UNIVERSITAT DE BARCELONA

The Role of L2 Speaking Anxiety in L2 Speaking Fluency, Pronunciation Accuracy and Comprehensibility in Instructed English

Gisela Sosa López

The Role of L2 Speaking Anxiety in L2 Speaking Fluency, Pronunciation Accuracy and Comprehensibility in Instructed English

Gisela Sosa López

PhD thesis

Supervised by: Dr. Joan Carles Mora

Programa de Doctorat en Ciència Cognitiva i Llenguatge

Departament de Llengües i Literatures Modernes i Estudis Anglesos

Universitat de Barcelona

2025

Acknowledgements

This dissertation would not have been possible without the help and support of many people to whom I would like to dedicate the following words.

First and foremost, I would like to thank my supervisor, Dr. Joan Carles Mora. Your support and feedback have been very valuable to both my academic and personal growth. It all began with that "beca de col·laboració" in my 4th year of English Studies, which offered me the opportunity to see firsthand how research studies are designed, conducted, and their data analyzed.. When I began English Studies, I was fascinated with literature; little did I know that courses such as Phonetics, Phonology, and Applied Linguistics along with the FPU grant would shape my future so profoundly. Thank you, Joan Carles, for allowing me to take an active role in your projects from the very beginning, for your patience in teaching me from the very beginning (how to use Praat for analyzing VOT, how the eye-tracker works) and for guiding me through this dissertation. I have learnt priceless lessons and invaluable skills that have contributed to this thesis and my growth as a researcher.

My second acknowledgement goes to Gosia, without whom my research stay in Wroclaw would not have been as enjoyable, nor would the analysis of my qualitative data be as thorough as it is. Thank you for welcoming me in Poland, introducing me to your family, and for all the inspiring "food for thought" conversations in your office. A part of this dissertation is thanks to you as well.

A special thanks to all the English Studies teachers in the *Department of Modern Languages and Literatures and English Studies*, which has been my academic home for so long. I have learnt so much from your classes throughout my degree and master's program, it brings me joy whenever we cross paths in the hallways. My gratitude also extends to the project meetings with Miren Adrian, Dr. Cristina Aliaga, Dr. Núria Gavaldà, Valeria Galimberti, Dr. Ingrid Mora-Plaza, Dr. Natalia Fullana, Dr. Mireia Ortega and Josh Frank, and the *Grup de Recerca en Adquisicó de Llengües* (GRAL) at the University of Barcelona. I am grateful for all that I have learned through our discussions, your advice, the countless hours shared in the L2 speech lab, and the opportunities to present at numerous conferences.

I am deeply grateful to Valeria, Gonzalo and Radha. Without your help, it would not have been possible to collect such a large dataset. Furthermore, I would also like to express my gratitude to the other *Llengua Anglesa II* teachers (Dr. Mayya Levkina and Marc Miret) and Dr. Cristina Aliaga for allowing me to go to their classes for participant recruitment. It goes without saying thanks to all my students from *Llengua Anglesa II* (2021-2022 & 2022-2023), the audience members (Isa, Berta, Dani, Sònia, Joan, Clara, Bridget, Gonzalo, Paulina, Josh, Radha, Neus and Jordi), and the native speakers who contributed to the value of this dissertation. This dissertation would not have been possible without all of you.

Last but not least, I extend my heartfelt thanks to my dear family and friends, who have always believed in me, encouraged me through challenging moments, celebrated each milestone with me, and brought laughter into even the most stressful times. Your friendship has made this experience even more meaningful. To my colleagues at work, who I also consider my dear friends, Valeria, Radha, Bridget, Katherine, Miren and Josh; and my friends Natalia, Marta, Elena, Júlia, Eli, Clara and Mercè. Completing this dissertation has been a challenging journey at times, and you have listened to me patiently as I discussed (and probably over-discussed) every detail of this project for so many years. Your patience, advice, and genuine interest have kept me motivated and reminded me of the bigger picture when I needed it most. I am so lucky to have you in my life!

A special thank you to Alex, my partner, whose support has also been invaluable throughout this journey. Whenever something did not work or when I needed help automating parts of the data analysis, you saved me countless hours with your expertise and "magical" solutions. Beyond the practical help, your belief in me and your endless patience, listening to my repeated concerns, have been a constant source of encouragement. Thank you for being by my side, both in this project and in life.

I would especially like to thank my parents, who have always believed in me, no matter the circumstances. They have taught me that with effort, patience, and determination, so much is possible. Your encouragement has been my foundation and knowing that you are always there cheering me on has been a source of strength throughout this journey. I am profoundly grateful for everything you have done for me, and I love you to the moon and back. - *Quisiera agradecer especialmente a mis padres, quienes siempre han creído en mí, sin importar las circunstancias. Me habéis enseñado que, con esfuerzo, paciencia y determinación, todo es posible. Vuestro apoyo ha sido mis cimientos en todos mis logros, y saber que siempre esteréis animándome me ha dado fuerzas a lo largo de este camino. Estoy profundamente agradecida por todo lo que habéis hecho por mí, os quiero hasta el infinito y más allá.*

The completion of this PhD thesis was made possible thank to the support of the Spanish Ministerio de Educación y Formación Porfesional, through a four-year predoctoral research grant (FPU19/00279) and the three-month stay-abroad grant (EST22/00559) in the University of Wroclaw, which was crucial in the successful completion of this work. The research experience I have acquired throughout these years is thanks to the participation in the research projects directed by Dr. Joan C. Mora (PID2019-107814GB-I00 and PID2022-138129NB-I00) from the Spanish Ministry of Science, Innovation and Universities (MCIN/AEI/10.13039/501100011033), and the grant 2023SGR00303 from the GRAL Research Group by the Catalan Agency for Management of University and Research Grants (AGAUR).

Barcelona, 1st March 2025

Abstract

Socio-psychological factors such as Foreign Language Anxiety (FLA) are an important source of individual differences in foreign language acquisition of (Horwitz et al., 1986). FLA is known to hinder language achievement by impairing learners' ability to perform in a FL class successfully (Baran-Łucarz, 2022; Horwitz et al., 1986; MacIntyre, 2017; Teimouri et al., 2019), and negatively interacting with cognitive processes underlying speaking, such as attention control (Kormos, 2015).

Despite extensive previous research on FLA, L2 speaking anxiety (defined as the feeling of nervousness when speaking the target FL) and the reason for its fluctuations have been under-researched, even in instructed FL classrooms where learners report speaking-oriented activities, especially oral presentations, to be highly anxiety-evoking (Tóth, 2017; Price, 1991; Young, 1990). Furthermore, few studies have examined the effects of task type and context, and the role of L2 speaking anxiety on L2 speaking fluency (e.g., Mora et al., 2023), comprehensibility, accentedness, and lexical and structural complexity. Moreover, proficiency and cognitive factors such as working memory and attention also contribute to individual differences in foreign language acquisition (Mora, 2022). However, it remains an under-researched empirical question whether higher proficiency, high working memory capacity and more efficient attention control provide an advantage to anxious learners when performing speaking tasks, resulting in more fluent speaking performance.

