabulary** incremented a lot.
	S43	Especially **vocabulary** and pronunciation
	S50	Pronunciation and learning **new words**
	S52	Pronunciation and I have learned a lot of **new words**
	S55	Principally my pronounciation and I also learnt **some words.**
	S69	The pronunciation in some words and I also learned some **vocabulary**

		S72	I know how to pronounce better and I have learned many **new words**
		S67	I've learned lots of **new words** and I've learned to pronounce words that are written different but sound “the same”.
SPEAKING		S04	my pronunciation and **speaking** in general
		S53	**speaking** in the "scenario" in English
		S17	I **don't spend** so much time **thinking** about the word I want to say, I just **jump**
		S37	Saying many words **without stopping** and good pronunciation
		S56	I speak **faster** and with better pronunciation
		S62	My **response time** while **speaking** and pronunciation.
		S10	I speak with more "confianza" [self-confidence] and **faster**
LISTENING		S32	**Understanding** people in the listenings with **different pronunciation**
ALL SKILLS		S25	my **English** and my motivations to learn English
		S58	My **pronunciation**, my **vocabulary**, my **speaking**, my **grammar**, my **listening**
		S41	My **english** and my pronunciation, also my vocabulary incremented a lot.
COMMUNICATION		S45	**Comunication** and pronunciation
		S20	I can **communicate** with my friend better in English
		S27	The fact of **communicating** in English. I felt very confident and comfortable in the last classes
AFFECTIVE FACTORS	MOTIVATION	S25	my English and my **motivations** to learn English
		S15	I am more **motivated** to speak English now
	SELF-CONFIDENCE	S10	I speak with **more "confianza"** [self-confidence] and faster
		S27	The fact of communicating in English. I felt very **confident and comfortable** in the last classes
		S71	The pronunciation of new words now i feel more **confident** talking in english.
		S09	Pronunciation and **less "vergonya"** [shyness]
		S12	i am not so **shy** when I speak English
		S61	participating in speaking activities **without trembling**
	SOCIAL SKILLS	S68	My pronunciation and relationship with my classmates
PROBLEM-SOLVING		S44	Pronunciation and solving problems
		S64	I think that I have improved my pronunciation and my solving-things skills.
NO RESPONSE		S47	I **don't know**