The Introduction of Rhythm Instruction in the English as a Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students.

Leticia Quesada Vázquez

UNIVERSITAT ROVIRA i VIRGILI

The Introduction of Rhythm Instruction in the English as a Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students

---

LETICIA QUESADA VÁZQUEZ

TESI DOCTORAL - TESIS DOCTORAL - DOCTORAL THESIS
2019

UNIVERSITAT ROVIRA I VIRGILI
The Introduction of Rhythm Instruction in the English as a Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students.
Leticia Quesada Vázquez

Leticia Quesada Vázquez

The Introduction of Rhythm Instruction in the English as a Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students

DOCTORAL THESIS

Supervised by Dr. Joaquín Romero Gallego

Submitted to the
Department of English and German Studies

UNIVERSITAT ROVIRA I VIRGILI
Tarragona, 2019

The research leading to these results has received funding from a *Martí i Franquès* doctoral grant from Rovira i Virgili University and project FFI2017-84479-P from the Spanish Ministry of Economy and Business

UNIVERSITAT ROVIRA i VIRGILI

FAIG CONSTAR que aquest treball, titulat "The Introduction of Rhythm Instruction in the English as Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students", que presenta Leticia Quesada Vázquez per a l'obtenció del títol de Doctor Internacional, ha estat realitzat sota la meva direcció al Departament d'Estudis Anglesos i Alemanys d'aquesta universitat.

HAGO CONSTAR que el presente trabajo, titulado "The Introduction of Rhythm Instruction in the English as Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students", que presenta Leticia Quesada Vázquez para la obtención del título de Doctor Internacional, ha sido realizado bajo mi dirección en el Departamento d'Estudis Anglesos i Alemanys de esta universidad.

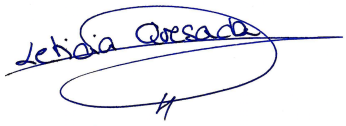
I STATE that the present study, entitled "The Introduction of Rhythm Instruction in the English as Foreign Language Classroom to Improve the Comprehensibility and Fluency of English for Specific Purposes Students", presented by Leticia Quesada Vázquez for the award of the degree of International Doctor, has been carried out under my supervision at the Department d'Estudis Anglesos i Alemanys of this university.

Dr. Joaquín Romero Gallego
Doctoral Dissertation Advisor

Tarragona, September 3, 2019

Declaration

I, Leticia Quesada Vázquez, hereby declare that this thesis is entirely my own work, carried out at Rovira i Virgili University for the degree of Doctor of Philosophy, and that it has not been submitted as an exercise for a degree at any other university. Where other sources of information have been used, they have been acknowledged. Some parts of this thesis have been accepted for publication in Quesada Vázquez, L. (2019). Pronunciation instruction in ESP teaching to enhance students' prosody. In S. Papadima-Sophocleous, E. Kakoulli Constantinou & C. N. Giannikas (Eds.), *ESP teaching and teacher education: Current theories and practices* (pp. 163-176). Researchpublishing.net. https://doi.org/10.14705/ rpnet.2019.33.932; and Quesada Vázquez, L., Romero, J. (2018). *The improvement of Spanish/Catalan EFL students' prosody by means of explicit rhythm instruction*. Paper presented at the International Symposium on Applied Phonetics (ISAPh) (pp. 104-109). DOI: 10.21437/ISAPh.2018-19.



Leticia Quesada Vázquez

Tarragona, September 3, 2019

Acknowledgments

Three years have gone by since I started this journey and from beginning to end I have felt loved, supported and encouraged. It is still pleasantly shocking with what confidence people have always looked at me and told me I was going to successfully complete this project. Therefore, I cannot write a word without thanking all of them.

First of all, I would like to thank the Department of English and German Studies at Rovira i Virgili University, which has seen me grow academically, professionally and personally. Not only have I completed my bachelor's, my master's and my PhD studies at this division, but I have also had the opportunity to build my teaching career as part of its team. Those who were first my professors have ended up being my colleagues, and hence they are part of what I am as a professional today.

A very special thankyou to my thesis supervisor, Joaquín Romero, to whom it feels almost impossible to express my most sincere gratitude and admiration. You have always been a mentor, you offered me my first job at university, and you insisted on me accepting this challenge. You have always believed in me and, thanks to that, I have been able to achieve each of my professional goals little by little. I am forever in your debt.

I strongly appreciate the economic support provided by both the university's *Martí i Franquès* doctoral grant and project FFI2017-84479-P from the Spanish Ministry of Economy and Business, which provided funding to guarantee the execution and dissemination of this study.

A word of gratitude to Carmen Rueda, who let me carry out my experiment in the course she is coordinating and who has always encouraged me to find my place at university; and to Irene and Nadia for their unconditional support inside and outside the

lab. I really hope we can keep working together as a consolidated research team in the near future. I also want to thank Anca, Kasia, Renata and Nune for acting as advisors with their experience, and Cai, who started this adventure with me.

I would like to express my appreciation to the members of the *laboratoire* LIDILEM at the University of Grenoble-Alpes, in France for the warm welcome and valuable support. Dan Frost, thank you so much for accepting me and supporting me in my journey in the mountains. Alice Henderson and Jean-Pierre Chevrot, thank you for providing alternatives to the assessment and statistical problems of my study. A very special word to *ma colocataire* Françoise Boch, with whom I learned how to fumble through French and live as a real *grenobloise*. And, of course, *un gros merci* to Luca, Émile, Roxanne, Célia, Érica, Roland, Sibely, Laurence, Xinjiletu, Samuel, Johanna and Yujing, my family there, who made me feel as part of a real research team.

I could not finish these words without showing my eternal gratitude to the ones who have really put up with my fears, my insecurities and my tantrums: my family. Dad and mum, you have always been my role models. This work is yours, not mine. Helena, Alejandro and Sergio, you have always believed in your sister and I am grateful for that, and my little Laia and Marc, you have brought joy to my life. I also want to thank my external family, the one that you create, especially Helena, Marta, Paola, Carla, Marina, Sergio and Dani, with whom I have shared all my joys and sorrows from childhood to this day, and Dayana, Sandra, Judith, Cristina, Kiko and Víctor, whom I met at university and remain being real friends with today.

Thank you all, my favorite people.

"Lend me your ears and I'll sing you a song
and I'll try not to sign out of key
Oh, I get by with a little help from my friends" (The Beatles, 1967)

Abstract

For the last few decades, teaching pronunciation has regained the interest of researchers as a tool to improve oral production in a second language. Its purpose, however, is not longer focused on training speakers to sound native, but on helping them express themselves effectively. Along these lines, the present study looks into the effectiveness of introducing explicit rhythm instruction in the language classroom to improve the students' comprehensibility and fluency in English. First-year engineering students attending a technical English course participated in a weekly pronunciation module of ten thirty-minute sessions based on a communicative framework. Participants were divided into an experimental group, which received explicit rhythm instruction, and a control group, which did not. Students were recorded before and after the treatment and their performances were compared to examine their progress.

A mixed-method assessment was conducted. First, the VarcoV values and the pauses of the sentences were acoustically analyzed. Second, native English speakers rated the comprehensibility and fluency of the students' extemporaneous productions. Also, a student satisfaction survey, the teacher's journal and the video-recorded sessions provided evidence of the implementation of the experiment.

The findings reveal differences between group means and significant effect sizes for the acoustic measurements despite general results not always being significant. Also, a ceiling effect is detected in the ratings, which poses some interesting questions regarding the validity of this kind of data. Despite this, the combination of quantitative and qualitative results suggest that explicit instruction in suprasegmental aspects of speech, like rhythm, can be beneficial for all learners and that pronunciation instruction has to be taken into consideration in any English as a Foreign Language context.

Resumen

En las últimas décadas se ha recuperado el interés por la enseñanza de la pronunciación para mejorar la producción oral en segundas lenguas, aunque su propósito ya no es entrenar hablantes que suenen como nativos, sino ayudarles a expresarse eficazmente. En este sentido, el presente estudio examina la eficacia de introducir instrucción explícita en ritmo en la clase de inglés para mejorar la comprensibilidad y la fluidez. Los participantes eran estudiantes de primer año de ingeniería en un curso obligatorio de inglés técnico. Se diseñó un módulo de pronunciación semanal de sesiones de treinta minutos con marco comunicativo. Se dividió a los estudiantes en un grupo experimental, con instrucción explícita en ritmo, y un grupo control, sin ella. Los estudiantes se grabaron antes y después del tratamiento y sus intervenciones se compararon para examinar su progreso.

El método de evaluación fue mixto. Primero, los valores de VarcoV y las pausas de las frases se analizaron acústicamente. Segundo, nativos calificaron la comprensibilidad y fluidez de los discursos. Además, una encuesta de satisfacción, el diario de la profesora y las sesiones videograbadas informaron sobre la implementación del experimento.

Las medidas acústicas revelan diferencias entre las medias de los grupos y un tamaño de los efectos significativo aunque los resultados generales no siempre lo fueran. Por otro lado, se detecta un efecto de techo en las calificaciones, el cual suscita preguntas interesantes a cerca de la fiabilidad de este tipo de datos. A pesar de ello, la combinación de resultados cuantitativos y cualitativos sugiere que la instrucción en aspectos suprasegmentales del discurso, como el ritmo, puede beneficiar a los estudiantes y que la instrucción de la pronunciación debe considerarse en cualquier contexto de inglés como lengua extranjera.

Resum

Durant les últimes dècades, s'ha recuperat l'interès per l'ensenyament de la pronunciació per millorar la producció oral en segones llengües tot i que el seu propòsit ja no és entrenar parlants que sonin com nadius, sinó ajudar-los a expressar-se eficaçment. En aquest sentit, el present estudi examina l'eficàcia d'introduir instrucció explícita en ritme a la classe d'anglès per millorar la comprensibilitat i fluïdesa. Els participants eren estudiants de primer any d'enginyeria en un curs obligatori d'anglès tècnic. Es va dissenyar un mòdul de pronunciació setmanal de sessions de trenta minuts amb un marc comunicatiu. Els estudiants es van dividir en un grup experimental, amb instrucció explícita en ritme, i un grup control, sense aquesta. Els estudiants es van enregistrar abans i després del tractament i les seves intervencions es van comparar per examinar el seu progrés.

El mètode d'avaluació va ser mixt. Primer, els valors de VarcoV i les pauses de les frases es van analitzar acústicament. Segon, nadius van qualificar la comprensibilitat i fluïdesa dels discursos. A més, una enquesta de satisfacció, el diari de la professora i les sessions enregistrades en vídeo van informar sobre la implementació de l'experiment.

Les mesures acústiques van revelar diferències entre les mitges dels grups i una mida dels efectes significativa tot i que els resultats generals no sempre ho eren. D'altra banda, es va detectar un efecte de sostre a les qualificacions, el qual planteja preguntes interessants sobre la fiabilitat d'aquest tipus de dades. Tot i així, la combinació de resultats quantitatius i qualitatius suggereix que la instrucció en aspectes suprasegmentals del discurs, com el ritme, pot beneficiar els estudiants i que la instrucció de la pronunciació s'ha de considerar en qualsevol context d'anglès com a llengua estrangera.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

In today's globalized world, learning languages becomes essential in order to apply for a good job position or survive abroad. On the one hand, it is common to find people who migrate looking for different life opportunities. On the other hand, it is likely for people staying in their country to work for international companies or with workers from all around the world. Hence, today it is not enough to be good at your job, but you are also required to be able to communicate, at least, in the world's lingua franca, English (Levis, 2018).

Both enterprises and educational institutions have had to take clear steps to meet the demands of internationalization within the labor market (Wilkinson, 2008). For this reason, they often offer different courses that focus on second language acquisition, going from subjects taught in English (i.e. content and language integrated learning) to all sorts of second and foreign language courses. These courses all share a main goal: providing students with skills and knowledge to successfully communicate in the target language.

However, oral communication continues to be problematic. As examined in MacDonald (2002) and Henderson et al. (2012), although both students and teachers are aware of the importance of practicing speaking, not much training tends to be done in class. Limited time in class or a lack of training are some of the struggles that instructors need to face. When it comes to pronunciation, many teachers confess that they completely avoid teaching it because they do not know how or what to teach. Besides, they admit to not being able to identify relevant improvement among students when working on

pronunciation. As a result, its potential to enhance successful communication is, indeed, questioned.

Second language pronunciation teaching remains a hot topic nowadays. After long periods of neglect, pronunciation stepped forward again within education research at the beginning of the century. Pronunciation had been traditionally misconceived as a synonym of native accent. Such a demanding goal had led to countless failed attempts that caused frustration and even rejection among students, especially adult learners. A shift in focus towards communication has fostered new research approaches to pronunciation instruction that aim at improving students' global intelligibility, regardless of how native they sound. However, these proposals do not usually filter down to the second language classroom (Derwing & Munro, 2015). Therefore, more class-based studies are demanded in order to examine the applicability of the research advances on pronunciation teaching and guide practitioners on how to deal with pronunciation effectively in their classes.

This dissertation responds to the call and seeks to provide more evidence on the importance of pronunciation teaching by trying to address some of the current concerns and the limitations associated with the field. The study seeks to bridge the gap between research and the classroom by testing the effectiveness of language rhythm instruction to improve students' comprehensibility and fluency in English at a university in Tarragona, Catalonia (Spain).

1.1. Aims of the study

In previous years it had been observed that engineering students at Rovira i Virgili University struggled when communicating orally in English. By the end of the course,

students had to deliver an oral presentation. Students often sounded chopped, they made pauses at the wrong places, stress was not properly placed and, in general terms, they were not keeping the natural flow of the discourse. The same type of difficulties arose in conversations and debates initiated in class. Those problems made it difficult to understand what they wanted to say regardless of their level of English. Experts such as Wong (1987), Gilbert (2008) or Derwing and Munro (2015) have argued that an appropriate adaptation of language rhythm can tackle the issue and improve students' intelligibility, particularly in this type of cases where the rhythm of the mother tongue and the target language clearly differ (for a complete review, see section 2.6.). Therefore, the introduction of rhythm instruction was thought to be beneficial for the students' communicative skills.

To this end, the present study attempts to address the following research questions:

1. To what extent does explicit rhythm instruction help reduce the mother tongue's "negative transfer" (Celce-Murcia, Brinton and Goodwin, 1996, p. 36) when speaking in the target language?
2. To what extent does explicit rhythm instruction improve students' comprehensibility and fluency when communicating orally in English?
3. To what extent does the introduction of Celce-Murcia et al.'s (1996) steps to teach communicatively (p. 36) within the classroom enhance students' comprehensibility and fluency when communicating in English?

The corresponding hypotheses are as follows:

H1: Students receiving explicit rhythm instruction reduce more their L1 negative transfer when speaking in English than those who do not.

H2: Students receiving explicit rhythm instruction are more comprehensible and fluent when communicating orally in English than those who do not.

H3: The introduction of Celce-Murcia et al.'s (1996) steps to teach communicatively (p. 36) within the classroom enhances students' comprehensibility and fluency when communicating in English.

In order to test these hypotheses, a classroom-based study was carried out in an English for Specific Purposes (ESP) course taken by first-year undergraduate Spanish/Catalan engineering students at Rovira i Virgili University. In short, students were separated into two pronunciation groups, one that received explicit rhythm instruction and one that did not. After participating in ten weekly pronunciation sessions integrated in the course, students' improvement in their oral competence was tested by means of pre- and posttests. During the process, sessions were video-recorded, the teacher wrote a journal, and students had to fill in a satisfaction survey so as to further investigate questions of methodology and assessment.

1.2. Structure of the thesis

This thesis is divided into six chapters. After introducing the main theme in this chapter, Chapter 2 starts with a historical overview of pronunciation teaching and defines the concepts of intelligibility, comprehensibility, fluency and nativeness to clearly establish the different goals of this discipline. It continues by examining which pronunciation features have been claimed to be more effective to teach, distinguishing between the

potential of segmentals and suprasegmentals, and putting emphasis on those studies focused on improving intelligibility in general, particularly those carried out within a classroom setting. Then, it focuses on the teaching methods that have been proven to meet the communicative expectations better and those group and individual factors from students that can affect instruction. After establishing the background, it examines language rhythm in depth: its nature, its features, its applicability within the field of second language teaching and acquisition, and the resulting controversies.

Chapter 3 expands on the methodology applied for the experiment. It provides detailed information about the course, the participants of the study, the instruction provided during the experiment, how data was gathered, and the assessment tools and the statistical analyses applied.

In Chapter 4 results are presented according to the assessment procedure used. Starting with the quantitative results, acoustic measurements of rhythm and pausing are introduced first. Then, scores for comprehensibility and fluency improvement are provided. Finally, students' responses to the multiple choice and rating scales questions extracted from the satisfaction survey are presented, followed by qualitative results from students' comments to teachers' observations, reinforced by the video recordings of the sessions.

A summary of the main findings leads to discussion in Chapter 5, where results are assessed in the light of the hypotheses postulated. Also, further insights derived from the design of the study and the implementation of the experiment are presented, which respond to some of the principal areas of friction between research and practice.

Finally, Chapter 6 presents the main conclusions reached, paying special attention to the limitations of the study and making suggestions for future research in the field.

Chapter 2. Literature review

2.1. Pronunciation teaching: A change of focus, from nativeness to intelligibility

The interest in the instruction of the pronunciation of a second language (L2) is relatively recent, especially if its study and implementation is compared to other linguistic aspects, such as grammar and vocabulary. The methods inherited from teaching ancient languages like Latin and Greek, which were no longer spoken, dictated the traditional approach to second language teaching, based on practicing writing production and comprehension through translations and grammar rules, instead of enhancing the oral competence. For this reason, working on syntax or the lexicon was prioritized. In fact, as Celce-Murcia et al. (1996) pointed out, in the twentieth century methods like grammar translation and reading-based contributions were still popular and functioning.

Pronunciation did not come into play within the second language teaching scenario until the late 1800s, not even two centuries ago. The first approach, known as the Direct Method, was grounded on intuition and imitation: By listening to a model, either the teacher's or a recording, students had to try their best to emulate the speech, repeating and imitating the speaker. A lack of an initial awareness of how to teach pronunciation, thus, made its instruction rely on instinct and be reduced to the learner's ability to listen and adopt the sounds of the sample provided.

The emergence of the *Reform Movement* and the consequent development of the International Phonetic Association and its International Phonetic Alphabet (IPA) at the end of the nineteenth century paved the way to an analytic-linguistic approach, which advocated for explicit instruction thought to complement accurate perception and

imitation (Celce-Murcia et al., 1996). Analytical contributions to language teaching, such as Audiolinguism during the 1940s and 1950s, or the Silent Way in the 1970s, gained popularity, although the Direct Method also evolved into influencing methods like the Community Language Learning in the seventies. Hence, both the intuitive-imitative and the analytical-linguistic approaches coexisted during most part of the twentieth century.

It was during the 1980s when the Communicative Approach arose to dominate most of the pronunciation scene of the last decades of the twentieth century and onwards. Its main goal was communication, i.e. it focused on training L2 learners that could be understood and intelligible, and who could manage communicative situations effectively. Until then, most of the pronunciation teaching techniques aimed at students speaking as native speakers. To name a few, listen and imitate activities, minimal pair drills, or phonetic transcription were common exercises whose target was to reproduce sounds and speeches as accurately as possible. The Communicative Approach considered sounding native an overambitious goal and, despite supporting the introduction of techniques that enabled learners to communicate (Celce-Murcia et al., 1996), it got by without pronunciation instruction. As a consequence, pronunciation teaching became mostly neglected during the last part of the twentieth century.

However, several researchers, faced with the denial of pronunciation teaching, have been advocating for its importance in the L2 classroom until today (Busà, 2012; Chela de Rodríguez, 1976; Derwing, 2008; Derwing & Munro, 2015; Gilbert, 2008; Levis, 2005; Munro, 2008, among others). They, besides, have maintained the perspective towards being intelligible rather than sounding native, illustrated by the inclination towards the intelligibility principle and its teaching implications over the nativeness

principle (Levis, 2005). As Munro (2008) noted, nowadays a large amount of second language speakers communicate effectively worldwide in spite of having a foreign accent, so getting rid of it does not seem to be the most suitable aim to reach within pronunciation teaching.

Nevertheless, the concepts under the umbrella of the intelligibility and the nativeness principles have not always been clearly defined, blurring the real purpose of pronunciation studies. Hence, it is essential to establish what is really understood by intelligibility, comprehensibility, fuency and accentedness so as to identify the final purpose of both the research already conducted in the field and the present study.

2.2. Defining the purpose: The difference between intelligibility, comprehensibility, accentedness and fluency

The study of pronunciation from its different approaches (i.e. acquisition, teaching, and so on) has experienced a controversy when providing definitions to some important concepts, such as intelligibility and comprehensibility. For this reason, it is important to clarify the interpretations used within this dissertation.

Derwing and Munro's (1995) attempt to explain intelligibility, comprehensibility and accentedness from a teaching perspective has led to different connotations and misunderstandings. To start with, they broadly defined intelligibility as "the extent to which the speaker's message is actually understood by a listener" (p. 76). As pointed out by Thomson (2018), an indistinct reading of their words may lead to the conclusion that intelligibility equals to "listening comprehension" (p. 14). However, what they really tested was the extent to which the listener could literally decode the speakers' words. In this dissertation intelligibility alludes to two different nuances: When

speaking from an abstract and theoretical point of view, intelligibility makes reference to the general "capability of being understood" (WordReference.com, 2019); when talking about concrete aims of measurement and assessment, the narrower and more specific description on accurate message decoding is used.

The latter definition, though related, must not be confused with the concept of comprehensibility, which makes reference to how difficult or easy a listener finds the speakers' intended message to be understood (Thomson, 2018). Accentedness, on the other hand, has not generated much discussion among experts. As Thomson (2018) summarized, there is a more general agreement of foreign accent being interpreted as the extent to which an L2 speech sounds native-like.

On the other hand, the definition of fluency tends to be imprecise despite much existing discussion (for a review, see Wood, 2010). Largely associated with proficiency and even nativeness, this dissertation opts for a more teaching-oriented interpretation, seeing it as "a naturalness of flow of speech" (Wood, 2010, p. 9), needed not to speak like a native, but to be understandable. As Levis (2018) emphasized, disfluent speech can break the native listener's train of thought when interpreting the message, so it can hinder comprehension.

In short, while intelligibility, comprehensibility and fluency build up Levis' (2005) intelligibility principle, which enhances L2 students' ability to be understood, accentedness is related to the nativeness principle, which advocates for reaching a native command of the target language. The items seen and the teaching techniques used in class will differ depending on the purpose established. As the main object of this study is to help students be more understandable, research on intelligibility, comprehensibility and fluency is going to be highlighted in the following sections.

2.3. What to teach

Determining the pronunciation features that are more detrimental to intelligibility and the way to teach those to L2 learners is like walking on thin ice. In spite of the fact that pronunciation teaching researchers have studied different possibilities to answer this question in the last couple of decades (for more information, see Derwing & Munro, 2015), no solid conclusions have been reached. A lack of agreement among experts has led the teacher to single-handedly decide which are those aspects and techniques to deal with in the L2 classroom based just on experience and intuition (Derwing & Munro, 2005, 2015; Derwing, Munro & Wiebe, 1998; Henderson et al., 2012; Levis, 2008; MacDonald, 2002; Munro & Derwing, 1995). In fact, many teachers still mostly rely on traditional intuitive-imitative activities, which continue reinforcing the presence of nativeness goals, due to insufficient training and guidance (Derwing & Munro, 2015; Henderson et al., 2012; MacDonald, 2002; Munro, 2008). Therefore, there is an urgent call within the field to conduct more research studies that prove the prominence and effectiveness of different pronunciation items, which help guarantee the acquisition of successful communicative skills and guide pronunciation teaching instruction (Munro, 2008).

2.3.1. Error assessment.

Second language acquisition researchers have come up with several error theories that try to shed light on the issue of determining the pronunciation features to teach, outlining certain guidelines that could help teachers decode the aspects that best suit their students' problems.

The contrastive analysis hypothesis (CAH) proposed by Lado in 1957 (as cited in Celce-Murcia, 1996, p. 19) is the most enduring theory on how learners acquire a second language and establishes the foundations for subsequent approaches on phonological acquisition. It postulates that the native language fixes the way we acquire the target language by easing the learning process of those items that are similar between them, and negatively interfering when the items are different or even inexistent in the target language, more commonly known as *negative transfer* (Celce-Murcia et al., 1996, pp. 19-20). Hence, a Spanish learner of English is supposed to struggle more when pronouncing /ɪ/ rather than /i/ because the former does not exist in Spanish while the latter does. Nonetheless, the theory does not explain why L2 speakers are found to mispronounce both sounds depending on the context in real situations, and to what extent this type of mispronunciation affects intelligibility. Therefore, it fails to delimitate the degree of difficulty and prominence that language interference and observable errors have in second language users' ability to make themselves understandable.

As cited in Celce-Murcia et al (1996), the error analysis theory supported by, among others, Banathy and Mandaraz (1969) and Richards (1971), was born in the late sixties / early seventies to further explore the relationship between the nature of the errors made and acquiring a second language. It focused on analyzing the different potential mistakes that an L2 speaker may commit and delimiting those aspects that are more difficult to acquire (p. 20). Despite offering an error classification, a hierarchy of gravity for communicative purposes was still missing.

Flege (1987, 1995), on the other hand, differed from the CAH by claiming that those aspects in the second language that are perceived similar to the ones in the first

language are more difficult to acquire than those which are the same or even completely different. The perspective changed: While CAH and error analysis theories focus on production, Flege's (1995) speech learning model (SLM) focuses on perception. The SLM theorizes that L2 speakers have problems to perceive the difference between similar pronunciation features, not being able to avoid negative transfer from the mother tongue (L1) to the L2. Consequently, it gets more complicated to modify something that you cannot perceive. However, the theory does not directly point at concrete mistakes that affect L2 speakers' intelligibility.

Another influential hypothesis is the marked theory defended by Trubetzkoy (1939), Jackobson (1941) and Eckman (1977), together with other linguists from the Prague School. This theory distinguishes between unmarked elements, i.e. more common and basic items which, hence, are easier to acquire, and their counterpart marked elements, i.e. more rare and specific items which are more difficult to acquire (as cited in Celce-Murcia, 1996, p. 22). Again, the frequency and simplicity of an element implies a certain degree of difficulty when acquiring it, but it does not tell anything about its importance in being understandable.

Derwing and Munro (2015) or Levis (2018) seem to lean towards the functional load theory, based on the importance of an error according to "the amount of work the contrast does in the language" (Levis, 2018, p. 4). Levis (2018) expanded on the topic and compared it to the *Pareto* principle, also known as the *80-20* rule:

> The principle may suggest, for example, that 20 percent of the workers create 80 percent of the wealth, or even that 20 percent of the students take up 80 percent of a teacher's time and energy. Applied to pronunciation teaching, then, we can guess that the largest improvement in a learner's intelligibility may come from improvement in a minority of the learner's errors. (Levis, 2018, p. 15)

The theory, hence, suggests that there are certain aspects which, if pronounced wrongly, may lead to serious misunderstandings, so they need to be prioritized in the L2 classroom to improve intelligibility. By contrast, there are others that may not affect the comprehension of the message, so there is no need to put them first when teaching. However, supporting findings are restricted to the segmental level so far (see Derwing and Munro, 2015), so more research should be conducted.

Despite the fact that pronunciation errors have been widely discussed, the different theories on the topic fail to point at specific aspects to teach in order to avoid a communication breakdown. Something seems clear, though: The more rooted in the mother tongue a linguistic element is, the more difficult to modify it to suit the target language. When there is something that feels innate, it tends to be more difficult to explain and understand, so it is more complicated to consciously and willingly manipulate or adapt it whatever the final purpose is. For example, when there is a person who can twist their tongue in the shape of a flower and is asked how to do it, the answer tends to be: "I do not know, it just comes natural to me," and they are not able to explain or teach the movement. Similarly, if a Spanish speaker asks native speakers of English how they *swallow* their words when speaking and still get the message through, they probably will not be aware of how they use language rhythm, failing to give an explanation. L2 speakers, meanwhile, perceive something unfamiliar to them that they struggle to reproduce with the same success because it is alien to their native language.

This does not mean that no similar mechanism exists in their mother tongue (every language has its rhythm), but they have complete ignorance or little knowledge of its existence and how to produce it. As Busà (2012) pointed out, for instance, people do not tend to have a lot of knowledge about prosody in their mother tongue; therefore, it can

be really difficult for them to consciously distinguish and use prosodic features, despite being able to do it at an unconscious level and succeed in daily communication. Within the field of L2 pronunciation instruction, teachers need to bridge the gap by introducing, teaching and practicing this sort of intrinsic elements in the classroom when proved to affect mutual understanding (Munro, 2008).

Errors at the segmental (i.e. sounds) and the suprasegmental (i.e. prosodic elements such as stress, rhythm and intonation) levels have been investigated in the field so as to detect which are more detrimental to communication. As Derwing and Munro (2015) pointed out, "it is possible to speak with a strong foreign accent, yet be perfectly intelligible. However, some heavily-accented speakers are highly unintelligible" (p. 72). The studies that have compared segmental and prosodic errors in order to examine which of them are more severe lead to inconclusive findings. While Anderson-Hsieh, Johnson and Koehler (1992), Derwing and Rossiter (2003), Johansson (1978) and Palmer (1976) concluded that suprasegmental errors where more serious than segmental ones, Fayer and Krasinksi (1987) and Koster and Koet (1993) reached the exact opposite conclusion (as cited in Derwing and Munro, 2015, p. 73). There are some reasons that justify this contradiction. Derwing and Munro (2015) explain that there are many other factors which play a role when analyzing the gravity of an error, such as the first and the second languages under study, or the methods used. Also, the problems of an L2 speaker's intelligibility may not rely just on one of the two pronunciation levels, but on both of them. Hence, they concluded that what is important is to assess the variables at stake so as to determine those concrete pronunciation items that are crucial for the learners' intelligibility according to their particular context.

The influence of concrete pronunciation aspects in L2 speakers' intelligibility has been brought under analysis by artificially manipulating both first and second language speakers' utterances. A good example is Tajima, Port and Dalby's (1997) experiment on timing patterns: They resynthesized English productions of Mandarin speakers to match native timing patterns, and the other way around, they manipulated native speaker productions in English to get closer to Mandarin timing patterns. Then, English listeners had to assess both the original and the edited versions by identifying what they were listening to. As a result, they observed that those versions containing the English timing patterns were more intelligible than those with Mandarin ones, no matter their first language. Acquiring English timing patterns, thus, was revealed to be key for Mandarin speakers. Nevertheless, it is not enough to analyze contrasts between language behaviors in the laboratory, but it is essential to test whether those pronunciation features influencing communication are learnable in class (Levis, 2005). For this reason, instruction research is needed to reach reliable conclusions (Baker & Murphy, 2011; Derwing & Munro, 2005, 2015; Gordon, Darcy and Ewert, 2013).

Derwing et al. (1998) carried out a ten-week study with forty-eight adult L2 learners of English as participants. Subjects were divided into three different groups of sixteen students each: one group was taught suprasegmentals for twenty minutes a day; another one received segmentals instruction for twenty minutes a day; and the control group received no explicit pronunciation instruction. Students recorded themselves reading aloud sentences and doing a picture narrative exercise before and after the instruction. On the one hand, forty-eight native listeners assessed the sentences in terms of comprehensibility and accentedness using a 9-point Likert scale. Results showed improvement of both the segmental and the suprasegmental groups after instruction,

whereas the control group remained the same. On the other hand, 45-second excerpts were extracted from the picture narrative exercise and assessed by six trained ESL teachers who rated the samples' comprehensibility, fluency and accentedness using the same type of scale. The only group to show improvement in any of the dimensions under study was the suprasegmental group. As a conclusion, Derwing et al. (1998) highlighted the importance of suprasegmental instruction under extemporaneous circumstances, but also remarked the improvement of segmental instruction in controlled situations since, if there was a problem during communication, this group would be likely to fix it addressing individual segments.

Gordon et al. (2013) also examined the weight of suprasegmentals and segmentals in comprehensibility improvement by means of the research of explicit instruction in pronunciation communicative approaches. They compared recorded sentences from twelve leaners who received one of the three treatments under study: four learners came from the group trained with suprasegmentals, four from the group working on segmentals, and four from the group receiving no explicit instruction. Twelve L1 English speakers rated the sentences according to their comprehensibility using a 9-point Likert scale. Only the suprasegmental group showed significant improvement. However, the authors stated that segmental instruction was based on the learning of four vowel sounds only, whereas the suprasegmental instruction was more complete.

Although much research needs to be conducted to reach definite conclusions, these studies provide support for the effectiveness of suprasegmental training to improve the students' ability to be better understood, especially when improvised speeches are tested. Segmentals are also important in second language learners' intelligibility, but the role they play in extemporaneous circumstances seems not to be so crucial.

2.3.2. The rise of suprasegmentals.

Suprasegmentals started gaining popularity within the research field of pronunciation teaching when attention started to be centered on communicative purposes. For instance, McNerney and Mendelsohn (1992) claimed that suprasegmentals instruction enhances comprehensibility because it implies noticeable improvement in a short period of time, which is encouraging for students (as cited in Celce-Murcia et al., 1996, p. 10).

Suprasegmental features have a central function at the discourse level. They shape the speech by harmonically and coherently linking the different elements that form it, giving complete meaning and intention to the discourse. Segmental features, on the other hand, center on isolated sounds in very controlled contexts. It is definitely true that a good pronunciation of the sounds of the second language is needed to understand speakers, but there are some cases in which students are able to pronounce sounds properly and they are still struggling when communicating. In the end, as Henry Sweet (1900) claimed, "although language is made up of words, we do not speak in words, but in sentences" (p. 98).

For example, L2 speakers of English can know how to pronounce all the sounds of the word 'record' but they may not be aware of the difference in stress between the noun (i.e. REcord) and the verb (i.e. reCORD), so they might mispronounce it and use the wrong pronunciation in a sentence, leading to syntactical problems that affect the listener's comprehension. Conversely, there are cases in which sounds are not accurately pronounced, but students manage to make themselves understood. Taking the same sample words, students may not pronounce the *schwa* in unstressed syllables perfectly, but if they place stress properly and reduce the vowel sound to some extent,

they ease the perception of the listener, who will end up being able to deduce the right word.

Different exercises to teach suprasegmentals have been designed, presented both in research articles (Anderson-Hsieh, 1990; Chela-Flores, 1997b, 2003; Chela de Rodríguez, 1976; Gilbert, 2008; Goodwin, 2013; Stevens, 1989; Tuan & An, 2010; Wong, 1987) and textbooks (Beisbier, 1995; Earle-Carlin, 2011; Gilbert, 2005a, 2005b). Taking into account Derwing and Munro's (2015) compilation of intervention studies together with further research on the topic, however, few research studies working on suprasegmentals were carried out directly in class during the last three decades.

One of what could be considered early intervention-research on suprasegmentals teaching is Chela de Rodríguez (1981) ten-lesson module on rhythmic patterns included in a teaching-training program taken by students of English at the university of Zulia in Venezuela (as cited in Chela-Flores, 1997b, p. 122). Although the module was mainly designed to improve students' perception of rhythmic patterns, the participants reported that, thanks to an improvement in their perceptive skills, they felt more confident when detecting and self-correcting their production mistakes.

Couper (2006) investigated the epenthesis and absence phenomena. Despite being segmental elements *per se*, they affect prosody by increasing or reducing the number of syllables found in an utterance (Derwing & Munro, 2015). The subjects of the experiment were twenty-one ESOL immigrants, mostly from Asia, who attended an English course from July to November in New Zealand. They took a general diagnostic test at the beginning of the course and common mistakes with rhythm and stress were detected. Besides, adding extra sounds or omitting essential ones created problems with word boundaries both in isolation and as a consequence of connected speech. Learners

participated in twelve sessions of thirty minutes each of different natures but mainly focused on syllable instruction. They took the diagnostic test and also a specific test on epenthesis and absence consisting on sentences recorded in three times: before, directly after, and three months after the instruction. Results revealed that the error rate in the elements under study decreased after training both in the direct and delayed tests. Besides, when comparing their performance with the following year's group used as baseline, the latter did not show signs of improvement.

Tsiartoni (2011) examined the effectiveness of rhythm instruction in three different age groups (10, 12 and 16 years old respectively). For each age set, an experimental group, which received rhythm instruction within regular classes, and a control group, which did not, were established. After measuring the vocalic and consonant variability of all the groups before and after the instruction, signs of a more English-like rhythm were observed for the experimental groups at the different age levels, while the control groups did not show signs of improvement.

Sardegna and McGregor (2013) conducted a study in which fifteen international graduate teaching students with different L1s participated in a fifteen-week scaffolded pronunciation program based on primary stress, intonation, vowel reduction, and linking. Participants recorded a reading-aloud passage at the beginning and the end of the course. After comparison, results revealed that students significantly improved their accuracy in all four aspects.

These studies have one thing in common: Although they test the power of suprasegmentals to decrease error probability, they do not directly examine intelligibility. It seems logical that the fewer mistakes an L2 speaker makes when communicating in the target language, the more intelligible he/she sounds, but it is

necessary to count on a listener's perspective to verify this correlation (Derwing & Munro, 2015: Munro & Derwing, 2015).

2.3.3. Suprasegmentals teaching to improve students' communicative ability.

Derwing, Munro and Wiebe (1997) conducted a study focused on how explicit pronunciation instruction could help fossilized adult learners of English improve their production in the target language. Thirteen learners of different mother tongues who had lived in Canada for an average of ten years and suffered from a standstill in their learning process attended a *Clear Speaking* course two evenings a week for twelve weeks. Since it was difficult to establish common segmental problems among the participants due to their different L1 backgrounds, the course focused on suprasegmentals and related aspects such as voice quality or speech rate. Native speakers assessed a series of true and false sentences that the subjects recorded before and after the instruction by means of an intelligibility task (i.e. transcription), and comprehensibility and accentedness tests (i.e. 9-point Likert scales). Results revealed that all the sentences were significantly more intelligible and true sentences were, besides, significantly more comprehensible and less accented after instruction. False sentences did not show signs of improvement in terms of comprehensibility and accentedness, but the researchers discussed that it was probably because their negative content was hard to predict.

Ramírez Verdugo (2006) conducted a study based on intonation instruction. Two groups of ten Spanish speakers each took a teacher training course offered at the Universidad Autónoma de Madrid. One of the groups received instruction on intonation for ten weeks. The sessions were of fifty minutes each and took place in a laboratory

setting. The other group attended regular English classes with no intonation training. Both groups took a pre- and a posttest based on the conversations recorded the first and the last day of instruction respectively. Ten English listeners assessed their performances with a 6-point Likert scale looking for signs of intelligibility improvement. Results showed a significant improvement for the experimental group while the control group remained stable. However, results are highly questionable. On the one hand, the extreme points of the scale were described as 1 being difficult to comprehend and 6 being a native-like performance. Hence, both intelligibility and nativeness principles were mixed up, misguiding the raters' criteria. On the other hand, the author admitted that the general English classes attended by the control group suffered from a lack of speaking practice. Therefore, improvement could be associated just to the introduction of pronunciation training within the class rather than the effectiveness of intonation instruction.

Kennedy and Trofimovich (2010) designed a thirteen-week English pronunciation course based on suprasegmentals in Canada. Each week ten adult learners with different mother tongues attended a three-hour session. They took a pre-test in week 1 and a post-test in week 11 based on a reading-aloud passage. Ten English teachers and teachers-to-be rated the tests accentedness, comprehensibility and fluency using a 9-point Likert scale. No significant improvement was found in any of the variables. However, students wrote a journal thanks to which it was possible to know a little bit more about their learning progress. Apparently, those students who reported greater awareness of the pronunciation items under study were those who obtained better results in the tests, a correlation that was shown significant.

In spite of the scarcity of classroom-based studies, the instruction of suprasegmental features has been found effective. However, not only the final purposes of the studies differ, but also their teaching approaches: Some of them opted for stand alone classes (Derwing et al., 1997; Kennedy & Trofimovich, 2010; Ramírez Verdugo, 2006; Sardegna & McGregor, 2013), while others preferred to incorporate lessons in a broader course (Chela de Rodríguez, 1981, as cited in Chela-Flores, 1997b; Couper, 2006; Derwing et al., 1998); The organization of the sessions could follow a presentation-practice-production sequence (Gordon et al. 2013; Kennedy & Trofimovich, 2010), a perception-practice-presentation sequence (Chela de Rodríguez, 1981, as cited in Chela-Flores, 1997b), or even a mixture of the two (Couper, 2006; Ramírez Verdugo, 2006); Also, there were studies focusing only on explicit instruction (Chela de Rodríguez, 1981 as cited in Chela-Flores, 1997b; Couper, 2006; Kennedy & Trofimovich, 2010) whereas there were others which tested differences between explicit, implicit and no instruction (Derwing et al., 1997, 1998; Gordon et al. 2013; Ramírez Verdugo, 2006). Although there is no consensus among studies, it is fundamental to also analyze which teaching approaches are more likely to improve students' pronunciation of the target language for communicative purposes.

2.4. How to teach

The methodologies related to second language teaching go along with the historical evolution of pronunciation instruction. As mentioned in section 2.1, traditional approaches taught a decontextualized language through imitation and repetition. Therefore, the emphasis fell on the acquisition of the language itself, no matter the context. The goal of the communicative approach, on the other hand, was to learn to be

communicative when using the language. The focus of the instruction, thus, shifted to content, making language the tool needed to talk about different subjects (Lightbown & Spada, 1999). Consequently, explicit pronunciation instruction was generally avoided in class. Nowadays this early version of the communicative approach is more likely to be found in content language integrated learning (CLIL) classes, where subjects like maths or physical education are taught in a second language, rather than in second language classes. In fact, what is brought into focus is one of the main differences between CLIL and ESP settings: it is not the same to teach English for engineering students (focus=language) than to teach engineering in English (focus=content).

Regarding second and foreign language instruction nowadays, the focus has moved towards achieving language through different communicative contexts, considered one of the most effective techniques to teach a second language (Lightbown & Spada, 1999). This change in perspective somehow favors the integration of pronunciation for communicative purposes into the second language curriculum, a much welcome practice for pronunciation teaching researchers (Celce-Murcia et al., 1996; Chela-Flores, 2001; Hahn, 2004; Iruela, 2007; Levis & Grant, 2003; Munro, 2008).

2.4.1. Why to teach pronunciation explicitly: Form- vs. meaning-based instruction.

As shown throughout the examination of classroom-based studies carried out in section 2.3.3, explicit pronunciation instruction is indeed effective to improve students' intelligibility and comprehensibility. In fact, it is especially useful when teaching adult learners since it appeals to the "adults' cognitive maturity, metalinguistic awareness and world knowledge" (Lightbrown & Spada, 1999, pp. 32-33). However, it is useless to

learn how to pronounce colors in English if students cannot describe a landscape they saw or the outfit they wore to a wedding, for instance. For this reason, form-based instruction should always be contextualized.