FLA has traditionally been studied as a static construct and has been assessed subjectively through questionnaires. However, it varies in intensity and duration and fluctuates over time. Consequently, more recent research includes new subjective tools (e.g., the idiodynamic task; MacInyre, 2012) and physiological measures of emotional arousal (e.g., heart rate). Nonetheless, the relationship between subjective and physiological measures of speaking anxiety is still not well understood. Similarly, studies investigating whether self-reported anxiety and others' perceptions align or the specific non-verbal cues used to detect nervousness are scarce.

The present study employs a mixed-methods approach to investigate five key areas related to L2 speaking anxiety: (1) the effects of task type, context and language; (2) the alignment between self-reported anxiety and others' perceptions, and the most prominent non-verbal anxiety cues; (3) the alignment between subjective and physiological measures; (4) anxiety effects on L2 speech; and (5) L2 speaking anxiety

sources. 86 EFL learners performed eight oral presentations, in counterbalanced order, under two presentation contexts (one-to-one vs. group, N = 7) and two task types (descriptive vs. argumentative). Four of these presentations were performed in their L2 (English) and the remaining four in their L1. In the descriptive tasks, learners had to explain past experiences (e.g., a tough challenge), while the argumentative tasks consisted of discussing three different pictures in relation to a topic (e.g., three modes of transportation within the city).

After each presentation, audience members perceptually assessed the level of nervousness in the speakers, while speakers self-assessed their distress using questionnaires and the idiodynamic task. In the latter, learners had to watch videos of their oral presentations and self-assess their moment-to-moment distress levels. Additionally, stimulated recalls were also conducted for the L2 oral presentations, allowing participants to reflect on anxiety episodes in their speeches. To capture multiple perspectives of L2 speaking anxiety, self-reports were accompanied with physiological measures, including heart rate and electrodermal activity (Boucsein, 1992). Apart from computing objective L2 speech measures (e.g., breakdown fluency), accentedness and comprehensibility were assessed through native speakers' ratings. In addition, individual differences controlled for included L2 proficiency (elicited Imitation; Wu et al., 2021), vocabulary size (yes/no vocabulary size test; Meara & Miralpeix, 2017), attention (trail making test; Reitan, 1958) and working memory (forward and backward digit span task; Woods et al., 2011).

Results indicated that physiological measures were largely unaffected by task type, context, and language, while subjective measures revealed significant main effects across all three speaking conditions. Anxiety was higher in argumentative tasks, particularly in group settings in the L2. Discrepancies were found between self-reported anxiety and external evaluations, especially during descriptive tasks. Additionally, the audience members, particularly the helper, perceived participants as more anxious than participants reported. While the audience identified many non-verbal cues of L2 speaking anxiety, the lack of eye contact with them was the most frequently reported sign. Regarding subjective and physiological measures of L2 anxiety, weak significant correlations were found between self-assessments and physiological measures of anxiety, suggesting a certain relationship between the measures. In this study negative relationships were found between anxiety and L2 speech. Hence, anxiety had a detrimental effect on fluency, comprehensibility, accentedness, accuracy, and lexical

variety. However, proficiency emerged as the strongest predictor of speaking performance. After controlling for L2 proficiency and cognitive factors (working memory and attention), pause frequency measures were detrimentally affected by FLA. Lastly, cognitive-related concerns emerged as the main sources of L2 speaking anxiety, with high-anxiety learners focusing on cognitive challenges, while low-anxiety learners emphasizing linguistic issues, such as pronunciation. Globally, the present study suggests an influence of L2 speaking anxiety on task performance and speech quality and paves the way for further research in L2 speaking anxiety to promote pedagogical practices aimed at reducing anxiety and its effects within foreign language classrooms.

Key words: L2 speaking anxiety; L2 speech; fluency; comprehensibility; accentedness; self-assessment; physiological measures; non-verbal cues; L2 speaking anxiety sources.

Resum

Els factors sociopsicològics, com l'ansietat en l'aprenentatge d'una llengua estrangera (FLA, per les seves sigles en anglès), representen una important font de diferències individuals en l'adquisició de llengües estrangeres (Horwitz et al., 1986). La FLA afecta negativament el rendiment en l'aprenentatge, ja que impedeix als estudiants desenvolupar-se amb èxit en una classe de llengua estrangera (Baran-Łucarz, 2022; Horwitz et al., 1986; MacIntyre, 2017; Teimouri et al., 2019) i interactua negativament amb processos cognitius subjacents a l’expressió oral, com el control de l'atenció (Kormos, 2015).

Malgrat l’extensa investigació prèvia sobre la FLA, l’ansietat al parlar en la L2 (*L2 speaking anxiety*; definida com la sensació de nerviosisme en parlar la llengua estrangera) i les raons de les seves fluctuacions han estat poc investigades, fins i tot en les aules on els estudiants consideren les activitats orals, especialment les presentacions orals, com el hi generen alts nivells d’ansietat (Tóth, 2017; Price, 1991; Young, 1990). A més, pocs estudis han examinat els efectes del tipus de tasca i el context, així com el paper de l’ansietat en la fluïdesa oral en L2 (e.g., Mora et al., 2023), la comprensibilitat, l’accent, i la complexitat lèxica i estructural. El nivell de competència i factors cognitius com la memòria operativa i l’atenció també contribueixen a les diferències individuals en l’adquisició de llengües estrangeres. No obstant, continua sent una pregunta empírica si un major nivell de competència i capacitat de memòria operativa, així com un control de l’atenció més eficient beneficien els estudiants més ansiosos durant les tasques orals, expressant-se així amb més fluïdesa.

La FLA s’ha estudiat tradicionalment com un constructe estàtic i s’ha avaluat subjectivament mitjançant qüestionaris. Tanmateix, aquesta varia en intensitat i durada, i fluctua amb el temps. En conseqüència, investigacions més recents inclouen noves eines subjectives (p. ex., la tasca idiodinàmica; MacIntyre, 2012) i mesures fisiològiques (p. ex., la freqüència cardíaca; Gallego et al., 2021). Tot i així, la relació entre les mesures subjectives i fisiològiques de l’ansietat en la parla segueix sense comprendre’s del tot. De la mateixa manera, hi ha escassos estudis que explorin si l’ansietat autoavaluada i la percepció dels altres coincideixen o els indicadors no verbals específics de l’ansietat que poden fer-se servir per detectar el nerviosisme.

Aquest estudi utilitza un enfocament de mètodes mixtos per investigar cinc àrees clau relacionades amb l’ansietat en la parla de la L2: (1) els efectes del tipus de tasca,

context i idioma; (2) l'alineació entre l'ansietat autoavaluada i la percepció dels altres, i els indicadors no verbals d'ansietat més prominents; (3) la relació entre les mesures subjectives i fisiològiques; (4) els efectes de l'ansietat en la parla de la L2; i (5) les causes d'ansietat en la parla de la L2. Un total de 86 estudiants d'anglès com a llengua estrangera van realitzar vuit presentacions orals en un ordre compensat, en dos contextos de presentació (un a un vs. davant d'un grup, *N* = 7) i dos tipus de tasca (descriptiva vs. argumentativa). Quatre d'aquestes presentacions es van fer en la L2 (anglès) i les quatre restants en la seva L1. En les tasques descriptives, els estudiants havien d'explicar experiències passades (p. ex., un repte difícil), mentre que les tasques argumentatives consistien en discutir tres imatges diferents en relació amb un tema (p. ex., tres formes de transport per moure's a la ciutat).

Després de cada presentació, els membres de l'audiència avaluaven el nivell de nerviosisme percebut, mentre que els participants s'autoavaluaven el seu nivell d'ansietat utilitzant qüestionaris i la tasca idiodinàmica (MacIntyre, 2012; Gregersen et al., 2014). En aquesta última, els estudiants veien els vídeos de les seves presentacions orals i autoavaluaven els seus nivells d'ansietat moment a moment. A més, es van fer entrevistes d'estimulació de la memòria en les presentacions de la L2, cosa que va permetre als participants reflexionar sobre els seus episodis d'ansietat durant les presentacions. Per capturar les múltiples perspectives de l'ansietat en la parla de la L2, aquest estudi inclou autoavaluacions juntament amb mesures fisiològiques, com la freqüència cardíaca i l'activitat electro-dèrmica (Boucsein, 1992). A més de calcular les mesures objectives de la parlar (ex., la fluïdesa), l'accent i la comprensibilitat han estat avaluades per parlants nadius. També es van controlar diferències individuals en el nivell de competència en la L2 (*elicited imitation*; Wu et al., 2021), el nivell de vocabulari (*yes/no vocabulary size test*; Meara & Miralpeix, 2017), l'atenció (*trail making test*; Reitan, 1958) i la memòria operativa (*forward and backward digit span task*; Woods et al., 2011).

Els resultats han mostrat que les mesures fisiològiques són majoritàriament insensibles al tipus de tasca, context i idioma, mentre que les mesures subjectives han revelat efectes significatius en les tres variables. L'ansietat va ser més alta en tasques argumentatives, especialment en davant l'audiència grupal en la L2. A més, es van trobar discrepàncies entre l'ansietat autoavaluada i les avaluacions externes, especialment a la tasca descriptiva. Així mateix, l'audiència, en concret l'ajudant d'aquest estudi, van percebre els participants més ansiosos del que ells van reportar. Tot i que l'audiència va identificar molts indicadors no verbals de l'ansietat en la parla de la L2, la falta de contacte visual va ser el més freqüentment comentat. Pel que fa a la relació entre les mesures subjectives i fisiològiques, aquestes són febles però suggereixen un cert nivell d'alineament entre les mesures. A més, en aquest estudi s'han trobat relacions negatives entre l'ansietat i les mesures de la parla. De fet, l'ansietat va afectar negativament la fluïdesa, comprensibilitat, accent, i varietat lèxica. Tot i això, el nivell de competència explica una major variabilitat interindividual en la producció oral. Després de controlar la competència en L2 i els factors cognitius (memòria operativa i atenció), la freqüència de pauses es va veure afectada negativament per la FLA. Per últim, entre les principals fonts d'ansietat, els estudiants van manifestar preocupacions cognitives com els bloqueigs mentals. Els estudiants més ansiosos es van enfocar en els desafiaments cognitius, mentre que aquells amb menys ansietat van destacar qüestions lingüístiques, com la pronunciació. En conjunt, aquest estudi suggereix que l'ansietat en la parla de la L2 influeix en el rendiment en les tasques i en la qualitat de la parla. A més, aquest estudi posa les bases per a futures investigacions en l'ansietat en la parla de la L2 amb l'objectiu de fomentar pràctiques pedagògiques que la mitiguin, així com els seus efectes en les classes de llengües estrangeres.

Paraules clau: Ansietat en la parla de la L2; producció en la L2; fluïdesa; comprensibilitat; accent; autoavaluacions; mesures fisiològiques; indicadors no verbals; causes de l'ansietat en la parla de la L2.

Table of contents

List of abbreviations

Analysis of Speech Units (AS-unit or ASU)

Analysis of Speech (AS)

Age of Onset of Learning (AOL)

Articulation rate (AR)

Autonomic nervous system (ANS)

Beats Per Minute (BPM)

Blood volume pulse (BVP)

Common European Framework of Reference (CEFR)

Communication apprehension (CA)

Digit Span Test (DST)

English as a Foreign Language (EFL)

Electrocardiogram (ECG)

Electrodermal activity (EDA)

Elicited Imitation Task (EIT)

First Language (L1)

Foreign Language (FL)

Foreign Language Anxiety (FLA)

Foreign Language Classroom Anxiety Scale (FLCAS)

Galvanic Skin Response (GSR)

Generalized Linear Mixed Models (GLMM)

Grup de Recerca en Adquisicó de Llengües (GRAL)

Heart Rate (HR)

Heart Rate Variability (HRV)

Individual Differences (IDs)

Language Background Questionnaire (LBQ)

Mean length of runs (MLoR)

Native Speaker (NS)

Non-Native Speaker (NNS)

Personal Report of Communication Apprehension (PRCA)

Phonation-time ratio (PTR)

Research Question (RQ)

Second Language (L2)

Second Language Acquisition (SLA)

Skin Conductance (SC)

Skin Conductance Level (SCL)

Skin Conductance Response (SCR)

Subjective Units of Distress Scale (SUDS)

Speech Rate (SR)

Trail Making Test (TMT)

Trier Social stress Test (TSST)

Working Memory (WM)

List of figures

List of tables

Introduction

Anxiety is one of those topics on which significant differences of opinion can be found. Some people believe that anxiety is a minor inconvenience for a language student, perhaps an excuse for not participating in class or a guise to hide a lack of study. Others seem to feel that anxiety may be the linchpin of the entire affective reaction to language learning…
(MacIntyre, 1999, p. 24).

Empirical investigation in second language acquisition (SLA) has demonstrated that second language (L2) learning, either in naturalistic language learning (L2 immersion contexts) or in classroom language instruction contexts (instructed SLA), is conditioned by a series of factors that can account for variance in learners' L2 attainment and performance (Mora, 2022). These factors include linguistic experience in terms of use, quantity and quality of input in real communicative contexts (Gass & Mackey, 2006), age of onset (Muñoz & Singleton, 2011), and variables of cognitive and psychological nature such as aptitude (Granena & Long, 2013; Granena et al., 2016), motivation (Saito et al., 2017), and cognitive individual differences in executive functions that influence SLA (Juffs & Harrington, 2011; Robinson, 2001).