Saito (2012) extensively examined fifteen quasi-experimental studies focusing on the instruction received: form-focus instruction in controlled circumstances only (FonFS), form-focus instruction within both controlled and extemporaneous communicative contexts (FonF), and meaning instruction (FonM). He highlighted the effectiveness of FonF instruction over the other two: On the one hand, FonFS instruction only showed improvement at structured situations, failing to provide knowledge that transcends the classroom setting due to a lack of improvised practice. On the other hand, FonM did not show signs of improvement, failing to provide language training to solve communication breakdowns. FonF, thus, was the only sort of instruction that revealed a positive evolution at both controlled and extemporaneous circumstances, helping students to learn linguistic concepts that they could consciously apply in communicative situations. This study agrees with those assumptions that advocate for a focus on both form and meaning to effectively teach pronunciation (Celce-Murcia et al., 1996; Iruela, 2007; Levis & Grant, 2003; Lightbrown & Spada, 1999) and, at the same time, highlights the importance of practicing speaking under both structured and unstructured conditions.

2.4.1.1. Celce-Murcia et al.'s steps to teach communicatively.

In an attempt to provide an action plan that ensures pronunciation teaching within a communicative framework, Celce-Murcia et al. (1996) outlined one of the most popular set of steps to follow, which are still used, studied and recommended within the second

language classroom nowadays (Gordon et al, 2013; Sardegna & McGregor, 2013). Related to the learning process described by Chela de Rodríguez (1976), the framework consists of the steps shown in Table 1.

Table 1. Celce-Murcia et al.'s steps to teach communicatively (1996, p. 36)

1. Description and analysis (e.g. oral and written illustrations of when and how the feature occurs to raise learner consciousness)
2. Listening discrimination (focused listening practice with feedback)
3. Controlled practice and feedback (e.g. oral reading of minimal pair sentences, short dialogues, etc., with special attention paid to the highlighted feature)
4. Guided practice and feedback (e.g. structured communication exercises that enable the learner to monitor for the specified feature, such as information gap activities, cued dialogues)
5. Communicative practice and feedback (e.g. less structured activities that require the learner to attend to both form and content of utterances)

Celce-Murcia et al. (1996) supported a presentation-practice-production sequence and encouraged the introduction of activities that are heavily structured first, and progressively evolve to more extemporaneous exercises. Also, it is important to highlight the role of listening discrimination: Students work on distinguishing and recognizing pronunciation features before practicing them, firstly training perception to enhance production, a practice related to Flege's (1995) approach of error gravity.

It must be said, though, that it is not always easy to build this communicative framework on the foundations of a given curriculum. When working with ESP students it is important to bear in mind the technical content of the subject to foster successful and useful communication, for example. Therefore, this and other variables should be

examined in detail to guarantee a complete integration of pronunciation into the curriculum of the L2 class.

2.4.2. How to integrate pronunciation in a curriculum.

In spite of the fact that the choice whether to include pronunciation within the general curriculum of second language teaching or to devote a separate program to this topic tends to be made by the institutions themselves (Derwing & Munro, 2015), many experts agree that it is good to deal with pronunciation within regular classes as another linguistic skill (Celce-Murcia et al., 1996; Chela-Flores, 2001; Hahn, 2004; Iruela, 2007; Levis & Grant, 2003; Munro, 2008). In the end, pronunciation, as grammar or vocabulary, is part of the language, and as so it needs to be practiced over time to obtain positive results. Nevertheless, studies reporting the teachers' dedication to pronunciation in the classroom revealed that there is a general tendency to avoid it (Henderson et al. 2012; MacDonald, 2002). Among the different reasons mentioned (lack of time, no previous training, an so on), practitioners claimed that pronunciation did not tend to be well integrated within the curriculum, which did not provide precise guidelines on what to teach and how to do it (MacDonald, 2002). The results of Henderson et al.'(2012) s survey showed that Spanish teachers, for instance, felt conditioned by schools' demands to reach concrete results, so they preferred to prioritize other skills and vaguely deal with some pronunciation issues that happen to appear, at best, while practicing speaking.

In order to include pronunciation within the curriculum, a series of steps have to be followed: the immediate needs of the students should be analyzed, learning goals should be set, materials should be adapted to the course to meet the established goals, a plan for

implementation should be structured, and both the course and students' results should be assessed with the proper means (Derwing & Munro, 2015). These preliminary and follow-up steps have also been recommended to plan other courses such as ESP ones to ensure and test their effectiveness (Dudley-Evans & St John, 1998). This way, problems can be predicted and errors can be avoided according to the circumstances of the group.

Within the different types of existing English courses (ESL, EFL or EAP, among others), ESP is possibly one in which pronunciation instruction is highly neglected. Students tend to focus more on writing and reading material and, if speaking is practiced, pronunciation does not tend to be explicitly addressed despite the fact that features such as phrasing, pausing, and intonation are considered important for the students' comprehensibility (Dudley-Evans & St John, 1998). However, it is possible to include it in the classroom as in any other second language context. A good example of it is Anderson-Hsieh (1990), which provided science-oriented material to work on suprasegmentals within a course addressed to ITAs who were teaching within a chemistry degree. The activities and materials were not only revealed to improve students' intelligibility, but also to be more motivating for them. Therefore, knowing the audience you are addressing becomes also a key factor for the teacher to succeed in the classroom.

2.5. Who to teach

Apart from linguistic (suprasegmentals vs. segmentals), institutional (curriculum issues), methodological (types of instruction), and setting (EFL, ESL or EAP/ESP) variables, it is also important to examine the learners' features that are at play (i.e. students' age, first and second languages, aptitude and motivation, among others) for an

efficient pronunciation instruction (Celce-Murcia et al., 1996). These factors, together with the needs of the students and the goals of the instruction, should be taken into consideration to decide on the materials and activities to implement. Although not all of them can be a hundred percent controlled, they must be examined to face potential flaws in the teaching that could affect results.

2.5.1. Age.

Despite disagreements in determining which is the exact critical period from which a learner finds it more difficult to improve and acquire a second language, what is clear is that young and adult learners do not acquire a second language in the same way, a fact that influences their learning process. Hence, it is important to examine the type of students at hand so as to offer an instruction that suits them better.

Regarding pronunciation, there seems to be a general understanding that adult learners often fail to achieve a native-like accent (For a review, see Derwing & Munro, 2015). In many cases, in fact, their pronunciation is branded "fossilized" (Chela-Flores, 1997a; Derwing et al., 1997). They can be intelligible when communicating, though. As shown in previous sections, this type of students can improve their global intelligibility after receiving appropriate pronunciation training (Derwing et al., 1997, 1998; Gordon et al., 2013). Therefore, focusing on introducing features and techniques that improve their intelligibility seems to be more beneficial for these students.

However, it is true that the problems to understand, for example, German adult speakers when speaking French, do not tend to be the same than those found when speaking English, and, at the same time, these tend to differ from those of Spanish learners of English. Despite the possibility of sharing some of the difficulties, some others are

completely unrelated. Hence, the combination of the languages under study also influences the intelligibility of the speakers.

2.5.2. Languages.

Although age is an important conditioning factor as far as second language acquisition is concerned, it is possible to find adult learners who reach high proficiency levels. Thus, other factors should be taken into account as well. As mentioned in section 2.3.1, both the students' first language and the target language influence the extent to which certain elements are more or less important to acquire. Derwing and Munro (2015) observed that a Dutch learner of English will probably experience more problems with segmentals than suprasegmentals, because Dutch and English do not differ much form each other in terms of prosody. By contrast, a speaker of a syllable-timed language such as French or Spanish is predicted to struggle more when learning the stress-timed rhythm of English. Therefore, the proximity between the first and the second languages affects the learning process and should be taken into account when deciding what is more productive to teach in class.

Not only the linguistic, but also the spatial closeness to the target language plays a role on the learning process. Students may be learning the language of the country where they reside (i.e. second language), or they may be taking classes of a language that is rarely spoken in their own country of residence (i.e. foreign language). The opportunities to practice the language outside the classroom are different in these two scenarios: while the former can communicate in the target language on a daily basis, the chances of the latter to do so tend to be reduced to the class setting. It is true that nowadays it is easier to have contact with the target language thanks to the Internet, by

watching videos or playing videogames online, for instance. However, students often make use of these resources in a passive way. For this reason, providing activities and materials that help foreign language learners practice in class is of utmost importance.

Nonetheless, it is true that, although second language learners have the chance to take advantage of their situation, a lot of them do not. Many second language students show slight signs of improvement despite living in the country of the target language maybe because they just interact with people who speak their mother tongue or they spend a lot of time on their own, among other reasons. As a consequence, practice in the L2 class becomes also essential to guarantee an improvement in their language skills.

Providing opportunities to practice the target language should be fundamental for L2 instructors, but the students need to be willing to take them to really improve. Other personal factors such as motivation, hence, also play an important role on how learners acquire a second language.

2.5.3. Motivation and aptitude.

In addition to general characteristics of the learner group, there are individual traits that also influence the student's pronunciation acquisition. Derwing and Munro (2015) pointed out that there is "a widespread view that some L2 speakers are simply better pronouncers than others because of an underlying talent for pronunciation itself" (p. 46). There might be some learners who stand out from the group because they easily assimilate and acquire certain concepts. In other words, these students show a remarkable aptitude to acquire the pronunciation of the language. According to Lightbrown and Spada (1999), aptitude for second language acquisition is made of different abilities, which go from distinguishing and memorizing new sounds to

understanding grammatical functions. Regarding pronunciation, studies have linked aptitude to learners' musical skills or their capacity to accurately imitate new sounds (for more information, see Derwing & Munro, 2015), but that does not mean that students who do not have a role model aptitude cannot improve their pronunciation. There are different abilities that may put students in a privileged position when acquiring a second language, especially pronunciation, but a more communicative approach towards second language teaching has challenged the idea that these are vital to be intelligible (Lightbrown & Spada, 1999).

What really seems to defy pronunciation teaching is the students' will to learn, that is, their attitude and motivation. Lightbrown and Spada (1999) claimed that the students' motivation is mainly defined by the learners' need to communicate and their attitude towards the language community. If students have to speak the target language to succeed in their career, for instance, they will probably get more involved in the second language classroom. Likewise, if learners are keen on American movies or Britpop music, they are likely to be more engaged in acquiring the language to understand the meaning of the dialogs or the lyrics. Then, motivation is built up by the learners' personal enrichment, known as integrative motivation, and their goals, coined instrumental motivation (Lightbrown & Spada, 1999). These intrinsic factors, together with their self-confidence and personality, boost students will to communicate (Derwing & Munro, 2015) and consequently trigger students' concern for acquiring pronunciation, making them feel more attracted and eager to achieve a good command.

Both aptitude and motivation can be controlled to certain extent in the L2 classroom by, for example, including a wide variety of activities that appeal both to low and high level students, or exercises that are linked to their professional field. However, it is not

possible to fully control the individual demands of each student. Apart from the student's talent and attitude towards pronunciation, each individual learner shows other personal circumstances that could also influence their learning process, such as experience abroad or outside the classroom, or a different degree of cognitive development (i.e. attention, memory, or thinking skills). The teacher's responsibility is mainly to respond to the general needs of the group to ensure that all, not just a few, of the members improve despite their individual conditions.

2.6. Teaching language rhythm as an intelligibility enhancement

Introducing rhythm within the L2 classroom can be beneficial to enhance the students' communicative skills (Anderson-Hsieh, 1990; Chela-Flores, 1997b, 2003; Derwing & Munro, 2015; Frost & Picavet, 2014; Gilbert, 2008; Goodwin et al., 2013; Ordin & Polyanskaya, 2015; Quené & Van Delf, 2010; Tajima et al., 1997; Tuan & An, 2010; Wong, 1987). Rhythm anticipates and delimitates crucial information for the successful comprehension of a message. For instance, the listener can recognize words just by trusting rhythmic patterns (Sweet, 1900). Indeed, a misuse of the right rhythm in the target language can lead to serious communication issues that impede mutual understanding (Gilbert, 2008). This is particularly true when the L1 and the L2 clearly differ in rhythmic patterns, such as the case of Spanish speakers learning English.

Tajima et al. (1997) were not the only ones to prove the influence of rhythmic patterns on L2 speakers' intelligibility. Anderson (1993) found a certain correlation between the length of interstress intervals and the speakers' comprehensibility in English: Those speakers who produced shorter interstress intervals and fewer stressed syllables were considered more comprehensible (as cited in Chela-Flores, 1997b, p. 115). Hahn

(2004), one the other hand, conducted an experiment with different versions of an ITA's speech and proved that native listeners thought it was more intelligible when primary stress was correctly placed.

Whether rhythm should be at the forefront of the features to teach has been questioned, though. Jenkins (2000) did not encourage teaching how to produce an English-like rhythm for intelligibility purposes when English is used as a lingua franca, as nonnative speakers are the ones involved in the communicative context and they can adopt a different rhythm without jeopardizing communication (as cited in Levis, 2018, p. 143). Levis (2018) argued, however, that in this globalized world L2 speakers are in contact with native Englishes either because they live in an English speaking country, or because they have access to means of English origin, such as TV series or broadcasting news, so its important that they are aware of how natives produce rhythm. He also maintained that even when rhythm does not interfere with the correct interpretation of the message, it is likely to ease its understanding, so it affects comprehensibility and, consequently, global intelligibility.

When integrating rhythm within the general curriculum, it is necessary to put emphasis on those rhythmic aspects that are easier for L2 learners to understand, acquire, and apply when speaking and that, at the same time, make a greater difference for effective communication. Second language courses have always suffered from timing restrictions, since all the competences need to be taught and practice in class, so the time spent on teaching rhythm is definitely limited. Currently, there is no a general agreement on which rhythmic aspects are the most suitable to teach, but the choice of one or another has to rely on both the students' mother tongue and target language. In Chela-Flores (1997b), Couper (2006) or Tsiartioni (2011), rhythm was introduced in the

classroom, reaching encouraging results. However, while the Spanish learner-oriented teaching program described by Chela-Flores (1997b) was based on early perception of rhythmic patterns, Couper (2006) concentrated on epenthesis and absence because it was found to be especially detrimental for immigrants whose origin was largely Asian.

In the case of Spanish/Catalan speakers learning English, rhythm can be tackled through vowel lengthening (see section 2.6.5). Not only has vowel duration been pointed at as one of the most remarkable indicators of stress in English, but it is also a quite simple concept that can be easily introduce in the ESL/EFL classroom. Also, the prosodic hierarchy of rhythm can help the teacher decide the order in which rhythmic elements should be introduced to the students. These suggestions have influenced the design of the pronunciation module of the study here presented. However, to fully understand the role of rhythm within pronunciation teaching, it is important to further elaborate on the concept of speech rhythm itself, its components, and the studies conducted about it.

2.6.1. Defining language rhythm.

Despite several attempts to define what language rhythm is, researchers do not agree on a universal definition. While Derwing and Munro (2015) described rhythm as "patterns of stress within phrases, sentences, and longer utterances" (p. 60), Patel (2007) understood it as "systematic patterning of sound in terms of timing, accent, and group" (p. 96). In general terms, rhythm has always been an organizational tool traditionally associated with the artistic concept of timing both in music and speech: The speed, the pulses, and the pauses of a musical piece or a discourse are adjusted to guarantee the harmonious transmission of a message (Patel, 2007).

By means of the rhythmic organization, the listener can anticipate information that helps understand the intended message. When listening to music, rhythm helps people adjust their dance movements in the right way as a response to the pattern; when listening to a speech, on the other hand, rhythm encourages the listener to predict important syntactic and lexical information needed for successful communication (Chela de Rodríguez, 1976; Derwing & Munro, 2015; Patel, 2007; Sweet, 1900).

However, the regularity of timing in language is not as consistent as in music, at least not in a straightforward way. While it seems fairly easy to distinguish the beats that establish the foundations of a song, the rhythmic patterns of languages are not as easily recognizable due to their more flexible nature. In other words, people would not show too much trouble to identify and mentally follow the three-beat rhythm of different waltz songs, but they may struggle when identifying the rhythmic elements of different English speakers. Even so, they might be able to recognize a similar *melody* among them. A good example are people who do not understand English but can still guess when someone is speaking in English, or imitate the English *tonality* when speaking another language [e.g. Spanish speakers imitating English natives speaking Spanish (Feist, 2018)].

Although several studies have scientifically shown a lack of regular linguistic division of time (i.e. *isochrony*) in speech rhythm, others still believe that there is some sort of perceptible temporal systematicity in the organization of speech (for a review, see Dauer, 1983; Lehiste, 1977). This phenomenon occurs thanks to the different existing combinations of rhythmic features such as duration, tone or pitch, which will be further discussed in the following section. In fact, teaching this perceptible pattern could become a helpful mechanism to enhance L2 students' intelligibility in their target

language. As previously mentioned, studies like Tajima et al. (1997) examined the effect of certain rhythmic cues in the intelligibility of nonnative speakers obtaining encouraging results. Hence, its introduction in the L2 classroom may be beneficial for the learner's communicative skills.

2.6.2. The prosodic hierarchy of rhythm.

As indicated by the definitions quoted at the beginning of the previous section, the prosodic elements of the oral discourse are intertwined and structured at different levels, creating, among other suprasegmentals, the rhythm of the language. Patel (2007) cognitively relates the organization of music with the organization of the discourse. As a piece of music can be arranged in beats, note values and bars, the speech is divided into different elements which are assembled according to perceptual boundaries that go from simple (i.e. the syllable) to more complex (i.e. the utterance) connections. In both cases, the boundaries are mainly established by the use of pausing.

The hierarchical structure of the discourse helps us to understand the links among syllables within words, phrases, and full sentences, which constitute the complete meaning of speech. Although all these connections play a role, those that create thought groups (also known as intonational units) are considered key to build the rhythm of a language. Thought groups are semantically and syntactically coherent and meaningful segments into which an utterance is divided. In spite of the fact that there are no strict rules on how to segment the discourse into intonational units and the division can vary depending on the speed, people have an intuitive notion of when it is right to pause (Celce-Murcia et al., 1996). For example, if in a sentence like " the teacher told me to

close the door", the speaker pauses like "the teacher told/ me to/close the/door", the sense, intention and meaning of the sentence are lost.

On the other hand, the perception of prominence and rhythm show some commonalities (Levis, 2018). If we listen to a drummer play a three-beat rhythm striking the drumsticks strongly on the first beat, our brain connects the pattern to a waltz rhythm. However, when the drummer strikes strongly on the last beat, our brain connects the pattern to a more rock-like song. Languages also differ in rhythmic prominence. English, for instance, emphasizes stressed syllables in such a way that an interstress interval pattern becomes clearly perceptible, while other languages like Spanish or Catalan do not use stress in the same way (see section 2.6.4). Prominence, hence, also outlines the rhythm of a language at the different prosodic levels.

This concept of the prosodic hierarchy has been adapted through different pronunciation teaching perspectives so as to help students adopt the rhythm of the target language (Chela-Flores, 1997b; Gilbert, 2008). However, there are many rhythmic features that come into play at each level, which need to be examined in depth in order to fully understand the concept of speech rhythm.

2.6.3. Rhythmic features.

Prominence is obtained by stressing certain elements to stand out from others that do not receive so much attention (Derwing & Munro, 2015). In languages, prominence can vary according to the application of certain prosodic features. It is precisely the combination of these prosodic features that make up the rhythm of a language.

Delattre (1966) distinguished three main prosodic cues that shape speech rhythm: length, loudness and pitch. Dauer (1987) and Patel (2007) included vowel quality as a

forth cue. These elements are conditioned by the combination of phonological features such as syllable duration and structure, full and reduced vowels and consonants, or the intonation contour (Dauer, 1983, 1987), which shape the rhythmic pattern of a language. For instance, while English reduces vowels in unstressed position and has a wide range of syllable structures, Spanish uses just full vowels and has quite simple syllable structures (usually, CV and CVC). The pitch contour or the F0 amplitude, on the other hand, can ease the identification of prosodic boundaries and, hence, recognize thought groups (Gilbert, 2008; Patel, 2007). Hence, it is the existence, intensity or even absence of these phonological components that determine the rhythm of a language (Celce-Murcia, 1996; Derwing & Munro, 2015), giving rise to rhythmic phenomena like sentence stress, linking or blending (Wong, 1987).

Stress in English is characterized by considerably increasing the length, loudness and pitch of a syllable within words, or words within longer utterances. Knowing that is essential to understand and speak English because, not only does it clarify lexical and grammatical misunderstandings, but it also points at the meaningful information of speech. By contrast, in spite of the fact that each word has its stressed syllable and there are some well-defined rules on how to use syllable stress in Spanish, stressed syllables are not as long, loud and high in pitch as the English ones. Besides, sentence stress is not used to highlight important content; in Spanish a more flexible syntax allows pushing this information towards the end of a sentence in order to make it stand out (Chela de Rodríguez, 1976).

These rhythmic cues affect timing. It is logical to think that a long stressed syllable will take more time to be uttered than unstressed syllables in the same language, but also than shorter stressed syllables of other languages. Timing, however, has created

controversy when relating it to language rhythm due to a lack of consistency, leading to different views on how rhythm is produced, perceived, and classified.

2.6.4. A question of time: Controversy around isochrony.

As explained in the previous section, the combination and degree of participation of the different rhythmic features described above constitute one language rhythm or another. Inevitably, this combination affects duration and, consequently, timing. As in music, speech rhythms have been traditionally classified according to temporal patterns. This attempt to rhythmically organize speech in similar timing intervals has been traditionally referred to as "isochrony" (Lehiste, 1977, p. 254). Pike (1945) and Abercrombie (1967) considered isochrony to determine a fixed classification of rhythm: Languages were considered to have either a syllable-timed rhythm if the duration between syllables was approximately the same (e.g. Spanish, Italian or French), or a stress-timed rhythm if the duration between stresses was almost equal (e.g. English, Dutch and German). Perception wise, Lloyd James (1940) had previously described them as similar to a machine gun shot and a morse code message respectively (as cited in Ramus, Nespor and Mehler, 1999, p. 266). Ladefoged (1975) included a third category defined by the duration of moras, such as in Japanese (as cited in Nespor, Shukla and Mehler, 2011, p. 1149). However, this research study is going to concentrate on the first two categories within which the languages under study have long been placed.

By empirically measuring interludes, rhythm has been proved not to fall on exact intervals of syllables or stresses (for a review, see Dauer, 1983; Lehiste, 1977). Dauer (1983, 1987) claimed that duration differs depending on certain phonological aspects,

such as vowel reduction or syllable structure, which may be present in a language or not. Hence, a strong presence or absence of certain phonological features determines whether the rhythm is more syllable-timed or stress-timed. As a consequence, rhythm stopped being seen as a fixed two-way classification and it started being considered as a continuum between syllable-timed and stress-timed patterns over which the languages of the world were placed (see following section for further explanation).

Despite the fact that a lack of a complete isochrony has extensively been assumed in the last decades, the belief in an isochronic tendency of language rhythm is still dormant. Although openly taking a stance against isochrony, Dauer (1983) raised an interesting observation: "The problem is how to interpret data: two intervals of 50 cs [centiseconds] and 57 cs are objectively unequal, but are they to be considered 'the same' in production, or at least in perception?" (p. 52). Lehiste (1977), among others, considered that sometimes the difference in length among intervals is so small that the listener does not perceive it. Besides, the listener would tend to modulate the regularity of intervals systematically so as to structure sounds. This unconscious strategy would allow the listener to organize concepts and consequently ease comprehension.

The perception of a certain degree of isochrony allows the speaker to adjust the rhythm of the target language into quasi-equal intervals of stress when this is stress-timed, and into syllables when it is syllable-timed, which makes rhythm teachable (Frost & Picavet, 2014). L2 learners whose mother tongue differs from the target language in terms of rhythm could benefit from rhythm instruction in order to avoid negative transfer and, thus, ease their comprehensibility when speaking the second language. For this reason, it is important to know the rhythmic features of both the L1 and the L2 so as to learn how to adapt to them and ensure communication.

2.6.5. Differences in rhythm between English and Spanish/Catalan: The role of vowel duration.

Although there is an extensive acceptance of the dichotomy between syllable- and stress-timed languages, the debate on rhythm classification remains open nowadays (Patel 2007). As mentioned in the previous section, English and Spanish have been considered as the prototypical languages for stress- and syllable-timed rhythms respectively, but it has been seen that not all the languages can be strictly classified under one of the two categories. Catalan is a clear example: Despite its syllable structure being relatively simple as in Spanish, it also suffers from vowel reduction, as in English. For this reason, Catalan has been seen as an intermediate language placed somewhere in between stress- and syllable-timed languages (Ramus et al., 1999).

However, its rhythm seems to be closer to the Spanish rather than the English one. Prieto, Vanrell, Astruc, Payne and Post (2012) discussed that, in fact, vowel reduction does not always affect duration. In standard Catalan, vowels get reduced by mainly changing their quality. For instance, an /a/ sound becomes *schwa* or an /o/ sound is pronounced as /u/ in unstressed position. Therefore, the vowel gets a little bit shorter but the contrast with full vowels is not as evident as in English, whose reduced vowels are drastically shortened to the point that sometimes it is very difficult to hear them in conversation. Apparently, the emphasis that falls on stressed syllables in English causes a lengthening of the vowel (for a review, see Patel, 2007). Catalan and Spanish stress, by contrast, do not have such strength on stressed syllables, so the difference in duration between stressed and unstressed syllables is not as marked. Besides, Catalan variations such as Western Catalan do not always reduce the quality of the vowel when unstressed.

On account of this, Catalan rhythm is still considered more syllable-timed than stress-timed.

Vowel duration affects syllable duration, which not only distinguishes between the two groups at the end of the rhythm continuum (Dauer, 1983, 1987; Delattre, 1966, among others), but also affects the comprehension of nonnative speakers' oral production (for more information, see Chela-Flores, 2003). Consequently, showing native speakers of syllable-timed languages such as Spanish or Catalan how and when to lengthen stressed syllables when speaking stress-timed languages like English by means of vowel duration may help them be more comprehensible in the target language.

2.6.6. Rhythmic measures: VarcoV as an assessment tool in rhythm instruction.

Measuring the duration of rhythm has contributed to refuting the existence of isochrony within languages, challenging its traditional classification into the syllable- and stress-timed categories. However, the emergence of the analysis of rhythmic phonological features has provided a new dimension to measurements of duration, shifting them into potential tools to correlate phonetic cues, such as vowels (Patel, 2007). As a consequence, a durational difference between stress- and syllable-timed languages remains plausible if duration is considered from an acoustic perspective.

Grabe and Low (2002) measured the vocalic and intervocalic duration of eighteen different languages traditionally classified as either syllable-, stress- or mora-timed in order to test the validity of this classification. Since a characteristic of stress-timed languages is that their use of stress significantly increases the duration of stressed vowels in comparison to syllable-timed languages, a higher variability in vowel duration was expected in the former group. Grabe and Low (2002) measured the

pairwise variability indices of the different languages and correlated the results among them. Findings showed that vowel duration variability is, indeed, bigger for stressed-timed languages. Hence, results supported the traditional classification of rhythm, although the heterogeneity among languages suggests that this categorization is not static, but that languages can be more stress- or syllable-timed depending on their phonological and phonetic features.

Ordin and Polianskaya (2014, 2015), Prieto et al. (2012), Ramus et al. (1999), and White and Mattys (2007a, 2007b), who studied the use of rhythmic measures to classify speech rhythm, also supported the classic classification. White and Mattys (2007a) compared different rhythmic measures and concluded that the most reliable ones to distinguish among languages are %V, VarcoV and nPVI-V. Still, there is not a consensus on the single most reliable rhythmic measure.

Despite the fact that the overall reliability of rhythmic metrics to classify speech rhythm has been questioned due to a lack of consistency (Arvaniti, 2012; Ross, Ferjan & Arvaniti 2008), evidence from the abovementioned studies contributes to an enduring belief in the traditional categorization. However, not many studies have examined the veracity of these metrics to measure the evolution of L2 speakers. White and Mattys (2007a) took a step further and investigated the metrics that capture the acquisition progress of L2 speakers of English and Spanish. In this case, %V and VarcoV were shown to be the most credible. In a further study, they compared the results obtained with these measures to listeners' nativeness ratings and they interestingly found that listeners rated Spanish speakers of English as more nonnative-like when their VarcoV values were lower (White & Mattys, 2007b). Hence, VarcoV seems to be able to measure the acquisition progress of English rhythm by Spanish speakers to some extent,

a fact that may help to provide instrumental evidence to studies of rhythm instruction, at least, when this group of nonnatives is brought under analysis. In the end, as Celce-Murcia et al. (1996) claimed, "if the stress and rhythm patterns are too nonnative like, the speakers who produce them may not be understood at all" (p. 131).

The effectiveness of rhythmic metrics to measure intelligibility has been questioned in pronunciation teaching studies because of the lack of listener corroboration (Derwing & Munro, 2015), leading to its rejection as an assessment tool of the learning progress. However, metrics could provide some instrumental estimation of the evolution that could alleviate the subjectivity load of which popular intelligibility assessment tools have been accused. For instance, Tsiartoni (2011) applied both Grabe and Low's (2002) vocalic and consonant variability measures and observed an adaptation of a more English-like rhythm by those having received rhythm instruction. Unfortunately, no intelligibility tests were conducted to additionally correlate the adaptation to the target rhythm with an improvement on their communicative skills, which could have provided more information about the influence of rhythm on the students' global intelligibility.

2.6.7. Issues with assessment: Rhythmic measures vs. listeners' judgments.

How to assess intelligibility, comprehensibility and fluency has caused a heated debate. As mentioned in section 2.3.2, the communicative dimension of this type of assessment leads us to use evaluative tools that imply a human judgment. For this reason, the most popular technique applied is the listeners' appraisal by means of either forced choice identification tasks and transcription exercises to test intelligibility, or scalar ratings to test comprehensibility (for a complete list, see Thomson, 2018). These assessment tools are rather simple and user-friendly so, in principle, any person can use them without

much trouble, even if not linguistically trained, offering veracity and transparency (Isbell, 2018).

However, the use of human evaluators implies a certain degree of subjectivity in the judgments. First, listeners may have a different idea of what intelligibility is about and rely on different pronunciation cues when rating, or they may be more or less familiar with the subjects' native language, which may lead to biased assumptions impaired by social prejudices or a previous awareness of the common mistakes that a specific group of L1 speakers make (Derwing & Munro, 2015). Therefore, the knowledge and relationship the raters have with both the L1 and the L2 may influence whether they apply the scale in a more rigorous or flexible way (Isbell, 2018; Munro & Derwing, 1995, 2015; Southwood & Flege, 1999; Thomson, 2018). Second, other extrinsic agents such as fatigue, environmental noise, or family duties can also influence the listeners' opinions, since they act as distractions and alter the conditions under which the test is performed (Munro & Derwing, 2015; Thomson, 2018). Although conducting statistical analyses of reliability is necessary to assess the consistency of the judgments (Munro & Derwing, 2015), it is still not enough to fully control the human factor.

On the other hand, the test itself may undermine the listeners' appraisal as well. A 9-point scale is Derwing and Munro's favorite and has become the standard scale in comprehensibility studies (Thomson, 2018). Southwood and Flege (1999) turned down the use of small numerical scales such as 5- or 7-point ones in an accetedness study because variance cannot be estimated above a certain level in these scales and, thus, its use may lead to a ceiling effect. In contrast, Isaacs and Thomson (2013) examined the raters evaluation process in depth and found out that some raters considered a 9-point scale confusing due to the wide range of possibilities it offers, while they thought that

the 5-point scale did not give rise to as many doubts. In a recent study, Nagle (2019) applied a many-facet Rasch analysis to examine the severity of the raters evaluating comprehensibility, fluency and accentedness via a 9-point scale. Among the findings obtained, Nagle (2019) concludes that the scale steps were compressed for all three items tested, suggesting that a 5- or 7-point scale would have been more convenient due to the homogeneity of the speakers. Other possibilities beyond the use of Likert scales have been also explored. For instance, Saito, Trofimovich and Isaacs (2014) effectively applied a 1,000-point continuous sliding scale for pronunciation assessment purposes (as cited in Saito, Trofimovich and Isaacs, 2016, p. 220).

An additional problem in regards to listeners is how to recruit them. In EFL contexts, where generally the target language is not spoken, it is difficult to obtain a big enough number of listeners to provide reliable estimations. Nagle (2019) introduced the possibility to use online tools such as Amazon's *Mechanical Turk* to crowdsource raters, an option that is worth exploring in detail.

A thorough testing of the validity of rater-based assessment tools and their application help alleviate potential misconceptions and misguiding results (Harding, 2018), but full neutrality cannot be guaranteed. Ghanem and Kang (2018) and Van Moere and Suzuki (2018) pointed out that one of today's challenges in second language pronunciation research is to conciliate the communicative veracity obtained from listeners' judgments with the scientific objectivity of instrumental measures. Technological advances like speech recognition software are a potential solution. However, more research needs to be conducted to prove its reliability (See Kang & Ginther, 2018). Meanwhile, a handy solution is to conduct studies based on a mixed-method assessment that uses already-tested instrumental tools, such as acoustic measurements, combined with listeners'

perceptions, which have already helped to provide more evidence on detecting the linguistic aspects that influence comprehensibility and fluency (Derwing, Rossiter, Munro & Thomson, 2004; Isaacs & Trofimovich, 2012; Kang, Rubin & Pickering, 2010; Lennon, 1990; Riggenbach, 1991; Saito, Trofimovich, Isaacs and Webb, 2017). For the present study, it was chosen to introduce rhythmic metrics as an assessment tool together with raters' judgments to examine the effectiveness of rhythm instruction in depth and thus offer a more complete picture of what happens during and after the learning process.

Chapter 3. Methodology

This chapter presents the design of the study that was developed to test the hypotheses presented in section 1.1. First of all, the framework of action will be outlined, defining the setting where the experiment was performed (i.e. the course). Next, key information about the participants will be provided and the two treatments under study will be described. Then, the materials selected and created for the purpose of the experiment will be analyzed in detail, both the assessment tasks and the teaching resources. Finally, the assessment mechanism applied and the statistical analysis carried out to examine the outcomes will be presented.

3.1. Design

The present study is enclosed in Derwing and Munro's (2015) framework of a classroom-based pronunciation research study. It consists of a quasi-experimental longitudinal study that sought to test the effectiveness of explicit rhythm instruction within the EFL classroom to improve students' comprehensibility and fluency. To this end, a ten-week pronunciation module was created and embedded within a technical English course. The experiment followed a pretest-posttest design to assess the progress of the subjects. Although all the participants took pronunciation instruction, some of them took explicit rhythm training (experimental group) while the others did not (control group).

Both the course and the academic year schedules were used to assign treatment to the groups. On the one hand, the existence of breaks during the year limited the choices. Since breaks implied that classes would be cancelled either on Monday or Friday, it was decided not to use the lectures scheduled for those days as part of the study. On the

other hand, treatment was assigned depending on the time of the day the lecture was taking place in order to avoid the time difference affecting the study (students that took the course at eight o'clock in the morning could be less participative than those at noon, for instance), trying to create "schedule-balanced" groups. Table 2 shows the distribution of the treatment groups.

Table 2. Distribution of treatment groups within the course schedule

Class Time	Monday	Tuesday	Wednesday	Thursday	Friday
8-10	G1	G1	G5	G4	G4
12-14	G2	G2	G3	G3	G5
15-17		G6			
17-19	G6				

Note. G = Group. Red = Experimental group. Blue = Control group.

Groups 1, 2 and 5 became the experimental groups, while groups 3, 4 and 6 became the control groups. Three different teachers taught the regular classes: the course coordinator taught G2, G3 and G6, a teacher from the English department taught G1 and G4, and the researcher taught G5. Nevertheless, the researcher carried out the pronunciation sessions with all the groups, so when the pronunciation sessions were scheduled, the teacher teaching the regular lecture left the classroom and the researcher entered to start the pronunciation instruction. Students were told when the pronunciation sessions were scheduled to avoid absenteeism.

All the students were recorded the first week of class. During this week, students attended the course presentation and it was during that session when they were told about the project in general terms (i.e. they were informed that they would participate in an experiment to improve their communicative skills, but the difference in treatment was not mentioned), they signed the consent form to participate in it, and they recorded

the pretest. The English version of the consent form is attached in Appendix 1, but the correspondent Spanish and Catalan versions were also delivered to the students for them to fully understand the implications of the experiment.

The posttest was recorded one week after the pronunciation sessions had finished, and just one week before the course's final oral presentations. This time, a sheet of paper was hanged on the door of the teacher's office where students had to sign up for a day and time to record outside regular lectures so as not to interfere with the proper functioning of the course. The recordings were included as part of the course assessment to ensure participation: Students were given 0.25 points for each test recorded as long as the recording was complete (i.e. all the exercises were done) regardless of their speaking skills.

3.2. The course

Rovira i Virgili University offered the technical English course as a first-year compulsory subject in the following degrees:

- Degree in Industrial and Automatic Electronics Engineering
- Degree in Electrical Engineering
- Degree in Computer Engineering
- Degree in Telematics Engineering
- Degree in Telecommunications Engineering
- Double-degree in Electrical / Industrial and Automatic Electronics Engineering
- Double-degree in Computers and Biotechnology

The course lasted one semester, from February to May, equivalent to six ECTS credits (i.e. a student's workload of 150 hours). The target level of English to achieve was a B 2.1 level of the Common European Framework of Reference for Languages (CEFR). At that time, students were required to have a B1 level certificate of a foreign language in order to obtain the degree diploma of any university discipline according to article 1.7 of the university's policy (available on http://awww.urv.cat/la_urv/3_organs_govern/secretaria_general/legislacio/2_propia/auniversitaria/docencia/nacad_grau_master_2017_18.pdf.). In spite of the fact that the course did not provide an official certificate owing to its technical nature, it was aimed to help students reach that level so that they could pass the official exams. Since students were supposed to come from high school having reached a B1 level, no language requirements were asked in order to take the course. However, the reality showed mixed-level classes.

Although the distribution of groups was based on the general overall schedule of each of the disciplines, learners could choose to attend the group with the class time that suited them better. Therefore, students of different engineering disciplines attended the same session.

3.3. Participants

Despite all studying engineering degrees, the students had different profiles. Regarding their educational background, most of them came from high school, but there were some who gained access to university after vocational training, and others started university some years after working in the private sector, which directly affected their initial level of English. As previously mentioned, at high school the target level of

English is established at B1, but it is true that there were some students who had not reached that level by the time the course started, and others clearly surpassed it thanks to their extracurricular activities (language schools, online videogames, original version films, and so on[1]). Hence, there was a wide range of initial English levels within each group, which could affect the outcome of their learning process. As the study sought to be adjusted to the reality of the ESP classroom, where there is a real tendency towards a disparity of levels, all students participated in the same activities and explanations regardless of their level. However, their initial level was taken into account for the analysis of the data and the interpretation of the results in case it made a difference in the output.

On the other hand, although students were mainly Catalan L1 speakers, there were some students who came from other Spanish regions where Catalan is not spoken, or even other countries. Thus, there was some variety of mother tongues as well. However, all of them pursued primary and secondary education in Catalonia and took the exam to access Catalan universities, which includes Catalan and Spanish tests, so they reported a good command of both languages.

Classes, in general terms, were overcrowded. A high number of students was expected since the technical English course was a compulsory subject for all the engineering disciplines, so sessions were designed for large classes. Despite the fact that 72 students did not attend lectures at all, there were 226 students who were regularly coming to class. Trying to respect everybody's schedules, the distribution of students per class was as follows.

[1] The complementary information was extracted from both the placement test and the "introduce yourself" exercise of the recordings.

Table 3. Distribution of students per class

Group	**Number of students**
G1	39
G2	41
G3	20
G4	44
G5	40
G6	43

As displayed in Table 3, most of the classes had around forty students (M = 37,86). G3 was the only group with considerably fewer students. Overcrowded classes negatively affect both the class and the learning processes due to less control and focus on individual issues, an interesting factor that will be considered when analyzing and discussing results.

Initially, only those students who completed the treatment (both attending the sessions and taking the pre- and posttests) were considered as the subjects of the analysis due to the longitudinal nature of the study. There were twenty-four students from the experimental groups and twenty-eight from the control groups who did not take either the pretest, the posttest or both. Thus, those needed to be discarded because it was not possible to compare their performances before and after treatment. Table 4 summarizes the attendance of those students who, to some extent, attended the pronunciation sessions and took the two tests.

Unfortunately, the number of participants decreased dramatically owing to absenteeism. Attendance sheets and the videotapes of each of the sessions were revised in depth in order to carry out a reliable head count. Only twenty-four out of a hundred students from the experimental groups and fourteen out of seventy-four from the control groups attended all the sessions (just 21.83% of the total number of the potential participants).

Table 4. Potential participants' attendance per group

Sessions attended	Experimental group	Control group	Total percentage
10	24	14	21.83%
9	20	12	18.39%
8	12	10	12.64%
7	18	17	20.11%
6	10	8	10.34%
5	5	2	4.02%
4	4	5	5.17%
3	1	0	0.57%
2	3	3	3.44%
1	3	3	3.44%

Despite trying to control truancy by including the recordings within the course assessment and insisting on the importance of attending the pronunciation sessions for their final oral presentation, attendance is not compulsory in Spanish universities and, hence, a lot of the students either dropped out or skipped some of the sessions.

In order to balance the groups, students who had attended nine out of the ten sessions from the control group were taken into account, since just missing one class should not be decisive for the outcome of the learning process. This way groups got balanced: twenty-four students for the experimental group and twenty-six for the control group.

The next step was to check their recordings to verify whether it was possible to compare them. After segmenting the sentences and the opinions of both the pre- and the posttests using PRAAT (Boersma, P & Weenik, D., 2015, version 6.0.25), the researcher had to discard a few more participants. Within the control group there were four people who were almost incomprehensible when giving their opinion and committed innumerable mistakes when reading the sentences; and there was also one student who provided too short (just fifteen-second long) and incomplete opinions for both the pre- and the posttest. Within the experimental group there was one student who suffered from dyslexia and, consequently, showed problems when reading aloud; another student who

also showed serious trouble when reading despite not suffering from any recognized disorder; and one more student who was playing with his pen, which negatively affected the recording of the speech. As a consequence, the researcher ended up with two groups of twenty-one students each, which constituted the final number of students to analyze: The experimental group was formed by six students from G1, ten from G2, and five from G5, and the control group by four students from G3, twelve from G4, and five from G6.

The profile of these students is summarized in Table 5. The information about their age and language background was extracted from one of the exercises of the assessment test, which consisted in introducing themselves. On the other hand, the data related to their English level was taken from the course placement test, made of grammar and listening activities of several English level tests available online and two open questions that asked students to reflect on the skills they needed to improve and what they could do to improve them.

Table 5. Final participants' profiles

	Age		**Mother tongue**			**English level**		
Group	18	18+	Sp/Cat	Sp	Other	A2-B1	B1-B2	B2-C1
Experimental	21	0	20	1	0	10	6	5
Control	16	5	16	1	4	4	11	6

Note. Sp = Spanish. Cat = Catalan. English level is classified according to the Common European Framework of Reference for Languages (CEFR): A2-B1 = basic user. B1-B2 = independent user (intermediate). B2-C1 = independent/proficient user (upper-intermediate/advanced).

All the students from the experimental group were eighteen-year-old first-year undergraduates who accessed university from high school. All of them came from cities within the region of Catalonia and were balanced bilinguals in Spanish and Catalan,

except for one student who had arrived in Spain from Venezuela in 2001, but pursued his primary and secondary education in Catalonia, so his command of both languages was high. Two students reported to have studied French, three other students German, and one of those also Chinese. Only five students reported to have attended private lessons at a language school. After taking the placement test, ten of the students were shown to start from a basic level (from now on, beginners), six from an intermediate level, and five of them from an upper-intermediate/advanced level.

Regarding the control group, most of the students were also eighteen, but there were two students who reported being nineteen, two being twenty, and another one twenty-two. One of the nineteen-year-old and the two twenty-year-old students were taking the course for the second or third time. Most of the participants came from Catalonia and were balanced bilinguals of Spanish and Catalan, except for one student coming from Madrid, two from Morocco, one from Bulgaria, and another one from Brazil. They reported to have a high command of the Catalan language, having pursued their previous education in Catalonia. As for second languages, one student reported to have studied Basque, one student French, one student German, and one student to have self-studied both German and Japanese. Regarding their levels of English, four were classified as beginners, eleven as intermediate students, and six as upper-intermediate/advanced students.

3.4. The instruction

For ten weeks, all the students received pronunciation training for thirty minutes a week. Every session was video recorded: The camera was always fixed and placed strategically to show as much of the classroom and the students' interaction as possible.

Instruction was "scaffolded": the first sessions dealt with features influencing the pronunciation of the basic unit of the speech (the word); then, the following sessions worked on the pronunciation of longer utterances (the sentence); and finally, the last sessions focused on complex units (the discourse).

On the other hand, each individual session followed Celce-Murcia et al.'s (1996) steps to teach communicatively (see section 2.4.1.1.). Each session, hence, started with a description and analysis of the item to work on. The teacher's explanations were complemented with videos, images, and kinesthetic activities, explained in detail within Appendix 2, which shows the structure and content of the sessions. For example, in session 6 the teacher provided rubber bands to the experimental students in order to explain the difference in length between stressed and unstressed syllables. They had to place their hands within the rubber band and separate the hands when pronouncing a stressed syllable and get them closer when they pronounced an unstressed one. At the same time, students were asked to lengthen stressed syllables while separating the hands and shortened them when putting the hands together. This way, students could feel the difference in length between the two types of syllables.

Then, students listened to recordings or watched videos to test whether they could recognize the feature and get familiar with it. For instance, when students where learning how to pronounce the regular past tense properly in session 2, they had to listen to sentences and decide whether they were past or present actions.

Next, students started practicing the item with controlled exercises such as reading aloud or "listen and imitate" activities. A recurrent exercise was to make students listen to sentences or speeches and make them read the transcript aloud trying to imitate the speaker.

Later on, they practiced with guided activities, which were not so controlled. A good example of it is a pair work exercise carried out in session 7 where each student had half of a sales graph and they had to describe their part to their partner so that they could draw the part that was missing.

Finally, sessions finished with a communicative activity to practice extemporaneous speech. For instance, students participated in debates (session 9) or were asked to sell a product within the context of a technology fair (session 10).

Feedback was provided while the students were carrying out their activities, individual when possible and as a group when different people shared the same mistake. Nevertheless, individual feedback could not always be guaranteed owing to the high number of students in class, so they were asked to provide feedback to their partners when working together.

3.4.1. Experimental treatment.

As previously mentioned, experimental groups received explicit rhythm instruction. The training included the explanation and practice of the following rhythmic phenomena: word stress, sentence stress, linking and sentence focus. Rhythm instruction was mainly based on vowel lengthening for different reasons. First, the scarce time for pronunciation instruction limited the quantity of rhythmic cues to make teaching doable and effective. Second, as explained in detail in section 2.6.5, vowel duration has been proved to be a prominent distinguishing cue among different rhythms, especially between English and Spanish/Catalan. Third, the participants of this study were not language-oriented, so it was considered that the simpler an alien concept was taught, the easier the concept would be assimilated: Indeed, length is a much more common

concept than pitch or tone, so its assimilation to understand and acquire language rhythm was thought to be less problematic.

Relating it to the scaffolding structure of the units of speech, the introduction of the rhythmic items in class was done progressively, from the simplest to the most complex ones. One new concept was introduced in each session but its practice was not limited to that session: They were further practiced in the subsequent ones. Its distribution within the sessions is summarized in Figure 1.

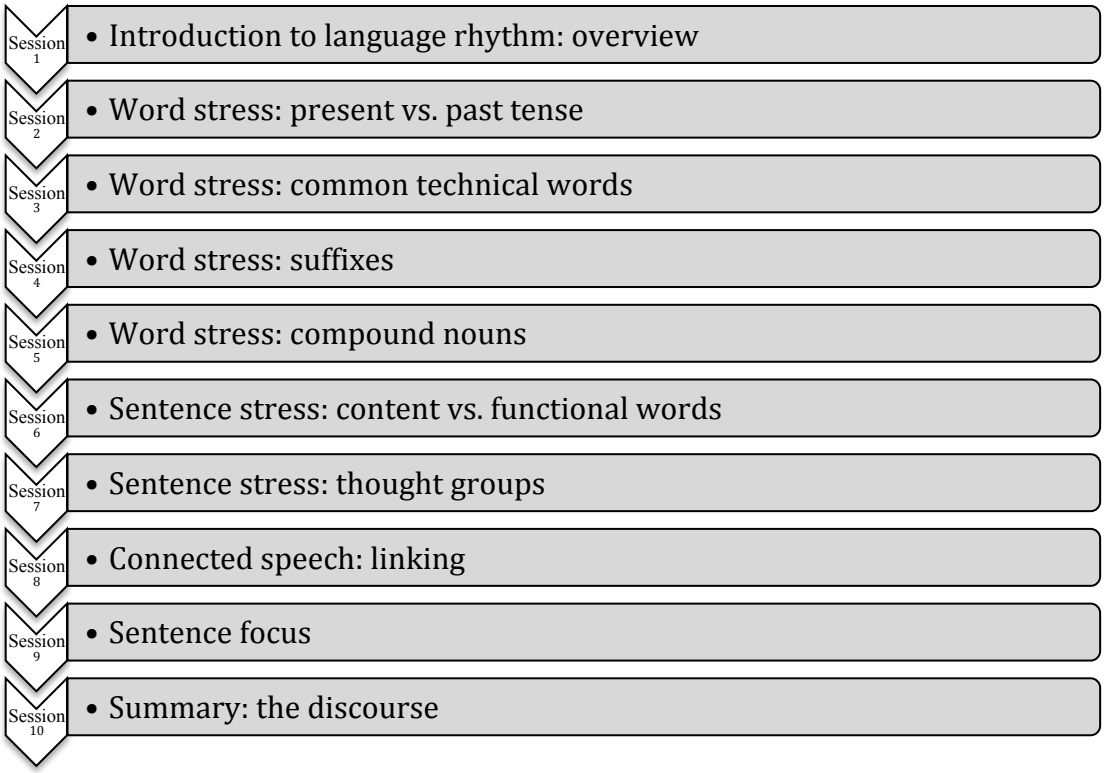


Figure 1. Distribution of rhythm instruction per session

An example of an activity that implied rhythm training is the following: When teaching the pronunciation of common technical words in session 4, the syllables of the word were replaced by *LA* when they carried the main stress and *la* for the rest of them (e.g.

electrician became *la la LA la*, and *chemical* became *LA la la*), so students got familiarized and practiced the pronunciation of the words through their rhythmic patterns. Secondary stress was excluded in order to focus on the acquisition and practice of primary stress. This technique of using rhythmic patterns to explain word stress, adapted from Chela-Flores (1997b, p. 127), was further used in other sessions to explain other items, such as the difference in pronunciation between compound nouns and adjective + noun combinations in session 5.

3.4.2. Control treatment.

Control groups did not take rhythm training, so the pronunciation features were explained and practiced with traditional methods. For example, the pronunciation of technical words was only explained through examples and rules according to the suffix they carry (i.e., words ending in *–ian* stress the second from last syllable).

3.5. Recording sessions

Both sessions (before and after training) took place in the library of the School of Engineering. Three semi-sound-proof isolated rooms were booked to carry the recordings out. Three different rooms were arranged to speed the sessions up, since students were recorded individually. Each student was expected to record themselves for approximately ten minutes. Two Sony PCM-M1O and a Zoom H4nsp recorders were used. Only two students recorded under different conditions at the beginning of the course because they skipped the course presentation session. Hence, they had to record on a different day at the teacher's office. However, neither of those students were taken into account for the final analysis. When recording the posttest, students also had

to fill in a questionnaire on their opinion about the course to check how satisfied they were with it.

3.6. Materials

A wide range of materials were compiled to put the experiment into practice. On the one hand, the design of the pronunciation module implied the search and creation of different learning resources. On the other hand, the assessment test and the satisfaction survey were also customized to suit the purpose of the study.

3.6.1. Pronunciation module.

The pronunciation module was created and included as part of the technical English course. It was designed bearing in mind the technical and the linguistic content for several reasons: First, to harmonize it with the rest of the course. Second, to guarantee students' motivation (Anderson-Hsieh, 1990). Third, to ensure the applicability of the knowledge acquired to the students' direct reality, meeting the students' needs (Derwing & Munro, 2015; Dudley-Evans &St John, 1998).

Both the course syllabus and the course book were analyzed in order to outline the content of the pronunciation module. A copy of the complete syllabus of that academic year is included in Appendix 3. The course content was divided into seven units plus the pronunciation module. Activities and materials to perform the pronunciation sessions were attached at the end of the regular course book for organizational purposes. Each of the sessions of the pronunciation module was related to one of the units. Figure 2 summarizes the relationship between the pronunciation outline and the course syllabus.

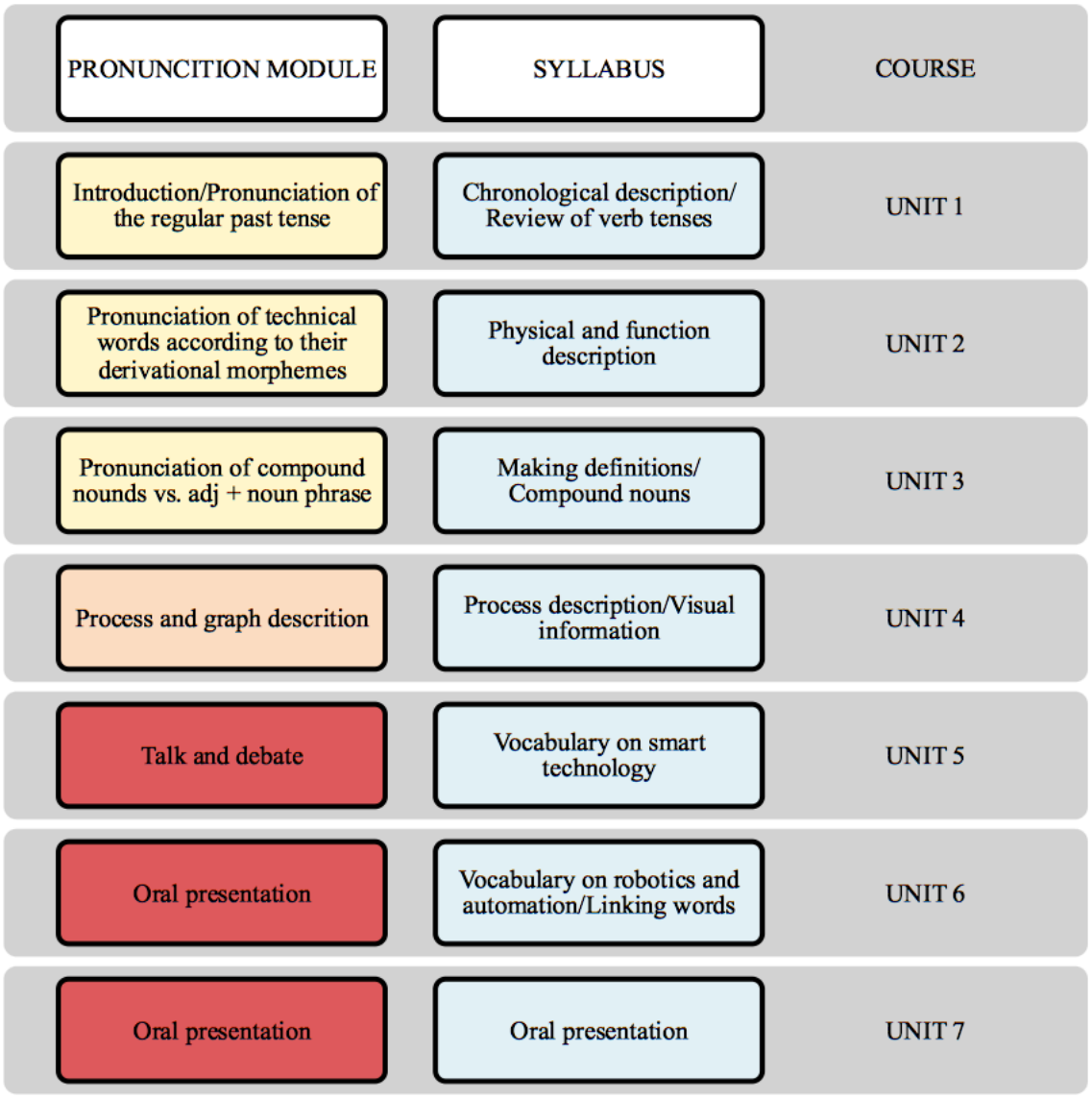


Figure 2. Relationship between the course syllabus and the pronunciation module

Colors represent the units of speech that were dealt with when working on each particular pronunciation item (i.e. yellow represents the word; orange, the sentence; and red, the discourse). The main points described at the pronunciation plan presented to the students are related to one or more of the course units. However, the module was divided into ten sessions, so some of the units, which tended to be the ones that took longer to teach, occupied more than one pronunciation session. For example, although sentence instruction was based just on Unit 4, there were two separate pronunciation

sessions related to that unit: one worked on process description and the other one on graph description, concepts included within the same regular unit.

3.6.2. PowerPoint presentation and course book materials.

Sessions were mainly guided by means of a PowerPoint presentation, supporting explanations and providing the required materials to do the activities proposed. PowerPoint presentations were modified according to the treatment group since the sessions for the experimental group included rhythm explanations. The slides, as the sessions themselves, were organized following the structure of Celce-Murcia' et al. (1996)'s communicative framework.

When necessary, some complementary materials were incorporated within the course book as part of the pronunciation section. Both the experimental and the control complementary materials were included in the course book, so it was the teacher's job to decide which ones to use depending on the treatment group. At certain points handouts were also delivered to carry out some of the exercises.

As further detailed in Appendix 2, explicit rhythm activities were mainly adapted from textbooks and research papers to suit the technical background of the course (Chela-Flores, 1997b; Earle-Carlin, 2011; Gilbert, 2005a, 2005b, 2008). Those other activities not focusing on rhythm explicitly were created from scratch from online resources (YouTube videos, images, texts, etc.) or adapted from activities for academic purposes and pronunciation exercises (Aquascript, 2017; Beisbier, 1995; Université Franche-Comté, 2004).

3.6.3. Test and questionnaire.

The pretest and the posttest were the same in order to guarantee comparability. They consisted of four exercises, including both controlled and extemporaneous activities. The test started with a reading aloud exercise inspired on Couper's (2006) speaking test (p. 61). It was made up of ten sentences, both simple and complex, that contained technical vocabulary or everyday situations that students could face in their real lives. The second exercise was a diagnostic passage proposed by Celce-Murcia et al.'s (1996, p. 398) to check learners' pronunciation, which students also had to read aloud. The third exercise consisted in introducing themselves, also inspired by Couper's (2006) test. Finally, the fourth exercise was about giving their opinions on a hot topic that was appealing to their immediate reality and their engineering background. The topic was meant to be simple so as to ease students' potential speaking anxiety.

The decision to introduce both controlled and extemporaneous activities was deliberate to examine the effectiveness of rhythm at different stages. The sentence exercise was used to analyze controlled performances and the opinions to analyze extemporaneous performances. The "introduce yourselves" exercise helped to compile useful information about the students' profiles. The text had to be discarded from analysis due to lack of time.

The questionnaire was a satisfaction survey to get to know the students' opinion about the course. It was made up of three statements with which they had to agree or disagree by using several agreement scales, another statement assessed by using a percentage scale, a multiple-choice question, and two open questions. In general terms, this questionnaire was used to examine whether students found the course useful and

whether there was a difference in opinion between control and experimental students. A copy of both the test and the questionnaire is attached in Appendix 4.

3.7. Baseline group: Native English speakers

Six volunteering native American students (three men and three women) were recorded for comparison purposes while they were doing a one-semester stay at Rovira i Virgili University in 2017. They were between nineteen and twenty years old. They came from different states in the United States of America and they were studying Spanish at that moment, some of them as minors or majors of their university studies.

3.8. Listeners

Seven native speakers of English (two men and five women) were recruited to judge the students' extemporaneous opinions. They came from different English-speaking countries: four from the United Kingdom, one from Canada, one from the United States of America and one from Singapore. They were all considered experienced listeners because they had experience teaching English as a second language in Catalonia and to Spanish/Catalan speakers. Nevertheless, the length of teaching experience varied considerably, between one and thirty-five years (M = 17.43; SD = 14.8). All of them were living in Catalonia at the time they took the test, so they were familiar with the Catalan and Spanish accent to some extent, although they also highly differ on their length of residence, from moths to years (M = 15.86; SD = 15.32). They participated voluntarily in the project.

3.9. Data analysis

A mixed-method assessment consisting of both acoustic measures and listeners' scores was used so as to provide a detailed study of the learning process (for a discussion, see section 2.6.7). On the one hand, the degree of rhythm acquisition at the end of the course was examined by acoustically measuring the rhythm of the sentences before and after the instruction. Two aspects were taken into account: VarcoV values, based on vowel duration variability, and pauses. On the other hand, comprehensibility and fluency improvements were assessed by means of a scalar-rating task performed with the students' extemporaneous productions. The analyses sought to investigate whether there was a difference in the learning process among groups depending on the instruction received and whether there were further correlations among the results of both types of tests, thus offering a more detailed report on the students' evolution both in controlled and extemporaneous circumstances.

3.9.1. Acoustic analysis.

A total of 900 sentences were segmented and acoustically analyzed using PRAAT. Both the students and the natives recorded the ten uncontrolled sentences displayed in the first exercise of the test sample in Appendix 4. Sentence structure was intentionally varied to cover a broad range of rhythmic patterns, and the vocabulary and grammar used was chosen according to the course requirements. The main focus of the analysis was the 840 utterances recorded by the students before and after instruction. The remaining 60 sentences, uttered by the native speakers, were used as baseline for comparison.

An individual sound file and its corresponding text grid were created for each of the sentences (see Figure 3). Sound identification followed Olive, Greenwood and Coleman's (1993) guidelines and the segmentation was carried out manually following Ordin and Polyanskaya's (2014) criteria (p. 5).

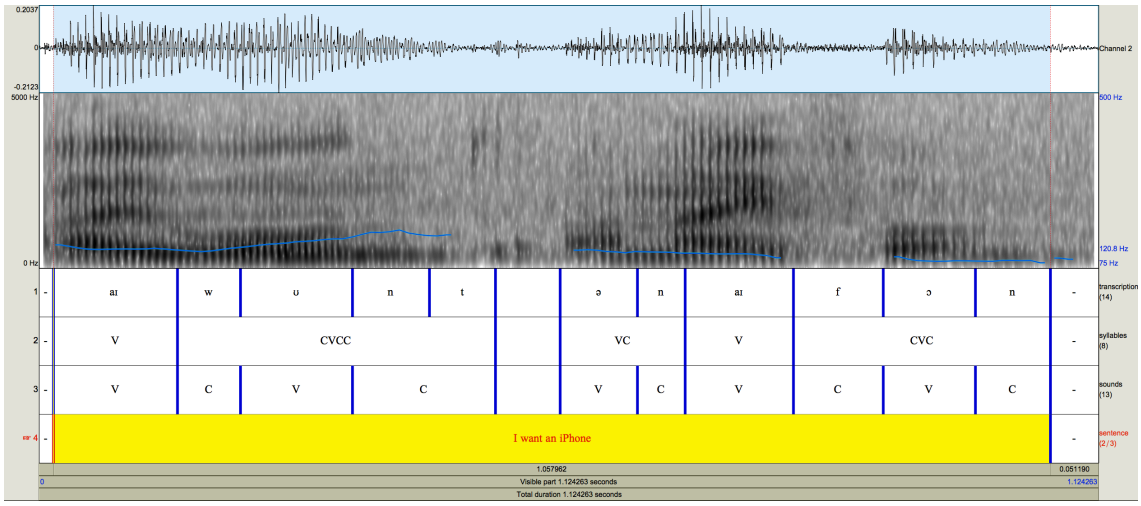


Figure 3. PRAAT sample of the acoustic analysis of sentence 1

The text grid contained four different tiers: From top to bottom, the first tier displayed an approximate transcription of the sounds uttered by the student; the second tier showed the combination of vocalic and consonant sounds for the syllables in each word; the third tier presented vocalic and consonant intervals in the sentence, grouped within the same cluster when there were no pauses in between; finally, the fourth tier showed the sentence to pronounce. Blanks represented both filled (i.e. hesitations, repetitions, etc.) and unfilled pauses.

In order to analyze the rhythm of the utterances, Varco-V values were obtained. Varco-V is defined as "the standard deviation of vocalic interval duration divided by mean vocalic duration, multiplied by 100" (White & Mattys, 2007b, p. 242). There were

different reasons to use this rhythmic metric. As explained in section 2.6.6, Varco-V has shown higher effectiveness than other metrics when measuring rhythm within second language acquisition (White & Mattys, 2007, 2007b). Furthermore, Prieto et al. (2012) noticed a prominent difference in Varco-V values between English and Spanish/Catalan, where the former has higher values than the latter. On the other hand, as previously mentioned in this chapter, rhythm instruction was based on vowel lengthening, so students were expected to adjust vocalic duration after training. Ordin and Polyanskaya's script (2014, 2015) was run on the third tier to measure the values. For this analysis, pauses were excluded and final-syllable lengthening was kept.

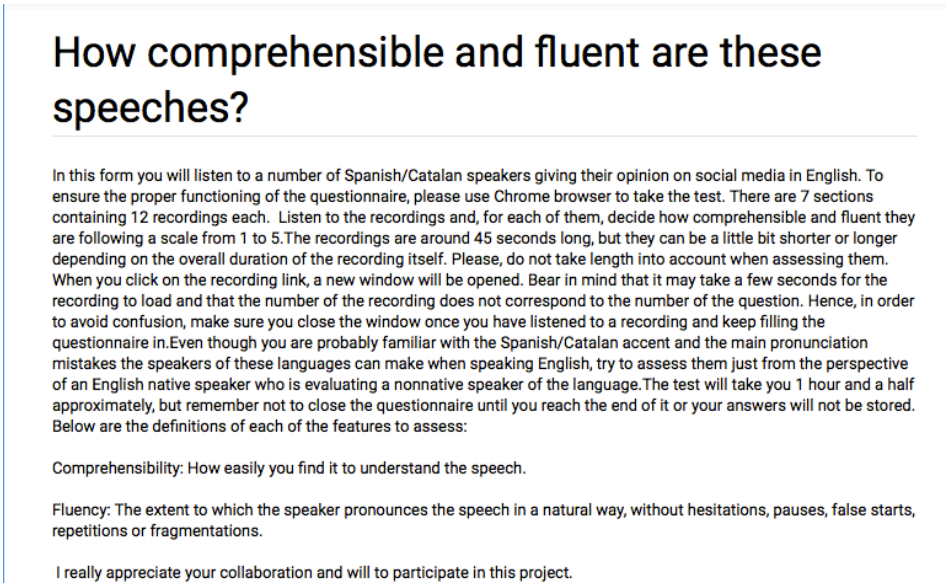
Pauses were analyzed separately based on frequency. They were manually counted and two analyses were performed: one took into account only the unfilled pauses while the other took into consideration the total number of pauses (both filled and unfilled). The reason for performing two analyses was that, despite the general frequency of pauses being an important indicator of fluency, unfilled pauses seem to be a more reliable indicator than filled pauses (Wood, 2010), so it was interesting to examine the frequency of just unfilled pauses independently.

3.9.2. Listener-based analysis.

An online Google form was designed with the eighty-four opinions, forty-two of the control group and forty-two of the experimental group before and after instruction, to assess students' progress on comprehensibility and fluency. An individual audio file was created for each of the opinions. By using Audacity software Version 2.2.2 (2018), audios were normalized at -1 dB. There were a few audios that were still too low and volume needed to be further manipulated to ensure a confortable hearing level without

distorting. On the other hand, files were cut into forty-five seconds when the contributions were too long (Derwing et al., 1998). When possible, the beginning of the opinion was taken for assessment and, when that was not possible, the part with fewer hesitations, pauses, etc. The end point was established at a natural pause in the argumentation.

The next step was to randomly distribute the audios throughout the form. As shown in Figure 4, at the beginning of the form there were instructions for the listeners on how to carry out the test and the explanations of the concepts to analyze, based on Derwing and Munro's (2015) and Wood's (2010) definitions of comprehensibility and fluency. Also, listeners had to give their age and nationality, the years of residence in Catalonia, and the years of teaching English as a foreign language.



How comprehensible and fluent are these speeches?

In this form you will listen to a number of Spanish/Catalan speakers giving their opinion on social media in English. To ensure the proper functioning of the questionnaire, please use Chrome browser to take the test. There are 7 sections containing 12 recordings each. Listen to the recordings and, for each of them, decide how comprehensible and fluent they are following a scale from 1 to 5.The recordings are around 45 seconds long, but they can be a little bit shorter or longer depending on the overall duration of the recording itself. Please, do not take length into account when assessing them. When you click on the recording link, a new window will be opened. Bear in mind that it may take a few seconds for the recording to load and that the number of the recording does not correspond to the number of the question. Hence, in order to avoid confusion, make sure you close the window once you have listened to a recording and keep filling the questionnaire in.Even though you are probably familiar with the Spanish/Catalan accent and the main pronunciation mistakes the speakers of these languages can make when speaking English, try to assess them just from the perspective of an English native speaker who is evaluating a nonnative speaker of the language.The test will take you 1 hour and a half approximately, but remember not to close the questionnaire until you reach the end of it or your answers will not be stored. Below are the definitions of each of the features to assess:

Comprehensibility: How easily you find it to understand the speech.

Fluency: The extent to which the speaker pronounces the speech in a natural way, without hesitations, pauses, false starts, repetitions or fragmentations.

I really appreciate your collaboration and will to participate in this project.

Figure 4. Guidelines for raters

After accessing the application, listeners were presented with a serious of sections. Each of the sections contained twelve audios and there was a total of seven sections. For each

file, listeners opened a link were they listened to the audio and they had to decide how comprehensible and fluent that student was by using a 5-point Likert scale that went from 1 "not at all" to 5 "totally" (see figure 5). This type of scale was preferable as the test implied a large number of audios of a wide range of speakers and the listeners' experience varied considerably, so it was estimated that it would be more suitable to use a simple short scale and, hence, ensure user-fiendliness.

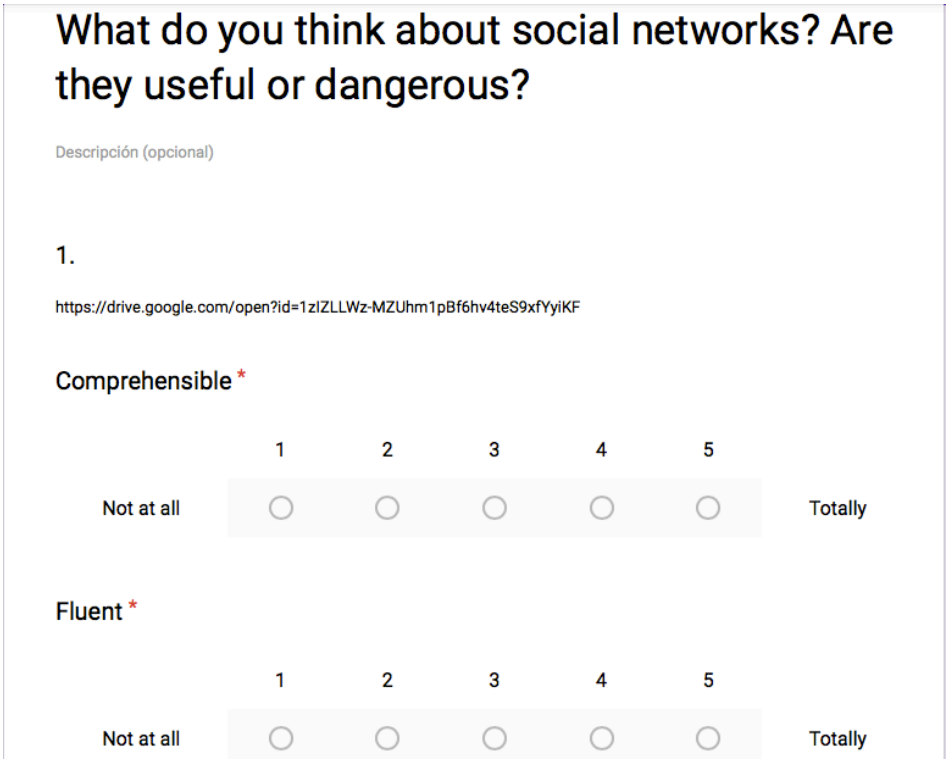


Figure 5. Example of the Likert scale used for the comprehensibility and fluency tests

Listeners had to take the test online due to the inability to bring all of them together at the same place on the same day. The limitations of the online form, which did not allow saving the answers to take a break, forced listeners to take the test in a straight run. The test lasted approximately one hour and a half. Therefore, the fatigue factor could not be controlled for and will be considered when analyzing and discussing the results.

3.10. Statistical analysis

The statistical analysis of both the acoustic data (VarcoV values, total number of pauses and unfilled pauses) and the listener-based data (comprehensibility and fluency scores) included descriptive statistics, repeated measures ANOVAs, and both paired-samples and independent-samples t-tests.

Descriptive statistics provided an overview of the results of each group in each of the scenarios. Both repeated-measures ANOVAs and t-tests studied the general impact of the instruction and the difference in performance before and after training. Apart from this main analysis, two further tests were performed. On the one hand, correlation tests were conducted so as to examine the relation between VarcoV values, pausing, and comprehensibility and fluency scores. On the other hand, intra-class correlation coefficient (ICC) tests were run on the listeners' scores in order to guarantee the reliability of the results.

The data from the descriptive analysis that will be brought under study are the number of learners (*n*), the mean (*M*) and the standard deviation (*SD*). Skewness and kurtosis values helped to explain the data obtained. The significance of the correlation tests will also be taken into consideration to examine the link between the changes made in VarcoV values and pauses after instruction. Finally, from ICC estimate, 95% confidence interval will be studied to determine the degree of reliability of the listeners' scoring.

Two different applications were used for the statistical analysis:

Microsoft Excel: This software was used to obtain the descriptive analysis.

SPSS: This software was used to carry out the most complex statistical tests (i.e. t-tests, ANOVAs, correlation and ICC estimations).

Chapter 4. Results

This chapter examines the results obtained from the different types of data collected, previously explained as part of the methodology of the study. The results are divided into two main subgroups. On the one hand, the quantitative results, which include the descriptive and statistics tests conducted to analyze the students' recordings and their responses to the satisfaction survey. On the other hand, the qualitative results, which explore the implementation of the module in class by examining the video recordings, the comments of the satisfaction survey, and the teacher's journal.

The quantitative results will be mainly displayed by means of summary tables, bar and line graphs, and pie charts. The summary tables will show the descriptive analysis of the data and the statistical results, while the graphs and the charts will provide further evidence of the students' learning progress according to the treatment received, and pool students' opinions on the pronunciation module. As for the qualitative results, these will be presented through the students' and the teacher's remarks.

4.1. Quantitative results

In order to investigate the students' learning progress, several aspects were taken into account. Firstly, an acoustic analysis was conducted for the recorded sentences. VarcoV values were measured and both overall and unfilled pauses were counted before and after training. Progress was compared after treatment for each of the aspects under analysis by means of mixed ANOVAs and t-tests, but also correlations between the two groups of pauses and VarcoV values were examined.

Secondly, a listener-based analysis was performed with the students' extemporaneous speeches. The scores obtained from the scalar rating task were investigated for signs of

improvement in students' comprehensibility and fluency. Raters' reliability was validated via ICC tests and several mixed ANOVAs were run so as to analyze students' progress depending on the training received. Due to a high disparity of levels and the nature of the scale, the possibility of the existence of a ceiling effect was contemplated. For this reason, mixed ANOVAs were conducted to address not only the performances of the total number of students, but also students with a high speaking competence (from now on, high-level students) and a low speaking competence (low-level students) individually. Correlations were also computed between, on the one hand, comprehensibility and fluency ratings and, on the other hand, each of the items scored with each of the items analyzed acoustically.

4.1.1. Acoustic analysis.

As mentioned in the methodology section, sentences were segmented and analyzed using PRAAT. The main item to investigate was VarcoV, as the rhythmic metric chosen to measure rhythm. However, the total number of pauses and the number of unfilled pauses were also examined as additional aspects that could influence the pace of the utterances. The three items under analysis will be presented separately below, followed by the correlations between VarcoV values and each type of pauses.

4.1.1.1. VarcoV values.

This section focuses on the results obtained from the VarcoV measurements. The following subsections will present the descriptive analysis and the results generated from the mixed ANOVAs and the t-tests performed.

4.1.1.1.1. Descriptive statistics.

The performances of the forty-two participants who completed the treatment (experimental group, *n* = 21; control group, *n* = 21) were analyzed. The means and standard deviations of VarcoV values for each sentence were calculated and rearranged according to time (T1 = pretest; T2 = posttest) and group (experimental; control; natives), as displayed in Table 6. The control group scored higher in more sentences of the pretest than the experimental group (VarcoV values were higher in sentences 2, 3, 4, 5, 6, 7 and 8, lower in sentences 1 and 9, and the same in sentence 10). However, when examining their progress after instruction (highlighted in light gray), it is observed that the experimental group reached higher VarcoV values in eight out of ten sentences (1, 2, 3, 4, 5, 7, 8 and 10) while the control group increased in five out of ten sentences (5, 6, 8, 9 and 10). There are only two sentences in which the control group improved their scores while the experimental group did not. In sentence 6 the experimental group's values remained the same at 45, but in sentence 9 the mean decreased from 52 to 48.

Table 6. VarcoV means (*M*) and standard deviations (*SD*) per sentence

	Control group				Experimental group				Natives	
	T1		T2		T1		T2		T1	
Sentence	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	37	0.08	36	0.11	39	0.11	40	0.08	33	0.08
2	40	0.09	40	0.09	38	0.07	39	0.08	48	0.03
3	39	0.12	38	0.10	38	0.10	39	0.13	36	0.06
4	54	0.09	52	0.08	50	0.09	51	0.09	56	0.07
5	45	0.09	47	0.14	41	0.09	50	0.12	57	0.06
6	48	0.12	49	0.10	45	0.10	45	0.09	53	0.06
7	55	0.10	53	0.09	53	0.13	56	0.09	64	0.05
8	42	0.07	43	0.08	41	0.08	43	0.06	50	0.05
9	49	0.10	51	0.13	52	0.11	48	0.09	47	0.07
10	48	0.11	52	0.11	48	0.05	51	0.08	52	0.11
Total	**45.8**		**46**		**44.4**		**46.1**		**49.7**	

On the other hand, there are three sentences in which both groups improved (5, 8 and 10). In sentence 5, the experimental group outperformed the control group results after treatment even though its score at the beginning of the course was lower. In sentence 8, both groups obtained the same results after instruction, but the experimental group scored lower in the pretest, so the improvement was higher. By contrast, in sentence 10, both groups scored the same before treatment but the control group scored higher in the posttest.

Overall, the descriptive analysis shows that the experimental group improved more after instruction, but the results did not always follow the same direction. As shown in the individual values presented in Appendix 5, beginners of both groups were the ones who tended to increase VarcoV values after instruction. However, they were supposed to be the ones with the lowest values as well, due to their low level of English, and this was not always the case.

Regarding the results for the native speakers, scores that were higher than those of the students are highlighted in dark gray. Native VarcoV means were always higher except for those in sentences 1, 3 and 9, which suggested potential problems in their performance. In fact, their individual values were not always higher as compared to those of the students (see Appendix 6). When recording the sentences, native speakers were told about the experiment, so some of them did not read the sentence in a natural way, but they articulated in excess, which could have affected results. Issues of level and individual performances will be brought up again in the discussion section.

Means for each group values of all the sentences together were also calculated. As shown in Figure 6, the control group's values remained almost stable over time while the experimental group increased its values after treatment. When further comparing the

students' overall performance to the natives, it could be seen that the natives' VarcoV mean value was higher than both the control and the experimental groups values before and after treatment.

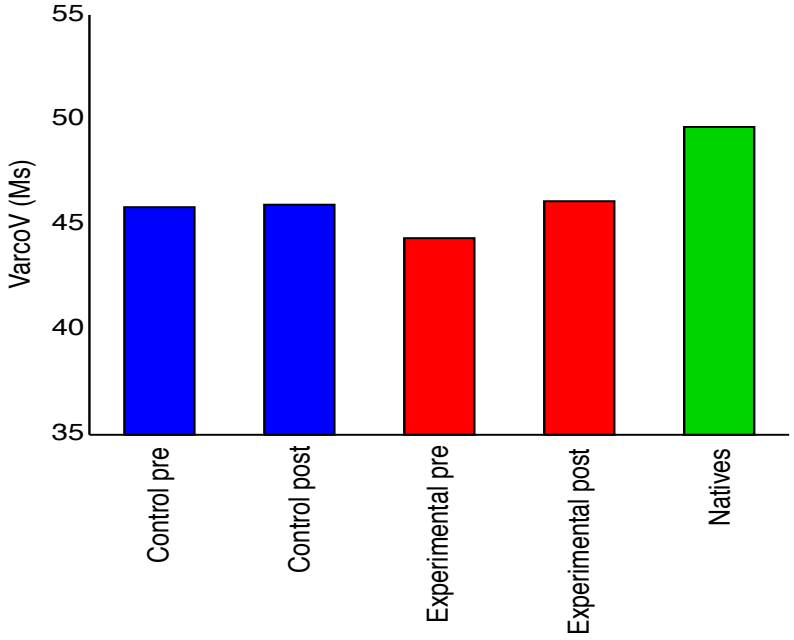


Figure 6. VarcoV means of the ten sentences together

4.1.1.1.2. Mixed ANOVAs.

Several repeated-measures mixed ANOVAs were run to analyze the improvement of the different groups. In order to investigate the general impact of treatment on each groups' performance, an ANOVA was carried out with time and sentence as the within-subjects factors and group as the between-subjects factor. VarcoV values were the dependent variable (see Table 7).

Table 7. Mixed repeated-measures ANOVA with VarcoV as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	2.006	.164
	sentence	9	32	29.172	.01**
Interaction	time*sentence	9	32	1.185	.313
Between-subjects	group	1	40	.267	.608

Note. ** $p < .01$.

Results were not significant for either time $F(1,40) = 2.006$, $p = .164$, or group $F(1,40) = .267$, $p = .608$, but they were for sentence, $F(9,32) = 29.172$, $p < .01$. However, the time*sentence interaction showed no significance, $F(9,32) = 1.185$, $p = .313$. In spite of the lack of statistical significance, the VarcoV mean of the experimental group increased over the control group's, which remained almost stable, as observed in the descriptive analysis (see Figure 7).

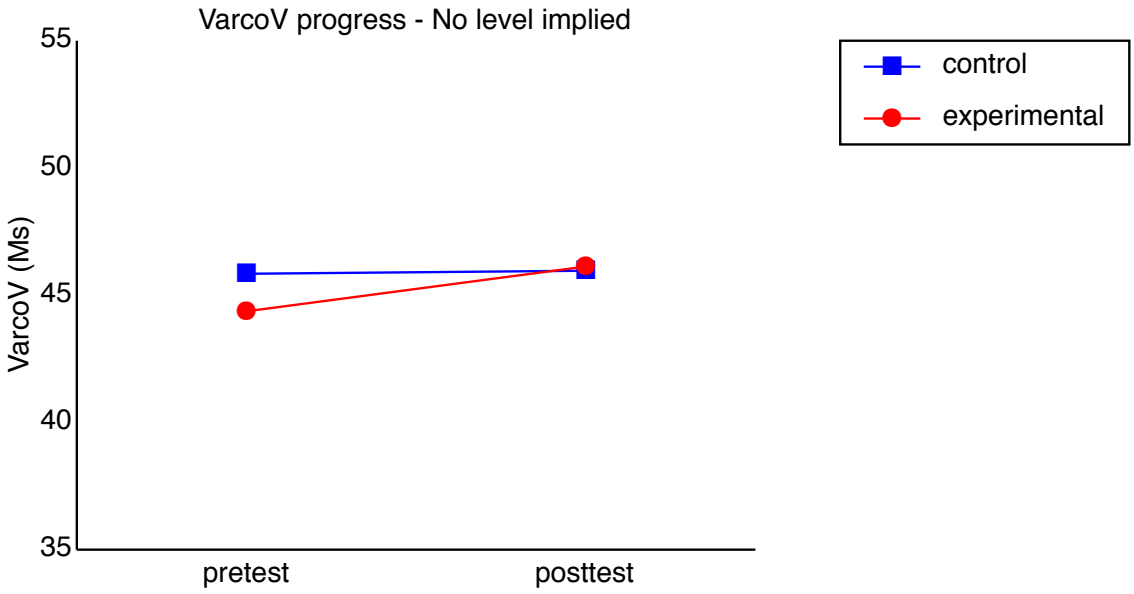


Figure 7. General group progress on VarcoV values according to treatment

In order to check to what extent a higher initial level could have influenced results, the repeated-mesures ANOVAs was repeated without high-intermediate/advanced students

from the control group. Consequently, the number of students from the control group decreased to fifteen while the experimental group kept its twenty-one students. Again, the results showed no significance for time, $F(1,34) = 1.405$, $p = .244$, or group, $F(1,34) = .136$, $p = .715$, but they reached significance for sentence, $F(9,26) = 23.187$, $p < .01$. However, the time*sentence interaction was not significant again, $F(9,26) = 1.009$, $p = .421$ (see Table 8).

Table 8. Mixed repeated-measures ANOVA with VarcoV as the dependent variable without control high-intermediate/advanced students

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	34	1.405	.244
	sentence	9	26	23.187	.01**
Interaction	time*sentence	9	26	1.009	.421
Between-subjects	group	1	34	.136	.715

Note. ** $p < .01$.

As displayed in Figure 8, students from the control group were performing better at the beginning of the course and they hardly improved after instruction. Hence, initial level was not revealed to influence results.