In recent years, there has been a growing trend towards examining another important source of individual differences: affective factors. Evidence shows that affective factors, such as anxiety (Horwitz et al., 1986), enjoyment (Dewaele & MacIntyre, 2014; MacIntyre et al., 2016), boredom (Dewaele et al., 2023) or flow (Csíkszentmihályi, 2008; Horwitz, 2017) can significantly affect language learning, development and achievement in instructed SLA (Baran-Łucarz, 2022; Teimouri et al., 2019). This shift highlights the understanding that language acquisition is not solely a cognitive process but is deeply intertwined with learners' emotional experiences and perceptions, which can significantly impact learners' motivation and engagement. Additionally, although affective factors have traditionally been studied as a static construct and assessed subjectively through questionnaires, they vary in intensity and duration and fluctuate over time (Gregersen et al. 2014).

The goal of many foreign language (FL) teachers is to prepare learners for successful communication outside the classroom. However, learning conditions in classroom instruction take place under low input conditions for an optimal linguistic development, especially regarding oral competence (i.e., fluency and pronunciation). For that reason, affective factors (such as enjoyment) can play a larger role in linguistic development through classroom instruction than age or input (Muñoz, 2008). Despite the prevalence of positive emotions (e.g., enjoyment and flow) in the L2 classroom (Dewaele & MacIntyre, 2014; MacIntyre et al., 2016), L2 speakers can also experience negative feelings as they are learning a second language. One of these negative psychological factors that have a decisive influence on SLA is anxiety. Although anxiety in second language acquisition (known as *Foreign Language Anxiety* or FLA) has been extensively studied as a construct in relation to achievement in a FL (Teimouri et al., 2019), it still remains as one of the more difficult affective variables to characterize and operationalize, which affects both young and adult English as a foreign language (EFL) learners (Arnaiz & Pérez-Luzardo, 2014; MacIntyre & Gardner, 1994b; Young, 1991).

Moreover, despite the growing interest in FLA, L2 speaking anxiety has received comparatively little attention. Although teachers frequently prioritized skills such as grammar, reading, and writing over communicative competence, there is an increasing emphasis on the communicative language teaching approach in instructed SLA nowadays. This approach highlights the importance of interaction and the practical use of language in real-life contexts, aiming to enhance learners' ability to communicate effectively and confidently. Therefore, integrating speaking tasks into the curriculum has become essential in classroom language learning contexts. However, FL learners who experience speaking anxiety may not benefit as much from speaking tasks since anxiety disrupts learners' ability to perform in a FL class successfully (Horwitz et al., 1986).

Considering that learners often report experiencing distress during speaking-oriented activities, especially oral presentations (Tóth, 2017; Price, 1991; Young, 1990), there is an urgent need to investigate which task types (Sosa-López & Mora, 2022a; Tóth, 2017) alongside contexts (Beatty & Behnke, 1991) can be more detrimental for EFL learners' speaking performance and overall language learning experience. For instance, when speaking activities such argumentative tasks are included in our classrooms, they may unintentionally be anxiety-inducing (Tóth, 2017). Hence, these activities risk overwhelming our students, limiting their ability to actively engage, practice, and internalize the language.

Furthermore, the impact of L2 speaking anxiety on L2 speech acquisition and language learning is still under-researched in instructed SLA, both short and long-term. However, empirical studies on FLA have found that L2 speaking anxiety impacts L2 speaking fluency (e.g., Bielak, 2022; Pérez-Castillejo, 2019; Mora et al., 2023; Szyszka

& Lintunen, 2023), comprehensibility (Saito et al., 2018; Mora et al., 2023; Nagle et al., 2022), accentedness (Mora et al., 2023; Suzukida & Saito, 2023), and lexical and structural complexity (Hewitt & Stephenson, 2012; Kormos & Dörnyei, 2004; Phillips, 1992). Therefore, understanding the role of speaking anxiety on L2 speaking performance is important not only as a contribution to explaining SLA, but also in terms of its pedagogical implications. For example, educators who identify anxious students can adjust assessments and adapt instruction to better support anxious learners' needs.

This thesis seeks to contribute to the growing FLA line of research by examining the effects of L2 speaking anxiety on L2 speech (more specifically speaking fluency, accentedness and comprehensibility) under two different communicative situations, which vary in the extent to which they can potentially lead to speaking anxiety levels. First, the findings from this study intend to provide a better understanding of the potential detrimental effects of anxiety (MacIntyre, 2017), while contemplating the role of individual differences such as proficiency and cognitive factors in L2 speech (Mora, 2022; Mora et al., 2023). Second, this study follows up on Gallego et al. (2021) and Gregersen et al. (2014) by incorporating physiological and subjective anxiety measures to provide an in-depth perspective on L2 speaking anxiety and expands previous research on the alignment between self-reported anxiety and others' perceptions (Gregersen, 2007a, 2009) and non-verbal cues of L2 speaking anxiety (Gregersen, 2005; Gregersen, 2007b; Gregersen et al., 2017). Last but not least, this dissertation seeks to complement quantitative results by offering insights into learners' sources of L2 speaking anxiety when performing oral presentations through qualitative reports (Tóth, 2009).

Outline of the dissertation

The current dissertation is organized into three main parts: the background to the study, the study itself and the discussion and conclusions. *Part I: Background to the study* comprises two chapters that present the theoretical framework for the dissertation. Chapter 1 introduces the theoretical background on Foreign Language Anxiety (FLA) and L2 speaking anxiety, exploring the former's role as an individual difference. This chapter also reviews FLA's relation to overall L2 achievement, socio-biographical variables and cognitive factors, its non-verbal cues, causes, and its measurement. Chapter 2 focuses on L2 speaking anxiety and its effects on L2 speech, examining fluency, accentedness and comprehensibility, as well as lexical and structural complexity. *Part II: The study* outlines the objectives and the research questions (Chapter 3), details the methodology and the data collection procedure (Chapter 4), and presents the results (Chapter 5). Finally, *Part III: Discussion and conclusions* consists of Chapter 6 on discussion and Chapter 7 on concluding remarks. Chapter 6 provides the reader with the discussion and interpretation of the study's findings, comparing them with previously published empirical evidence. Finally, Chapter 7 summarizes the main findings while proposing pedagogical implications, acknowledging some study limitations, suggesting further research and offering concluding remarks.

PART I

BACKGROUND TO THE STUDY

CHAPTER 1: Literature review on FLA and L2 speaking anxiety

"Aptitude can influence anxiety,
anxiety can influence performance,
and performance can influence anxiety"
(MacIntyre, 1995, p. 95).

This chapter explores the multifaceted phenomenon of anxiety, beginning with a broad overview in Section 1.1. The focus narrows to foreign language anxiety (FLA) in Section 1.2, which is further divided into several subsections: FLA as an individual difference (1.2.1), FLA and L2 achievement (1.2.2,) the interplay between FLA and socio-biographical variables (1.2.3) and the FLA and cognitive factors (1.2.4). Section 1.3 offers an in-depth examination of L2 speaking anxiety. Section 1.4 provides a comprehensive overview of the non-verbal cues associated with anxiety, while Section 1.5 focuses on the different sources of anxiety. Finally, Section 1.6 is dedicated to the measurement of FLA, which includes subsections that address both subjective measurements and objective measurements.

1.1. The multiple facets of anxiety

The term anxiety is a complex multifaceted psychological and social construct which encompasses different dimensions including clinical anxiety, psychological constructs such as social anxiety, trait anxiety and state anxiety, as well as language-related anxieties such as communication apprehension and foreign language anxiety. Thus, attempting to encapsulate its complex nature into a single definition is a challenging task.

However, research defines anxiety as the "subjective feeling of tension, nervousness, and worry associated with an arousal of the autonomic nervous system (ANS)" (Spielberger, 1983, p.15). This definition captures the combination of subjectiveness (i.e., the feeling) and objectiveness (i.e., the arousal) that belongs to anxiety. Before moving forward, it is important to grasp the fundamentals of the ANS which is comprised of the sympathetic and the parasympathetic nervous systems. On one hand, the sympathetic nervous system is in charge of activating the physical anxiety reaction by presenting physiological symptoms such as sweaty palms, gastrointestinal discomfort, trembling, and increases in heart rate (Croft et al., 2004; MacIntyre et al., 2010; Young, 1992). On the other hand, the parasympathetic nervous system is responsible for the body's recovery after periods of stress (e.g., slowing heart and respiratory rate). Besides the physiological responses mentioned, anxiety is also connected with our mind and beliefs. That is, experiencing an overwhelming sense of apprehension is closely connected to self-esteem issues (Horwitz, 2017; Sellers, 2000; Tóth 2008, 2017). To illustrate, when learning a new language, a learner's self-esteem can be vulnerable when realizing that that their communicative choices are restricted.

Over the years, many differentiations related to the term "anxiety" have emerged to reflect the specific characteristics or contexts in which the nervous arousal takes place. In this section, we will define psychological and learning-related anxieties, including *trait anxiety, state anxiety, situation-specific anxiety*, as well as *facilitative and debilitative anxiety*. Foreign language anxiety will be defined in *Section 1.2*.

A well-known differentiation is that between *trait anxiety* and *state anxiety* (Spielberger 1966, 1983). *Trait anxiety* can be simply defined as an individual's tendency to be anxious in any situation. Trait anxiety is considered as a stable emotional state that affects a wide range of situations, which may vary in intensity, but is part of the individual's personality and may affect cognitive, affective and behavioral functioning in general (MacIntyre & Gardner, 1991a). In relation to FL learning, people with trait anxiety are likely to feel nervous in the language learning context because of their tendency to interpret every situation as dangerous and threatening (MacIntyre & Gardner, 1989). *State anxiety* refers to a transient reaction closely related to a situation in a specific moment. Additionally, state anxiety changes in intensity and duration, and fluctuates over time depending on how an individual interprets the stressful situation (Spielberger, 1966). Thus, an arousal or burst of anxiety would be provoked by a stimulus in a particular situation at a particular point in time.

The term *situation-specific anxiety* is typically considered a subtype of state anxiety and refers to temporary feelings of anxiety that arise in response to a particular context or event. However, unlike state anxiety, situation-specific anxiety happens consistently over time in this specific context (MacIntyre & Gardner 1991a; Horwitz, 2017). For example, when students are learning a new language, they might experience situation-specific anxiety when they are required to speak in public in the FL. This anxiety is specific to the context of public speaking in the FL and may not be present in other situations, such as during oral tasks with peers in the FL.

The terms *facilitative anxiety* and *debilitative anxiety* were conceptualized by Alpert & Haber (1960) early in the 60s when they asserted that anxiety per se may not always be detrimental to the learners' performance or learning process. These researchers were the first to propose a positive effect of anxiety adding bidirectionality to this multi-faceted construct. As the name indicates, facilitative anxiety influences learners positively, and as Alpert & Haber (1960) believed, would emotionally stimulate L2 learners to face, for instance, challenges such as new learning tasks, and accept risks as part of their process of learning. Debilitative anxiety, by contrast, induces unpleasant feelings (apprehension, dread, etc.) that interfere with the learning process and may ultimately hinder successful performance. To illustrate, FL learners might avoid a particular learning task, withdraw from FL classes or feel insecure when participating. Although a slight degree of arousal may be beneficial for certain FL learners in maintaining their engagement and preventing boredom, it is challenging to express an unequivocal belief that anxiety is facilitative, especially when FLA has predominantly been identified as a negative predictor of learners' FL achievement (Horwitz et al., 1986; MacIntyre, 2017; Teimouri et al., 2019).

1.2. Foreign Language Anxiety (FLA)

The emergence of the concept of foreign language anxiety (FLA) marked a significant development in the field of SLA to specifically address the sensation of apprehension experienced in foreign language contexts. The construct of FLA was developed by Horwitz et al., (1986), and is defined as "a distinct complex of self-perceptions, beliefs, feelings and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (Horwitz et al., 1986, p.128). In research, FLA has also been referred to as "foreign language classroom anxiety" or just "language anxiety" and has been measured through the Foreign Language Classroom Anxiety Scale (FLCAS; see *Section 1.6.1.1* for further details).

According to Horwitz et al. (2009), FLA has been conceptualized in two ways. Firstly, it has been seen as a manifestation or transfer of other types of anxiety, such as trait anxiety, test anxiety, or public speaking anxiety into the language learning context. However, FLA does not correlate with other constructs such as general trait anxiety and other situation-specific anxieties such as test anxiety (Horwitz, 2000, 2001). This indicates that FLA is not merely a transfer of anxiety from another domains such as test anxiety or communication apprehension but is a unique L2-related type of anxiety.

Alternatively, FLA can be understood as a *situation-specific anxiety*, specific to language learning situations. According to MacIntyre & Garnder (1991a, 1991c), anxiety starts as an undifferentiated, negative affective response to some experience in the language class. However, when this negative response is repeated over time, it leads to the association of this negative feeling with the language class, becoming differentiated from other contexts. Consequently, FLA as a *situation-specific anxiety* emerges over time due to its repeated experience in the L2 classroom. Moreover, FLA is conceptually similar to other situation-specific anxieties, such as social anxiety, communicative apprehension and test anxiety (Botes et al. 2020; Horwitz et al. 2009; Gkonou, 2017). The author of this dissertation understands FLA as the latter, meaning that it is a distinct type of anxiety (i.e., not a transfer) specifically related to L2 learning, which arises in situations specific to the FL learning process.