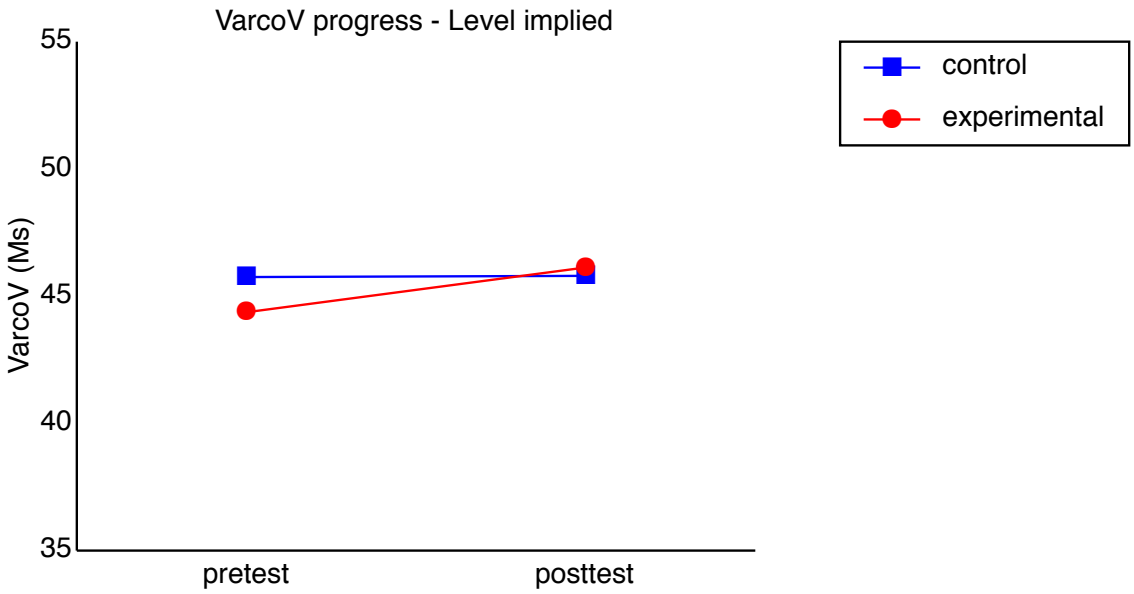


Figure 8. Group progress on VarcoV values according to treatment without the high-intermediate/advanced students from the control group

On the other hand, the relevance of the difference in production before and after treatment was also investigated. An ANOVA was performed with the difference in VarcoV figures before and after training for each sentence as the dependent variable, sentence as the within-subjects factor and group as the between-subjects factor (see Table 9).

Table 9. Mixed repeated-measures ANOVA with VarcoV difference over time as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	sentence	9	32	1.185	.313
Between-subjects	group	1	40	1.532	.223

No significance was found for either group, $F(1,40) = 1.532$, $p = .223$, or sentence, $F(9,32) = 1.185$, $p = .313$. The same ANOVA was performed without the high-intermediate/advanced students from the control group and, again, it did not show any

significant results: sentence, $F(9,26) = 1.009$, $p = .421$, group, $F(1,34) = 1.262$, $p = .269$ (see Table 10). At first sight, hence, the difference in production between the groups does not seem substantial.

Table 10. Mixed repeated-measures ANOVA with VarcoV difference over time as the dependent variable without control high-intermediate/advanced students

	Effect	df	Error df	*F*	*p*-value
Within-subjects	sentence	9	26	1.009	.421
Between-subjects	group	1	34	1.262	.269

4.1.1.1.3. T-tests.

Two paired-samples (control T2 vs. control T1; experimental T2 vs. experimental T1) and two independent-samples t-tests (control T1 vs. experimental group T1; control T2 vs. experimental group T2) per sentence were calculated to further compare the learning progress of the different groups (see Appendix 7). Only the paired-samples t-test run with the results of the experimental group for sentence 5 was found significant $T(20) = 2.988$, $p < .007$.

However, some observations in the progress between groups could be made when analyzing all the effect sizes of the paired-samples t-tests. Firstly, the size of the effect of both groups tended to be positive: the control group showed a positive effect in six sentences and the experimental group in eight. However, the size of the effect was always bigger in the experimental group's performance. Secondly, when the effect size was negative, it was always greater for the control group (see Figure 9).

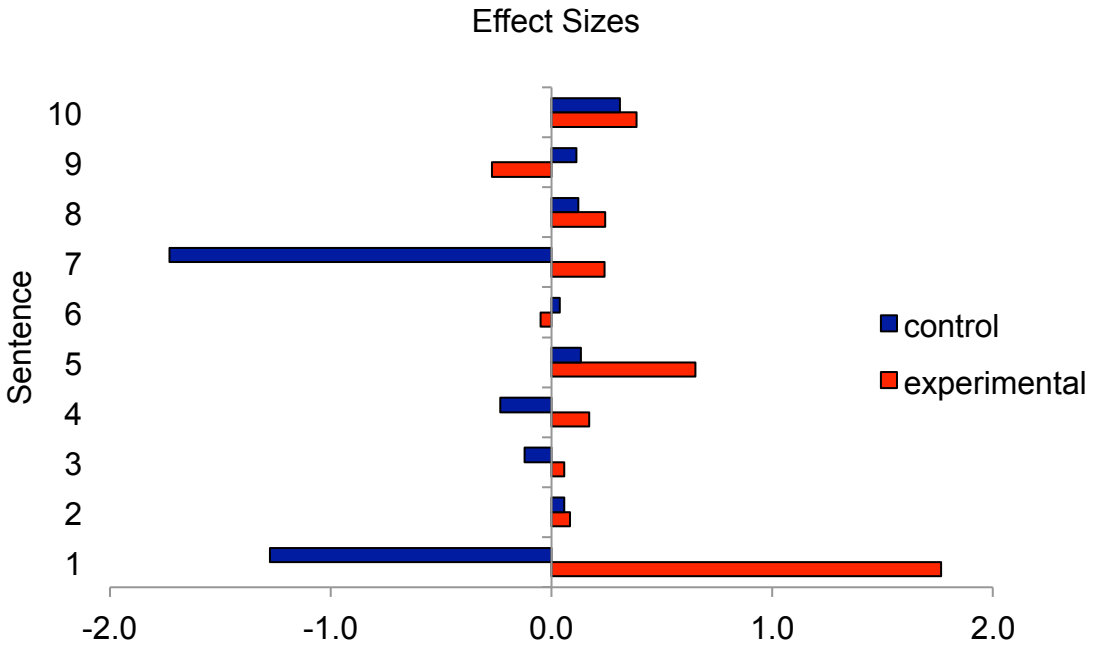


Figure 9. Effect sizes for VarcoV paired-samples t-tests

Therefore, the results obtained from the effect sizes of the paired-samples t-tests showed more improvement for the experimental group despite the lack of significance, agreeing with the comparison of the means. However, the sentence factor seemed to make a difference in students' performance. Bearing in mind the variability of the sentence structure and the vocabulary used, it seemed logical that students should perform better in some of the sentences than others. An independent-samples t-test was run with the effect sizes of all the paired-sample t-tests in order to analyze the overall importance of the difference. This time, results reached significance $T(18) = -2.102$, $p < .05$.

4.1.1.2. Pauses.

As mentioned in section 3.9.1, the frequency of pauses is considered as an indicator of fluency and, consequently, is believed to influence prosody. For this reason, pausing within the sentences was also examined to provide more acoustical evidence on

students' evolution. Two factors were analyzed: the overall number of pauses and unfilled pauses particularly. Analyses followed a similar sequence as the one carried out for VarcoV values. No native scores were used as baseline this time since natives were not expected to make inter-sentence pauses.

4.1.1.2.1. Total number of pauses.

The overall number of pauses included both filled (hesitations, fall starts, repetitions, and so on) and unfilled pauses. Once the pauses were counted, the descriptive and statistical analyses were performed, and the results are presented in the following subsections.

4.1.1.2.1.1. Descriptive statistics.

In order to study general group performance, the means and standard deviations of the pauses for each sentence were estimated and distributed according to time (T1 = pretest; T2 = posttest) and group (experimental; control). The interest fell in examining whether students pause less after treatment. Table 11 summarizes the descriptive analysis outlined above, with the decrease in values after instruction highlighted in light gray. The experimental group only increased its mean in sentence 3, while the control group increased it in sentences 1 and 10 and remained unchanged in sentence 2. On average, both groups made fewer pauses after instruction.

Table 11. Total number of pauses means (*M*) and standard deviations (*SD*) per sentence

	Control group				Experimental group			
	Time 1		Time 2		Time 1		Time 2	
Sentence	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.57	0.51	0.62	0.74	0.52	0.6	0.33	0.48
2	1.86	1.39	1.86	1.15	1.86	1.49	1.52	1.36
3	1.76	1.48	1.14	1.01	0.9	0.77	1.1	1.09
4	3.29	1.38	2.62	1.69	3	1.7	2.48	1.57
5	3.76	2.33	3	1.51	4.12	2.66	3.86	1.55
6	3.26	1.38	2.9	1.3	3.1	1.87	2.71	1.38
7	2.95	1.69	2.52	1.78	2.81	1.6	2.29	1.45
8	4.33	2.73	4.24	2.14	3.86	2.08	3.05	2.22
9	2.43	1.4	2.26	1.18	3.21	1.33	2.56	1.07
10	2.53	0.8	2.9	1.33	2.48	1.21	2.14	1.11
Total	**2.7**		**2.5**		**2.6**		**2.2**	

As shown in Appendix 8, this time intermediate students from the control group and high-intermediate/advanced students from the experimental group were the ones to show more improvement, unlike with VarcoV values.

4.1.1.2.1.2. Mixed ANOVAs.

Two repeated-measures ANOVAs were performed. First, an ANOVA was performed with time and sentence as the within-subjects factors and group as the between-subjects factor. The total number of pauses was the dependent variable. Results did not show significance for group, $F(1,40) = .427$, $p = .517$, but they did for time, $F(1,40) = 8.995$, $p < .005$, and sentence $F(9,32) = 46.743$, $p < .01$. However, the time*sentence interaction did not reach significance $F(9,32) = .578$, $p = .739$ (see Table 12).

Table 12. Mixed repeated-measures ANOVA with the total number of pauses as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	8.995	.005**
	sentence	9	32	46.743	.01**
Interaction	time*sentence	9	32	.578	.739
Between-subjects	group	1	40	.427	.517

Note. ** $p < .01$.

As displayed in Figure 10, the number of pauses made dropped more after treatment for the experimental group than the control group. This tendency was also found in the descriptive analysis when comparing group means.

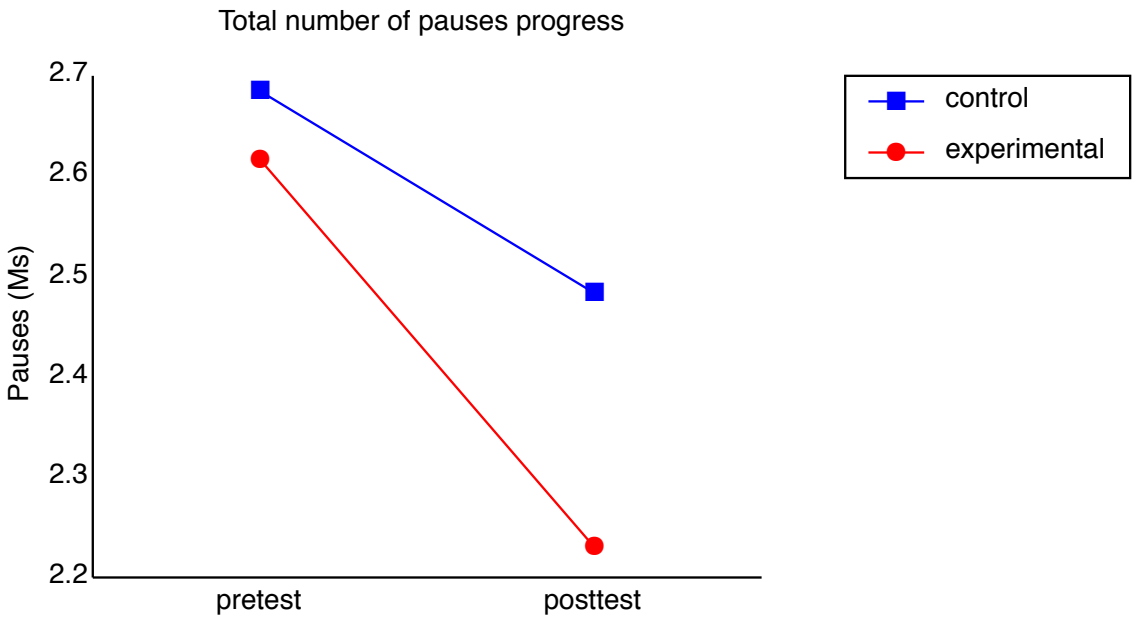


Figure 10. Group progress on the total number of pauses according to treatment

Second, another ANOVA was run with the difference in the number of pauses for each sentence before and after training as the dependent variable, sentence as the within-subjects factor and group as the between-subjects factor. Neither sentence, $F(9,32)$

= .827, p = .546, nor group, $F(1,40) = 1.567$, p = .218, showed significance (see Table 13).

Table 13. Mixed repeated-measures ANOVA with the total number of pauses difference over time as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	sentence	9	32	.827	.546
Between-subjects	group	1	40	1.567	.218

4.1.1.2.1.3. T-tests.

Two paired-samples (control T2 vs. control T1; experimental T2 vs. experimental T1) and two independent-samples t-tests (control T1 vs. experimental group T1; control T2 vs. experimental group T2) per sentence were calculated, which reached significance only in a few specific cases (see Appendix 9). Independent-samples t-tests were significant when comparing both groups before treatment in sentence 3, $T(40) = 2.355$, $p < .023$, and after treatment in sentence 10, $T(40) = 2.044$, $p < .048$. As for paired-samples t-tests, they showed significance for the control group performance in sentences 3, $T(20) = -2.444$, $p < .024$, and 4, $T(20) = -2.092$, $p < .049$, and for the experimental group performance in sentence 9, $T(20) = -2.103$, $p < .048$.

Unlike when analyzing VarcoV effect sizes, a difference between the control and the experimental group's progress was not obvious, since both groups showed mostly negative effect sizes, which were translated as a reduction of the number of pauses made (see Figure 11). Even so, an independent t-test comparing the effect sizes of all

the paired-samples t-tests was also carried out to examine the relevance of the improvement. This time, results did not reveal significance $T(18) = 1.229$, $p = .235$.

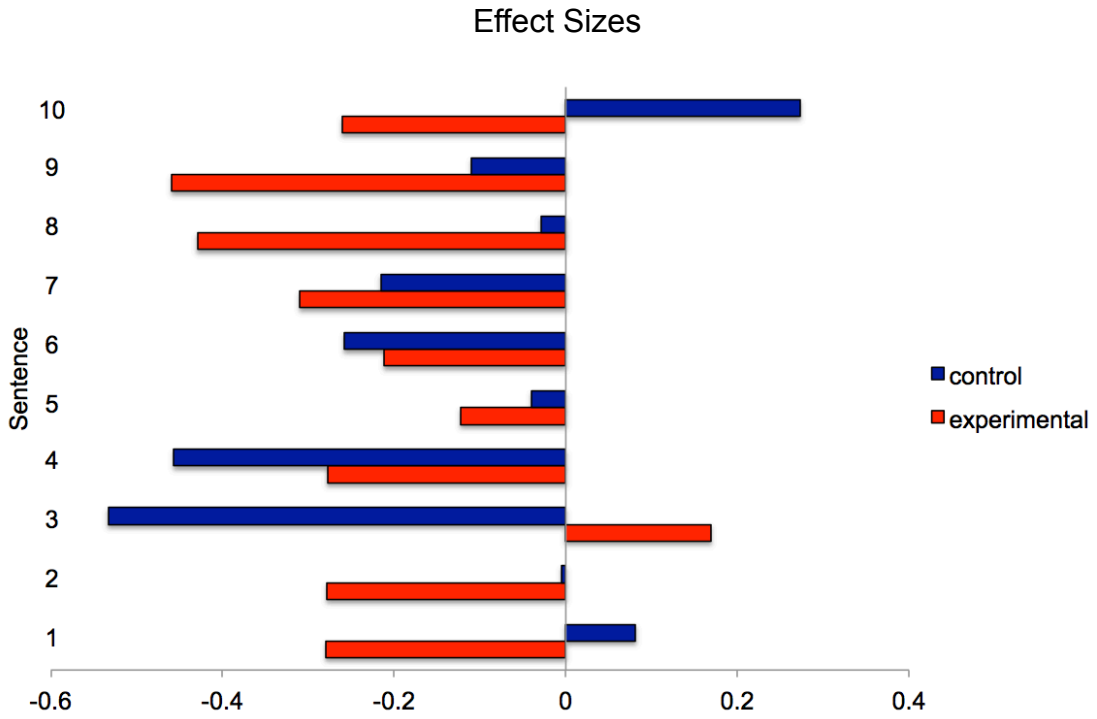


Figure 11. Effect sizes for the total number of pauses paired-sample t-tests

4.1.1.2.2. Unfilled pauses.

After measuring the total number of pauses, it was decided to put filled pauses aside and concentrate in unfilled pauses, since they have been considered a better indicator of fluency (see section 3.9.1). The same analytical process as the one used to analyze the overall number of pauses was followed to determine whether participants made fewer unfilled pauses after treatment.

4.1.1.2.2.1. Descriptive statistics.

Unfilled pauses per sentence were also arranged according to time and group and the corresponding means and standard deviations were calculated (see Table 14). The values that showed a decrease in the number of unfilled pauses made after instruction are highlighted in light gray. On average, the experimental group reduced the number of unfilled pauses in eight out of ten sentences, made the same amount in sentence 3, and made more in sentence 5. The control group, on the other hand, made fewer pauses in six out of ten sentences, the same amount in sentence 8, and more in sentences 1, 9 and 10. Hence, the experimental group seemed to perform better after instruction.

Table 14. Unfilled pauses means (*M*) and standard deviations (*SD*) per sentence

	Control group				Experimental group			
	Time 1		Time 2		Time 1		Time 2	
Sentence	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.52	0.51	0.57	0.75	0.48	0.51	0.33	0.48
2	1.95	1.47	1.81	1.03	1.86	1.49	1.52	1.36
3	1.43	1.03	1.1	0.89	0.9	0.77	0.9	0.83
4	2.76	1.04	2.52	1.63	2.81	1.66	2.29	1.62
5	3.43	1.8	3.33	1.35	3.52	1.97	3.79	0.89
6	2.72	0.79	2.71	1.1	2.81	1.54	2.43	1.16
7	2.57	1.36	2.19	1.5	2.29	1.19	1.86	1.06
8	4	2.41	4	2.05	3.71	2.03	2.86	1.93
9	2.19	1.08	2.24	1.37	3.05	1.16	2.33	1.02
10	2.33	1.06	2.71	1.1	2.38	1.07	2.1	1.04
Total	**2.4**		**2.3**		**2.4**		**2**	

Again, the overall mean of both groups decreased after instruction, but individual performance differs depending on the sentence examined (see Appendix 10). Even so, progress seemed more obvious at any level for both control and experimental groups except for beginners from the control group.

4.1.1.2.2.2. Mixed ANOVAs.

A repeated-measures ANOVA was run with time and sentence as the within-subjects factors and group as the between-subjects factor. The number of unfilled pauses was the dependent variable. As displayed in Table 15, results were not significant for group, $F(1,40) = .349$, $p = .558$, but they were for sentence, $F(9,32) = 56.725$, $p < .01$, and for time, $F(1,40) = 5.478$, $p < .024$. However, the time*sentence interaction was not significant, $F(9,32) = .773$, $p = .589$. Therefore, results seemed to correlate with the ones obtained when analyzing the total number of pauses.

Table 15. Mixed repeated-measures ANOVA with unfilled pauses as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	5.478	.024*
	sentence	9	32	56.725	.01**
Interaction	time*sentence	9	32	.773	.589
Between-subjects	group	1	40	.349	.558

Note. * $p = .05$. ** $p = .01$.

Nevertheless, Figure 12, which displays the progress of the different groups in regard of the number of unfilled pauses, shows a slight different scenario from the one presented regarding the total number of pauses in Figure 10. Firstly, while the control group made more pauses than the experimental group on average at the beginning of the course, the number of unfilled pauses was almost the same for both groups. Secondly, when analyzing all the pauses, the progress of both groups was similar, since the control group made 0.2 pauses and the experimental group 0.4 pauses less than the first time they took the test. However, when taking into account just the unfilled pauses the impact of the progress changed. The control group remained almost the same at 2.4

pauses, while the experimental group made 0.4 pauses less after instruction. Thus, the improvement of the control group seemed to result from making fewer hesitations and repetitions rather than silent pauses (for more information, see the discussion chapter).

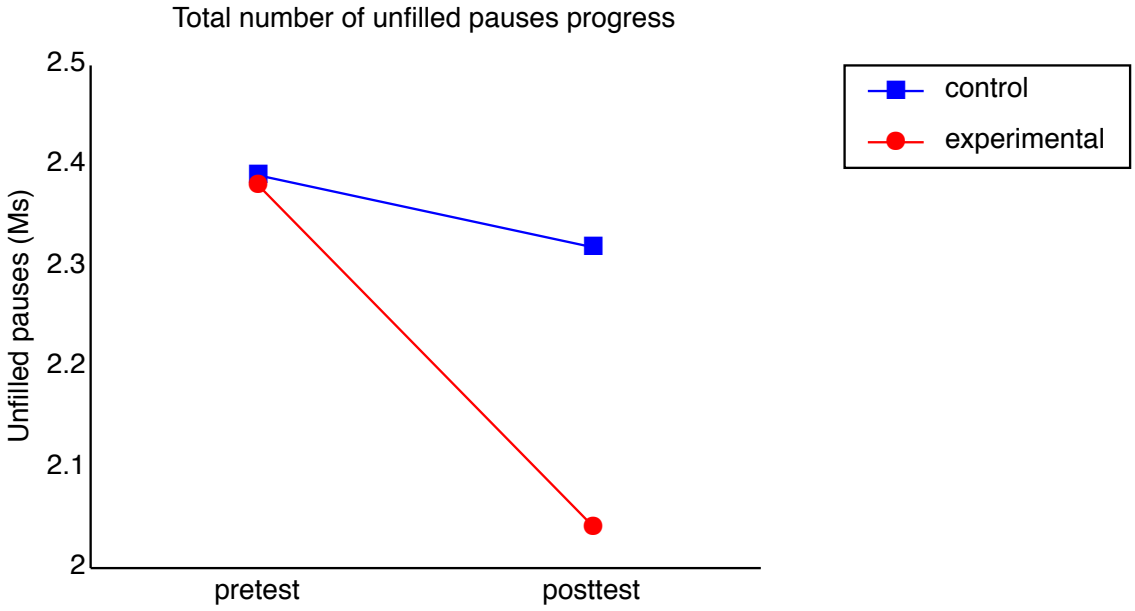


Figure 12. Group progress on the unfilled pauses according to treatment

A second ANOVA was carried out with the difference in times as the dependent variable, sentence as the within-subjects factor and group as the between-subjects factor. As shown in Table 16, this time results did not research significance for any of the factors under study: sentence, $F(9,32) = .708$, $p = .643$; group, $F(1,40) = 2.470$, $p = .124$.

Table 16. Mixed repeated-measures ANOVA with the unfilled pauses difference over time as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	sentence	9	32	.708	.643
Between-subjects	group	1	40	2.470	.124

4.1.1.2.2.3. T-tests.

Two paired-samples (control T2 vs. control T1; experimental T2 vs. experimental T1) and two independent-samples t-tests (control T1 vs. experimental group T1; control T2 vs. experimental group T2) per sentence were run. As with the total number of pauses, t-tests reached significance in only some of the cases (see Appendix 11). This time, only one independent-samples t-test showed significance: groups before treatment in sentence 9, $T(40) = -2.480$, $p < .017$. As for the paired-samples t-tests, the experimental group showed significance in sentences 8, $T(20) = -2.186$, $p < .041$, and 9, $T(20) = -2.855$, $p < .010$.

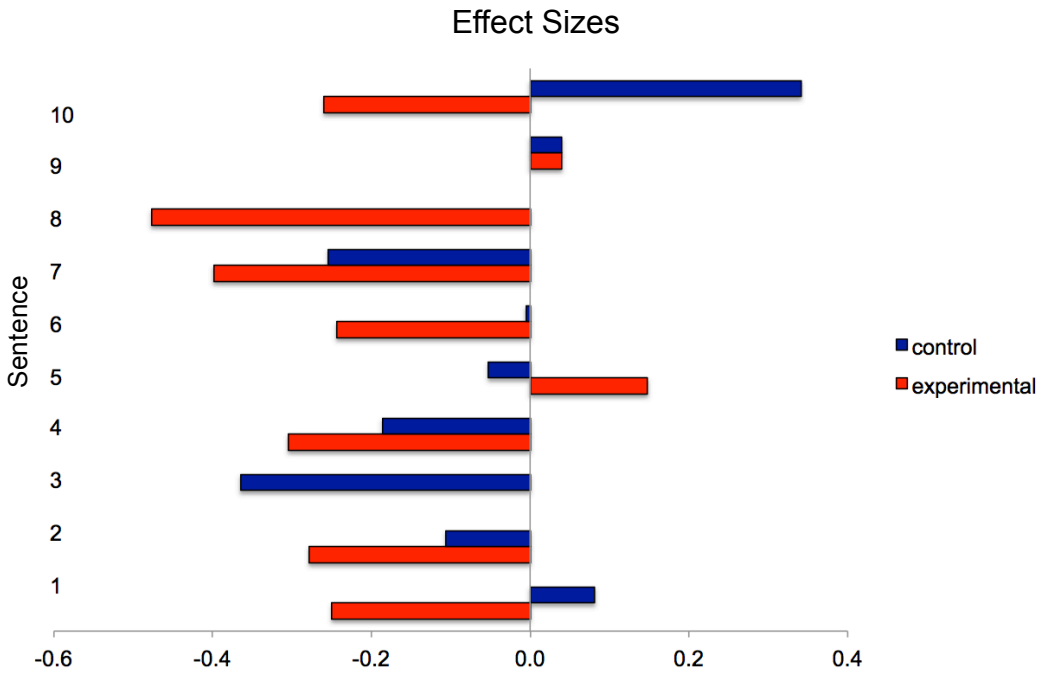


Figure 13. Effect sizes for unfilled pauses paired-samples t-tests

As displayed in Figure 13, the effect sizes of all the paired-samples t-tests showed a greater improvement for the experimental group performances in most of the cases. In order to examine the relevance of this difference in the effect size depending on the treatment, an independent-samples t-test with these effect sizes was run, which showed significance $T(18) = 2.343$, $p < .031$.

To sum up, the different analyses carried out for the three items (VarcoV values, total number of pauses and unfilled pauses) suggested favorable effects on students' prosody when pronunciation was introduced in class, especially when rhythm was explicitly taught, even though results were not always significant. For this reason, it was considered interesting to examine whether these features correlate with each other. The following section will present these correlations.

4.1.1.3. Correlations between acoustic measurements.

Several Pearson product-moment correlation tests were conducted to evaluate the relationship between VarcoV values and the total number of pauses, on the one hand, and VarcoV values and unfilled pauses, on the other hand. The items were compared according to time and group so as to test effects on treatment.

4.1.1.3.1. VarcoV - total pauses.

In total, 40 Pearson product-moment correlation coefficients were computed to assess to what extent measurements of VarcoV were related to the total number of pauses obtained. The items were correlated for each of the ten sentences before and after instruction and according to the treatment received. Hence, four different Pearson product-moment correlation coefficients (i.e. correlations for the control before and after instruction, and for the experimental group before and after instruction) were run for each of the sentences (see Appendix 12).

There was a positive correlation between values in very specific cases. Regarding the control group, correlations were significant for sentence 8, $r = .460$, $n = 21$, $p < .036$, and sentence 10, $r = -.454$ $n = 21$, $p < .039$, before treatment, and sentence 1, $r = -.497$, $n = 21$, $p < .022$, after treatment. As for the experimental group, results showed a significant correlation only in sentence 3, $r = .481$, $n = 21$, $p < .027$, and sentence 9, $r = .449$, $n = 21$, $p < .041$, before treatment.

4.1.1.3.2. VarcoV - unfilled pauses.

As performed with VarcoV figures and overall pauses, 40 Pearson product-moment correlation coefficients were computed to assess to what extent VarcoV values were

related to unfilled pauses. The items, again, were correlated for each of the ten sentences before and after training and according to the treatment received (see Appendix 13).

Correlation was also positive only in some of the sentences. Regarding the control group, this time correlation was only significant before treatment for sentence 8, r = .521, n = 21, p < .015, but after treatment there was another time significant correlation for sentence 1, r = -.558, n = 21, p < .009, and also for sentence 4, r = .449, n = 21, p < .041. As for the experimental group, this time results showed a significant correlation only in sentence 3 before treatment, r = .481, n = 21, p < .027.

To conclude, VarcoV values do not correlate either with the total number of pauses or just the unfilled pauses most of the time. Hence, no evidence was provided of a relationship between this rhythm metric and pausing.

4.1.2. Listener-based analysis.

This section will present the results obtained from the online assessment test taken by the seven experienced listeners. As mentioned in Chapter 3, the raters listened to eighty-four 45-second recordings extracted from the opinion exercise of both the pre- and the posttests. All the scores of each of the listeners were grouped by recording, thus each contribution had a total of seven different judgments. The following sections will present results on comprehensibility and fluency separately and the corresponding correlation between these two aspects.

4.1.2.1. Comprehensibility.

This section focuses on comprehensibility scores. The following subsections will present the ICC test performed, the descriptive analysis, and the results generated from the Mixed ANOVAs and the t-tests.

4.1.2.1.1. Reliability test.

An intraclass correlation coefficient (ICC) test was computed to examine the reliability of the raters. A two-way random effect model based on mean ratings and consistency assessed the uniformity of the scoring. Results reported a good mean estimation of .873 along with a good to excellent 95% confidence interval (IC = .826-.910). Hence, scores were considered reliable.

4.1.2.1.2. Descriptive analysis.

Table 17 illustrates the means of the raters' scores for each of the students' speeches. Marks were distributed according to time (T1 = pretest; T2 = posttest), group (experimental; control), and the level of English (1 = beginners; 2 = intermediate; 3 = high-intermediate/advanced). Highlighted in light gray, twelve students from the control group (mainly beginners and intermediate students) and eight from the experimental group were rated higher for the after-treatment production. As for group performance, the control group was considered to improve slightly, while the experimental group showed certain signs of deterioration. However, in both cases results remained almost stable. Therefore, the comparison of the means suggested that the control group were considered more comprehensible after instruction than the experimental group even though scores were approximately equal, oscillating between 3 and 3.3 scores.

Table 17. Raters' mean scores (*M*) and standard deviations (*SD*) per student for comprehensibility

	Control group					Experimental group				
		T1		T2			T1		T2	
Student	Level	*M*	*SD*	*M*	*SD*	Level	*M*	*SD*	*M*	*SD*
1	1	2.0	1.00	2.6	0.53	1	4.0	0.58	3.4	0.98
2	1	2.6	0.79	3.6	0.79	1	3.6	0.79	3.6	0.53
3	1	2.1	0.90	1.9	0.69	1	3.4	0.79	2.3	0.95
4	1	2.6	0.53	3.3	0.76	1	3.0	0.58	2.1	0.69
5	2	3.4	0.98	2.9	1.07	1	3.0	0.82	3.4	0.53
6	2	2.9	0.38	3.3	0.76	1	3.0	0.58	3.6	0.53
7	2	1.6	0.53	1.9	0.69	1	1.7	0.49	1.7	0.76
8	2	2.9	0.69	3.0	0.58	1	3.1	0.90	2.7	0.49
9	2	2.7	0.49	3.1	1.07	1	4.0	0.58	3.4	0.53
10	2	2.4	0.79	2.9	1.21	1	2.1	0.90	2.4	0.79
11	2	3.7	0.76	3.9	0.69	2	3.3	0.76	3.7	0.76
12	2	3.0	1.15	3.4	0.79	2	4.1	0.90	3.6	0.53
13	2	3.3	0.49	2.9	0.69	2	3.0	0.82	3.7	0.49
14	2	2.9	0.69	2.4	0.79	2	3.3	0.76	3.0	0.00
15	2	2.6	0.53	1.9	0.69	2	3.9	0.38	3.9	0.69
16	3	3.4	0.79	3.3	0.49	2	3.0	0.00	4.0	0.82
17	3	4.1	0.38	3.7	0.49	3	3.1	0.38	3.0	0.58
18	3	3.7	0.49	3.9	0.69	3	2.9	0.69	3.6	0.53
19	3	3.9	0.38	3.6	0.53	3	4.4	0.53	4.1	0.38
20	3	3.9	0.38	3.6	0.98	3	3.4	0.98	3.7	0.49
21	3	4.4	0.53	4.6	0.53	3	3.6	0.53	3.6	0.79
Total		**3.05**		**3.11**			**3.29**		**3.27**	

4.1.2.1.3. Mixed ANOVAs.

Three repeated-measures ANOVAs were conducted to study the students' comprehensibility progress according to the type of treatment they received. First, an ANOVA analyzing all the participants was performed. Second, two more ANOVAs analyzing high-level students and low-level students respectively were carried out in order to examine a potential ceiling effect.

4.1.2.1.3.1. Total students.

The first repeated-measures ANOVA was run with time as the within-subjects factor and group as the between-subjects factor (see Table 18). The means of the raters' scores were the dependent variable. Results were not significant for either time, $F(1,40) = .066$, $p = .798$, or group, $F(1,40) = 1.007$, $p = .322$.

Table 18. Mixed repeated-measures ANOVA with the means of the raters' comprehensibility scores as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	.066	.798
Between-subjects	group	1	40	1.007	.322

Figure 14 shows that, despite having been marked as higher before the instruction, the experimental group was considered less comprehensible after training. However, this deterioration was minimal. By contrast, the control group showed improvement, although it kept being scored lower than the experimental group.

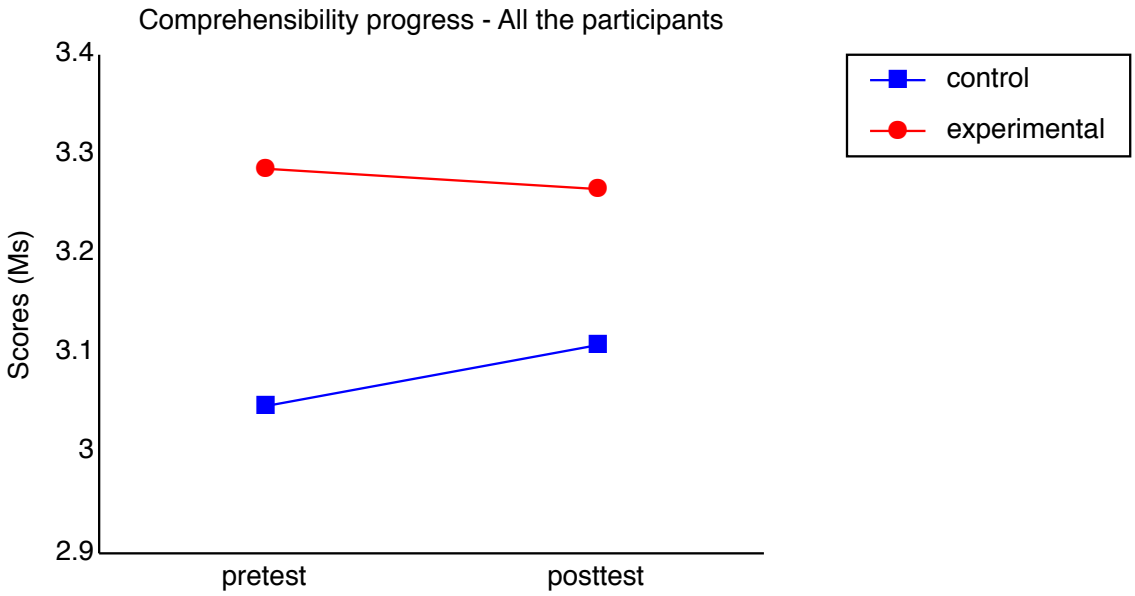


Figure 14. All the participants' comprehensibility progress according to treatment

Results contradicted the tendencies observed in the acoustic analyses: The VarcoV values and pauses tests supported the hypothesis that the control group started with a higher level of English, matching the placement test results (see Table 5 on the participants' profiles), and that the experimental group tended to positively improve more than the control group. By contrast, this time the experimental group seemed to have a higher level of English at the beginning of the course and the control group was the one to improve after treatment while the experimental group did worse. Hence, the speaking skills of each group did not seem to correspond to the levels extracted from the placement test (see the following chapter for further discussion).

For this reason, two separate groups were established according to their initial speaking skills instead of their initial level in the placement test and analyzed separately: students who achieved a 3.5 score or higher in the opinions' ratings at the beginning of the course, considered high-level students, and those who achieved a lower mark than 3.5

on that same test, considered low-level students. Each level group of students was examined according to the treatment received by conducting a mixed ANOVA.

4.1.2.1.3.2. High-level students.

A repeated-measures ANOVA was run with only high-level students. The scores of six students from the control group and seven from the experimental group were compared. As displayed in Table 19, results were significant for time, $F(1,11) = 6.114$, $p < .031$, but not for group, $F(1,11) = .515$, $p = .488$. The time*group interaction was not significant either, $F(1,11) = 1.529$, $p = .242$.

Table 19. Mixed repeated-measures ANOVA with the means of the raters' comprehensibility scores as the dependent variable for high-level students only

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	11	6.114	.031*
Between-subjects	group	1	11	.515	.488
Interaction	time*group	1	11	1.529	.242

Note. * $p < .05$.

Figure 15 shows that high-level students from the control group were approximately as comprehensible as the ones from the experimental group at the beginning of the course. However, the latter showed a greater decrease after instruction.

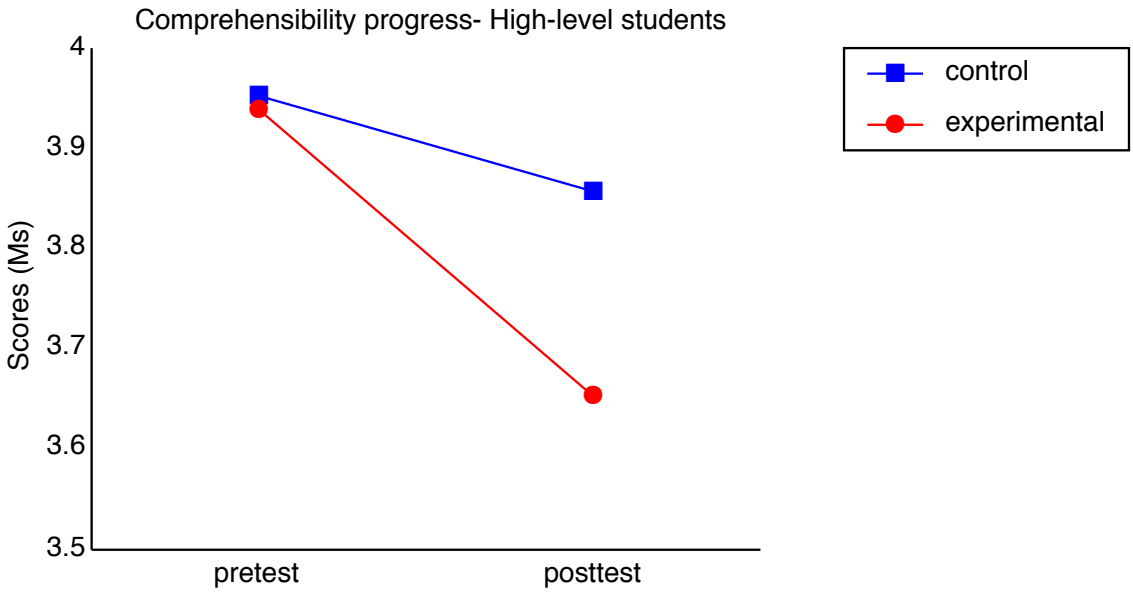


Figure 15. High-level students' comprehensibility progress according to treatment

4.1.2.1.3.3. Low-level students.

Another repeated-measures ANOVA was performed with low-level students. The mean of the scores of fifteen students from the control group and fourteen from the experimental group were brought under analysis as the dependent variable, time as the within-subjects factor and group as the between-subjects factor. As presented in table 20, results did not show significance for either time, $F(1,27) = 1.241$, $p = .275$, or group, $F(1,27) = 2.052$, $p = .163$.

Table 20. Mixed repeated-measures ANOVA with the means of the raters' comprehensibility scores as the dependent variable for low-level students only

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	27	1.241	.275
Between-subjects	group	1	27	2.052	.163

When analyzing the groups' evolution throughout time in Figure 16, it can be observed that both groups scored higher after instruction and showed a steady improvement. The experimental group performed better at the beginning of the course, questioning the placement test results (see Table 5), which showed a considerably higher amount of beginners for the experimental group than for the control group. Taking into account the results of all the ANOVAs computed, some degree of ceiling effect was suggested for the comprehensibility test.

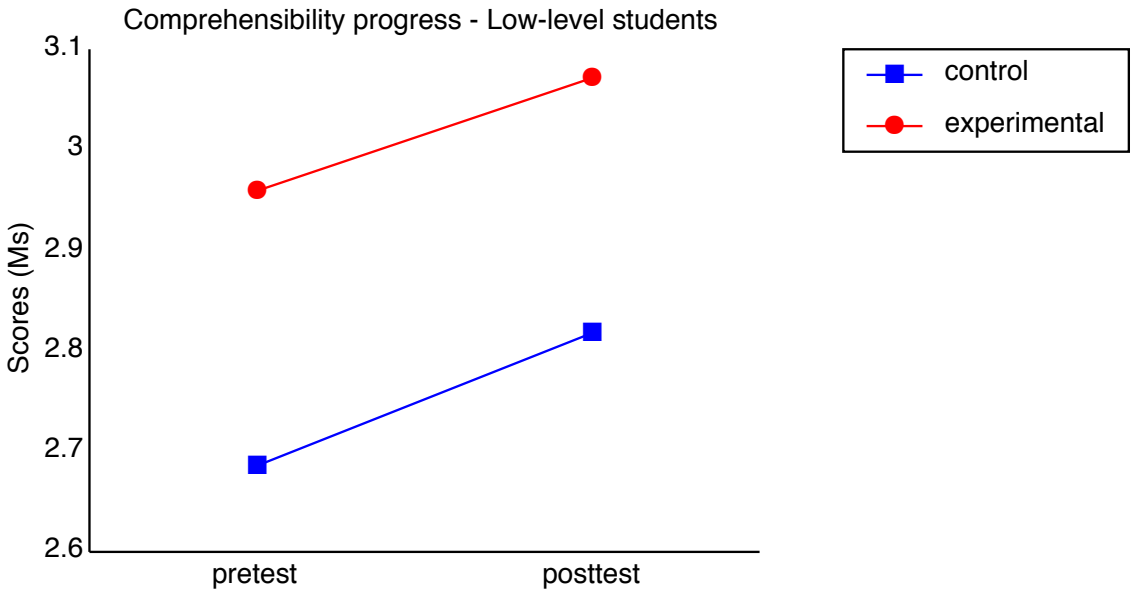


Figure 16. Low-level students' comprehensibility progress according to treatment

4.1.2.1.4. T-tests.

Two paired-samples t-tests (control T2 vs. control T1; experimental T2 vs. experimental T1) and two independent-samples t-tests (control T1 vs. experimental group T1; control T2 vs. experimental group T2) were conducted with the means of each student's speech. As displayed in Table 21, none of the tests showed statistical significance ($p > 0.5$). A

negative effect size is detected for the experimental group progress, which agrees on the general decrease of experimental students' comprehensibility after instruction shown by the ANOVAs and the descriptive analysis.

Table 21. Comprehensibility scores t-tests

T-test		df	*T*	*p*-value	Effect size (Cohen's *d*)
Paired	C T2-T1	20	.604	.552	1.32
	E T2-T1	20	-.167	.869	-.365
Independent	C T1-E T1	40	-1.118	.270	.345
	C T2-E T2	40	-.741	.463	.229

4.1.2.2. Fluency.

This section focuses on the fluency ratings. The following subsections will present the descriptive analysis and the results generated from the ICC test, the mixed ANOVAs and the t-tests computed.

4.1.2.2.1. Reliability test.

An intraclass correlation coefficient (ICC) test was computed to examine the reliability of the raters. A two-way random effect model based on mean ratings and consistency assessed the uniformity of the scoring. Results reported a good mean estimation of .895 along with a good to excellent 95% confidence interval (IC = .857-.926). Hence, scores were considered reliable.

4.1.2.2.2. Descriptive analysis.

Table 22 displays the results from the fluency ratings. Again, the mean of the raters' scores for each of the students were distributed according to time (T1 = pretest; T2 =

posttest), group (experimental; control) and the level of English (1 = beginners; 2 = intermediate; 3 = high-intermediate/advanced). Highlighted in light gray, twelve students from the control group and eleven from the experimental group improved their performance after treatment. For the control group, beginners and intermediate students were the ones who tended to improve, while the improvement in the experimental group was more obvious for intermediate and high/intermediate learners. On average, both treatment groups improved after instruction: The control group improved more than the experimental group, but the difference was very small.

Table 22. Raters' mean scores (*M*) and standard deviations (*SD*) per student for fluency

	Control group					Experimental group				
		T1		T2			T1		T2	
Student	Level	*M*	*SD*	*M*	*SD*	Level	*M*	*SD*	*M*	*SD*
1	1	2.0	0.82	2.1	0.38	1	3.9	0.69	3.3	0.76
2	1	2.7	0.49	3.4	0.53	1	3.4	0.53	3.3	0.49
3	1	2.0	0.58	1.6	0.53	1	3.4	0.79	1.4	0.53
4	1	2.3	0.49	3.3	0.95	1	2.4	0.53	2.4	0.98
5	2	2.6	0.53	2.1	0.69	1	3.3	0.49	3.3	0.76
6	2	3.1	0.38	3.7	0.49	1	3.0	0.58	3.1	0.38
7	2	1.3	0.49	1.6	0.53	1	1.7	0.76	1.4	0.53
8	2	2.7	0.76	3.0	0.58	1	2.1	0.38	2.9	0.38
9	2	1.9	0.69	2.1	0.69	1	3.3	0.76	4.0	0.58
10	2	2.4	0.53	2.7	0.76	1	1.9	0.90	2.6	0.53
11	2	3.4	0.53	3.6	0.53	2	3.3	0.49	3.6	0.53
12	2	3.0	0.82	3.4	0.53	2	3.3	0.76	3.3	0.76
13	2	2.6	0.53	2.7	0.76	2	2.9	0.90	3.7	0.49
14	2	2.9	0.69	2.3	0.49	2	2.7	0.49	2.9	0.38
15	2	2.1	0.69	2.0	0.58	2	3.6	0.79	3.6	0.53
16	3	3.0	0.58	3.0	0.82	2	3.3	0.76	3.6	0.53
17	3	3.9	0.69	3.3	0.76	3	3.3	0.76	2.1	0.69
18	3	3.1	0.38	3.4	0.53	3	2.6	0.79	3.3	0.49
19	3	3.6	0.53	3.3	0.76	3	3.6	0.53	3.7	0.49
20	3	4.0	0.00	3.4	0.53	3	3.3	0.49	2.7	0.49
21	3	4.4	0.53	4.0	0.00	3	2.4	0.79	3.0	0.58
Total		**2.81**		**2.86**			**2.98**		**3.01**	

4.1.2.2.3.Mixed ANOVAs.

As proceeded with the analysis of comprehensibility, fluency progress was further assessed conducting three different repeated-measures ANOVAs: an ANOVA analyzing all the participants' scores, and two more ANOVAs analyzing high-level students and low-level students respectively to study a potential ceiling effect.

4.1.2.2.3.1. Total students.

The first ANOVA took into account all the participants' performances as the dependent variable, time as the within-subject factors and group as the between-subject factors. Table 23 presents the results obtained, which did not show significance for either time, $F(1,40) = .208$, $p = .651$, or group, $F(1,40) = .637$, $p = .430$.

Table 23. Mixed repeated-measures ANOVA with the means of the raters' fluency scores as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	.208	.651
Between-subjects	group	1	40	.637	.430

As it can be seen from Figure 17, both groups were rated more fluent after treatment. The experimental group scored higher at the beginning of the course, but the control group showed slightly greater improvement. Again, the experimental group was rated as more fluent than the control group at the beginning of the course, which challenges the results of both the placement test and the acoustic analysis.

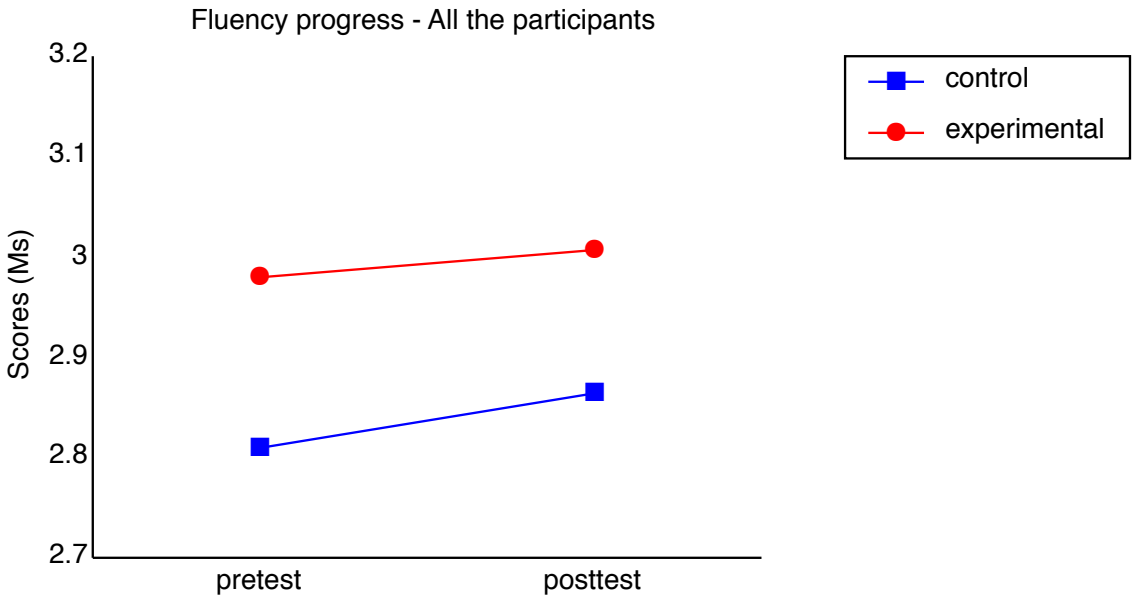


Figure 17. All the participants' fluency progress according to treatment

4.1.2.2.3.2. High-level students.

A second repeated-measures ANOVA was computed with only those who obtained a 3.5 or higher mark in the pretest. In this case, the scores of four students from the control group and three from the experimental group were compared. As displayed in Table 24, results were significant for time, $F(1,5) = 9.243$, $p < .029$, but not for group, $F(1,5) = .454$, $p = .530$. The time*group interaction was not significant either, $F(1,5) = 2.591$, $p = .168$.

Table 24. Mixed repeated-measures ANOVA with the means of the raters' fluency scores as the dependent variable for high-level students only

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	5	9.243	.029*
Between-subjects	group	1	5	.454	.530
Interaction	time*group	1	5	2.591	.168

Note. * $p < .05$.

Figure 18 shows that high-level students from the control group were considered more fluent than the ones from the experimental group at the beginning of the course. However, the scores of both groups decreased to practically the same level after instruction.

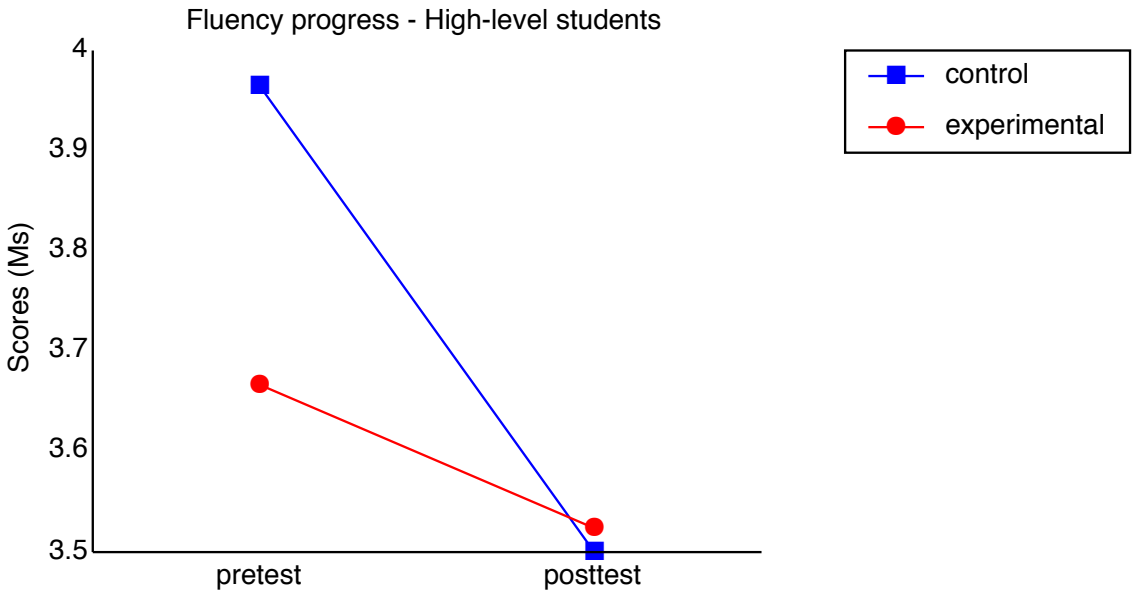


Figure 18. High-level students' fluency progress according to treatment

4.1.2.2.3.3. Low-level students.

A third repeated-measures ANOVA was run with those who scored less than 3.5 in the pretest. In this case, the dependent variable consisted of the means of the scores of seventeen students from the control group and eighteen from the experimental group. Again, time acted as the within-subjects factor and group as the between-subjects factor. As shown in Table 25, results did not show significance for either time, $F(1,33) = 1.332$, $p = .257$, or group, $F(1,33) = 1.945$, $p = .172$.

Table 25. Mixed repeated-measures ANOVA with the means of the raters' fluency scores as the dependent variable for low-level students only

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	33	1.332	.257
Between-subjects	group	1	33	1.945	.172

Both groups showed higher scores after instruction (see Figure 19). Although the control group improvement looked greater, the difference between the two groups' performances before and after treatment was very small. However, the experimental group was scored as more fluent at the beginning of the course again despite having more beginners.

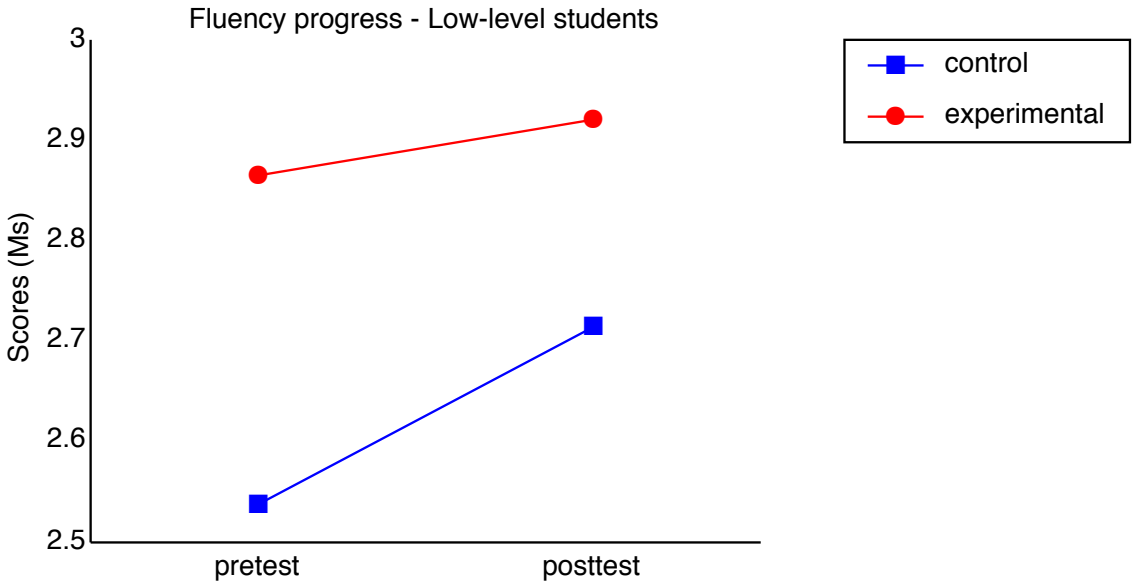


Figure 19. Low-level students' fluency progress according to treatment

As found in the comprehensibility analysis, the time variable was only revealed significant in the test run with high-level students, which suggested that they were rated so high at the beginning that there was no room to detect an improvement. Besides, the

experimental low-level students scored higher than the control ones for the pretest despite having a lower general level of English, which could be a consequence of overrating their pretest performances (see subsequent chapters for further discussion).

4.1.2.2.4. T-tests.

As displayed in Table 26, the means of the ratings provided for each student's speech were used to compute two paired-samples t-tests (i.e. control T2 vs. control T1; experimental T2 vs. experimental T1) and two independent-samples t-tests (i.e. control T1 vs. experimental group T1; control T2 vs. experimental group T2), which did not show statistical significance ($p > 0.5$).

Table 26. Fluency scores t-tests

T-test		df	*T*	*p*-value	Effect size (Cohen's *d*)
Paired	C T2-T1	20	.554	.586	1.21
	E T2-T1	20	.182	.857	.040
Independent	C T1-E T1	40	-.802	.427	.248
	C T2-E T2	40	-.653	.518	.201

To sum up, both comprehensibility and fluency results showed similar tendencies. The following subsection will investigate a potential correlation between these two items.

4.1.2.3. Correlation between comprehensibility and fluency.

The comprehensibility and fluency means of the scores obtained by each of the students were used to compute four Pearson product-moment correlation tests (one per group and time). As displayed in Table 27, although the correlation coefficients of the control group tests were higher than the ones of the experimental group, all four tests showed

statistical significance. Hence, the tests suggested that the students' level of comprehensibility was related to their level of fluency.

Table 27. Comprehensibility - fluency Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	.911	21	.01**
	2	.910	21	.01**
Experimental	1	.769	21	.01**
	2	.844	21	.01**

Note. ** $p < .01$.

However, the opposite tendencies observed between the analyses of the acoustic measurements and the raters' judgments challenged the idea of a mutual relation. For this reason, further correlations between the different items of both sorts of analyses will be also brought under study.

4.1.3. Listener-based –acoustic analyses: correlations.

Correlations between the listener-based and the acoustic aspects will be presented in the subsequent sections in order to examine a potential relationship. First, comprehensibility scores will be correlated with VarcoV values, the total number of pauses and the unfilled pauses respectively. Then, the correlations between fluency ratings and each of the items analyzed acoustically will be investigated.

4.1.3.1. Comprehensibility - acoustic measurements.

The students' means of the comprehensibility ratings were further correlated to the students' means of the different items analyzed acoustically. Firstly, the students' VarcoV means taking all the sentences together were used to compute the correlations

with comprehensibility scores (see Table 28). Again, four Pearson product-moment correlation tests (one per group and time) were performed. In this case, correlations were only significant for the control group after instruction and the experimental group before instruction. Their coefficients, however, were not really high ($r < .5$).

Table 28. Comprehensibility - VarcoV Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	.020	21	.931
	2	.474	21	.030*
Experimental	1	.490	21	.024*
	2	.304	21	.180

Note. * $p < .05$.

As for the correlations between comprehensibility figures and the students' mean pauses (either when comparing the total number of pauses or the unfilled pauses only), the control group, both before and after instruction, showed significance while the experimental group did not (see Table 29 and 30).

Table 29. Comprehensibility - total number of pauses Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	-.611	21	.003**
	2	-.527	21	.014*
Experimental	1	-.123	21	.595
	2	.108	21	.640

Note. * $p < .05$. ** $p < .01$.

Table 30. Comprehensibility - unfilled pauses Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	-.666	21	.001**
	2	-.547	21	.010*
Experimental	1	-.133	21	.565
	2	.112	21	.627

Note. * $p < .05$. ** $p < .01$.

In short, comprehensibility ratings tended to correlate with the acoustic measurements for the control group but not for the experimental group. Bearing into account that the experimental group was unexpectedly found to be less comprehensible after training and that results did not correlate before instruction either for this group, a potential limitation of the test itself is suggested. This interpretation will be further developed in the discussion section.

4.1.3.2. Fluency - acoustic measurements.

The same analytical procedure was followed with fluency ratings. Table 31 displays the results obtained from the four Pearson correlation tests (one per group and time) computed between fluency scores and VarcoV means. No significant correlation was observed for any of the tests. Hence, results suggested that fluency and VarcoV were not correlated.

Table 31. Fluency - VarcoV Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	.028	21	.905
	2	.354	21	.115
Experimental	1	.183	21	.427
	2	.195	21	.397

Regarding pausing, only the values from the control group revealed significance, both when taking into account the total number of pauses (see Table 32) and the unfilled pauses (see Table 33).

Table 32. Fluency - total number of pauses Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	-.551	21	.010*
	2	-.510	21	.018*
Experimental	1	-.261	21	.254
	2	.709	21	.087

Note. * $p < .05$.

Table 33. Fluency - unfilled pauses Pearson product-moment correlation coefficients

Group	Time	*r*	*n*	*p*
Control	1	-.611	21	.003**
	2	-.528	21	.014*
Experimental	1	-.255	21	.266
	2	.065	21	.780

Note. * $p < .05$. ** $p < .01$.

Hence, a lack of significance again between ratings and acoustic measurements for the experimental group but not for the control group further supported the possibility of problems in the tests performed by the listeners.

4.1.4. Students' satisfaction survey.

The quantitative measures used so far do not capture factors such as the effort, the investment, the motivation or the attitude of the students towards the course, which undoubtedly could influence the outcome of the students' learning progress. For this reason, the following section analyzes the students' impressions on the pronunciation module to shed more light on the issue.

In order to further assess the benefits of the pronunciation module, the satisfaction survey that the students filled in at the end of the course was also analyzed. In total, 79 students from the control group and 104 students from the experimental group answered the questionnaire. Graphs will illustrate the results obtained from items 1, 2, 3, 4 and 6 (see Appendix 4). Answers to items 5 and 7 had to be dismissed: Item 5 asked students to determine whether they practiced what was taught during the pronunciation sessions out of the classroom. Some students understood the question as related to the English classes in general, or previous practice before starting the studies at university, and reported consequently. Therefore, answers were mixed up and did not reflect the aim of the question.

On the other hand, item 7 asked students whether they would like to take the course again. Although the purpose of the question was to check if the students liked the course enough so as to include it within the regular classes, many students interpreted the question as if they would take the exact same course another time in the future. Hence, several students answered that they would not take it because they had already taken it, which did not address the real purpose of the question.

The first three items were assessed via a 5-point agreement scale, which ranged from "strongly agree" on the left to "strongly disagree" on the right. Answers to item 1 are illustrated by means of the pie charts in Figure 20. The statement formulated was: "The module has met the expectations regarding an improvement in your oral skills." The pie chart on the left represents the opinions of the control group while the pie chart on the right summarizes the opinions of the experimental group. 11% of the students from the control group strongly agreed with the statement, 79% agreed, 9% were undecided, 1% did not agree, and nobody strongly disagreed. Regarding the experimental students,

10% strongly agreed with the statement, 78% agreed, 11% were undecided, 1% disagreed, and nobody strongly disagreed. Students from both groups, thus, mostly believed in the positive impact of the module in their speaking skills.

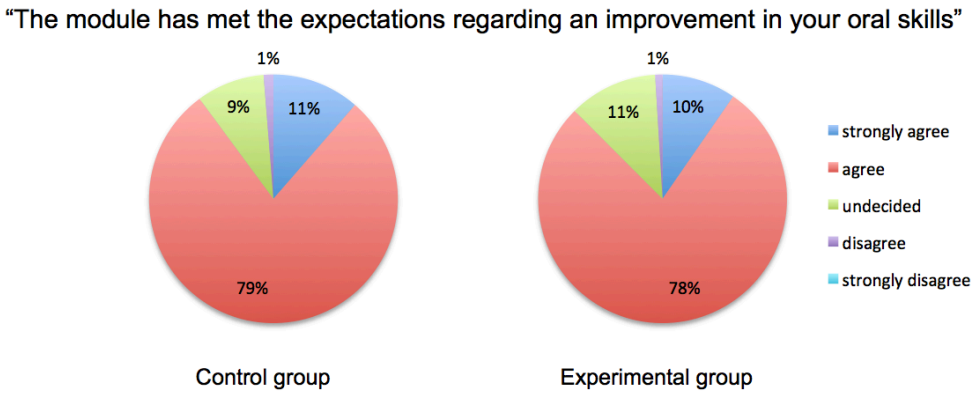


Figure 20. Students' answers to item 1 of the satisfaction survey

Item 2 was aimed to assess the following statement: "The activities proposed have helped you practice your oral skills." Again, two pie charts summarize the answers of the students of each group (see Figure 21). Regarding the control group, 30% of the students strongly agreed with the statement, 62% agreed, and 8% were undecided. Nobody disagreed with this statement to any extent. As for the experimental group, 17% of the subjects strongly agreed, 72% agreed, 10% were undecided, 1% disagreed, and none of them strongly disagreed. Hence, most of the participants found the activities useful regardless of the treatment received.

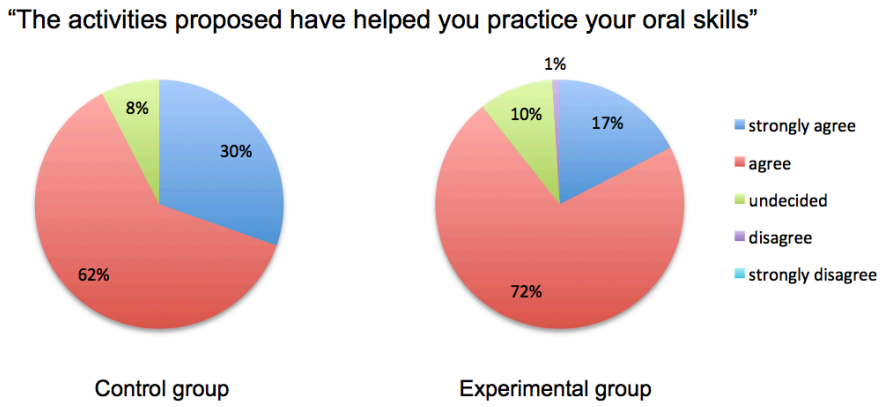


Figure 21. Students' answers to item 2 of the satisfaction survey

Item 3 focused on the explanations. Students had to give their opinion on the following statement: "The explanations provided were useful." Pie charts in Figure 22 illustrate the results. The pie chart on the left shows that 38% of the control students strongly agreed with the statement, 51% agreed, 10% were undecided, nobody disagreed to any extent; and one student did not answer the question. On the other hand, the pie chart on the right illustrates that 42% of the experimental students strongly agreed, 53% agreed, 4% were undecided, 1% disagreed, and nobody strongly disagreed. Therefore, a wide majority of the students of both groups thought that the explanations were beneficial.

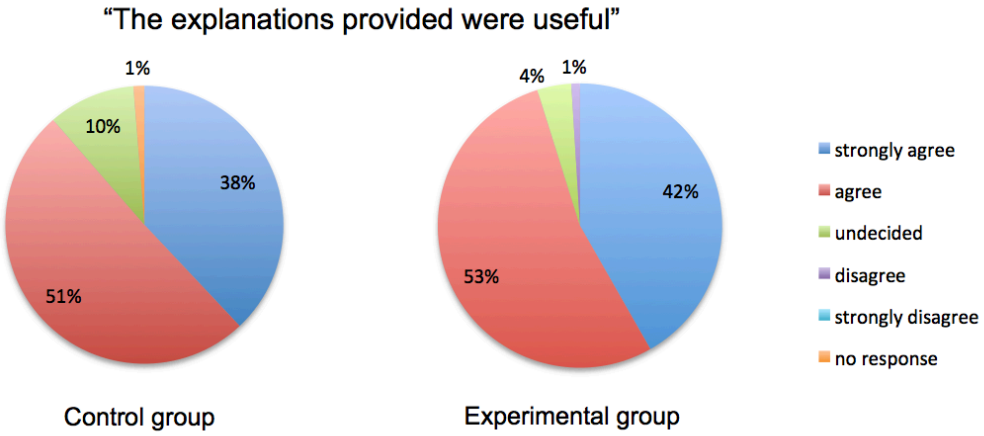


Figure 22. Students' answers to item 3 of the satisfaction survey

Item 4 asked students to self-assess the effort they had invested in the pronunciation sessions. Participants had to answer the question "How much have you participated/worked in class?" by means of a 5-point scale based on percentages, which ranged from 100% on the left to 0% on the right. To avoid ambiguity, percentages were translated into words as follows: 100% = A lot; 75% = Quite a lot; 50% = Average; 25% = Not so much; and 0% = Not at all. Figure 23 shows a summary of the results per group. As observed from the pie chart on the left, 16% of the students from the control group thought to have made the most of the sessions, 32% of them claimed to have worked hard; 42% considered to have worked at half speed, and 10% confessed to have not invested much effort on the sessions. As for the students coming from the experimental group, 8% considered to have made the most of the sessions, 50% of them claimed to have worked hard, 33% considered to have worked at half speed, and 9% confessed to have not invested much effort on the sessions.

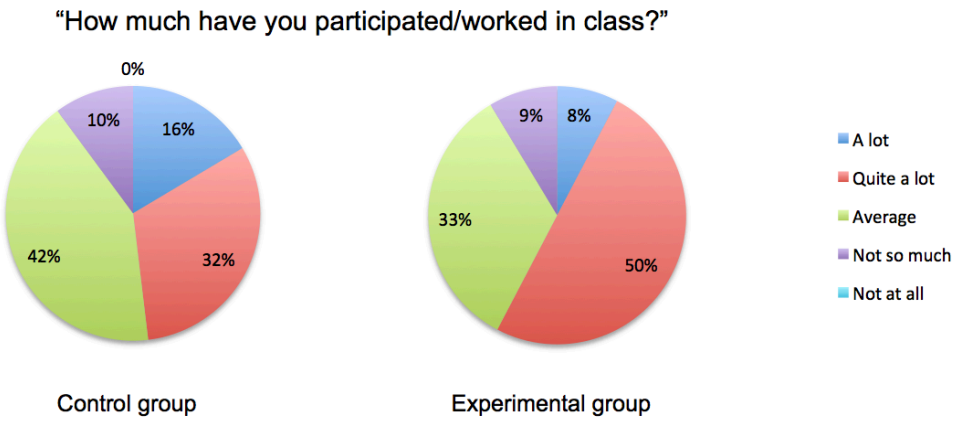


Figure 23. Students' answers to item 4 of the satisfaction survey

Finally, item 6 inquired about the students' opinions on the benefits derived from the pronunciation module. Students were given five different options and they could choose more than one. Also, they were given the possibility to write down any other benefits that they could have noticed, which were assembled under the "other" category.

The bar graphs in Figure 24 summarize the outcome obtained from each group. Results from the control group are presented in the top graph. Of the 79 participating students, 46 students considered they pronounced words more clearly thanks to the pronunciation sessions, 37 thought that they could understand people speaking English better, 29 felt themselves more fluent, 24 pointed out that their classmates could understand them better, and 23 considered that they had reduced their Spanish/Catalan accent. One person did not answer the question.

Five people also claimed having experienced other benefits: One student believed that sessions also helped to learn more vocabulary and grammar; two students pointed at the benefits to speak in public, one of them specifying that the course has helped him/her lose the fear of speaking in front of an audience; another student talked about learning

different types of speech; and finally, one student claimed that he/she could undertand every TV series after taking the course.

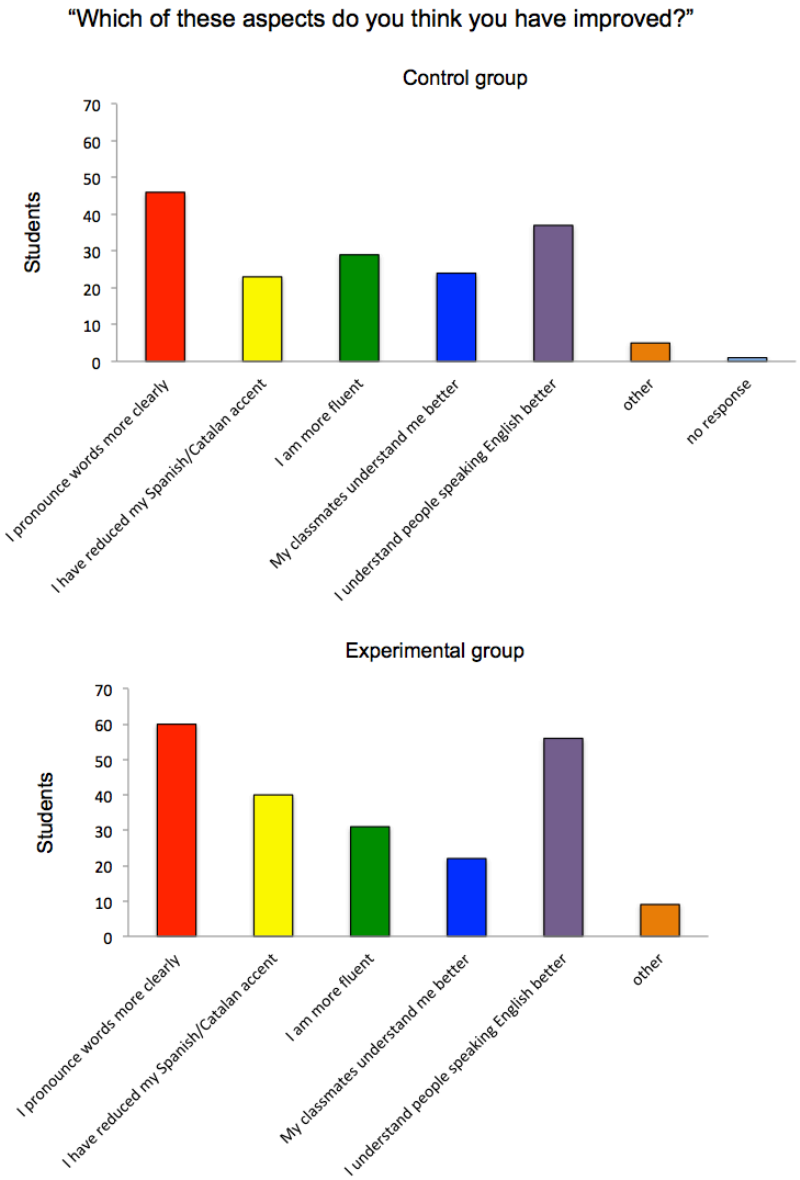


Figure 24. Students' answers to item 6 of the satisfaction survey

The bottom graph displays the results obtained from the experimental group. Of the 104 students who took the questionnaire, 60 thought that the pronunciation module helped them pronounce words more clearly, 56 considered they understood people better, 40

felt that they had reduced their native accent, 31 claimed to be more fluent, and 22 thought their classmates could understand them better.

Eight of them also pointed out other benefits. A couple of students highlighted specific language advantages: One student identified that he/she could pronounce the regular past participle properly and another one that he/she could emphasize words correctly; also two students mentioned an improvement in their English rhythm; one student thought that he/she was more understandable; two students made reference to feeling more confident when speaking; and another student considered that he had obtained some extra tips to speak but he/she did not explain which ones.

To sum up, students' reported an overall positive impression of the pronunciation module and they considered that the sessions provided both production and perception benefits. However, as reported in previous subsections, their improvement is not statistically supported by the previous quantitative analyses in most of the cases, so a qualitative analysis of what happened in class could help understand the factors which could have influenced the results.

4.2. Qualitative results

This section presents the results of the qualitative data, assembled from different sources, which provides a bigger picture on the implementation of the project. Since the experiment was not carried out in a laboratory setting, but was included within regular classes, certain factors that went beyond the researcher's control should be born in mind in order to understand and interpret the results. For this purpose, the personal impressions from both the students and the teacher will be brought under analysis, together with a further look at the sessions themselves. The following subsections will

discuss the comments made by the students in the open questions of the satisfaction survey, the observations written by the teacher in her teaching journal, and the video recordings of the pronunciation sessions.

4.2.1. Students' comments.

Although item 7 of the satisfaction survey could not be included for quantitative analysis due to the fact that several students misinterpreted the question, there were several interesting comments that revealed some of the limitations and some of the gains of the study. The opinions included in this dissertation have been directly extracted from the satisfaction surveys trying to be faithful to the original, so grammar, spelling and punctuation were only corrected when comprehension was seriously jeopardized.

A frequent criticism among those students who provide justified answers to the question "would you like to take the course again?" was the lack of time:

> *G1 student 1:* I would like, but I think it should be longer; we had no time to learn/practice all the things explained during the course.
>
> *G2 student 1*: Pronunciation classes should last longer than 25/30 minutes.
>
> *G2 student 2:* Yes but more than 30 minutes because I think it is more important this than the other things that we do in class.
>
> *G2 student 3:* Yes I find this course so useful and I hope we can have more than just 30 minutes if we have it again.
>
> *G5 student 1:* No because in this pronunciation course we only have ten minutes and you can't improve pronunciation in a few minutes.

These students took the experimental treatment and, thus, rhythm explanations consumed some of the time students had to practice. However, students from the control group also complained about the short duration of the pronunciation sessions:

> *G4 student 1:* No because I think we will learn more English if we do more hours of pronunciation than grammar classes but it is the other way around.
>
> *G4 student 2:* I want to do more hours of pronunciation, not just forty-five/thirty minutes a week.
>
> *G4 student 3*: Yes but dedicating more time to it. It should last longer so we could talk more with our classmates and improve our pronunciation skills.
>
> *G3 student 1*: I would take the course again if it was longer, because in such a short period is difficult to improve.
>
> *G3 student 2:* Yes because I think that the best way to improve the English skills is talking like we did in class.
>
> *G3 student 3:* I think it is important to speak in classroom and in my opinion is necessary to speak more hours because it is the method of learn English.

Another recurring thought in both groups that emerged together with the time issue was the large amount of students per class:

> *G2 student 4:* If it means with fewer people and longer sessions I would take another course.

G4 student 5: Yes because I think that we need to talk more in English and it is difficult in the class. Maybe we are many students, but in my case I do not talk every class in English.

G6 student 1: Too much people in class.

G6 student 2: If the group is more reduced and exercises are different, yes.

There were students who even specified how overcrowded classes affected their improvement and the development of the sessions. Some of them pointed at the need for individualized attention:

G1 student 2: Yes because of the time (it was too short!) and I wish someone could listen to me when I spoke and tell me personally my weak points.

G2 student 4: Yes but I think that in a big classroom is very difficult to teach pronunciation. It's more useful student by student.

G2 student 5: Yes but with less people in class and more personal attention.

Other students related it to attitude:

G5 student 2: Yes but with less people with more interest. I think that the strategy would have been more effective if people would have tried harder, but 2 hours straight at 8 a.m. are hard to stand.

G6 student 3: If some people in the class did not want to pay attention and work, the whole group didn't function correctly.

Nevertheless, students also provided positive criticism. A very important point that some students made is that lessons offered them opportunities to speak and changed their impressions on the subject and English itself:

> *G5 student 3:* I like to have a reason for speak English. That not only English for pass the subject and exam.
>
> *G4 student 6:* Yes because it gives you freedom in speaking.
>
> *G4 student 7:* Yes because we improved things that in a normal lectures doesn't appear. I think that we've learned more in this module than in my whole life.

Besides, students found sessions entertaining:

> *G2 student 6:* Of course, the classes were fun.
>
> *G6 student 4:* It made classes more interesting.
>
> *G6 student 5:* It is a fun part of the course and we also learn English.
>
> *G4 student 8:* Yes, the classes are very dynamic and the teacher tries to help us to lose our fear to speak to other people.

As mentioned in the last comment, some students also thought the sessions helped them feel more confident when speaking:

> *G1 student 3*: Yes because you can see you are not the only one who speaks in a wrong way.
>
> *G3 student 4:* Yes I think it is a good way don't be afraid of talking in English.

The students' opinions on the pronunciation module helped understand the context where sessions took place better and spot some of the limitations faced while implementing the experiment. Some of the concerns mentioned by the students go along the lines of the teacher's observations, which will be discussed in detail in the following subsection.

4.2.2. Teacher's journal.

The teacher took note of any observation she considered interesting to bear in mind when analyzing the students' learning process. After every session with each of the groups, the teacher wrote down any situation that influenced its functioning: whether there was time to carry out all the planned activities, students' attitudes towards the exercises and how they worked on them, how the classroom setting affected practice, and so on. The instructor also took some notes on the recording sessions.

As recurrently mentioned by students, the teacher also pointed out the time restrictions. Despite trying to fit all the sessions in the time given, on certain occasions she claimed that there was not enough time to carry out all the activities planned, especially with the experimental group, whose explanations took longer:

> *Teacher:* First session on rhythm at the sentence level. It took me a bit longer because I wanted all of them to participate individually at least once and also do the extemporaneous activity. I think that for this session the high number of students really harms its efficiency because they do not have enough time to practice, so I do not know to which point it's going to have a good effect on them. [G1, session 6]

> I did not have time to do the extemporaneous activity (i.e. the interview). Actually, I could have lengthened the class 10 more minutes to do it, but that would have been ten extra minutes making the class last longer (40 minutes). Also, I decided not to do that because I saw they were tired and already disconnecting, so I did not think it would have really helped. In this case, I think that controlled and guided practice was much more needed since it is a complicated new concept. [G1, session 8]
>
> After seeing the success of the interview today, I'm sad the experimental group could not have that extemporaneous practice, but with the time given it was impossible (5-10 extra minutes would have been strictly necessary). [G6, session 8]

As highlighted in the first comment, overcrowded classes were one of the factors that influenced the pace of the sessions. However, it was not the only inconvenient. Also, a high number of students implied disruptions and scarce individual feedback:

> *Teacher:* Quite a crowded group, so it was a bit more difficult to keep them quiet and focused. [G5, session 1]
>
> It is extremely difficult to control who is speaking and try to give some feedback with so many people since you cannot listen to them properly. Class arrangement is a huge disadvantage. [G2, session 2]
>
> I think that the way students sit influences feedback as well. In overcrowded classes I hardly reach people who are working in the middle of the class. In not so crowded groups, sometimes it is easier to listen to everybody and provide some

> feedback because of the way they sit, with some space for me to walk around. Sometimes that's not possible because they sit in a such a way that I cannot comfortably walk towards everybody. [G6, session 2]
>
> More than 30 students, so it is more difficult to guarantee that everybody is speaking during the group activity: groups in the front row work really well, but the ones in the back tend to speak in Catalan instead of English. [G5, session 3]
>
> More feedback could be provided but the class was so noisy that sometimes it was difficult to understand what they were saying. [G4, session 3]

Class arrangement was also pointed out as a disadvantage, especially when performing group work. On the one hand, it was difficult for the teacher to access everybody and listen carefully to what they were saying and, on the other hand, students wasted time when rearranging the space to work in groups:

> *Teacher:* It takes ages for them to move and start an activity and classroom arrangement doesn't help. [G4, session 4]

In fact, when fewer students came to class, the sessions tended to be reported more positively:

> *Teacher:* Very few people - class easy to master, they could practice a lot. [G3, session 1]
>
> Less than 15 students and we could work perfectly well. I think that the amount of people in class really influences the class performance. With few students, people

> participated more actively, even those with a low level, and I could provide feedback and even interact with them. [G3, session 4]
>
> It worked really well. I think that the fact that they were fewer people than usual (25 students) helped the class work better. [G5, session 6]

The teacher was convinced that group 3 would improve more than the rest of the groups not only due to the fact that there were fewer students in class, but also because they seemed much more motivated:

> *Teacher:* I'm expecting them to improve more than other groups because:
>
> 1. They're fewer people, so the class is not so disruptive and more feedback can be given.
> 2. They take it seriously and take advantage of the opportunity by speaking English most of the time.
>
> Therefore, I won't be surprised if their results by the end of the course are better than the ones from other control groups and even some experimental groups. [G3, session 3]

Indeed, attitude and motivation were constantly mentioned throughout the journal as a factor that could play an important role in success:

> *Teacher:* Many different types of students that really influence the rhythm of the class. Some of them are lazy and not motivated, others too shy, others do not have the level. In general, they are not taking advantage of the speaking activities so

> much: they tend to speak Catalan or not speak at all. I have caught a student even playing chess and others were doing other things - No attitude. [G6, session 4]
>
> This group is quite lazy and it's very difficult to make them participate actively. They speak a lot but in Catalan. [G4, session 4]
>
> They looked quite skeptical and a lot of them did not seem willing to practice (maybe because they were sleepy or maybe because they don't see the point). I got a bit disappointed and I'm not sure now about their improvement. [G5, session 7]
>
> I'm not expecting to get great results with this group: I think that both a lack of motivation and a passive attitude have negatively affected their learning process. [G5, session 10]
>
> Now the experiment is done and after listening to their interventions during the last session, I think that both groups have improved, but those students who took classes seriously would improve more. In fact, I think that attitude is going to make a difference since I'm just expecting those who worked hard to improve. [Final thought]

The teacher also commented a lot on the mixed-level nature of the groups as a handicap:

> *Teacher:* Very crowded group with a lot of people with a low level. Therefore, speaking activities were difficult to master and many people were not really willing to speak: they tried, but they did not put too much effort into it. [G6, session 1]

> I have also observed that those who have a lower level tend to have more problems to follow the rhythm during the explicit instruction and to apply it in the guided and extemporaneous practice than those who have a higher one. [G1, session 7]
>
> Low-level students relied too much on high-level ones when discussing, so the latter were speaking all the time while the others did not speak as much. [G3, session 9]

Whether it was a question of attitude, level, or any other reason, the teacher also observed that there were students who found it difficult to become actively engaged in the activities:

> *Teacher:* Some are not used to doing speaking activities and they feel intimidated when I walk around to listen to them. [G4, session 1]
>
> I'm starting to think that clapping and tapping, which is the different activity included in the experimental group, is not really understood, makes them feel self-conscious and may handicap their performance in the class. [General thought after session 2]
>
> I have realized that they tend to have serious difficulties when creating stories: The times they have to use their imagination, a lot of them get blocked and there is a higher tendency to start speaking in their mother tongue. [G1, session 3]
>
> It seems they understand the visual aid better than the kinesthetic one (clapping) to distinguish stressed/unstressed syllables; maybe they feel more comfortable

> since it is a passive aid instead of an active one that they're not used to. [G2, session 3]
>
> Some are still shy and stop talking when I get close because they feel observed. [G4, session 8]

However, they adopted an active role when activities were presented as games:

> *Teacher:* Very crowded group. However, they worked well today: maybe the fact that the main exercise was a game helped. [G6, session 5]
>
> The more game-like the activities are, the more active the students are and the more positive their attitude is. [General thought after session 6]

In short, the teacher considered that there were many different aspects that could influence the learning process of the students and, hence, the results. At some point she even wrote:

> *Teacher:* I'm not sure if they will improve in extemporaneous situations since I think that they need more individual practice, but I think that the fact that they understand and are aware of the phenomenon will help them analyze their speech and try to do it at least when reading aloud and do conscious practice [General thought after performing session 6 with experimental groups]

4.2.3. Video recordings.

Every session was video-recorded as a support to follow up the implementation of the experiment. It was not possible to transcribe students' interactions when working in groups due to the high amount of participants. Therefore, transcription was discarded as part of the data to analyze. However, recordings could be used to verify and complement the students and the teacher's observations of the classroom setting, attitude, and other impressions of the groups' performances in general to have a bigger picture of the situation in class. An example of it is Figure 25, which exemplifies the arrangement and size of the classroom. The screenshot, extracted from one of the session with group 6, shows a classroom setting based on traditional rows, which seriously limited both students' interaction and the teacher's observation, especially bearing in mind the high number of students in the group.