FLA has been extensively studied in SLA. In order to trace the evolution of this construct over time, it is helpful to divide the literature into three broad approaches, which reflect the historical trends and underlying assumptions about FLA (MacIntyre, 2017). The initial phase is called the *Confounded Approach*, with its focal point being Scovel's (1978) review of language anxiety literature. Since the 1960s, scholars have shown interest in the impact of anxiety on second language learning and performance. However, during this period, research on FLA and its effects drew from various sources without meticulous consideration of the construct of anxiety *for* language learners.

It was not until the 1970s that a comprehensive review of all literature available on language and anxiety was published (Scovel, 1978), pointing out important avenues for future research. The review found mixed results, as some studies reported a negative relationship between anxiety and course grades, which were taken as measure of FL achievement, whereas other studies found no relationship (Tucker et al., 1976) or even positive relationships between anxiety and language achievement (Chastain, 1975; Kleinmann, 1977). In response to these findings, Scovel argued that the discrepancies could be attributed to the lack of reliable and valid measurement instruments, as well as problematic definitions of anxiety. Scovel's arguments were entirely reasoned, particularly when considering that anxiety is multifaceted, and the concept of FLA had not yet been coined. In fact, Scovel (1978) observed that some of the measures available (e.g., indicators of physiological arousal and self-report questionnaires) for anxiety were predominantly derived from psychology research and held little relevance to language studies. Consequently, he emphasized the need for future research to specify the type of anxiety researchers intended to measure.

Undoubtedly, the introduction of FLA as a construct and subsequent development of the Foreign Language Classroom Anxiety Scale or FLCAS to measure it (Horwitz et al., 1986) solved Scovel's concerns. This marked the beginning of the second trend, called the *Specialized Approach*. In this period anxiety experiences that were specifically liked to language were identified, defined and investigated. Furthermore, research during this phase focused on (1) the sources of anxiety (for learners MacIntyre, 2017; and FL teachers Horwitz, 1996; Goetze, 2013), (2) the development of skill-specific anxiety constructs (later in this Section), and (3) its effect on language learning (e.g., MacIntyre & Gardner, 1991b; Steinberg & Horwitz, 1986), cognitive processing (MacIntyre & Gardner, 1994b) and social behavior in class (Ely, 1986).

Nonetheless, it is important to acknowledge that during this phase not all researchers endorsed the notion of FLA as a unique type of anxiety, separated from difficulties in the learner's native language. Richard Sparks, Leonore Ganschow, and their colleagues took a systematic line of research focusing on what they labeled the Linguistic Coding Differences Hypothesis (LCDH). The LCDH posits that the ability to learn an L2 is closely linked to a person's skills in their L1, and challenges in learning an L2 arise from difficulties in the native language (L1) (Sparks & Ganschow, 1991). Thus, these researchers relegated FLA to the status of an unfortunate side effect caused by the learning difficulties present in the learners' L1 (Sparks & Ganschow 1991, 1996; Sparks, Ganschow & Javorsky, 2000).

However, these arguments have been refuted repeatedly by MacIntyre (1995a, 1995b; MacIntyre & Gardner, 1991a, 1991b) and Horwitz (2000, 2001), who assert that FLA arises from the specific challenges associated with learning and using a FL rather than being a mere side effect. Thus, they assert that FLA is independent of L1 learning difficulties and highlight the fact that it is not only L2 learners with learning difficulties that experience anxiety in foreign language learning (MacIntyre and Gardner, 1994a). Two experimental studies conducted by MacIntyre and Gardner (1994a, 1991b) hold significance in rejecting the LCDH. Building on the theory proposed by Tobias (1979, 1986), which suggests that anxiety impacts input processing and output production, MacIntyre and Gardner (1994a) divided seventy-two learners into four groups with FLA intentionally aroused by introducing a video camera at the three key stages (input, processing, and output) in a vocabulary learning task: (1) during their initial exposure to the stimuli, (2) when beginning to learn the meanings of the words, (3) when producing the words, and (4) a control group that did not experience anxiety arousal. The three experimental groups showed significant increases in anxiety when the video camera was introduced, and those increases were associated with deficits in vocabulary acquisition. However, as the learners adapted to the presence of the video camera, some recovery from the effect of anxiety were observed. In MacIntyre & Gardner (1991b), the researchers found that anxiety limited short and long-term memory functions such as recalling number sequences and vocabulary items in the L2, yet it did not impede performance in L1 versions of the same tasks. These findings were used to refute the idea that FLA is a result of difficulties in the L1, suggesting instead that anxiety itself is an independent factor that contributes to performance deficits.

During the *Specialized Approach* phase, skill-specific anxieties were identified within the broad spectrum of FLA, and subsequently, specific tools were developed taking the FLCAS as a point of departure. For example, Elkhafaifi (2005) found foreign language listening anxiety and FLA were two distinguishable constructs, and that L2 listening anxiety affected the students' listening comprehension scores as well as their final grades as a measure of overall achievement. Similarly, foreign language reading anxiety was distinguished from FLA by Saito et al. (1999), who found that foreign language reading anxiety arose from the inability to create a 'sound-symbol' correspondence with an unfamiliar writing system or the inability to fully understand the contextual meaning of the text, despite knowing the individual words.

Regarding writing, Cheng et al. (1999) differentiated FLA from foreign language writing anxiety and found that foreign language writing anxiety is explained mostly by learners' confidence in English writing (Cheng 2002, 2004). Concerning speaking, Woodrow's (2003, 2006) results revealed FLA and foreign language speaking anxiety were two distinguishable constructs. In fact, her findings further supported the viability of differentiating between in-class and out-of-class foreign language speaking anxiety. While Woodrow highlighted this distinction, it is worth noting that when Horwitz et al. (1986) developed the FLCAS (explained in *Section 1.6.1.1*), the scale predominantly included speaking-related items (Gkonou, 2014). This emphasis on speaking suggests that FLA was conceptualized mainly bearing in mind the anxiety experienced by learners during speaking activities.

Finally, the third phase, the *Dynamic Approach*, is a relatively recent line of research in which FLA is studied from a dynamic perspective. This approach is in line with the current theories and developments in the field of SLA following the Complex Dynamic Systems Theory, which encompasses four key features (see Larsen-Freeman, 1997; Larsen-Freeman & Cameron, 2008 for detailed accounts):

1. Dynamic systems are inherently unstable, undergoing continual change over time, with each state representing a continuation of the preceding one.
2. Within such systems, variables are interconnected and exert mutual influence on each other.
3. The dynamic system tends to self-organize into two states. The states that are favored are known as *attractor states*, and the non-favored are called *repeller states* or *perturbations*.
4. Additionally, there exists a phenomenon called the *butterfly effect*, where even a minor alteration in one part of the system can lead to significant and unforeseen changes throughout the entire system, and conversely.

Following these key features, FLA studies focusing on this dynamic trend focus not only on the temporal evolution of a construct, but also its interconnections with other variables. FLA is regarded as a dynamic construct because it is an emotion that *fluctuates* over time, with rises and falls during an event such as a communicative situation. Moreover, these fluctuations can be observed across different timescales, ranging from seconds to longer periods such as months or even years (MacIntyre, 2017).