Figure 25. Group 6 performing a communicative activity

Both the analyses of quantitative and qualitative data raise some questions on the effectiveness of the experiment. On the one hand, the quantitative analyses of VarcoV values, pauses and students' opinions show some traces of effectiveness of pronunciation instruction and rhythm instruction in particular, but the lack of statistical significance in most of the tests run make findings inconclusive. On the other hand, the qualitative analyses of the students' and the teacher's comments together with a close observation of the sessions thanks to the video recordings display several limitations in the implementation of the experiment, which could have influenced results. The following chapter will attempt to explain, discuss and justify all these issues.

Chapter 5. Discussion

This chapter examines and interprets the results presented in Chapter 4. First, the hypotheses formulated at the beginning of the study will be assessed. Second, further results found during the process will be expanded on so as to provide a more complete picture of the reasons and causes affecting the outcomes.

5.1. Hypotheses testing

The main purpose of this study was to investigate to what extent explicit rhythm instruction could help EFL Spanish/Catalan tertiary level students sound more comprehensible and fluent in English. For this purpose, two different pronunciation groups were created within an ESP course and tested through sentences and opinions that the students recorded before and after the experiment. Both groups received pronunciation instruction based on Celce-Murcia et al.'s (1996) communicative framework, but one of them received explicit training on rhythm during the sessions while the other did not. The expected outcome was that students who were trained on rhythm would show bigger improvement in their prosody, translated into a decrease of the L1 negative transfer and an enhancement of their comprehensibility and fluency, and that explicit pronunciation instruction within a communicative context would also help students develop their intelligibility in its broad sense. With this idea in mind, three different hypotheses were formulated, which will be tested in the following subsections by examining the results obtained.

5.1.1. Hypothesis 1.

The first hypothesis (H1) postulates that those students who receive explicit rhythm instruction will reduce the L1 negative transfer when communicating in the L2 more than does who do not. The veracity of this statement was investigated mainly by comparing VarcoV measurements, and further complemented by counting the amount of pauses committed before and after instruction. The experimental group, hence, was expected to increase its VarcoV values and decrease the number of pauses more than the control group by the end of the experiment. The argumentation to test this hypothesis is going to focus on three different aspects: VarcoV results, pauses results, and the correlation between the two.

5.1.1.1. Negative transfer from the perspective of VarcoV results.

The descriptive analysis of the values compiled for the sentences of both the pre- and the posttests showed that means increased over time in more sentences for the experimental group as compared to the control group. Moreover, when examining the overall performance of the groups, Figures 6 along with the group means presented in Table 6 revealed almost no changes for the control group after instruction while the experimental group increased its VacoV mean (control group: T1 $M = 45.8$, T2 $M = 46$; experimental group: T1 $M = 44.4$, T2 $M = 46.1$). However, the statistical analysis did not show a significant difference between groups for any of the tests performed except from the independent-samples t-test carried out with the effect sizes of the sentences paired-samples t-tests. It is true that several of the effect sizes were small, but that seems to be the case when metrics are involved (Arvaniti, 2012). Even so, the fact that there was a significant difference depending on the group reveals that treatment

influenced students' learning progress, suggesting a better outcome when rhythm was explicitly taught.

These results could be partly compared to the ones obtained in Tsiationi (2011), in which the control groups did not show signs of improvement after instruction while the experimental groups revealed an increase in the rhythmic values even though these were not always statistically significant. However, Tsiartioni used normalized vocPVI and consPVI (vocalic and consonant pairwise variability indices respectively) to measure rhythm and obtained statistical improvement for vocalic variability. Hence, although White and Mattys (2007a, 2007b) considered VarcoV a highly reliable rhythmic metric to examine the acquisition progress of the L2 rhythm when Spanish speakers of English were analyzed, it would be interesting to apply other metrics to see if changes could be corroborated.

Individual values present a different scenario. As displayed in Appendix 5, eleven students in both the control and the experimental groups, i.e., half of the members in each group, increased their personal mean. There are no students who increased all their sentence values, or any sentences in which all the members of the group showed higher values after instruction. Besides, the means per sentence oscillated between 36 and 52 for the control group and between 38 and 56 for the experimental group, which is quite a noticeable difference. In fact, sentence nature is shown to really influence the outcome since it was found significant in the ANOVAs conducted with VarcoV as the dependent variable ($p < .01$).

When further examining native speaker scores for comparative purposes, it is observed that the variability between individual means (M = 49-51) and the group mean (M = 50) was small, but individual values for each of the sentence were, again, volatile. Sentence

means were also remarkably different for native utterances, oscillating from 33 to 64, so the wide range of sentences used definitely influenced the estimations of VarcoV values. Interestingly, despite the fact that in most of the cases the mean of the sentence was higher for the natives compared to the students', it was the other way around in three of the sentences.

Certain utterances caused problems to the students. Learners found it more difficult to read some of the sentences owing to the complexity of their grammar and vocabulary and, consequently, they mispronounced certain words, hesitated more, and repeated some parts to correct themselves, which could have affected rhythm. As for natives, they were aware of being recorded for pronunciation research purposes, so they unconsciously altered the rhythm of some sentences by slowing down or even pausing within thought groups. This seems to be especially the case of the shortest sentences, whose VarcoV means were found smaller than those of the students.

The variability of values is not uncommon in rhythmic metrics research. Previous studies tended to agree that VarcoV values are higher for English than for Spanish and Catalan, but no absolute values have been established. Table 34 summarizes the VarcoV means obtained in those studies that used sentence as stimuli. Values of the same language vary not only depending on the study, but also on the subtype of stimuli used within the same study, and the variation is sometimes quite large. However, with the exception of the means obtained for uncontrolled sentences in Arvaniti (2012), the rest of the results confirmed that English VarcoV is higher than that of Spanish and Catalan, and that Spanish speakers of English show values in between. Thus, the mean values obtained in the present study go along the lines of previous research.

Table 34. VarcoV means obtained in pronunciation research studies using sentence as a stimulus

Study	Stimuli	English	L2 English	Catalan	Spanish
White & Mattys (2007a)	Uncontrolled sentences	64	54		41
Prieto et al. (2012)	CV sentences	52[a]		44[a]	39[a]
	CVC sentences	56[a]		37[a]	38[a]
	Mixed sentences	54[a]		38[a]	36[a]
Arvaniti (2012)	"stress-timed" sentences	48			40
	"syllable-timed" sentences	46			43
	Uncontrolled sentences	50			57
Present study	**Uncontrolled sentences**	**50**	**44/46[b]**		

Note. L2 English=English spoken by Spanish speakers. C = Consonant sound. V = Vowel sound. [a]Approximate values taken from Figure 4 in Prieto et al. (2012). [b]Means of the control and the experimental group both before and after instruction.

Arvaniti (2012) and Ross et al. (2008) concluded that both inter-speaker variability and the chosen corpus affect the measurement of rhythm and hinder the possibility of obtaining statistical significance. Tsiartioni (2011) also remarked inter-speaker variability in her failure to find statistical significance between L1 English and L1 Greek. Despite the fact that the wide disparity of levels reported at the beginning of the course did not seem to have influenced the students' progress (see Figures 7 and 8), individual factors such as motivation, attitude and speaking competence, together with the uncontrolled nature of the sentences, could have affected inter-speaker and sentence variability, further explaining the impossibility of reaching statistical results in the present study.

In short, although VarcoV measurements have not confirmed hypothesis 1 statistically, means trends and the significance of the effect sizes support that rhythm instruction helps acquire L2 rhythm in the EFL classroom to a certain extent, consequently reducing the negative transfer from the L1 into the L2.

5.1.1.2. Negative transfer from the perspective of pausing results.

For both unfilled and the total number of pauses, the descriptive analysis did not reveal a clear difference in the treatment received although the experimental group showed a slightly greater improvement in both cases (see Tables 11 and 14). On the one hand, both the control and the experimental group reduced their sentence means after training and, on the other hand, the overall mean per group was a bit lower for the experimental group after instruction. As for individual values, there are more students from the experimental group whose means decreased after training (see Appendixes 8 and 10). However, as happened with VarcoV values, no students improved in all the sentences and there is no sentence in which all of them improved. In fact, the sentence standard deviations oscillated between one and two in most of the cases, so a wide variability among speakers seems to be confirmed. In addition, the sentence variable in the ANOVAs conducted with each type of pauses was, again, significant ($p < .01$). Hence, a wide inter-speaker and sentence variability could have been, again, detrimental to reach statistical significance.

The time variable was also found significant this time (total pauses: $p < .005$; unfilled pauses: $p < .024$), but the time*sentence interaction was not, so no statistical connection can be established between the two variables. The statistical results are a bit more encouraging when examining the data for just unfilled pauses, though. As suggested in

the previous chapter when comparing Figures 10 and 12 to examine the number of pauses made, the control group progress seems to be thanks to the fact that they made fewer hesitations and repetitions after instruction, which could be the consequence of working on certain words as part of the vocabulary of both the pronunciation module and the regular classes. However, they made approximately the same amount of silent pauses after instruction, which suggests that they kept pausing within thought groups. On the other hand, the experimental group reduced the number of both filled and unfilled pauses to the same extent. Besides, the analysis of the effect sizes of the paired-samples t-tests conducted for the unfilled pauses was significant ($p < .031$), while the one for the total number of pauses was not. These results reveal, hence, that the experimental group recognized thought groups better and avoided pausing in between them more than the control group, producing a more English-like rhythm.

Given the above, the total number of pauses cannot be said to hint at a decrease in the L1 negative transfer whereas unfilled pauses can. These results go along the lines of previous research summarized in Wood (2010), who concluded that:

> The findings of previous research about the importance of pause times and frequencies tells us a great deal about speech fluency, particularly that related to the value of unfilled pauses. It appears that analyzing filled pauses yields mixed and inconclusive results (p. 24)

To sum up, the analysis of the total number of pauses extracted from the sentences does not reassert the first hypothesis while the analysis of the total number of unfilled pauses complements the one of VarcoV measurements and provide more evidence of an improvement when rhythm instruction takes place. However, the study of pauses has been considered as an additional analysis in this investigation and, consequently, further

indicators such as pause location and length should be examined as well to take full advantage of what pausing can tell us about the students' prosody (for a review, see Wood, 2010).

5.1.1.3. Negative transfer: Correlations between VarcoV and pausing.

In general terms, VarcoV values correlated with neither the total number of pauses nor unfilled ones. Correlation tests showed statistically significance in such a specific and limited amount of cases that no relationship could be established regarding time or treatment.

Although some studies relating rhythmic features (e.g. syllable duration) and pausing have been conducted (Adams, 1979; Anderson-Hsieh & Venkatagiri, 1994), the attempt to correlate a rhythmic metric with pausing has not been contemplated as the focus of study in previous classroom-based research. However, it was thought to be worth examining so as to provide evidence of pausing as an important tool to delimit boundaries among thought groups (Patel, 2007) and, thus, keep the rhythm. Moreover, it was believed to agree with those rhythm teaching studies that highlight pausing as a mechanism to help the learners of English create coherent and comprehensible chunks of information (Adams, 1979; Anderson-Hsieh, 1990; Chela-Flores, 1997b, 2003; Gilbert, 2008; Goodwin, 2013; Wong, 1987). However, a lack of statistical significance does not support a correlation between rhythm and pausing in the present study. Further research will be needed to corroborate such a relationship, either by examining another type of stimuli (i.e. full speeches), or analyzing more subjects. Equally interesting would be to examine the correlation between pauses and other rhythmic metrics.

5.1.2. Hypothesis 2.

The second hypothesis (H2) states that those students who are trained in rhythm will be more comprehensible and fluent than does who are not. This hypothesis was mainly tested by means of raters' judgments. However, their potential relationship with the acoustic measures and students' impressions on their improvement will also be brought under analysis. Hence, the argumentation of H2 will analyze comprehensibility and fluency individually through, first, raters' judgments only; second, the correlations between the raters' scores and the different acoustic measurements; and third, the students' responses to the questionnaire.

5.1.2.1. Comprehensibility from the raters' perspective.

According to the statistical analysis performed with all the subjects, the control group was more comprehensible after training whereas the experimental group was less, so explicit rhythm instruction was found to be detrimental for the students' comprehensibility. However, no significance was found for any of the tests performed, thus the hypothesis could not be statistically rejected. Given such an unexpected outcome, there were reasons to believe that there could be flaws in the initial approach. Consequently, further investigation was conducted to try to elucidate the causes.

Although the experimental group was scored slightly higher at the beginning of the course, raters considered that both groups started from an intermediate level (control group T1 M = 3.05; experimental group T1 M = 3.29). When looking at the means of the raters' scores per group after instruction, the control group was rated higher after training while the experimental group was rated lower, showing, again, a negative effect of rhythm instruction. However, the difference between means before and after

treatment was minimal. The control group increased its mean by 0.06 points and the experimental group decreased by 0.02 points. Thus, both groups remained almost stable and, in fact, treatment did not seem to make a real difference in their comprehensibility.

A close look at the students' individual means spots certain incongruities (see Table 17). First, there is a dubious relationship between the English level reported by the placement test and the comprehensibility scores. Generally speaking, beginner students are expected to be less comprehensible while high-intermediate students are expected to be more comprehensible, at least at the beginning of the course. The individual means obtained for this study did not seem to agree with this statement all the time, though.

The individual means of the control students in the pretest performances and their level of English tended to correspond. The mean scores of beginners stood at 2 points, intermediate students' oscillated between 2.5 and 3.7 points (with the exception of student 7 who obtained a score of 1.6), and high-intermediate/advanced students' oscillated between 3.4 and 4.4 points. In contrast, most of the beginners in the experimental group obtained a score of 3 or higher for the pretest, indicating an intermediate/high-intermediate level of spoken English, while student 18, who was reported to have a B2-C1 level at the beginning of the course, did not even reach a score of 3.

Finding variation between the general level of English of students and their oral production might be predictable to some extent since the speaking competence tends not to be as often practiced in class as others, such as reading or writing (Henderson et al., 2012; MacDonald, 2002), and EFL students tend to limit their speaking practice to the classroom setting (see section 2.5.2). However, pretest scores for the experimental group differed considerably from the level distribution of the placement test, suggesting

that the negative effect found after instruction could be the consequence of overrating the performances of these students before instruction.

The mixed ANOVAs carried out with students who scored with a high-level speaking competence at the beginning of the course (i.e. score of 3.5 or higher) and those with a low-level (less than 3.5) respectively shed more light on the issue. On the one hand, the analysis of high-level scores showed a decrease in both groups after instruction (see Figure 15) and found significance for the time variable ($p < .031$). These results suggest that students were rated so high for their pretest intervention that it was unlikely for the listeners to discern an improvement for their posttest recording, increasing the chances of undervaluing their delayed performances. This is especially true for the experimental group, which showed a higher decrease. In other words, the results seem to reveal some sort of ceiling effect. The ceiling effect phenomenon has been considered before in comprehensibility and accentedness research studies such as Derwing et al. (1998) and Southwood and Flege (1999) to bring inconclusive and misleading results.

On the other hand, low-level scores steadily increased after instruction for both groups (see Figure 16), but results did not reveal significance. Therefore, a tendency towards improvement was detected regardless of the treatment received for this set of students, but this could not be statistically confirmed. Interestingly, the experimental group, which was formed by more beginners than intermediate students and had a lower level of English on average, was found more comprehensible at the beginning of the course. These results further support the assumptions made when investigating the individual means that suggested that the pretest performances of those with a lower level were overrated.

As far as this study is concerned, these outcomes cannot be directly compared to previous research because no classroom-based studies specifically testing the effectiveness of rhythm instruction to improve comprehensibility have been conducted. However, this battery of results does not support previous claims of many acquisition and teaching researchers who encourage rhythm instruction in the L2 classroom to enhance the students' communicative skills (Anderson-Hsieh, 1990; Chela-Flores, 1997b, 2003; Derwing & Munro, 2015; Frost & Picavet, 2014; Gilbert, 2008; Goodwin, 2013; Ordin & Polyanskaya, 2015; Quené & Van Delf, 2010; Tajima et al., 1997; Tuan & An, 2010; Wong, 1987).

On the basis of these findings, comprehensibility ratings do not confirm H2. However, the existence of a potential ceiling effect suggests possible inadequacies in the application of the test.

5.1.2.2. Comprehensibility through scores and acoustic measures correlations.

In an attempt to alleviate the potential subjectivity of the judges that could have affected the way they applied the scale (Derwing & Munro, 1995; Isbell, 2018; Thomson, 2018) and provide more scientific objectivity to the outcome (Ghanem & Kang, 2018; Van Moere & Suzuki, 2018), correlations of raters' judgments with VarcoV values and the number of pauses were computed.

Regarding comprehensibility scores and VarcoV values, a moderate correlation was found for the control group after instruction, $r = .474$, $p < .030$, and for the experimental group before instruction, $r = .490$, $p < .024$. Hence, those students from the control group that were considered more comprehensible after instruction showed higher VarcoV values. On the contrary, a significant correlation was shown for the

experimental group before the instruction but not after it, reaching the opposite conclusion. Although a certain degree of correlation is revealed between comprehensibility and VarcoV values, these findings suggest that the introduction of conscious rhythm instruction into the classroom did not help students sound more comprehensible. However, the ceiling effect detected in the ratings could have negatively affected the results.

As for the relationship between comprehensibility scores and pausing, tests did only show significance with a high negative correlation for the control group, both before and after instruction (correlation with total number of pauses T1: r = -.611, p < .003, T2, r = -.527, p < .014; with unfilled pauses T1, r = -.666, p < .001, T2, r = -.547, p < .010). Hence, those control students who were considered more comprehensible were the ones who made fewer pauses at both times. As for the experimental group, no significant correlations were found and, what is more, the correlation coefficient of the posttest was positive for both the total number of pauses and the unfilled pauses, which implies that those who were considered more comprehensible tended to be the ones pausing more. Again, a potential problem with the rating exercise could have affected the outcome.

Although there are researchers who have shown a significant relationship between comprehensibility ratings and rhythmic features such as stress (Kang et al., 2010), no previous research has been found that studies the relationship between a rhythmic metric such as VarcoV and comprehensibility ratings. However, the results can partly relate to Isaacs and Trofimovich's (2012) findings on correlations between comprehensibility ratings and pausing.

Despite the fact that the experimental group scores could have been negatively affected by the ceiling effect, the significant correlations found for the control group when analyzing the relationship between comprehensibility scores and the total number of pauses and unfilled pauses respectively go along the lines of Isaacs and Trofimovich's (2012) results, which revealed a relationship between comprehensibility and the use of pausing. Although the correlation with unfilled pauses is not as high as in the present study, it was found to be significant too in their investigation. They did not study the relationship with the total number of pauses though, but they correlated the scores with filled pauses and the pause error ratio, which were also found significant.

A correlation between pausing and comprehensibility is suggested, but it cannot be related to the instruction of rhythm. Consequently, the results of this study do not support previous studies such as Anderson (1993) as cited in Chela-Flores (1997b) or Hahn (2004), which proved that the correct application of rhythmic aspects such as shorter interstress intervals, fewer stressed syllables, or a correct placement of sentence stress, made students more understandable.

In short, despite the fact that the acoustic analysis glimpsed that VarcoV increased and pauses decreased for the experimental group, correlations with rater scores cannot corroborate an improvement in comprehensibility when that happens. Therefore, a decrease of the negative transfer cannot be anticipated to be more comprehensible when rhythm is explicitly taught and, as a consequence, the correlations of comprehensibility ratings and acoustic measures do not confirm H2.

5.1.2.3. Comprehensibility from the students' perspective.

As shown in Figure 24 the results obtained for the answers to item 6 from the questionnaire revealed that students of both groups did not really think they became more comprehensible. More than half of the students polled believed that they could pronounce words more clearly (58% of the students in both groups), but very few of them considered that their classmates understood them better (control group: 30%; experimental group: 21%). In fact, the latter was the least picked option for the experimental group. Although control students seemed more positive about it, no evidence of an influence of treatment in their comprehensibility could be deduced just from examining their impressions.

Previous research has asked raters about their application of assessment scales (Isaacs & Thomson, 2013; Isaacs & Trofimovich, 2012; Isbell, 2018), but few studies have asked students about their learning process. Kennedy and Trofimovich (2010) did it to correlate the subjects' language awareness with their comprehensibility scores. Despite comprehensibility improvement being non-significant, students' comments revealed that those students who were more comprehensible were the ones who reported a greater degree of language awareness. In this study, although not statistically verified, students' answers do not confirm H2.

5.1.2.4. Fluency from the raters' perspective.

Despite showing an improvement in both groups' fluency after training, the statistical analysis did not show significance for either the t-tests or the repeated-measures ANOVAs performed, initially rejecting H2. As proceeded with the interpretation of comprehensibility, the descriptive analysis and further statistical analyses according to

the speaking competence were investigated so as to detect potential inconsistencies in the process.

Students of both groups were considered to start at an intermediate level on average and improve their fluency after training, even though, again, the difference was almost inexistent (control group T1 $M = 2.81$, T2 $M = 2.86$; experimental group T1 $M = 2.98$, T2 $M = 3.01$). When having a close look at the individual means, it was observed that the experimental beginners were rated quite high for the pretest (see Table 22). Six out of ten of those students obtained a mean score between 3 and 3.9, which is really unusual for students who were reported to have an A1-A2 level. In addition, there are particular cases that rapidly catch the eye. For instance, experimental student 3 obtained a score of 3.4 in the pretest, which dramatically decreased to 1.4 in the posttest. Therefore, the possibility of a ceiling effect was, again, contemplated.

The mixed ANOVA analyzing those students who obtained a 3.5 or higher score for the pretest showed a decrease in fluency for both groups after instruction, with the time variable being significant, $p < .029$. This time, the control group dropped more dramatically. On the contrary, the analysis of the students who scored lower than 3.5 showed an improvement in both groups, slightly bigger for the control group, but no significance was found. Figure 19 shows experimental low-level students score higher than those in the control group when the experimental group is made up of a large number of beginners. Results, then, follow the same directions as the ones obtained in the statistical analysis of comprehensibility, and the existence of a ceiling effect seems to be reasserted.

Again, research mainly focusing on rhythm instruction to improve students' fluency has not been found. However, the findings reported in this subsection go hand in hand with

the ones observed for the comprehensibility ratings in section 5.1.2.1 and, as a consequence, these cannot sustain that rhythm instruction enhances students' communication either. Hence, fluency scores do not confirm H2.

5.1.2.5. Fluency through scores and acoustic measures correlations.

As opposed to comprehensibility, the correlation coefficients for VarcoV values with fluency scores were small and no significance was found. Those students who obtained higher VarcoV were not the ones scored as more fluent and, as a consequence, VarcoV cannot be considered a valid indicator of fluency.

By contrast, the correlation tests carried out with the fluency scores and the number of pauses (total and unfilled respectively) showed the same picture as in the comprehensibility analysis. The tests performed with the control group before and after instruction revealed a significant moderate correlation coefficient (correlation with total number of pauses T1: $r = -.551$, $p < .010$, T2, $r = -.510$, $p < .018$; with unfilled pauses T1, $r = -.611$, $p < .003$, T2, $r = -.528$, $p < .014$), while no significance was found for the experimental group tests, which also showed a positive correlation coefficient after instruction. Again, the existence of a potential ceiling effect seems to affect the correlations. Therefore, the inconsistency of the results obtained for the correlations between fluency ratings and the two types of pauses analyzed seem to point at test performance issues again.

As mentioned before, the complete absence of positive correlations between the ratings and VarcoV discards the latter as a useful indicator of fluency. Hence, the vowel duration variability does not directly affect the perception of whether a person is fluent or not, even though rhythm is thought to ease L2 output production. To the researcher's

knowledge, this type of correlation had never been done before, so it reveals new information related to the link between rhythm and fluency.

Pausing, on the other hand, has long been associated with fluency studies. Riggenbach (1991) found statistical significance for ratings and unfilled pauses when the latter were .5 seconds or greater, but not for short pauses. For this study pauses were counted only for sentences, so most of the pauses were short. Derwing et al. (2004) found the total number of pauses to be a good indicator of fluency, especially when assessing L2 beginners. The lack of significance in the correlation with the experimental group data in this study suggests the opposite, but the fact that low-level students were scored fairly high for the pretest recordings could explain this dissonance.

In a nutshell, the tendency towards a decrease in the negative transfer spotted by the acoustic analyses performed cannot be correlated with an improvement in fluency when rhythm is explicitly taught and, hence, H2 cannot be confirmed.

5.1.2.6. Fluency from the students' perspective.

Regarding the students' impressions on their fluency, very few of them considered they sounded more fluent after taking the course. Only 37% of the control students and 30% of the experimental students claimed to feel more fluent after instruction. Hence, students' opinions also drop a hint about a low improvement regardless of the training received. To the researcher's knowledge, there are no previous studies that have reported students' impressions on their development of fluency when pronunciation, namely rhythm, is trained in class, so results cannot be compared. In short, the students' self-assessment of their own fluency does not support H2 either.

5.1.3. Hypothesis 3.

The third hypothesis (H3) postulates that Celce-Murcia el at.'s (1996) communicative framework improves students' comprehensibility and fluency when communicating in English. The sessions of both treatment groups were based on this teaching approach. Therefore, both groups were expected to improve after training to some extent regardless of rhythm being explicitly taught or not. This hypothesis will be mainly addressed by going back to the quantitative data examined to verify H1 and H2.

The positive tendencies shown by the control group are worth examining in depth despite results not reaching significance in most of the analyses performed. Overall, the raters considered that the control group was more comprehensible and fluent after instruction. Looking at the descriptive analysis, it could be observed that there were twelve students, mainly with a beginner and intermediate level of English, who were scored higher for both comprehensibility and fluency after training, so a bit more than half of the group was found to progress in both skills.

When further investigating their advancement according to their speaking competence, control students who were scored lower for the pretest were the ones to show improvement for the posttest. This also happened for the experimental group, in spite of the fact that the overall group analysis showed a deterioration in the students' comprehensibility. On the contrary, those scored high for the pretest were considered less comprehensible and fluent in the posttest. Putting the negative consequences of the ceiling effect aside, the improvement of low-level students for both groups is a good sign of a relative effectiveness of the teaching method.

The control group also showed signs of improvement when sentences were acoustically measured. As the experimental group, the control group was found to pause less after

instruction, not only when examining the total number of pauses, but also when examining unfilled ones. It is true that the general group mean for the unfilled pauses hardly decreased, but when examining sentence means, the group reduced the number of unfilled pauses made in more than half of the sentences. Similarly, although VarcoV figures showed that the control group remained stable while the experimental group improved when all the sentences were analyzed together, the control group was observed to increase its VarcoV mean values in half of the sentences. Therefore, the generally positive drifts of both the experimental and the control groups point at benefits from the introduction of pronunciation teaching based on Celce-Murcia et al.'s (1996) communicative approach. However, it must be pointed out that the course did not offer any pronunciation training before the experiment was performed. Hence, positive tendencies could be just a consequence of introducing pronunciation instruction into the classroom, regardless of the teaching method applied (see supplementary findings for further discussion).

Finally, the positive feedback towards the sessions also suggests that students welcomed the introduction of this teaching method. For the questions of the satisfaction survey, there was always between 88 and 95% of both control and experimental students who agreed or strongly agreed with the effectiveness of the pronunciation module itself, its activities, and its explanations. However, not many students considered that they were more understandable and fluent after training so, despite finding the module beneficial in different aspects, they did not report clear gains in their comprehensibility and fluency skills.

After examining fifteen classroom-based studies on pronunciation, Saito (2012) concluded that a focus on form within a communicative context is the most effective

type of pronunciation instruction, and highlighted the fact that one of the few studies under analysis which failed to provide statistical evidence, viz. Saalfeld (2011), also encountered ceiling effect issues (as cited in Saito, 2012, p. 846).

Kennedy and Trofimovich's (2010) was one of the studies compiled in Saito's (2012) work whose teaching approach was based on a presentation-practice-production sequence, like the one proposed by the present study. As previously explained, an improvement in students' comprehensibility and fluency ratings was not statistically confirmed, but the learners who scored higher were statistically proven to have a better knowledge of morphosyntax, phonology and lexis. Therefore, high-level students were the ones scoring better, a relation that cannot be established in the present study owing to the ceiling effect, especially when experimental beginners are taken into consideration.

Another study to report the efficacy of Celce-Murcia et al.'s (1996) approach was Gordon et al.'s (2013) study on explicit instruction. Gordon et al's (2013) groups were trained with the same communicative method, but the control group did not receive any type of explicit instruction at the presentation stage while the experimental groups received either explanations on suprasegmentals or segmentals. Hence, the success of the method fell on the explicit explanations rather than on the teaching approach itself. In the present study, the control group benefited from explanations on the same pronunciation phenomena as the experimental group although no explicit training on rhythm was offered. For instance, pausing was not explained through thought groups, but it was explained through punctuation, or the past tense was addressed in a traditional way with no clapping practice to distinguish syllables (see Appendix 2).

As a whole, the lack of statistical significance does not provide robust evidence to confirm H3 even though positive tendencies for both groups are encouraging and support improvement to a certain extent.

5.2. Supplementary findings

The data gathered and analyzed in this study mainly aimed at testing the abovementioned hypotheses. However, side results were also obtained, which provide additional enlightening information linked to the main purpose of the study (i.e. the improvement in comprehensibility and fluency through rhythm instruction). In the following lines, supplementary findings obtained from the quantitative data will be addressed for further justification and understanding. In doing so, explanations will bring to the table different topics that indirectly relate to the central objective of this dissertation. The topics that will be approached in the next subsections are the following: The relationship between comprehensibility and fluency, the introduction of pronunciation training into the EFL curriculum, raters' judgments as an assessment tool in pronunciation teaching research, comparisons among different stimuli, and the benefits from the enhancement of perception to improve production.

5.2.1. The relationship between comprehensibility and fluency.

A strong positive correlation between comprehensibility and fluency was shown to be highly significant ($p < .01$) for all the Pearson coefficient tests performed. These results indicate that raters considered those students who were more comprehensible also to be more fluent at any time during the project implementation. Derwing et al. (2004) also obtained a highly significant correlation between comprehensibility and fluency ratings

not only when assessing monologues, but also when testing picture narratives and conversations. Hence, comprehensibility and fluency must be related, at least, at a perception level.

Finding those pronunciation features that simultaneously influence the way a speech is perceived as both fluent and understandable is key to determine those aspects to teach in class in order to improve the overall intelligibility of L2 students. Anderson (1993) as cited in Chela-Flores (1997b), Derwing et al. (2004), Hahn (2004), Isaacs and Trofimovich (2012), Kang et al., 2010, Lennon (1990), Quené and Van Delf (2010, Riggebanch (1991), Tajima et al., 1997, or Saito et al. (2017), have examined the relationship between specific pronunciation features making use of different instrumental tools to achieve this goal, but no final conclusions can be reached yet. The present study has also investigated how rhythm influences L2 learners' comprehensibility and fluency by means of quantitative measurements, namely VarcoV values and the total number of pauses and unfilled pauses, obtaining mixed results. Further research in this direction could shed more light on which features should and should not be taught to improve L2 students' intelligibility in its broad sense, with the ultimate objective of providing effective teaching guidelines to practitioners.

5.2.2. The introduction of pronunciation training into the EFL curriculum.

As already mentioned when examining Hypothesis 3, both groups showed signs of improvement in most of the tests conducted even though these were not statistically significant. Regarding the experimental group, only comprehensibility ratings revealed a negative effect of the treatment, which, mostly likely, had to do with the existence of a ceiling effect. As for the control group, the generalized tendency of its means towards

improvement both for the listener-based (comprehensibility and fluency ratings) and most of the acoustic analyses (total number of pauses and unfilled pauses) indicates a positive effect of the introduction of pronunciation training into the course. The general descriptive analysis of VacoV values was the only one that did not show changes after instruction for this group, which could be somehow understandable since VarcoV measures rhythm, and rhythm was not explicitly taught in the control sessions. Still, it did show signs of improvement when sentences were examined individually as its VarcoV values increased for five out of ten sentences. That means that non-rhythmic explanations on word stress or sentence focus, among others, did make slight changes in control students' rhythm indirectly.

On the other hand, students reported that the pronunciation module was useful, both in terms of explanations and practice, so they welcomed it as part of the course. In fact, several mentioned that the module offered them a good opportunity to practice speaking. Hence, the introduction of pronunciation instruction into the classroom improved students' speaking skills.

It cannot be concluded with certainty that this improvement is due to suprasegmental teaching or the use of a focus on both teaching and meaning with the data gathered, but results point at a positive effect of explicit pronunciation training within the EFL classroom. This study, thus, joins the list of research studies that support the integration of pronunciation within the second language curriculum (Celce-Murcia et al., 1996; Chela-Flores, 2001; Hahn, 2004; Iruela, 2007; Levis & Grant, 2003; Munro, 2008) despite the limitations faced during its implementation (see section 6.5).

5.2.3. Raters' judgments as an assessment tool in pronunciation teaching research.

As observed when examining Hypotheses 2 and 3, the rating test led to mixed results. The analyses of the scores showed that the overall performance of the experimental group slightly worsened after instruction and that both control and experimental high-level students suffered from a significant deterioration of their comprehensibility and fluency skills over time. These findings not only suggest that the treatment did not work, but also that it negatively affected those learners with a good command of English. The significant decrease in the high scores for both groups reveal the existence of a ceiling effect, which could have seriously affected the general analysis. Even so, a problem with either the test itself or its application is also contemplated, since these results contradicted the general positive tendencies of the rest of the analyses. For this reason, the reliability of the task needs to be reevaluated in depth.

The test suffered from certain limitations that could have influenced the results. As explained in the methodology section, the test consisted in an online form that was sent via email to the raters with a few guidelines on how to perform the task and how to interpret the aspects to assess. The main problem with this type of assessment test has to do with the impossibility of controlling the context in which it is performed: the quality of the audio may not have been appropriate due to the conditions of the equipment, external factors such as family or other duties could have interfered in the performance, and so on. In fact, one of the principal drawbacks of the form was the fact the scores could not be saved until the very end, which prevented the raters from taking breaks and, hence, there was a higher risk of suffering from fatigue.

Munro and Derwing (2015) already warned about the drawbacks of listeners carrying out assessment tests at home. Without a doubt, it would have been much more convenient to bring all the raters together in the same space and make them carry the test out under the same conditions. However, the incompatibility among their schedules impeded it and designing an online form was the only alternative left.

Other researchers in the field (Kang et al. 2010) have used an online test before, but not much about the constraints of the performance was reported. They did specify, though, that they used a trial sample that was not included in the analysis to create a reference that guided the raters' criterion. Other researchers also used checks by embedding native samples (Derwing et al., 1998; Derwing et al., 2004; Isaacs & Thomson, 2013). Trial or native samples to control the listeners' partiality were not included in this study's test, a choice that could explain why some speeches were overrated and, then, become the cause of the ceiling effect phenomenon.

Although the limitations of an online form affected the conditions of the performance, there are further factors that should also be taken into consideration. As mentioned above, fatigue could have played an important role on raters' judgments. The overall timing calculated was of one hour and a half due to the large amount of recordings under assessment and, since the form did not allow saving scores during the process, raters had to listen to all the recordings in a straight run. Munro and Derwing (2015) pointed at the threat that fatigue could pose in rating tasks and recommended that assessment sessions did not take more than an hour, with their corresponding breaks in between. Therefore, it is highly possible that fatigue influenced the scoring (Thomson, 2018). Again, an in-person assessment would have allowed dividing the task into different sessions and, hence, alleviating fatigue (Derwing et al., 1998; Munro &

Derwing, 2015; Saito et al., 2016), but the circumstances of the experiment did not permit it.

Another potential explanation for the ceiling effect would be the choice of raters. Listeners were ESL teachers that had lived in Catalonia, so they were familiar with the Spanish/Catalan language and the most common mistakes the speakers of those language make when speaking English. Although the instructions specified that they should rate as natives not as second language teachers, the familiarity with the language may have eased the degree to which they understood the speech and, as a consequence, their criterion could have become more lenient (Harding, 2018; Isaacs & Thomson, 2013). Southwood and Flege (1999) brought up this possibility when analyzing accentedness of Italian speakers of English. This fact, hence, could also explain why low-level students obtained higher scores than expected.

Finally, the type of scale chosen could also have influenced ratings. For this study, a 5-point rather than the standardized 9-point scale was preferred because it was expected to be user-friendlier in the sense that it simplifies the scope, with less room for doubts and hesitations. Isaacs & Thomson (2013) compared both scales and asked raters about their experience with them. Although some raters found the 5-point scale too restrictive, it was revealed that the 9-point scale offered too many possibilities and some of the levels were not used when rating. However, Southwood and Flege (1999) had already warned about the possibility of a ceiling effect to arise when a short scale like 7-point was performed, recommending the use of longer scales. Therefore, the choice of scale could partly explain the existence of a ceiling effect as well.

The online nature of the form, the type of listeners, fatigue, and the scale chosen, are assessment issues that could have affected the performance of the rating test. For this

reason, if the study had relied just on raters' judgments, results would have completely failed to support an improvement on comprehensibility and fluency, especially when rhythm was explicitly taught. However, further examination of the quantitative data has allowed not only to detect problems with the rating test, but also to discern certain tendencies that support a positive effectiveness of the module.

5.2.4. Comparisons among different stimuli.

It is important to bear in mind the difference in the stimuli used for each test when interpreting results. VarcoV and pauses were examined individually in reading aloud sentences whereas judges listened to extemporaneous speech samples. While rhythm instruction could have had an effect at a controlled level (significantly different effect sizes were detected), it may not have had it when improvising. In fact, students considered they pronounced words more clearly beyond sounding more intelligible or fluent (see Figure 24). The communicative framework guaranteed that students practiced speaking in spontaneous situations but, generally speaking, L2 students tend to stress more when they are asked to improvise (Walker & White, 2013), and focus more on aspects such as grammar rather than their pronunciation. Hence, it is conceivable that the experimental group showed more evident signs of improvement in the controlled than the uncontrolled exercise and, thus, that there was a positive correlation between fluency ratings and pauses after instruction as they might have managed pausing more efficiently when reading than when speaking freely.

On the other hand, the teacher highlighted that explicitly teaching rhythm shortened the time to carry out the communicative exercises (see section 5.3.1). Consequently, the control group could have enjoyed more time than the experimental group to practice

their communicative skills, affecting the outcome of their free samples. However, there were more factors intrinsic to the implementation of the experiment that could have had an effect on the learning process of both groups, which go beyond the time constraints caused by rhythm explanations.

Derwing et al. (1998) assessed both sentences and extemporaneous speech through listeners' ratings, but no acoustic analysis was carried out. On the other hand, other studies that performed a mixed-method assessment used the same stimulus to perform ratings and acoustic measures (Derwing et al., 2004; Isaacs & Trofimovich, 2012; Kang et al., 2010; Lennon, 1990; Riggenbach, 1991; Saito et al., 2017).

All in all, it is believed that the acoustic analysis of the sentences together with the rating of the free speeches provides more evidence of the learning process at its different stages, suggesting some improvement in controlled circumstances.

5.2.5. The benefits from the improvement in perception to improve production.

Although students did not perform perception tasks for their assessment, it is interesting to observe that several students considered they had improved their listening skills (see Figure 24). As part of the communicative method, students had to carry out discrimination tasks every session in order to distinguish the item under practice. Flege (1995) highlighted the importance of perception before production in his theory on error detection. Kennedy and Trofimovich (2010), who found a relationship between good comprehensibility and fluency and greater language awareness, concluded that:

> (…) the strong relationship between language awareness and learners' amount of L2 listening implies that for learners to become aware of L2 speech (and to improve in their L2 pronunciation), they need to reach a certain level of skill in L2 listening. (p. 183)

Hence, the analysis of students' thoughts helps understand that the learning process is positive but it is still ongoing. As Kennedy & Trofimovich (2010) also claimed, "(…) learning to use a language proficiently takes time. (…) Pronunciation of adult L2 learners is particularly resistant to change, even if those learners have received targeted pronunciation instruction" (p. 171).

Up to this point, the exploration of the quantitative results suggests that students may have needed more time of instruction to assimilate what was taught, put it into practice and, finally acquire it. As a result, the learning process could be considered partly successful because there are traces of improvement, but more time would be needed to obtain statistically significant results. In the following section, further research through qualitative data will be brought to analysis so as to have a wider perspective of the development of the study.

5.3. Qualitative findings

Apart from the data gathered by the recordings of the pre- and the posttests, additional qualitative data was also compiled to provide records of the implementation of the experiment. The students' comments on the satisfaction survey, the entries in the teacher's journal, and the videos of the sessions offered more information about how the sessions turned out, and revealed certain limitations of the module progress. There are a series of topics that get repeated over and over again and that depict some of the constraints of the experiment, such as the restrictions of time, class size and arrangement, motivation, and mixed levels. These topics will be discussed in depth in the following subsections.

5.3.1. Time.

Both students and the teacher found time to have a negative effect on instruction. As summarized in the qualitative results section, several students stated that, although useful, pronunciation sessions were too short for them to practice and improve. The teacher seemed to agree with them. Linking this with the quantitative results, it seems clear that more time would have been necessary to find statistically significant improvement.

The teacher pointed at timing management problems when carrying out the sessions, especially for the experimental group, whose explanations tended to be longer and more sophisticated. In fact, she confessed that sometimes the communicative practice had to be shortened or skipped to keep sessions on time. These findings go in line with Levis and Grant (2003)'s opinion about the constraints of the application of Celce-Murcia et al.'s (1996) communicative framework:

> In classes devoted to pronunciation, for instance, teachers apply the framework usually by moving from controlled pronunciation practice to less structured, communicative speaking practice. In this case, however, teachers often spend the majority of time on controlled or guided practice, and give short shift to the communicative end of the pronunciation spectrum. Actual speaking practice is usually unrelated to pronunciation or ignored altogether. (Levis & Grant, 2003, p. 13)

In the present study, the communicative practice of each session was designed thoroughly to practice the item explained during the session, but sometimes the controlled and guided practice took more time than previously scheduled. It must be understood that students were engaged in explanations and practice that were completely new to them, particularly to the experimental group, so they needed time to

understand what was going on and really get involved in the practice. In addition, the high number of students per class did not help maintain the timing for each activity (see following subsection).

Restrictions of time have delimited not only the implementation but also the design of the study. It must not be forgotten that this course was embedded into regular classes, so sessions had to strictly fit into the schedule in order not to interfere with the normal progress of the course. The ESP course in this case lasted 60 hours in total, divided into four hours a week during approximately four months. The sessions, as explained in the methodology section, took place once a week for ten weeks and lasted thirty minutes each. This does not seem much of an intensive training when previous classroom-based research that integrated pronunciation in the regular curriculum is examined. Derwing et al. (1998) performed their experiment in an ESL course of twenty hours a week during twelve weeks and those groups in which pronunciation was taught included instruction of a pronunciation component each day for twenty minutes. Couper (2006) introduced the pronunciation training within a twelve-week semester, but the thirty-minute pronunciation sessions were concentrated in two weeks. Chela de Rodríguez (1981) carried out a ten-week experiment within a teaching program but the duration of neither the sessions nor the regular classes was specified in Chela-Flores' (1997b) work.

The present study sought to truly integrate pronunciation as one more skill of the course, so it had to be connected to the grammar, listening or reading practice in good time. However, the fact that training was not concentrated in a short period of time but rather spread during a whole semester could have also affected the results.

5.3.2. Class size and arrangement.

Although the design of the module took into account the high number of students that every year enroll for the course, reality overcame the expectations. The teacher considered that the number of students could really have a negative effect on their learning process, since it interfered with the pace of the class and caused constant disruption, which the video recordings corroborated. Class arrangement did not help either. Classes were arranged in traditional rows and, as the teacher pointed out, these prevented her from organizing groups efficiently and to move smoothly among the learners in order to listen and provide proper feedback to what they were saying.

Frost and Picavet (2014) highlighted overcrowded groups as one of the common teaching conditions that need to be paid attention to in ESP courses in France, because these do not favor the introduction of pronunciation instruction within the classroom. Henderson et al. (2012), besides, pointed at the problems that large classes can cause to the student-teacher interaction in French universities nowadays. The situation was not different in the present study. There were around forty students per class, except for group 3, where there were twenty. It is true that attrition was one of the main limitations that the study faced (see the limitations section) and, consequently, not everybody came to class everyday, but around two thirds of them did. Very early in the study the teacher could already feel that group 3 would improve more than the rest of the groups, despite being a control group, just because there were fewer people in it and, hence, classes were easier to manage. In fact, students from all the groups but group 3 complained about the inconvenience of belonging to an overcrowded group because they could not practice properly and they could not receive individual feedback, which is also an aspect that the teacher highlighted.

Indeed, Celce-Murcia et al.'s (1996) communicative approach considers the importance of providing feedback in every stage of production (see Table 1). According to their guidelines, most of the feedback provided in this study took the form of self-monitoring and correction through language awareness, peer feedback during group/pair activities, and teacher correction when common mistakes were identified. However, the high number of students in class made individual feedback difficult. When discussing feedback, Celce-Murcia et al. (1996) claimed, "(…) in order for each learner to improve during the period of instruction, the teacher should provide continuous informal feedback on individual progress" (p. 348). The teacher tried to provide individual feedback when possible, but personal attention was impracticable. Unfortunately, students' individual needs could not be fully met in this sense, so a lack of personalized monitoring due to class size could have also affected the results of this study.

5.3.3. Motivation.

An attempt to positively contribute to students' motivation was carried out by engaging them in a wide range of activities that appealed to their different senses (i.e. aural, kinesthetic, and so on) and made sessions dynamic. Students, indeed, expressed their satisfaction with the activities and some of them said that classes were fun. On the other hand, the teacher observed that when activities were presented as games (i.e. *gamification*), students got more involved. These findings support Lighbrown and Spada's (1999) suggestions on how to enhance motivation in the second language classroom. However, these efforts did not seem enough to grab students' full attention and guarantee a significant outcome.

Both the teacher and some of the students reported that there were people who were not willing to participate actively in class and that adopted an idle attitude. This passivity was thought to affect not only their personal learning progress, but also their course mates', since it jeopardized both individual and cooperative work. Some of the potential reasons mentioned in the students and the teacher's comments were the early schedule of some of the sessions, laziness, and self-consciousness. Therefore, there were personal circumstances that influenced students' attitude towards the course no matter the effort to make sessions appealing.

Celce-Murcia et al. (1996) and Lighbrown and Spada (1999) made reference to the importance both integrative and instrumental motivations have in their success. The teacher tried to motivate learners in this sense by contextualizing teaching in real situations and emphasizing the importance of communicating properly in English when working within an international market, but it was not enough. As Derwing and Munro (2015) brings up, students rely on intrinsic factors such as their personality, communicative skills, or self-confidence to interact actively. It could also depend on the situation itself, the participants, or their particular mood. All these factors influence the students' will to communicate. Hence, although the experiment has tried hard to meet motivational needs, there are some aspects that go beyond the teacher's control which could also tell us something about their partial success in the experiment.

5.3.4. Mixed levels.

Sessions were outlined for ESP students and, therefore, the activities proposed were consciously designed to be manageable by students of any level. However, the teacher suggested that this was not the case. According to her, low-level students seemed to

participate less in the exercises, especially in cooperative activities, where they tended to let higher-level students guide the conversation. Although it was thought that mixing levels in group activities would contribute to making high-level students help low-level students communicate better as suggested by many teaching theories of different fields, it seemed that low-level students felt intimidated by their classmates' language competence and they decided to remain in the background and adopt a passive role in this sort of activities. However, the teacher's impressions may be mistaken. An improvement in high-level students, unfortunately, cannot be asserted due to ceiling effect issues, but quantitative results suggest that, generally speaking, low-level students improve both when acoustic and listener-based analyses are taking into consideration.

Most existing classroom-based research has paid attention to students who shared a similar level of language competence because the experiments were carried out in ESL/EFL courses, where a unique level is generally already established. ESP presents a completely different scenario where students of different levels are mingled, but who are also in demand of achieving good communication skills to succeed in their professional field. Frost and Picavet's (2014) project, known as THEMPPO, seems to be the only in-progress project that concerns ESP students improving their prosody. Therefore, more testing needs to be done with this type of students to conclude whether the choice of activities was beneficial to meet their needs.

Chapter 6. Conclusions

This chapter will present and summarize the remarks discussed in the previous sections so as to reach final conclusions on the effectiveness of rhythm instruction as a tool to enhance EFL students' comprehensibility and fluency. The limitations of the study will also be addressed to understand some of the decisions made, together with further recommendations and suggestions both in the research and teaching fields that look forward to what and how to teach and assess pronunciation depending on the context of instruction.

6.1. Rhythm instruction to reduce negative transfer

At first sight, none of the acoustic measures chosen for the assessment of this part of the experiment could statistically prove that there was a greater decrease of students' Spanish/Catalan rhythm interference in their English utterances when rhythm was explicitly taught. In general terms, the results of the statistical analyses run for the VarcoV values, the total number of pauses, or the unfilled pauses did not find significance. However, the study of the effect sizes for the paired-samples t-tests of VarcoV values and the unfilled pauses were significant. Larson-Hall (2010) or Munro and Derwing (2015) advocated for diverting the attention from general global *p*-values, whose relevance could be limited by the sample size, to effect sizes, which specifically show the magnitude of the change regardless of the number of subjects involved:

> Effect size gives researchers the information that they may have thought they were getting from a null hypothesis significance test (NHST), which is whether the difference between groups is important or negligible. An NHST merely tells the researcher and research consumer whether the study had sufficient power to find a difference that was greater, even if

> by a very small amount, than zero. An effect size gives the researcher insight into the size of this difference. (…) Whereas *p*-values and significance testing depend on the power of a test and thus, to a large measure, on group sizes, effect sizes do not change no matter how many participants there are (Maxwell & Delay, 2004, p. 100) [Larson-Hall, 2010, p. 114]

The number of participants that could be used for acoustic analysis decreased dramatically due to attrition. As a consequence, the power of the results inevitably got affected. It is true that the sizes of the effects of the t-tests varied depending on the sentence, but generally speaking these tended to be bigger for the experimental group. Since the difference in performance according to the group was shown significant, an effect of the treatment becomes evident. Therefore, explicit rhythm instruction helped students to increase their VarcoV values and decrease the number of unfilled pauses, which is translated into an approach to adopting an English-like rhythm with fewer unfilled pauses in thought groups, at least, at a controlled sentence level.

Bearing in mind that the teacher insisted on the explanation, discrimination and controlled practice stages of the sessions to guarantee the assimilation of the concept of rhythm in the experimental group, it makes sense that experimental students could apply rhythmic notions at this level better than control ones. Hence, and as suggested in the literature review, bringing metalinguistic awareness up with adults and explicitly training them on pronunciation aspects is effective.

Whether these findings could be extrapolated to the students' global prosody in English remains unknown. Timing and workforce constraints impeded performing acoustic analysis with long speeches. Besides, students had to read aloud a set of given sentences, not to improvise them. However, the sentences under study were different in terms of form and content, simulating different plausible prosodic contexts, so an effect

on their performance could be interpreted as a first step towards a general improvement in their prosody.

In short, the negative transfer caused by L1 rhythm interference is seen to decrease to a certain extent when explicit rhythm instruction takes place, although a wider population and stimuli should be investigated further to reach decisive conclusions.

6.2. Rhythm instruction to improve comprehensibility and fluency

The existence of a ceiling effect in the raters' judgments seriously affected statistical analysis. No significance was found for either of the tests, the difference of the means of both groups was almost inexistent, and, what is more, the experimental group was found to be less comprehensible after treatment. If general scorings had been the only analysis taken into account to assess students' comprehensibility and fluency, results would have led to misleading conclusions of rhythm instruction being disadvantageous. However, further research showed that some students were so highly rated for their pretest, especially for the experimental group, that it made it difficult to detect improvement when comparing it to the posttest.

It is logical to think that students with a high proficiency of the language are unlikely to show improvement, but two things need to be born in mind here: First, none of the students obtained a C2 level in the placement test; Second, the average level of experimental students and their values of VarcoV and unfilled pauses at the beginning of the course were worse than the control group's, so it seems dubious that, on average, they scored higher for the pretest.

Derwing and Munro (2015) claimed that the use of listeners to judge comprehensibility and fluency becomes essential due to the communicative dimension of these aspects.

Unfortunately, the discussion section uncovered some factors that could have influenced the listeners' scoring, like the familiarity listeners had with the language, their notion of what comprehensibility and fluency are, fatigue, or the use of a 5-point scale, which have long been discussed to bias ratings in this type of assessments (Harding, 2018; Isaacs & Thomson, 2013; Isbell, 2008; Munro & Derwing 1995, 2015; Southwood & Flege, 1999; Thomson, 2018).

Previous research mentioned both in the literature review and the discussion section have shown that there are linguistic aspects correlated with degrees of perceptual comprehensibility and fluency, so trying to use them to provide an impersonal perspective seems to be the answer to mitigate weak points of using human agents for assessment. In this study, correlations test with acoustic measures were run and the significant correlation between comprehensibility and the three acoustic measures and fluency with pausing for the control group suggest that there is a relationship between a command of these linguistic aspects and the parameters rated. However, the inconsistency of the correlations for the experimental group alludes to a greater presence of the ceiling effect for this group, which hinders the examination of the real impact of their treatment. All these correlations point to a call to redesigning the rating test, taking into account all the agents that could lead to a ceiling effect.

6.3. Celce-Murcia et al.'s communicative framework to improve comprehensibility and fluency

As explained in the discussion section, the small improvements in both groups in most of the statistical analyses, in spite of the ceiling effect and the lack of significance, suggests that the introduction of the pronunciation module within the course was

beneficial. Whether the improvement was tangibly thanks to the teaching approach applied seems hard to tell from these results only. The course offered no pronunciation training before the experiment was carried out and very few opportunities to practice speaking. Besides, no group was taught with a different approach, so there were no other techniques to compare to.

However, students' impressions seem to indicate that mingling form and content was a positive approach. First, students found that the module gave them the chance to speak in class and even ease their fear to speak in public, which would not have been possible without guided and communicative activities. Second, they also noticed that they could pronounce words more clearly, which should be related to form-focused explanations and controlled practice. Finally, students acknowledged an improvement in their listening skills, probably caused by the discrimination training and the engagement into pair and group work. Besides, none of them seemed to find sessions repetitive or boring, but rather fun and dynamic, probably due to the wide range of activities performed. This reassuring feedback reinforces Saito's (2012) conclusion that the use of the communicative teaching method with a focus on form was preferable to teach pronunciation, and that Celce-Murcia et al.'s (1996) approach can lead to encouraging gains in classroom-based studies, such as in Gordon et al. (2013) or Tsiartioni (2011).

It is true that being more comprehensible or fluent were not the benefits given more credit, though. However, all the little benefits abovementioned reveal progress. Big changes need time, which was scarce in this study. Improvements may not be noticed in a large scale, but they are grasped in a small scale, so the communicative framework applied helped students' improve their comprehensibility and fluency skills at least at the opening stages of the learning progress.

6.4. Supplementary findings

While the main focus of the study was to test the potential of rhythm instruction when teaching pronunciation, several side findings related to some of the main issues in the pronunciation teaching field came to the surface as well. An extensive examination of the quantitative data and the corresponding results has given rise to valuable insights about the purpose and the means to teach and assess pronunciation.

As discussed in the literature review, the focus of the field is moving from making students sound more native to making them be understandable so as to guarantee successful communication. Despite traditionally being related to notions of nativeness, fluency was found significantly correlated with comprehensibility in the present study, so the more fluent students were, the more understandable they sounded, and the other way around. Hence, the two parameters are revealed to go hand in hand, as in Derwing et al. (2004), and to be fit for the purpose of boosting students' global intelligibility in its broader sense.

On the other hand, studies such as Henderson et al. (2012) and MacDonald (2002) unveiled how reluctant teachers tend to be to teach pronunciation in class due to, among others, the restrictions linked to the curriculum. This study has shown that it is possible to introduce pronunciation in class in an efficient way by adapting it to the demands of the curriculum. The module was designed to meet the students' needs of an improvement in their communicative skills, bearing in mind their professional background. In order to do so, sessions were adapted to the context of the course by fostering feasible situations they could face during their career, and favoring the use of the language studied in regular classes. Hence, this module should be an inspiring

example for second language teachers who are willing to turn the tables on pronunciation instruction.

Regarding assessment, a detailed investigation of the use of raters highlighted the limitations of this methodology. Raters are humans and, as such, they are biased, they get tired, and they have different points of view about what being intelligible means. Despite the attempt to control as many of these factors as possible in the design of the test, there are situations that go beyond the researcher's control, like a scarcity of raters. For this reason, relying on more than one assessment tool is advisable. This study has added acoustic analysis and qualitative data to the equation, bringing additional information to the surface that would have been impossible to obtain just by examining the raters' scorings.

Equally important is to test different stimuli. Although it would have been ideal to evaluate just one stimulus with the same assessment tool for comparison reasons (i.e. assessing either sentences or full speeches with both acoustic measurements and raters' judgments), it is also true that bringing different stimuli into the analysis provides more information about the different stages of the learning process. As previously discussed, the analysis of the sentences in this study shows small improvements at an initial controlled level that could have been impossible to detect only by analyzing full speeches. Taking into account that it was the fist time that students were taught pronunciation, it seems logical to find small rather than big gains at an initial stage. Students take little steps to advance in their learning process and each milestone contributes to achieving their final goal. Therefore, benefits at the most basic level make cautious but definite progress on students' communicative competence.

6.5. Qualitative findings

Qualitative results reveal the intrinsic constraints of the implementation of the module and the teaching *per se*. Limitations of time, group size and class arrangement were detrimental to the development of the sessions. Students complained about the scarcity of time to practice and the lack of individual attention due to the large amount of students per class. The teacher, besides, had the impression of rushing through the different stages so as to stick to the teaching plan, and that she could not reach everybody's needs because classes were overcrowded.

In addition to that, there were personal differences among the students that also played a role in the development of the module. The teacher realized that some learners were not fully motivated: they were not active in class, they avoided speaking in English when possible, and they just seemed tired. Even so, most of the students claimed that sessions were interesting and fun. Both the teacher and the students' comments provide potential explanations: some sessions were early in the morning, so they felt sleepy; some students had never been asked to speak English in class, so they felt self-conscious or anxious about it and they preferred to keep quiet; and, as engineers-to-be, English was not the students' main worry at that time, so they may have felt it was not worth the effort. These, together with a mixture of English levels typical of ESP courses, are problems that cannot be avoided and undoubtedly affect the learning process. Indeed, these add up to the fears that compel teachers to avoid pronunciation practice in class.

6.6. Limitations of the study

Among all the limitations already mentioned, time, or better said, the scarcity of it, has been found to be the elephant in the room in many different ways. It affected teaching,

since it restricted the number of sessions, the sessions were found short, and some of the activities could not be performed or devoted as much time as wished; but it did also affect assessment, since the rating test was found long and exhausting.

Absenteeism also compromised the study. As explained in the methodology section, the experiment started with almost three hundred students, of which only forty-two could be quantitatively analyzed. That means that approximately 86% of the potential subjects were lost along the way. From the first pool of students, 72 of them dropped out the course straight away and 52 of them did not record themselves, but there were 174 remaining students who did participate in the experiment. Of those, 83% attended at least half of the sessions (see Table 4), so more than half of the learners were going to class regularly. However, only 40% of them completed the treatment. Attrition is difficult to control at a tertiary level, since students get to decide how often they go to class. Recordings were integrated into the assessment of the subject and leaners had to attend at least half of the regular classes to be continuously assessed, but that was not enough to guarantee full attendance, as it happens in any other university course.

The search for raters was also a complicated process. In a country where the target language is not daily spoken, it is difficult to find native speakers, especially in a small town such as Tarragona, but it is even more difficult to find native speakers that would like to participate in a study as such.

6.7. Implications for further research

Far from remaining a one-time reflection, this study offers a wide range of opportunities for future research. First of all, the study could be replicated by trying to fix the different shortcomings pointed at. The rating test could be retaken in person, more raters

could be recruited thanks to online tools such as *Mechanical Turk* (Nagle, 2019), and the scale could be adjusted to the field standards to see if that makes any difference in the result. It would also be advisable to design a placement test that includes a speaking exercise to control the level variable. In addition, more students could be analyzed despite having attended fewer than nine sessions; maybe those who attended at least eight or seven also present signs of improvement and would increase the pool of students to analyze. With the help of the new technologies, especially speech recognition software, sentences may also be rated and long running speeches could be acoustically analyzed to carry out solid comparisons at every prosodic level. Last, but not least, other rhythmic metrics could be computed and more information about pausing, such as length and location, could be taken into consideration as well to further test the relevance of language rhythm in pronunciation teaching.

Secondly, the study could be readapted in different language learning circumstances. The experiment here presented was carried out with ESP students, which present a very particular language learning context, with clear professional goals in mind and important restrictions of content and language level. Perhaps the study could lead to interesting outcomes if performed in other sort of EFL and even ESL courses, which would widen the scope to test effectiveness of rhythm.

Finally, the same teaching approach could be followed to introduce other pronunciation aspects into the classroom, or the same aspect could be taught by following a different teaching methodology. Indeed, any attempts to perform more classroom-based studies to reach decisive conclusions on what and how to teach pronunciation and, thus, bridge the gap between research and practice will be welcome contributions.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Adams, C. (1979). *English speech rhythm and the foreign learner*. Haage: Mouton.

American Society of Civil Engineer (ASCE). (2014, December 1). *Maria Gibbs: 2015 New Faces of Civil Engineering Professional* [YouTube]. Retrieved from https://www.youtube.com/watch?v=KwTW32TZ7-c

Anderson-Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field-specific materials. *English for Specific Purposes*, *9*, 195–214. https://doi.org/10.1016/0889-4906(90)90013-3

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of chinese ESL speakers. *TESOL Quarterly*, *28*(4), 807–812.

Aquascript. (2017). Line Graphs. Retrieved from https://aquascript.com/eap/descriptive-reports/line-graphs/

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*, 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Audacity [Computer software]. (2018). Retrieved from https://www.audacityteam.org/download/

Baker, A., & Murphy, J. M. (2011). Knowledge base of pronunciation teaching: Staking out the territory. *TESL Canada Journal*, *28*(2), 29–50. Retrieved from http://teslcanadajournal.ca/index.php/tesl/article/view/1071

Beisbier, B. (1995). *Sounds great: Intermediate pronunciation and speaking for learners of English*. Boston: Heinle, Cengage Learning.

Boersma, P., & Weenik, D. (2015). PRAAT: Doing phonetics by computer [Computer

software]. Retrieved from www.praat.org

Busà, M. G. (2012). The role of prosody in pronunciation teaching: A growing appreciation. In M. G. Busà & A. Stella (Eds.), *Methodological perspectives on second language prosody* (pp. 101–105). Padova: CLEUP.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambrige University Press.

Chela-Flores, B. (1997a). Fossilized pronunciation and the rhythm of English. *Lingua Americana*, *1*, 51–68.

Chela-Flores, B. (1997b). Rhythmic patterns as basic units in pronunciation teaching. *Onomázein*, *2*, 111–134.

Chela-Flores, B. (2001). Pronunciation and language learning : An integrative approach. *IRAL*, *39*, 85–101.

Chela-Flores, B. (2003). Optimizing the teaching of English suprasegmentals to Spanish speakers. *Lenguas Modernas*, *28–29*, 255–273.

Chela de Rodríguez, B. (1976). Teaching Engish suprasegmentals to Spanish speakers. *Kansas Working Papers in Linguistics*, *1*, 63–82.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*(1), 46–66.

Darvill, A. (2017). Energy resources: Wind power. Retrieved from http://www.darvill.clara.net/altenerg/wind.htm

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62.

Dauer, R. M. (1987). *Phonetic and phonological components of language rhythm*. In T.

Gamkrelidze (Ed.), Paper presented at the 11th International Congress of Phonetic Sciences (ICPhS) (pp. 447–450). Tallinn: Academy of Sciences of the Estonian S.S.R.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *IRAL*, *4*(3), 183–198.

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347–370). Amsterdam/Philadelphia: John Benjamins.

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*. Amsterdam/Philadelphia: John Benjamins.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for "fossilized" learners. Can it help?. *Applied Language Learning*, *8*(2), 217–235.

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. https://doi.org/10.1111/0023-8333.00047

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*(4), 655–679.

Dudley-Evans, T., & Jo St John, M. (1998). *Developments in English for specific purposes. A multi-disciplinary approach*. Cambridge: Cambridge University Press.

Earle-Carlin, S. (2011). *Q: Skills for success 5. Listening and speaking*. Oxford/New

York: Oxford University Press.

ElementalEnglish (2013, March 6). *Connected speech & linking/American English pronunciation* [YouTube]. Retrieved from https://www.youtube.com/watch?v=gAHUTKm_1n0

Feist, D. (2018). *El español es flamenco. El inglés es rock'n roll: Lo que nadie te ha enseñado del inglés y cómo puede mejorar tu comprensión y fluidez*. Kindle.

Flege, J. E. (1987). The production of " new " and " similar " phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. Retrieved from https://pdfs.semanticscholar.org/4a01/a0def4d4016c6fec441b3f2005087e0041e2.pdf

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.

Franche-Comté, U. (2004). English Online France. Retrieved from http://cla-tice2.univ-fcomte.fr/eolf/wp-content/uploads/pronunciation/02_-ed_endings/02_ed.htm

Frost, D., & Picavet, F. (2014). Putting prosody first - Some practical solutions to a perennial problem: The Innovalangues project. *Research in Language*, *12*(2), 1–11. https://doi.org/10.2478/rela-2014-0002

Ghanem, R., & Kang, O. (2018). Pronunciation features in rating criteria. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 115–136). London/New York: Routledge.

Gilbert, J. B. (2005a). *Clear speech: Pronunciation and listening comprehension in North American English. Student's book*. Cambridge: Cambrige University Press.

Gilbert, J. B. (2005b). *Clear speech: Pronunciation and listening comprehension in North American English. Teacher's resource book*. Cambridge: Cambrige University Press.

Gilbert, J. B. (2008). *Teaching pronunciation: Using the prosody pyramid*. New York: Cambrige University Press. Retrieved from www.cambridge.org

Gordon, J., Darcy, I., & Ewert, D. (2013). *Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom*. In J. M. Levis & K. Le Velle (Eds.), Paper presented at the 4th Pronunciation in Second Language Learning and Teaching Conference (pp. 194–206). Ames, IA: Iowa State University.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515–546). Berlin: Mouton de Gruyter. https://doi.org/10.1016/S0022-3476(75)80137-5

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201–223.

Harding, L. (2018). Validity in pronunciation assessment. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 30–48). London/New York: Routledge.

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). The English pronunciation teaching in Europe survey: Selected results. *Research in Language*, *10*(1), 5–27. https://doi.org/10.2478/v10015-011-0047-4

Intelligibility. (2019). In *WordReference.com.* Retrieved from https://www.wordreference.com/definition/intelligibility

Iruela, A. (2007). Principios didácticos para la enseñanza de la pronunciación en lenguas extranjeras. *MarcoELE. Revista de Didáctica*, *4*, 1–16.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*, 135–159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*, 475–505. https://doi.org/10.1017/S0272263112000150

Isbell, D. R. (2018). Assessing pronunciation for research purposes with listener-based numerical scales. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 89–112). London/New York: Routledge.

Kang, O., & Ginther, A (Eds.). (2018). *Assessment in second language pronunciation*. London/New York: Routledge.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness*, *19*(3), 171–185. https://doi.org/10.1080/09658416.2010.486439

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253–263.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377. https://doi.org/10.2307/3588485

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge: Cambrige University Press.

Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *Tesol Journal*, *12*(2), 13–19.

Lightbown, P. M., & Spada, N. (1999). *How languages are learned*. Oxford: Oxford University Press.

Macdonald, S. (2002). Pronunciation – Views and practices of reluctant teachers. *Prospect*, *17*(3), 3–18.

Midwestgeneration (2011, september 6). *How a power plant works* [YouTube]. Retrieved from https://www.youtube.com/watch?v=GI7AhajfhWE

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–2018). Amsterdam/Philadelphia: John Benjamins.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility , and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97.

Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, *1*(1), 11–42. https://doi.org/10.1075/jslp.1.1.01mun

Nagle, C. (2019). Developing and validating a methodology for crowdsourcing L2

speech ratings in Amazon Mechanical Turk. *Journal of Second Language Pronunciation*, 1–30. https://doi.org/https://doi.org/10.1075/jslp.18016.nag

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. Van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1147–1159). Wiley-Blackwell. https://doi.org/10.1002/9781444335262

Nerdwriter1 (2015, December 31). *How Donald Trump answers a question* [YouTube]. Retrieved from https://www.youtube.com/watch?v=_aFo_BV-UzI

Newsy (2015, October 7). *Is Smart Technology Making Us Dumb?* [YouTube]. Retrieved from https://www.youtube.com/watch?v=0uKxWcHU1D4

Olive, J. P., Greenwod, A., & Coleman, J. (1993). *Acoustics of American English speech: A dynamic approach*. New York: Springer-Verlag.

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*, 244–257. https://doi.org/10.1016/j.system.2013.12.004

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533–544. https://doi.org/10.1121/1.4923359

Patel, A. D. (2007). *Music, language, and the brain*. New York: Oxford University Press USA.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: The University of Michigan Press.

Prieto, Pilar, Vanrell, M. D. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic

and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*, 681–702. https://doi.org/10.1016/j.specom.2011.12.001

Quené, H., & Van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication*, *52*, 911–918. https://doi.org/10.1016/j.specom.2010.03.005

Ramírez Verdugo, D. (2006). A Study of intonation awareness and learning in non-native speakers of English. *Language Awareness*, *15*(3), 141–159. https://doi.org/10.2167/la404.0

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441. https://doi.org/10.1080/01638539109544795

Ross, T., Ferjan, N., & Arvaniti, A. (2008, March). *On quantifying speech rhythm.* Poster presented at the Workshop on Empirical Approaches to Speech Rhythm, UCL, London.

Rueda, C. (2017). *Technical English: Course workbook.* Tarragona: Universitat Rovira i Virgili.

Saito, K. (2012). Effects of instruction on L2 pronunciation experimental intervention studies. *TESOL Quarterly*, *46*(4), 842–854. https://doi.org/10.1002/tesq.67

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*, 217-240.

https://doi.org/10.1017/S0142716414000502 Second

Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2017). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment* (p. 209). Bristol: Multilingual Matters. https://doi.org/10.21832/ISAACS6848

Sardegna, V. G., & Mcgregor, A. (2013). *Scaffolding students' self-regulated efforts for effective pronunciation practice*. In J. M. Levis & K. Levelle (Eds.), Paper presented at the 4th Pronunciation in Second Language Learning and Teaching Conference (pp. 182–193). Ames, IA: Iowa State University.

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*, *13*(5), 335–349.

Stevens, S. G. (1989). A "Dramatic" approach to improving the intelligibility of ITAs. *English for Specific Purposes*, *8*, 181–194.

Sweet, H. (1900). *The practical study of languages: A guide for teachers and learners*. New York: H. Holt and Company.

Standford (2008, March 8). *Steve Jobs' 2005 Stanford Commencement Address* [YouTube]. Retrieved from https://www.youtube.com/watch?v=UF8uR6Z6KLc

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*, 1–24. https://doi.org/10.1006/jpho.1996.0031

The Beatles (1967). With a little help of my friends. On *Sgt. pepper's lonely hearts club band* [CD]. United Kingdom: EMI.

The DISCLAIMER.io. (2017). *Trump vs. Hillary: Debate Rap Battle* [Dailymotion]. Retrieved from https://www.dailymotion.com/video/x5ltag2

Thomson, R. I. (2018). Measument of accentedness, intelligibility, and comprehensibility. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). Abingdon/New York: Routledge.

Tsiartsioni, E. (2011). *Can pronunciation be taught? Teaching English speech rhythm to Greek students*. In E. Kitis, N. Lavidas, N. Tpointzi, & T. Tsangalidis (Eds.), Paper presented at the 19th International Symposium on Theoretical and Applied Linguistics (ISTAL) (pp. 447–458). Thessaloniki: Aristotle University of Thessaloniki.

Tuan, L. T., & An, V. T. P. (2010). Teaching English rhythm by using songs. *Studies in Literature and Language*, *1*(2), 13–29.

Van Moere, A., & Suzuki, M. (2018). Using speech processing technology in assessing pronunciation. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 137–152). London/New York: Routledge.

Walker, A., & White, G. (2013). *Technology enhanced language learning: Connecting theory and practice*. Oxford, UK: Oxford University Press.

White, L., & Mattys, S. L. (2007a). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*, 501–522. https://doi.org/10.1016/j.wocn.2007.02.003

White, L., & Mattys, S. L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaro, & M.-J. Sole (Eds.), *Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series)* (pp. 237–257). Amsterdam/Philadelphia: John Benjamins.

Wilkinson, R. (2008). Locating the ESP space in problem-based learning: English-medium degree programmes from a post-Bologna perspective. In I. Fortanet-Gómez & C. A. Räisänen (Eds.), *ESP in European higher education* (pp. 55–73). Amsterdam/Philadelphia: John Benjamins.

Wong, R. (1987). *Teaching pronunciation: Focus on English rhythm and intonation*. Englewood Cliffs, New Jersey: Regents Prentice Hall.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. London: Continuum.

Appendices

Appendix 1: The research consent form

Research Consent Form

I freely and voluntarily consent to be a participant in this research project conducted in the Spring semester of 2017 by researcher Leticia Quesada Vázquez, PhD student at Rovira i Virgili University. I understand I will be one of the people participating in this research and that my participation will form part of the Technical English course assessment. I have been told this study will be divided into 10 weekly sessions of approximately 30 minutes each, which will take place within the class schedule.

I understand that the purpose of this research is to investigate the use of communicative instruction strategies as discovery procedures in the English as a foreign language (EFL) classroom in order to enhance EFL students' speaking skills. To this aim, I will be asked to attend and participate actively in all the sessions, which will be directly connected to the course syllabus, and to take a test at the beginning and the end of the course, whose completion will count as one of the tasks included in the subject assessment. The tests will be audio-recorded and the final oral presentations and sessions could be video-recorded.

I understand that all my contributions to the project will be confidential (in the sense that my name will not appear in any public records or publications) and that only the researcher conducting the study and the supervisor/PhD dissertation committee members will have access to these data. The data will be used over the next three years although they will be retained indefinitely as records.

I have been told that I am free to ask questions concerning the procedure. I understand that if I would like more information about this research, I can reach Leticia Quesada Vázquez either at the "Professors campus extern" office on campus Sescelades, or at room 2.20 on campus Catalunya (Rovira i Virgili University). I can also contact her via email at leticia.quesada@urv.cat.

UNIVERSITAT ROVIRA i VIRGILI

I have read and I understand the informed consent form attached/He llegit i entès l'autorització adjunta/He leído y entiendo la autorización adjunta:

Name/Nom/Nombre	Signature/Signatura/Firma	Group and date/Grup i data/Grupo y fecha

Appendix 2: The outline of the pronunciation sessions

Session 1	*Experimental*	*Control*
Introduction	- Start the session asking students how they found the pretest (e.g. which parts were more difficult for them). - Ask for a volunteer student who reads aloud the transcript[a] of the first part of the video *Maria Gibbs: 2015 New Faces of Civil Engineering Professional* [American Society of Civil Engineer (ASCE), 2014, 0:00] in front of the class before playing it. - Play that part and discuss the differences between the students' and the natives' performance with the whole class. Make them think about why native speakers could have problems to understand them when they speak in English and which problems they face when they listen to native speakers. Try to guide the conversation towards features affecting intelligibility.	
Description and analysis	- Provide an overview on language rhythm following the PowerPoint explanations: Tell about the differences between Spanish (syllable-timed) and English (stress-timed) using Lloyd James' (1940) comparison of a machinegun shot and a Morse code message (as cited in Ramus et al., 1999), and the difference in length between musical notes.	
Listening discrimination	- Show the transcript of the video again with content words highlighted, and make them listen to the native speaker again.	- Play four clips of engineers speaking in English without showing the image. Ask students the speakers' nationality, if they can understand them, and if they think natives will understand them. - Play the videos again, this time showing the image. Discuss the answers provided and ask them about any surprises about the nationalities of the speakers.
Controlled practice	- Ask students to practice in pairs by reading aloud the transcript to their partners without the content words highlighted. Encourage them to give feedback to their partners while you walk around the classroom listening to them and trying to give feedback to as many students as you can.	- Play the first video again with subtitles and ask them to shadow the speech. - Ask them to practice in pairs by reading aloud the transcript to their partners. Encourage them to give feedback to their partners while you walk around the classroom listening to them and trying to give feedback to as many students as you can.
Communicative practice	- Arrange groups for discussion: Students have to talk about themselves specifying which brand of engineering they are studying and why.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 2	*Experimental*	*Control*
Description and analysis	- Warm up by remembering what *TA* and *ti* stand for. Make your students aware of the phonological nature of the syllable: We distinguish syllables by the way they are pronounced, not the way they are written. Make students clap to the syllables of the examples shown on the PowerPoint presentation so that they get familiar with this phenomenon. - Ask them if they know the different pronunciations of the regular past tense ending (*-ed*) in English. Make them try to pronounce the examples provided and show the answers. - Explain to students that those verbs whose infinitive ends in a /t/ or a /d/ sound add an extra syllable to the word while the other verbs do not. Help your explanation by clapping to the examples.	- Make your students aware of the phonological nature of the syllable: We distinguish syllables by the way they are pronounced, not the way they are written. - Ask them if they know the different pronunciations of the regular past tense ending (*-ed*) in English. Make them try to pronounce the examples provided and show the answers. - Explain to students that the *-ed* ending is pronounced when the infinitive of the verb ends in a /t/ or a /d/ sound, so they add an extra syllable.
Listening discrimination	- Ask students to decide whether the verbs shown on the screen change their number of syllables when pronounced in the past. Tell them to write their answers down on the table[a]. - Play the recordings available on the pronunciation section of the *English online France* website (Université de Franche Comté, 2014) and correct. - Play the recordings of Gilbert's (2005b) exercise B of *Unit 1 Quiz*: Students have to decide whether the sentences pronounced are in the present or in the past (p. 83). Then, correct.	
Controlled practice	- Make students repeat the verbs in the exercises from the listening discrimination section aloud after each activity.	
Guided practice	- (Pair work) Do the *Present or past?* exercise[a]: Students have to read the sentences aloud either in present or past to their partner and the partner needs to guess which version has been used.	
Communicative practice	- Do the *Enjuto Mojamuto: Stories in the past* exercise[a]: Students have to explain the story of the Spanish TV character Enjuto Mojamuto, a computer science freak, by looking at the comic strips. Pictures are screenshots extracted from different videos on Enjuto's stories available on YouTube. - If there is some extra time, students explain their own story.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 3	*Experimental*	*Control*
Description and analysis	- Based on the explanations of the previous sessions, students are aware now of the difference between spelling and pronunciation in English. Bearing this in mind, students have to decide which letters are not pronounced from the words given, which have been extracted from Gilbert's (2005a) exercise I (p. 6). The words are already classified according to the number of syllables they have, so students just have to cross out the *silent* letter. Also, ask them to draw a dot for unstressed syllables and a line for stressed syllables underneath the words to show the difference in length[a]. - Ask students if they have found any difficulties, if there were words that they considered not to have silent letters and, hence, to be wrongly classified. - Show on the screen the correction of the exercise with both the silent letters crossed out and the combination of dots and lines to exemplify stress.	- Based on the explanations of the previous sessions, students are aware now of the difference between spelling and pronunciation in English. Bearing this in mind, students have to decide which letters are not pronounced from the words given, which have been extracted from Gilbert's (2005a) exercise I (p. 6). The words are already classified according to the number of syllables they have, so students just have to cross out the *silent* letter[a]. - Ask students if they have found any difficulties, if there were words that they considered not to have silent letters and, hence, to be wrongly classified. - Show on the screen the correction of the exercise with the silent letters crossed out.
Listening discrimination	- Read words extracted from the course book for students to classify these according to their number of syllables in the table provided[a]. Show the answer with the stress makers on the screen. - Read five sentences containing words from the previous exercises. Students have to write them down and count the syllables of each sentence[a]. The exercise has been adapted from Gilbert's (2005a) exercise J (p. 6). Show the answers.	- Read words extracted from the course book for students to classify these according to their number of syllables in the table provided[a]. Show the answer on the screen. - Read five sentences containing words from the previous exercises. Students have to write them down and count the syllables of each sentence[a]. The exercise has been adapted from Gilbert's (2005a) exercise J (p. 6). Show the answers.
Controlled practice	- Ask students to repeat the words and sentences of the previous exercises aloud bearing in mind both the number of syllables and the stress markers.	- Ask students to repeat the words and sentences of the previous exercises aloud taking into account the number of syllables.
Guided/ Communicative practice	- Ask students to form groups of 4/5 people. Students create a story by using the words in the second table. Each student says one sentence with the first word and the next student has to continue the story by saying another sentence containing the next word, and so on.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 4	*Experimental*	*Control*
Description and analysis	- Ask students if they remember the difference between *TA* and *ti* regarding language rhythm. Introduce the *schwa* sound to shorten unstressed syllables and refer to its use in Catalan. Play a word sample in Spanish, Catalan and English to compare. - Substitute *TA* and *ti* for a long *LA* and a short *la* respectively to create abstract rhythmic patterns. Read several pairs of rhythmic patterns and ask students if they are different or the same. The exercise is adapted from Chela Flores (1997b, p. 127). - Play Earle-Carlin's (2010) recording pronouncing words that contain the most common rhythmic patterns in English (p. 241).	- Brainstorm technical words with students: Ask them to tell you words that contain the stems shown in the PowerPoint presentation. - Tell students about the existence of stressed syllables as syllables that are emphasized. Show them a simple example and then an example of two words that share the same stem but which have a different stressed syllable. Explain to them that the suffix influences the placement of the stressed syllable.
Listening discrimination	- Do Earle-Carlin's (2010) exercise A: Students listen to a recording where several words are pronounced while they have to circle the stress syllable (p. 242). Words are shown with the stressed syllables in red and their rhythmic patterns are included. - Students have to guess the rules of word stress according to common suffixes found in technical words. Show the answers.	- Do Earle-Carlin's (2010) exercise A: Students listen to a recording where several words are pronounced while they have to circle the stress syllable (p. 242). Words are shown with the stressed syllables in red. - Students have to guess the rules of word stress according to common suffixes found in technical words. Show the answers.
Controlled practice	- Students repeat the words of the previous exercises aloud emphasizing the length of the stressed syllable.	- Students repeat the words of the previous exercises aloud emphasizing the stressed syllable.
Guided practice	- (Pair work) Students do the exercise W*hich word is missing?*[a]. Each student has several sentences with a technical word that has been replaced by its rhythmic pattern. Students have to read the sentence and the partner has to guess which is the missing word.	- (Pair work) Students do the exercise *Answer with one word*[a] as a competition. In one minute, a student asks questions and the partner chooses a word from the ones given to answer them. They cannot change questions until they guess the answer. They switch roles.
Communicative practice	- Students work in groups of 4/5 people. They do the exercise *Discuss and report: how does engineering affect our lives?*[a] In this exercise students have to discuss the positive and negative effects of their branch of engineering related to the topics proposed.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 5	*Experimental*	*Control*
Description and analysis	- Explain the difference in stress between an adjective – noun combination and a compound noun and emphasize the difference in meaning. Show the difference in rhythmic patterns.	- Explain the difference in stress between an adjective – noun combination and a compound noun (2 words vs. 1 word) and emphasize the difference in meaning.
Listening discrimination	- Show some more examples and read them aloud. Ask them the difference in meaning depending on the way we pronounce them.	
Controlled practice	- Make students repeat the words from the previous exercise by lengthening the stressed syllable. - Show more examples of compound nouns, this time containing more than two syllables an ask students to read them aloud. Show their corresponding rhythmic pattern.	- Make students repeat the words from the previous exercise. - Show more examples of compound nouns, this time containing more than two syllables an ask students to read them aloud.
Guided practice	- (Pair work) Students do the exercise entitled *Compound or not?*[a]. Students have to read either the sentence containing the option of the compound or the adjective – noun combination to their partner and the latter needs to guess which option has been uttered.	
Communicative practice	- Students play a simple version of *Taboo* in groups of 4/5 people. Cards with vocabulary seen during the course are presented facedown on the table. A student chooses a card and describes the word to the rest of the people without mentioning the word. The rest of the members of the group have to guess which is the word. The person who guesses it receives one point.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 6	*Experimental*	*Control*
Description and analysis	- Exemplify language rhythm by showing a video of a hip-hop song created from speeches during the US election campaign in 2016 entitled *Trump vs. Hillary: Debate Rap Battle* (The DISCLAIMER.io, 2017). - Explain the difference between syllable-timed and stress-timed languages and make students feel the difference in duration using rubber bands. Tell them that content words are stressed while function words are unstressed in English.	- Explain the different grammatical features that we need to bear in mind to describe a process effectively. Then, draw their attention to pauses as a pronunciation element that also plays an important role in process description.
Listening discrimination	- Before telling them about the difference between content and function words, read different versions of a sentence aloud while clapping on the stressed words: Stressed words are the same in every version and unstressed ones are added in between. In order to keep the timing, unstressed words are pronounced shorter. The sentences are shown as part of the PowerPoint presentation.	- Watch the video *How a power plant works* (Midwestgeneration, 2011) and ask students to write down the main six steps described. Play the video twice and correct. Show the answers on the screen, which were directly transcribed from the video.
Controlled practice	- All the students repeat after the teacher clapping on the stressed syllables of the previous sentences. - Show some sentences on the screen to keep practicing clapping. First, rows marking the transitions between stressed words are not shown. Then, show the transitions and make them repeat if they got it wrong.	- Ask students if it was easy to follow the speaker and if they think that a lack of pauses could be the reason for bad comprehension. Ask them to work in pairs and read the steps aloud to their partner, pausing effectively after each step.
Guided practice	- (Pair work) Show the description of how wind power works extracted from *Andy Darvill's Science Site: Online Learning* (Darwill, 2017). Each group works on one of the sentences: They divide it into rhythmic units by using slashes and circling stressed words. Then, students read the sentences aloud by clapping to the sentences and pausing after rhythmic units. Show the correct answer with slashes delimitating rhythmic units and highlighting the content words in red. Ask them to repeat them aloud by clapping and pausing properly if they got it wrong the first time.	- Show the description of how wind power works extracted from *Andy Darvill's Science Site: Online Learning* (Darwill, 2017) as a unique paragraph without punctuation. Tell students to work in pairs and divide the text into different steps and to write a comma for little pauses and a full stop for big ones. Correct with them reading the sentences aloud and show the right answer on the screen.
Communicative practice	- Working in the same groups, ask students to describe some diagrams[a] that represent different processes related to renewable energies.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 7	*Experimental*	*Control*
Description and analysis	- Freeze the image of the graph shown in the video *How to Describe Diagrams – A Closer Look at Graphs and Charts* (English Language Learning, 2014, 1:14) and ask students which expressions we can use to describe it. - Play the video from minute 1:14 to minute 1:50 to show the expressions given, but make the point that there are others that they have already studied that can also be used to describe the same movement.	
Listening discrimination	- Play students four different graph descriptions retrieved from the line graphs section on *Aquascript.com* (Aquascript, 2017) and tell them to draw the line of each graph on the corresponding workbook section[a]. Ask for a volunteer to go in front of the class and do it on the blackboard: After each recording, ask students to compare their graph lines with the ones drawn by the volunteer. Then, show the transcript of the descriptions with the expressions to describe graphs highlighted in green and correct the drawings if necessary. - Give students a handout containing the transcripts of the descriptions. Listen to the recordings again and ask students to circle the content words that they would stress and draw a slash when they think a rhythmic unit finishes. Then, correct by showing the answers on the screen (content words in red and slashes for rhythmic units).	- Play students four different graph descriptions retrieved from the line graphs section on *Aquascript.com* (Aquascript, 2017) and ask them to draw the line of each graph on the corresponding workbook section[a]. Ask for a volunteer to go in front of the class and do the same on the graphs previously draw on the blackboard: After each recording, ask students to compare if their graph lines are the same as the ones drawn by the volunteer on the blackboard. - Give students a handout containing the transcripts of the descriptions without the main expressions to describe the different movements. Listen to the recordings again and ask students to fill in the gaps with the right expression. Correct by showing the answers on the screen highlighted in green.
Controlled practice	- Ask first a few students to read the descriptions aloud individually by clapping on the stressed syllables and pausing after rhythmic units and then, all together as a classroom. Guide the rhythm the first times but then let one of the students guide it.	- Ask students to read the descriptions aloud bearing in mind punctuation to pause properly.
Guided practice	- (Pair work) Students do the exercise on describing a graph extracted from the line graphs section on *Aquascript.com* (Aquascript, 2017)[a]. Each partner has the line of a half of the same graph and they have to describe their line to their partner by using the appropriate expressions for the partner to complete the drawing of the missing part.	
Communicative practice	- Students draw their own graphs and describe those to their partner.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 8	*Experimental*	*Control*
Description and analysis	- Refresh students on the use of the *schwa* sound to reduce unstressed syllables and tell them about the existence of other phenomena that help to speak faster and more fluently in English. - Show them the explanatory video *Connected speech & linking/American English pronunciation* (ElementalEnglish, 2013), where the consonant + vowel and consonant + consonant combinations are introduced. Students repeat after the speaker when pronouncing the sample sentences of the video about linking.	- Reflect on what a good talk is. Show students a word cloud containing adjectives used to describe a good talk and ask them to discuss in pairs or small groups why a good talk needs to have these features. Afterwards, share thoughts with the whole class.
Listening discrimination	- Play a small part of the video *Steve Jobs' 2005 Stanford Commencement Address* (Standford, 2008, 5:37) and ask students to pay attention to linking and the rhythm of the speech. In order to ease comprehension and help them follow the discourse, they can have a look at the transcript[a].	- Play a small part of the video *Steve Jobs' 2005 Stanford Commencement Address* (Standford, 2008, 5:37) and ask students to tell you if they consider it is a good talk based on the features previously discussed. In order to ease comprehension and help them follow the discourse, they can have a look at the transcript[a].
Controlled practice	- Students work in groups. Each group will analyze one paragraph. Students have to underline the sounds that are linked in the speech according to the rules previously explained. Then, show the answers and ask students if they have any questions or doubts about them. - After correcting, students work in pairs and they read the transcript to their partner, paying attention to linking and trying to keep the rhythm. Students can circle stressed words and draw slashes after rhythmic units if needed. The partner provides feedback about the performance bearing in mind everything seen in class until that day.	- (Pair work) Students have to read the transcript to their partner as if they were Steve Jobs. Emphasize the fact that they need to pause properly, trying to show the same emotion, and so on. The partner gives feedback on the performance based on the features of a good talk.
Communicative practice	- (Pair work) one partner plays the role of Steve Jobs and the other one the role of an interviewer. The interviewer asks questions and Steve Jobs has to answer them.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 9	*Experimental*	*Control*
Description and analysis	- Ask students the effect achieved by painting in gray the background of the picture of the butterfly (Gilbert, 2005a, p. 44). They need to deduce that the viewer's attention is drawn towards the butterfly. Tell them that speakers can also emphasize parts of the speech by changing the sentence focus. - Explain the difference in sentence focus between Spanish and English. Emphasize that Spanish uses grammar while English uses stress. Read the examples aloud. - Tell students that listeners' interpretation varies with the focus and read some examples. Ask students to read the sentence aloud lengthening the word to highlight. Examples have been extracted from Chela-Flores (2003, p. 264).	- Ask students what a debate is and show its short definition with a cartoon. - Show them the headings of the debating skills and brainstorm what these should be like for a debate to work. Then, show the answers.
Listening discrimination	- Play the video *Is Smart Technology Making Us Dumb?* (Newsy, 2015) and ask them to highlight the words they emphasize by circling or underlying them on the transcript[a].	- Play the video *Is Smart Technology Making Us Dumb?* (Newsy, 2015) and ask them to recognize these features in their arguments. They can follow the debate with the transcript[a].
Controlled practice	- (Pair work) Show the transcript with the emphasized words in red. One student reads the in favor arguments aloud and the other one the ones against, emphasizing the right words.	- (Pair work) One student has to read the in favor arguments aloud and the other one the ones against, trying to apply the debating skills previously discussed.
Guided practice	- Debate, part 1: (Pair work) Students are going to create a debate on the topic *Can we rely on robots?*. Two people will be in favor and two against. First, ask them to think with the partner on the same side which arguments they can use to defend their position and which parts they want to emphasize. Ask them to rehearse a little bit before the debate starts.	- Debate, part 1: (Pair work) Students are going to create a debate on the topic *Can we rely on robots?*. Two people will be in favor and two against. First, ask them to think with the partner on the same side which arguments they can use to defend their position. Ask them to rehearse a little bit before the debate starts.
Communicative practice	- Debate, part 2: (Group work) Students join their opponents and the debate starts. They have to discuss the topic trying to defend their point of view. Tell them to emphasize the concepts they want to draw their attention to.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Session 10	*Experimental*	*Control*
Description and analysis	- Watch the video *How Donald Trump answers a question* (Nerdwriter1, 2015) on how Donald Trump uses language to persuade citizens.	
Controlled/ Guided practice	- Oral presentation, part 1: Do the activity *Selling a Product: Presenting at a Robotics Fair*[a]. Students have to write a paragraph of approximately 120 words selling a robot or another device they need to present at a robotics fair. Give them time to write the paragraph and analyze its rhythm. They can use the symbols representing each of the rhythmic features seen in class shown on the PowerPoint presentation to make it. Then, they read their speech aloud to themselves to rehearse.	- Oral presentation, part 1: Do the activity *Selling a Product: Presenting at a Robotics Fair*[a]. Students have to write a paragraph of approximately 120 words selling a robot or another device they need to present at a robotics fair. Give them time to write the paragraph. Then, they read their speech aloud to themselves to rehearse.
Communicative practice	- Oral presentation, part 2: Students create groups of four/five people and they have to imagine they are at the fair trying to sell their product. Students have to present their product and the rest of the students have to provide feedback and decide who was the most persuasive.	
Note. [a] = Exercises included in the course book as part of the pronunciation section.		