In addition, FLA's dynamism is not only due to moment-by-moment fluctuations, but also to the fact that all the varieties of anxiety, including state and trait anxiety can interact (Gregersen, 2020), and contradictory elements can coexist, such as experiencing FLA and enjoyment (see Dewaele & MacIntyre, 2014; MacIntyre and Gregersen, 2012; Boudreau et al., 2018) or boredom (Dewaele et al., 2023).

These fluctuations can be due to several contextual variables along with L2 learners' IDs, which align with the second and fourth key aspects of the Complex Dynamic Systems Theory. In other words, an anxiety reaction can be caused by interactions of multiple factors, including the learner variables and the ongoing features of the communicative situation. Finally, in reference to the third key aspect, self-organization in dynamic systems, Gregersen (2020) exemplifies some FLA "perturbations" can be negative self-comparisons with other learners, fear of making mistakes and learners' desire for teacher approval, among others.

An example of a FLA study capturing this dynamicity is Gregersen et al. (2014) which analyzed the FLA levels in six learners while performing an oral presentation in their L2. After task performance, the learners evaluated their video recordings using the idiodynamic method (See Section *1.6.2.2* for further details) and results showed several interesting fluctuations in the data, which demonstrated the value of adopting a dynamic approach. Nevertheless, as MacIntyre (2017) acknowledges, there have not been many studies in this dynamic tradition, despite the initial promising results (see Gregersen et al., 2014; MacIntyre, 2012). Hence, the researcher indicates the need for more studies and a reevaluation of the notions of causality as the language learning experience unfolds, given that learners' experiences with language and communication are continuous.

1.2.1 FLA as individual difference

Even when exposed to identical classroom environments, with the same instructor, and using identical educational materials, individuals exhibit considerable variance in their L2 acquisition and performance. For this reason, SLA scholars have investigated the sources of individual differences (IDs) in FL learning since the 60s (Dörnyei, 2005). FLA is an affective ID, whose impact may vary depending on other IDs such as proficiency, working memory (WM), and attention control. For instance, individuals with greater WM capacity may be more effective at regulating anxiety while simultaneously processing and producing speech, as they can allocate cognitive resources more efficiently.

As the term suggests, IDs are characteristics that are assumed to apply to everybody, but individuals vary in the extent to which they manifest them (Dörnyei, 2005). To illustrate, these include personality, aptitude, motivation, learning styles and learning strategies, anxiety, and learner beliefs, among others (Dörnyei, 2005). IDs can be categorized into affective or psycho-social (i.e., motivation, personality), cognitive (i.e., aptitude, WM, attention, inhibition) and personality- and experiential-related factors (i.e., extraversion vs. introversion, age of onset, learning context) (Dörnyei, 2005; Gardner, 1985; Moyer, 2014a). As Mora (2022) points out, these factors may exert varying degrees of influence across different stages of acquisition.

Traditionally, IDs were considered to be stable, unchanging aspects of individual learners. However, Dörnyei (2009) reconceptualized IDs as part of interdependent dynamic systems, which means that they can change over time as learners interact with their learning environment. Hence, while some IDs are relatively stable over time (i.e., cognitive factors, aptitude), other factors such as proficiency or psycho-social factors (i.e., motivation and anxiety) can change over time depending on learners' experiences (see Figure 1). To illustrate, learners' motivation might fluctuate based on their experiences in the classroom, their progress, or their interactions with others. Similarly, learners' anxiety levels might decrease as they gain more confidence and experience in the language.

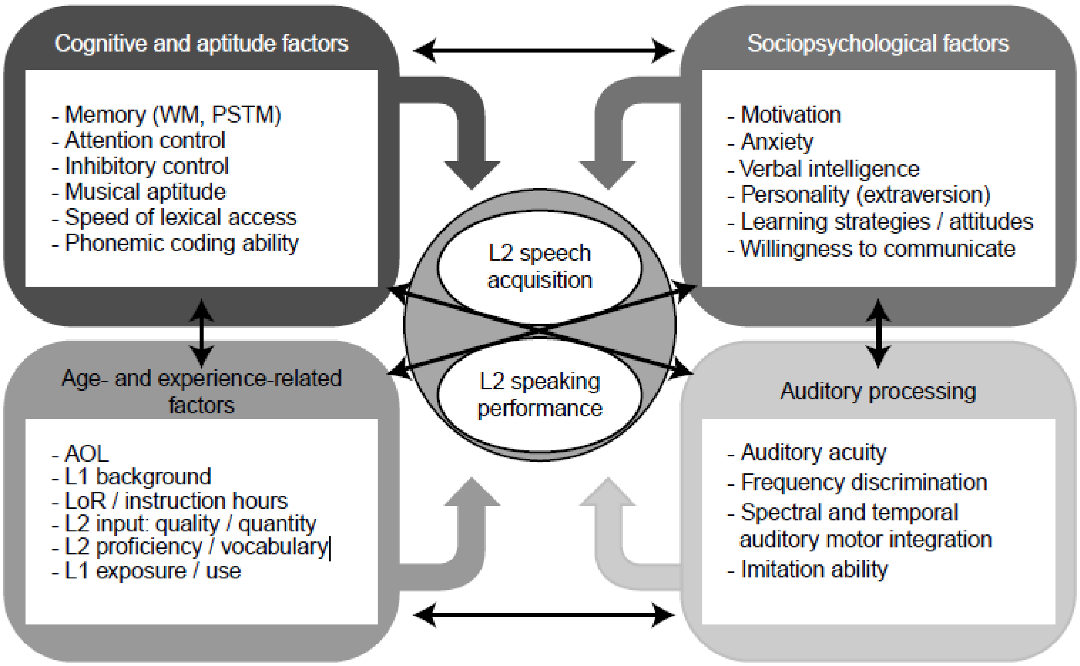


Figure 1: IDs in L2 speech acquisition and performance (from Mora, 2022, p. 72)

Concerning experiential factors, extensive research on L2-immersion settings has focused on age- and experience-related factors and has shown their significance in influencing L2 speech attainment (see Moyer, 2014a, 2014b). Such factors include, but are not limited to, proficiency, the age of onset of L2 learning (AOL), length of residence in the L2 context, frequency of L2 use, and the quantity and quality of L2 input received. However, as Mora (2022) reports, in the case of FL classrooms, age and experience-related factors appear to contribute to a lesser extent to L2 speech attainment due to the limited exposure and production opportunities (i.e., Larson-Hall, 2008; Mora, 2006; Saito, 2015; Suzukida & Saito, 2023). Nevertheless, research has found interesting results such as late starters outperforming early starters in oral fluency (Mora, 2006) or in a situation of minimal input, a younger starting age making a difference to phonological abilities measured through a phonemic discrimination test (Larson-Hall, 2008).