Appendix 3: The course syllabus

UNIT 1: HISTORICAL FACTS IN ENGINEERING
1.1 Chronological description: narrative texts
1.2 Review of verb tenses: ways to express time sequence
1.3 Genres: Encyclopaedias
1.4 Texts: *History of Engineering, Energy Supply, Electronics in the Home, Electromagnetism, Developments in Computing, History of Electricity* / Podcasts: *Studying in the UK, Global Workers.*

UNIT 2: A COMPUTERIZED WORLD
2.1 Physical and function description: ways to describe the physical properties, the function and parts of a device, and how something works.
2.2 Review of prepositions and Phrasal Verbs.
2.3 Genres: focus on audience and purpose.
2.4 Texts: *The IPad, Cloud Computing, Describing a Computer and How it Works: IDE Cables, What is a Tablet PC? Tech Habits that Need to Stop* / Podcasts: *Engineering Projects, Finding Information.*

UNIT 3: ELECTRONIC AND ELECTRICAL DEVICES
3.1 Making definitions: relative clauses and relative pronouns. Reduced relative clauses. Defining and Non-defining relative clauses.
3.2 Structure and use of compound nouns.
3.3 Paragraph writing: LEDs
3.4 Genres in Engineering: Magazine article and Research article.
3.5 Texts: *The Transistor, Graphene Electronics, Home Automation, What is a LED Screen? Flexible Home Care Automation* / Podcasts: *Effective Websites, The Future of English*.

UNIT 4: A GREENER WORLD: RENEWABLE ENERGIES
4.1 Expressions to indicate cause and effect.
4.2 Process description and sequence markers.
4.3 Comparison and contrast.
4.4 Visual information and in-text reference. Conditional sentences to express a hypothesis.
4.5 Genres and text layout: headings in texts
4.6 Texts: *Alternative Sources of Energy, Electricity Generation, Sources of Electricity, Nuclear Power, Nuclear Accidents, How Does a Nuclear Power Plant Work? How a Solar Cell Works* / Podcasts: *A Greener World; Climate Change; Hybrid Cars, Fuel from Garbage (video);*

UNIT 5: AN INTERCONNECTED CITY: MOBILE TECHNOLOGIES
5.1 Making predictions in English
5.2 E-mail writing and Netiquette in electronic communication.
5.3 Formal and informal language in electronic communication.
5.4 Email writing: Practice on common academic situations
6.5 Podcasts: *ESL Business Writing Video - Email Tune-up (video), Lifelong Learning.*

UNIT 6: ROBOTICS AND AUTOMATION
6.1 Expressions used in classifications. Ways of expressing criteria and giving examples.
6.2 Review of modal verbs expressing obligation, probability and deduction.
6.3 Review of linking words: coordination, subordination, and sentence adverbials.
6.5 Genres: textbooks and newspaper articles.
7.6 Texts: *Robotics, Types of Robots, Social Robots, The Three Laws of Robotics, Robots: The Future is Now, Miniature Robotics* / Podcasts: *Biologically-inspired robots (video).*

UNIT 7: ENGINEERING PROJECTS: ORAL PRESENTATIONS
7.1 Designing technical presentations
7.2 Stages for the design of an oral presentation: planning, delivery, evaluation
7.3 Delivering an oral presentation: parts
7.4 Online Videocasts: Extensive practice on oral presentations through activities and videos.

(Rueda, 2017, p.3)

Appendix 4: The recording test and the satisfaction survey

TECHNICAL ENGLISH 2016/2017: PRONUNCIATION TEST

1. **Record yourself reading these sentences aloud:**

   a) I want an iPhone.
   b) Smart cities consume less energy than conventional ones.
   c) I will search it on Google.
   d) The teacher told us to create a one-wire circuit.
   e) The technician measured the consumption of electricity of the factory.
   f) If you don't want to lose the information, upload it on the cloud.
   g) I would like to become an electrical engineer in the future.
   h) John is working on the oral presentation while Mark is looking at videos on Youtube.
   i) Greenhouse emissions have decreased dramatically.
   j) Yesterday I talked to my teacher about the project.

2. **Record yourself reading the following test aloud (Celce-Murcia, 1996:398):**

   If English is not your native language, people may have noticed that you come from another country because of your "foreign accent". Why do people usually have an accent when they speak a second language? Several theories address this issue. Many people believe that only young children can learn a second language without an accent, but applied linguists have reported cases of older individuals who have mastered a second language without an accent. (…) In the end, the path to learning to speak a second language without an accent appears to be a combination of hard work, a good ear, and a strong desire to sound like a native speaker. You also need accurate information about the English sound system and lots of exposure to the spoken language. Will you manage to make progress, or will you just give up? Only time will tell, I'm afraid. Good luck, and don't forget to work hard!

3. **Introduce yourself:** say your name, your age and where you are from; then, tell me which field of engineering you are studying and why you decided to study it; finally, tell me about your language background: which languages you speak, when you started studying English, which is/are the aspect/s of the English language that your have more difficulties with and why it is so.

4. **Give your opinion: what do you think about social networks? Are they useful or dangerous? Why do you think so? You have 1 minute to prepare** your answer: you can take notes about the points you want to talk about, but do not write the whole speech. After that, **you need to talk about the topic for 1 minute**.

SATISFACTION SURVEY (attached to the posttest):

1. The module has met the expectations regarding an improvement in your oral skills.

Strongly agree	Agree	Undecided	Disagree	Strongly disagree

2. The activities proposed have helped you practice your oral skills.

Strongly agree	Agree	Undecided	Disagree	Strongly disagree

3. The explanations provided were useful.

Strongly agree	Agree	Undecided	Disagree	Strongly disagree

4. How much have you participated/worked in class?

100%	75%	50%	25%	0%

5. Did you practice your speaking/pronunciation skills outside the classroom? If so, which type of exercises did you do?

6. Which of these aspects do you think you have improved?

a) I pronounce words more clearly.

b) I have reduced my Spanish/Catalan accent when speaking English.

c) I am more fluent in English.

d) My classmates understand me better.

e) I understand people speaking English better

f) Other: _____

7. Would you like to take this course again? Justify your answer.

Appendix 5: VarcoV values per student

VarcoV individual values and means for the control group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	53	41	56	73	36	34	44	40	43	50	**47**	53	40	54	52	36	48	62	37	57	54	**49**
2	1	24	57	28	68	47	44	75	43	47	58	**49**	27	60	50	52	73	55	69	49	78	80	**59**
3	1	28	21	33	46	46	41	46	35	35	32	**36**	17	29	26	44	44	42	49	38	46	40	**37**
4	1	44	61	38	46	43	66	52	46	60	65	**52**	40	45	41	52	46	62	58	39	63	57	**50**
5	2	29	38	34	67	39	49	51	42	27	32	**41**	35	46	20	57	55	44	44	50	61	53	**47**
6	2	32	38	40	44	47	48	51	39	38	48	**43**	47	38	30	41	41	33	35	45	33	48	**39**
7	2	45	32	49	68	45	64	49	49	37	38	**48**	35	36	39	35	24	37	61	33	51	37	**39**
8	2	37	39	37	47	33	56	42	26	50	35	**40**	40	22	41	60	59	49	56	40	38	50	**46**
9	2	37	53	65	61	55	47	66	47	66	61	**56**	37	51	43	53	70	55	38	41	71	52	**51**
10	2	34	40	32	51	60	51	66	42	54	57	**49**	13	39	28	65	32	46	57	42	43	43	**41**
11	2	46	29	36	42	35	36	46	33	57	52	**41**	36	45	32	37	36	54	49	39	59	75	**46**
12	2	38	38	31	59	43	73	67	56	57	46	**51**	48	43	25	56	43	76	59	60	50	42	**50**
13	2	59	44	38	44	46	33	51	35	54	51	**45**	35	28	49	51	32	45	43	44	38	41	**40**
14	2	37	47	36	39	67	33	48	39	63	37	**45**	34	38	34	60	59	39	38	47	43	51	**44**
15	2	28	37	36	53	48	60	54	41	42	40	**44**	47	45	34	57	62	52	52	39	46	45	**48**
16	3	35	32	24	51	33	31	66	38	41	44	**40**	45	48	45	50	32	38	61	40	36	41	**44**
17	3	38	39	41	53	45	45	70	49	54	62	**50**	33	48	35	48	64	47	49	33	48	64	**47**
18	3	31	36	20	57	51	47	60	31	39	66	**44**	35	34	24	63	54	47	61	49	33	55	**46**
19	3	38	35	44	56	35	41	46	49	54	55	**45**	26	32	58	44	40	48	60	39	64	55	**47**
20	3	32	28	34	60	48	67	62	43	53	39	**47**	18	32	42	53	34	62	62	33	44	46	**43**
21	3	41	44	69	56	50	51	52	52	56	45	**51**	53	42	41	55	58	48	56	63	61	52	**53**
Total												**46**											**46**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

VarcoV individual values and means for the experimental group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	45	45	24	68	52	55	48	49	43	50	**48**	40	32	64	70	79	39	55	37	48	68	**53**
2	1	38	39	49	49	29	37	49	31	53	45	**42**	46	41	34	60	35	49	62	42	60	50	**48**
3	1	42	35	37	32	52	28	62	32	30	50	**40**	54	35	25	43	40	41	62	39	40	57	**43**
4	1	40	26	50	51	41	35	34	36	65	52	**43**	41	35	39	51	59	46	47	43	51	52	**46**
5	1	41	43	35	48	36	34	58	39	40	62	**44**	39	29	42	45	61	31	61	49	38	63	**46**
6	1	39	28	25	55	37	61	44	48	31	41	**41**	48	44	55	62	41	43	49	39	51	53	**49**
7	1	23	41	47	49	48	46	52	43	48	44	**44**	49	38	44	50	52	45	54	41	44	41	**46**
8	1	45	25	51	44	36	39	35	32	51	43	**40**	38	54	33	43	41	39	39	36	45	33	**40**
9	1	37	49	42	31	48	54	42	46	52	46	**45**	40	32	39	43	47	54	56	47	43	49	**45**
10	1	50	34	26	50	25	46	61	36	45	46	**42**	48	31	22	43	45	32	48	41	46	49	**40**
11	2	25	20	31	51	26	42	53	52	72	39	**41**	28	42	25	37	35	33	42	42	37	50	**37**
12	2	36	41	60	52	46	49	49	44	72	55	**50**	45	50	58	61	59	43	54	43	35	44	**49**
13	2	30	24	30	58	29	39	36	35	51	45	**38**	36	41	32	50	72	36	63	50	40	51	**47**
14	2	43	32	34	71	40	49	88	48	53	51	**51**	42	46	22	49	54	62	61	51	52	47	**49**
15	2	56	58	36	55	49	56	47	54	63	49	**52**	42	31	68	70	40	47	71	53	60	48	**53**
16	2	24	34	41	48	37	45	41	40	44	46	**40**	23	40	39	51	56	57	48	43	48	62	**47**
17	3	16	54	27	36	44	55	57	42	58	49	**44**	26	27	40	53	50	38	59	31	56	50	**43**
18	3	55	30	37	47	49	37	56	30	53	45	**44**	38	35	35	45	41	44	46	37	46	43	**41**
19	3	58	50	47	47	34	35	65	39	61	43	**48**	51	42	37	40	44	54	57	44	45	54	**47**
20	3	29	42	46	49	45	45	57	31	54	50	**45**	31	33	36	56	47	56	65	38	65	51	**48**
21	3	41	41	28	51	59	68	69	49	53	50	**51**	45	53	31	56	51	49	70	48	64	48	**51**
Total												**44**											**46**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Appendix 6: VarcoV values per native speaker

Native speakers' scores (cleaned data)

St.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	M
1	37	43	32	58	53	55	58	45	57	55	**49**
2	34	51	32	45	64	50	64	47	55	66	**51**
3	18	49	31	62	56	51	73	46	47	59	**49**
4	42	49	44	64	64	47	60	52	41	51	**51**
5	35	47	35	53	50	51	64	52	42	49	**48**
6	34	51	43	55	52	65	66	58	39	34	**50**
Total											**50**

Note. St = Student. S = Sentence. M = Mean.

Appendix 7: VarcoV t-tests results

VarcoV t-tests

Sentence	T-test		df	*T*	*p*-value	Cohen's *d*
		C T2-T1	20	-.584	.565	-1.28
	Paired	E T2-T1	20	.810	.428	1.77
		C T1-E T1	40	-.475	.637	.147
1	Independent	C T2-E T2	40	-1.538	.132	.475
		C T2-T1	20	.265	.794	.058
	Paired	E T2-T1	20	.389	.702	.085
		C T1-E T1	40	.663	.511	.205
2	Independent	C T2-E T2	40	.580	.565	.179
		C T2-T1	20	-.559	.582	-.122
	Paired	E T2-T1	20	.271	.789	.059
		C T1-E T1	40	.272	.787	.084
3	Independent	C T2-E T2	40	-.427	.672	.132
		C T2-T1	20	-1.056	.303	-.231
	Paired	E T2-T1	20	.787	.440	.172
		C T1-E T1	40	1.615	.114	.498
4	Indepedent	C T2-E T2	40	.091	.928	.028
		C T2-T1	20	.608	.550	.133
	Paired	E T2-T1	20	2.988	.007**	.652
		C T1-E T1	40	1.588	.120	.490
5	Independent	C T2-E T2	40	-.657	.515	.203
		C T2-T1	20	.181	.858	.039
	Paired	E T2-T1	20	-.226	.823	-.049
		C T1-E T1	40	.877	.386	.271
6	Independent	C T2-E T2	40	1.417	.164	.437
		C T2-T1	20	-.793	.437	-1.73
	Paired	E T2-T1	20	1.102	.284	.240
		C T1-E T1	40	.794	.432	.012
7	Independent	C T2-E T2	40	-.804	.426	.029
		C T2-T1	20	.559	.582	.122
	Paired	E T2-T1	20	1.125	.274	.246
		C T1-E T1	40	.399	.692	.012
8	Independent	C T2-E T2	40	.094	.925	.029
		C T2-T1	20	.523	.607	.114
	Paired	E T2-T1	20	-1.230	.233	-.268
		C T1-E T1	40	-.891	.378	.275
9	Independent	C T2-E T2	40	.715	479	.221
		C T2-T1	20	1.422	.170	.310
	Paired	E T2-T1	20	1.764	.093	.385
		C T1-E T1	40	.201	.842	.062
10	Independent	C T2-E T2	40	.332	.742	.102

Note. C = control group. E = experimental group. T = time. Paired =paired-samples t-test. Independent = Independent-samples t-test. Cohen's *d* = Effect size. ** $p < .01$.

Appendix 8. Total number of pauses per student

Total number of pauses individual values and means for the control group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	1	2	2	5	5	4	4	5	1	4	**3.3**	1	4	2	5	6	4	5	8	4	4	**4.3**
2	1	1	2	5	4	3	3.52	3	3	3	3	**3.1**	1	3	4	1	3	5	1	5	4	2	**2.9**
3	1	1	1	2	4	1	5	3	1	4	3	**2.5**	2	2	2	4	4	6	3	5	3	3	**3.4**
4	1	1	1	3	3	5	4	4	6	2	2	**3.1**	0	3	1	2	4	3	4	5	2	5	**2.9**
5	2	0	3	0	3	2	3	2	3	5	3	**2.4**	0	3	0	4	6	4	6	3	2	3	**3.1**
6	2	1	1	2	3	3	3	4	2	2	3	**2.4**	1	1	2	2	4	3	1	4	1	4	**2.3**
7	2	1	2.14	3	2	9	2	3	6	2	4	**3.4**	1	2	1	1	4	3	1	5	2	3	**2.3**
8	2	1	2	1	4	4	5	4	3	2	3	**2.9**	0	0	1	2	4	4	1	2	1	4	**1.9**
9	2	0	4	5	5	5	2	5	4	4	1	**3.5**	0	1	1	2	6	2	5	4	3	3	**2.7**
10	2	1	5	2	4	5	7	6	4	6	2.57	**4.3**	2	3	2	6	6	3	5	8	2.52	4	**4.2**
11	2	0	0	1	2	2	2	0	2	1	1	**1.1**	0	0	0	2	1	1	0	2	0	3	**0.9**
12	2	1	2	2	4	4	2	4	12	3	3	**3.7**	1	2	2	3	5	3	1	5	3	1	**2.6**
13	2	1	2	3	4	10	3	3	6	2	2	**3.6**	1	2	1	4	4	3	3	3	2	3	**2.6**
14	2	0	3	2	7	5	4	3	5	3	3	**3.5**	0	3	2	6	3	3	2	8	2	6	**3.5**
15	2	1	1	1	2	5	2	2	2	1	3	**2.0**	0	1	1	2	3	2	5	6	2	3	**2.5**
16	3	0	2	0	2	2	2	2	0	2	2	**1.4**	1	0	1	1	3	2	1	2	2	2	**1.5**
17	3	0	0	2	3	2	2	5	7	0	2.57	**2.4**	0	1	0	1	1	0	2	0	0	1	**0.6**
18	3	0	0	0	2	3	2	0	8	2	2	**1.9**	0	2	0	1	4	3	1	5	2	2	**2.0**
19	3	0	0	0	3	1	3	4	6	2	2	**2.1**	0	1	0	0	4	2	2	2	4	2	**1.7**
20	3	1	4	1	2	3	5	1	2	2	2	**2.3**	2	3	1	3	2	2	2	3	4	2	**2.4**
21	3	0	2	0	1	2	3	0	4	2	2	**1.6**	0	2	0	3	2	3	2	4	2	1	**1.9**
Total												**2.7**											**2.5**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Total number of pauses individual values and means for the experimental group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	0	0	0	3	4	4	2	2	2	0	**2.1**	0	0	4	1	2	3	2	0	2	2	**1.6**
2	1	1	1	1	4	10	2	6	4	1	1	**3.3**	1	1	2	2	4.19	3	6	2	4	3	**2.8**
3	1	0	0	0	1	0	5	0	1	2	0	**0.9**	0	0	0	1	4.19	2	2	1	1	0	**1.1**
4	1	1	2	2	3	4	4	1	5	3	1	**2.7**	0	3	1	0	2	4	2	2	4	3	**2.1**
5	1	0	0	0	3	0	2	2	2	1	0	**1.2**	0	0	1	3	3	4	4	1	1	1	**1.8**
6	1	0	3	0	1	8	3	4	6	3	0	**3.1**	1	4	0	3	5	3	2	3	3	3	**2.7**
7	1	0	3	2	3	4	5	2	4	3	0	**2.9**	0	2	3	3	3	2	2	5	3	2	**2.5**
8	1	1	2	2	1	4	8	2	7	4	1	**3.2**	0	2	3	3	7	4	1	8	4	4	**3.6**
9	1	0	1	0	1	2	4	4	3	3	0	**2.1**	0	1	0	0	4.19	2	2	2	3	1	**1.5**
10	1	0	0	1	3	4	3	4	4	3	0	**2.4**	0	1	1	3	3	3	0	1	1	1	**1.4**
11	2	0	0	1	5	4	1	4	2	6	0	**2.4**	1	2	1	3	4	3	2	2	4	3	**2.5**
12	2	0	1	1	6	5	2	2	3	3	0	**2.5**	0	0	1	2	6	2	2	2	4	2	**2.1**
13	2	1	4	2	3	6	3	3	7	3	1	**3.5**	1	1	1	3	6	4	3	6	2	3	**3.0**
14	2	1	3	1	5	7	3	4	8	3.48	1	**3.8**	0	2	1	5	5	6	3	7	2.76	3	**3.5**
15	2	0	1	0	0	0	1	0	3	2	0	**0.8**	0	0	0	1	2	1	1	5	2	1	**1.3**
16	2	1	4	1	6	4	1	3	4	3	1	**3.2**	1	4	1	4	4	2	3	4	2	4	**2.9**
17	3	1	2	0	3	5	2	6	3	6	1	**3.1**	0	2	1	1	3	1	3	4	3	3	**2.1**
18	3	0	1	1	4	7	0	2	1	4	0	**2.3**	0	0	0	1	5	0	1	1	1	1	**1.0**
19	3	1	4	1	1	3	6	3	7	4	1	**3.4**	1	1	1	4	6	4	5	2	2	2	**2.8**
20	3	1	4	2	4	8	3	2	2	5	1	**3.4**	1	4	1	6	6	3	2	5	3	2	**3.3**
21	3	2	3	1	3	4	3	3	3	3	2	**2.6**	0	2	0	3	2	1	0	1	2	1	**1.2**
Total												**2.6**											**2.2**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Appendix 9: Total number of pauses t-tests results

Total number of pauses t-tests

Sentence	T-test		df	*T*	*p*-value	Cohen's *d*
		C T2-T1	20	.370	.715	.0814
	Paired	E T2-T1	20	-1.284	.214	-.279
		C T1-E T1	40	.277	.783	.090
1	Independent	C T2-E T2	40	1.482	.146	.464
		C T2-T1	20	-.022	.983	-.005
	Paired	E T2-T1	20	-1.276	.217	-.278
		C T1-E T1	40	.015	.988	.005
2	Independent	C T2-E T2	40	.855	.398	.269
		C T2-T1	20	-2.444	.024*	-.533
	Paired	E T2-T1	20	.777	.446	.169
		C T1-E T1	40	2.355	.023*	.729
3	Independent	C T2-E T2	40	.146	.884	.038
		C T2-T1	20	-2.092	.049*	-.457
	Paired	E T2-T1	20	-1.272	.218	-.278
		C T1-E T1	40	.597	.554	.187
4	Independent	C T2-E T2	40	.284	.778	.086
		C T2-T1	20	-.183	.857	-.040
	Paired	E T2-T1	20	-.561	.581	-.122
		C T1-E T1	40	-.741	.463	.228
5	Independent	C T2-E T2	40	-.764	.450	.236
		C T2-T1	20	-1.183	.251	-.258
	Paired	E T2-T1	20	-.969	.344	-.211
		C T1-E T1	40	.331	.742	.102
6	Independent	C T2-E T2	40	.460	.648	.141
		C T2-T1	20	-.987	.335	-.216
	Paired	E T2-T1	20	-1.419	.171	-.310
		C T1-E T1	40	.281	.780	.085
7	Independent	C T2-E T2	40	.475	.637	.142
		C T2-T1	20	-.133	.895	-.029
	Paired	E T2-T1	20	-1.966	.063	-.429
		C T1-E T1	40	.636	.528	.194
8	Independent	C T2-E T2	40	1.766	.085	.544
		C T2-T1	20	-.503	.620	-.110
	Paired	E T2-T1	20	-2.103	.048*	-.459
		C T1-E T1	40	-1.864	.070	.575
9	Independent	C T2-E T2	40	-.854	.398	.264
		C T2-T1	20	1.251	.225	.273
	Paired	E T2-T1	20	-1.195	.246	-.261
		C T1-E T1	40	.172	.865	.053
10	Independent	C T2-E T2	40	2.044	.048*	.629

Note. C = control group. E = experimental group. T = time. Paired =paired-samples t-test. Independent = Independent-samples t-test. Cohen's *d* = Effect size. * $p < .05$.

Appendix 10: Unfilled pauses per student

Unfilled pauses individual values and means for the control group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	1	2	2	5	4	4	4	5	1	4	**3.2**	1	4	2	5	6	4	5	8	4	4	**4.3**
2	1	1	2	2	2	3	3.05	3	3	3	2	**2.4**	1	2	3	1	3	4	1	5	3	2	**2.5**
3	1	1	1	2	4	1	4	3	1	4	2	**2.3**	2	2	2	4	4	5	3	5	3	3	**3.3**
4	1	0	1	2	3	4	3	3	6	2	2	**2.6**	0	2	1	2	4	3	3	5	2	4	**2.6**
5	2	0	3	0	3	2	3	2	3	4	2	**2.2**	0	3	0	4	5	4	5	3	2	3	**2.9**
6	2	1	1	2	3	3	3	3	2	2	3	**2.3**	0	1	2	2	3	3	1	4	1	4	**2.1**
7	2	1	5	2	2	6	2	3	6	2	3	**3.2**	1	2	1	1	3	3	1	5	2	3	**2.2**
8	2	1	2	1	3	3	4	3	2	2	3	**2.4**	0	2	1	2	3	3	1	1	1	4	**1.8**
9	2	0	4	4	3	5	2	4	4	3	1	**3.0**	0	1	1	2	5	2	4	4	3	3	**2.5**
10	2	1	4	2	4	5	3.05	5	4	4	5	**3.7**	2	3	2	5	5	3	5	7	6	3	**4.1**
11	2	0	0	1	1	2	2	0	2	1	1	**1.0**	0	0	0	1	1	1	0	2	0	2	**0.7**
12	2	1	2	2	4	4	2	3	10	3	3	**3.4**	1	2	2	3	5	3	1	5	3	1	**2.6**
13	2	1	2	2	4	8	3	3	6	2	2	**3.3**	1	2	1	4	4	3	2	3	2	3	**2.5**
14	2	0	3	2	2.95	5	3	3	5	3	3	**3.0**	0	3	2	6	3	3	2	7	2	5	**3.3**
15	2	1	1	1	2	5	2	2	2	1	3	**2.0**	0	1	1	2	3	2	3	6	2	3	**2.3**
16	3	0	2	0	2	2	2	2	0	2	2	**1.4**	1	0	1	1	3	2	1	2	2	2	**1.5**
17	3	0	0	2	3	2	2	4	7	0	0	**2.0**	0	1	0	1	1	0	2	0	0	1	**0.6**
18	3	0	0	0	2	3	2	0	6	2	2	**1.7**	0	2	0	1	2	2	1	3	1	2	**1.4**
19	3	0	0	0	2	1	2	3	5	2	2	**1.7**	0	1	0	0	3	2	1	2	2	2	**1.3**
20	3	1	4	1	2	3	4	1	1	2	2	**2.1**	2	3	1	3	2	2	2	3	4	2	**2.4**
21	3	0	2	0	1	1	2	0	4	1	2	**1.3**	0	1	0	3	2	3	2	4	2	1	**1.8**
Total												**2.4**											**2.3**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Unfilled pauses individual values and means for the experimental group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	0	0	0	3	3	4	1	2	2	3	**1.8**	0	0	3	1	3.38	2	1	0	2	2	**1.4**
2	1	1	1	1	4	7	2	4	4	1	3	**2.8**	1	1	2	2	3.38	3	4	2	3	3	**2.4**
3	1	0	0	0	1	0	4	0	1	2	0	**0.8**	0	0	0	1	3.38	2	1	1	1	0	**0.9**
4	1	1	2	2	3	4	3	1	5	3	2	**2.6**	0	3	1	0	3.38	3	2	2	3	3	**2.0**
5	1	0	0	0	2	0	1	2	1	1	2	**0.9**	0	0	0	2	3.38	3	2	1	1	1	**1.3**
6	1	0	3	0	1	6	3	3	6	3	3	**2.8**	1	4	0	3	5	2	2	3	3	3	**2.6**
7	1	0	3	2	3	4	5	2	4	3	3	**2.9**	0	2	2	3	3	2	2	5	3	2	**2.4**
8	1	1	2	2	1	3	6	2	5	3	1	**2.6**	0	2	2	2	5	3	1	5	3	3	**2.6**
9	1	0	1	0	1	2	4	3	3	3	3	**2.0**	0	1	0	0	3.38	2	1	2	3	1	**1.3**
10	1	0	0	1	3	3	3	3	4	3	2	**2.2**	0	1	1	3	3	3	0	1	1	1	**1.4**
11	2	0	0	1	4	4	1	2	2	5	1	**2.0**	1	2	1	3	4	3	2	2	3	3	**2.4**
12	2	0	1	1	6	3	2	2	3	3	2	**2.3**	0	0	1	2	3	2	2	2	3	2	**1.7**
13	2	1	4	2	3	6	3	3	7	3	3	**3.5**	1	1	1	3	5	4	3	6	2	3	**2.9**
14	2	1	3	1	5	5	3	3	8	6	3	**3.8**	0	2	1	5	5	5	3	7	5	3	**3.6**
15	2	0	1	0	0	0	1	0	3	2	1	**0.8**	0	0	0	0	3.38	1	1	4	2	1	**1.2**
16	2	1	4	1	6	4	1	3	4	3	4	**3.1**	1	4	1	4	4	2	3	4	2	4	**2.9**
17	3	1	2	0	3	5	2	5	3	4	3	**2.8**	0	2	1	1	3	1	3	4	2	3	**2.0**
18	3	0	1	1	3	3	0	2	1	4	3	**1.8**	0	0	0	1	4	0	1	1	1	1	**0.9**
19	3	1	4	1	1	3	5	3	7	4	4	**3.3**	1	1	1	4	5	4	3	2	2	2	**2.5**
20	3	1	4	2	4	6	3	2	2	3	3	**3.0**	1	4	1	6	5	3	2	5	3	2	**3.2**
21	3	1	3	1	2	3	3	2	3	3	1	**2.2**	0	2	0	2	2	1	0	1	1	1	**1.0**
Total												**2.4**											**2**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Appendix 11: Unfilled pauses t-test results

Unfilled pauses t-tests

Sentence	T-test		df	*T*	*p*-value	Cohen's *d*
		C T2-T1	20	.370	.715	.0814
	Paired	E T2-T1	20	-1.142	.267	-.250
		C T1-E T1	40	.302	.765	.078
1	Independent	C T2-E T2	40	1.227	.227	.382
		C T2-T1	20	-.484	.634	-.106
	Paired	E T2-T1	20	-1.276	.217	-.278
		C T1-E T1	40	.209	.836	.061
2	Indepndent	C T2-E T2	40	.766	.448	.240
		C T2-T1	20	-1.673	.110	-.365
	Paired	E T2-T1	20	.000	1	.000
		C T1-E T1	40	1.870	.069	.584
3	Independent	C T2-E T2	40	.717	.477	.232
		C T2-T1	20	-.852	.404	-.186
	Paired	E T2-T1	20	-1.395	.178	-.304
		C T1-E T1	40	-.117	.908	.036
4	Indepedent	C T2-E T2	40	.475	.637	.142
		C T2-T1	20	-.241	.812	-.052
	Paired	E T2-T1	20	.678	.505	.148
		C T1-E T1	40	-.164	.871	.048
5	Independent	C T2-E T2	40	-1.302	.201	.402
		C T2-T1	20	-.024	.981	-.005
	Paired	E T2-T1	20	-1.116	.278	-.244
		C T1-E T1	40	-.240	.811	.074
6	Independent	C T2-E T2	40	.816	.419	.247
		C T2-T1	20	-1.164	.258	-.254
	Paired	E T2-T1	20	-1.826	.083	-.400
		C T1-E T1	40	.724	.473	.219
7	Independent	C T2-E T2	40	.830	.412	.253
		C T2-T1	20	.000	1	.000
	Paired	E T2-T1	20	-2.186	.041*	-.477
		C T1-E T1	40	.416	.680	.130
8	Independent	C T2-E T2	40	1.860	.070	.573
		C T2-T1	20	.181	.858	.040
	Paired	E T2-T1	20	-2.855	.010*	-.623
		C T1-E T1	40	-2.480	.017*	.768
9	Independent	C T2-E T2	40	-.255	.800	.074
		C T2-T1	20	1.563	.134	.341
	Paired	E T2-T1	20	-1.188	.249	-.260
		C T1-E T1	40	-.144	.886	.047
10	Independent	C T2-E T2	40	1.869	.069	.568

Note. C = control group. E = experimental group. T = time. Paired =paired-samples t-test. Independent = Independent-samples t-test. Cohen's *d* = Effect size. * $p < .05$.

Appendix 12: Correlation between VarcoV and the total number of pauses results

Pearson product-moment correlation tests for VarcoV – Total pauses

Sentence	Group	Time	*r*	*n*	*p*-value
		1	.042	21	.856
	Control	2	-.497	21	.022*
		1	-.151	21	.513
1	Experimental	2	-.251	21	.272
		1	.186	21	.420
	Control	2	.074	21	.751
		1	-.179	21	.437
2	Experimental	2	.206	21	.369
		1	.205	21	.374
	Control	2	.109	21	.637
		1	.481	21	.027*
3	Experimental	2	.174	21	.452
		1	-.111	21	.631
	Control	2	.431	21	.051
		1	.328	21	.146
4	Experimental	2	-.149	21	.519
		1	.267	21	.242
	Control	2	-.105	21	.650
		1	-.256	21	.262
5	Experimental	2	.122	21	.600
		1	.055	21	.812
	Control	2	-.113	21	.627
		1	-.187	21	.417
6	Experimental	2	.076	21	.743
		1	.179	21	.438
	Control	2	-.223	21	.332
		1	.159	21	.491
7	Experimental	2	.192	21	.404
		1	.460	21	.036*
	Control	2	.174	21	.451
		1	.000	21	.999
8	Experimental	2	.131	21	.571
		1	-.083	21	.722
	Control	2	.379	21	.091
		1	.449	21	.041*
9	Experimental	2	.105	21	.650
		1	-.454	21	.039*
	Control	2	-.092	21	.692
		1	-.157	21	.496
10	Experimental	2	-.129	21	.576

Note. * $p < .05$.

Appendix 13: Correlation between VarcoV and the unfilled pauses results

Pearson product-moment correlation tests for VarcoV – Unfilled pauses

Sentence	Group	Time	*r*	*n*	*p*-value
		1	-.029	21	.900
	Control	2	-.558	21	.009**
		1	-.201	21	.383
1	Experimental	2	-.251	21	.272
		1	.097	21	.676
	Control	2	-.262	21	.251
		1	-.179	21	.437
2	Experimental	2	.206	21	.369
		1	.295	21	.194
	Control	2	.057	21	.807
		1	.481	21	.027*
3	Experimental	2	.104	21	.653
		1	.047	21	.838
	Control	2	.449	21	.041*
		1	.340	21	.131
4	Experimental	2	-.176	21	.445
		1	.400	21	.072
	Control	2	-.126	21	.587
		1	-.382	21	.087
5	Experimental	2	.237	21	.301
		1	-.034	21	.884
	Control	2	-.119	21	.607
		1	-.027	21	.909
6	Experimental	2	.131	21	.572
		1	.158	21	.493
	Control	2	-.172	21	.457
		1	.098	21	.673
7	Experimental	2	.086	21	.711
		1	.521	21	.015*
	Control	2	.142	21	.540
		1	.065	21	.779
8	Experimental	2	.193	21	.403
		1	-.165	21	.475
	Control	2	.142	21	.539
		1	.386	21	.084
9	Experimental	2	.093	21	.690
		1	-.259	21	.257
	Control	2	-.194	21	.398
		1	-.184	21	.423
10	Experimental	2	-.030	21	.897

Note. * $p < .05$. ** $p < .01$.

UNIVERSITAT
ROVIRA i VIRGILI