Besides age, experience-related factors also play a key role (Suzukida & Saito, 2023). Language proficiency significantly influences L2 speech production, encompassing various aspects such as fluency (see *Section 2.5.1*) and overall communicative effectiveness. L2 proficiency can be measured through different instruments (e.g., Oxford Placement Test, International English Language Testing System, etc.); however, Ortega et al. (2002) created an Elicited Imitation Task (EIT) as an effective measure of L2 proficiency, the simplicity and economical administration procedures of which makes it an attractive candidate to use effectively both in and out of the classroom context. The EIT involves decoding a sentence, which has to be parsed syntactically and semantically, retaining its meaning and processing it for production. According to Granena (2016), the hypothesis behind imitative performance is that it shows the learners' existing knowledge of grammar (i.e., implicit knowledge); hence, failing to imitate accurately would demonstrate that the structure is beyond their level. Furthermore, given that the EIT taps implicit knowledge, WM should not affect accuracy in EIT performance. In fact, research conducted by Kim et al. (2016) revealed a non-significant correlation between WM and EIT, suggesting that the EIT is a valid and reliable proficiency measure to test proficiency without reliance on WM.

In terms of psycho-social factors, personality traits (e.g., extraversion), affective factors (e.g., FLA, willingness to communicate) or learners' attitudes (e.g., motivation) may have an impact on L2 speaking and L2 development. For example, a large body of research has found FLA to be detrimental for L2 speech and L2 achievement (Teimouri et al., 2019; Botes et al., 2020), whereas motivation has been found to be one of the strongest factors for success (Moyer, 2014), for instance, improving comprehensibility (Saito et al., 2017) or reducing foreign accent (Moyer, 2014b).

Like all IDs, FLA varies considerably among language learners, with each person having a unique threshold. While some learners may feel comfortable engaging in language learning activities, others may exhibit heightened nervousness and avoidance behaviors. This variability is influenced by individuals' personality traits, particularly those outlined in the Big Five Personality Model (Goldberg, 1990). The Big Five Personality Model is a widely accepted model in psychology for understanding human personality, which includes openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Among these traits, neuroticism is especially relevant because anxiety is a key component of this dimension (Botes et al., 2024; Dörnyei, 2005).

Thus, individuals can either be emotionally stable or more prone to neuroticism (Gkonou, 2017).

Nonetheless, in general, FLA poses a substantial challenge to an individual's self-concept as competent communicator (Horwitz et al., 1986), directly affecting learners' self-confidence when using the target language (Jin & Dewaele, 2018). This, in turn, perpetuates and strengthens FLA, creating a self-reinforcing "vicious" circle (Boudreau et al., 2018). In other words, experiencing FLA is self-exacerbating: as learners experience anxiety-arousal, they may become concerned about how it will affect them in the future. This worry can, ironically, increase their levels of anxiety. For this reason, an anxious learner may hesitate to participate in certain activities, avoid speaking in the target language, or feel reluctant to interact with others. Moreover, correlational research has shown that FLA interacts with and can be related to other IDs (see Tóth, 2007 for a review) such as willingness to communicate (MacIntyre & Charos, 1996), motivation (MacIntyre, 2002; Papi, 2010) and personality traits such as perfectionism (Gregersen & Horwitz, 2002; Dewaele, 2017), introversion (Dewaele, 2002b; MacIntyre & Charos, 1996), extraversion (Dewaele, 2013; Botes et al., 2024) and self-esteem (Dörnyei, 2005).

Cognitive IDs are differences in aspects of mental functioning, such as memorizing and remembering; inhibiting and focusing attention; speed of information processing; and spatial and causal reasoning (Granena et al., 2016). In fact, language aptitude, WM and attention are considered the most important mediating factors in L2 acquisition (Granena et al., 2016). Indeed, WM and attention are crucial in the acquisition of language skills by allocating attention during speech processing or production (Kormos, 2015). According to Kormos (2015), IDs in attention control will be reflected in the speed and accuracy of L2 speech production.

As Juffs and Harrington (2011) review, WM is implicated in L2 aptitude and overall proficiency (see Linck et al., 2014 for a meta-analysis; Kormos & Sáfár, 2008); however, its role varies as a function of L2 learners' age, the WM task employed (i.e., digit span test, non-word repetition task), and the linguistic domain examined. WM research is based on Baddeley's work (Baddeley & Hitch, 1974). WM consists of a central executive system in control of cognitive processes and three subsystems: the phonological loop, the visuospatial sketchpad, and the episodic buffer. The phonological loop is vital in L1 and L2 acquisition because it comprises a short-term storage for verbal information known as phonological memory (PSTM). PSTM encodes phonological elements (sounds) and their serial temporal order in the form of auditory memory traces that decay quickly, and an articulatory rehearsal component prevents these auditory traces from vanishing through subvocal rehearsal (i.e., a silent articulation mechanism) (Mora, 2022). Although Kormos (2015) reports conflicting evidence for the role of WM in L2 speech production, research has found that PSTM plays a role in L2 acquisition (Kormos & Sáfár, 2008), and correlates with L2 speech production gains, specifically, oral fluency (O'Brien et al., 2007) and more accurate L2 pronunciation and perception (Darcy et al., 2015).

Attention control plays a crucial role in speech processing and perceptual learning (Mora, 2022). According to Simard (2022), WM and attention have a close and complex relationship, whether it is in the form of the central executive in Baddeley's WM model or as a supporting role in each other's functioning. In that sense, attention is necessary for retaining language information in WM. Attention's characteristics, such as its limited capacity and selectiveness, have been examined in research for its important role in L2 development and L2 proficiency (Kormos, 2006; Segalowitz & Frenkiel-Fishman, 2005). As Simard (2022) points out several instruments, such as the Trail Making Test (Reitan, 1958), are available to measure attention in L2 research, each measuring one or more aspects of attention (e.g., attention switching, selective attention).

Research findings suggest that learners with more efficient attentional control may be able to allocate their attention to select relevant stimuli (selective attention) while communicating, ignoring irrelevant ones (inhibition), and efficiently shifting attention between tasks (attention shifting) (Segalowitz, 2010; Miyake & Friedman, 2012). Moreover, attention control may also help learners develop more accurate representations of L2 sounds as it helps them notice L2-specific phonetic features or cross-language differences between L2 and L1 sounds (Mora, 2022). For example, Mora & Darcy (2023) revealed that learners with efficient attention switching could discriminate target vowel contrasts faster and, although the researchers did not find an association between attention switching and discrimination accuracy, the latter was found to be significantly related to the production of the same contrasting vowels. Thus, attention switching may be more implicated in speech production than in speech perception.

To conclude, understanding the contributions of various IDs is crucial for gaining a comprehensive and detailed picture of the factors affecting L2 speech performance and development. Therefore, this dissertation investigates and controls for the extent to which certain cognitive and affective factors, (such as WM, attention and anxiety) recognized as significant determinants of individual variation in L2 speech literature, influence L2 performance.

1.2.2. FLA and L2 achievement

The interference of FLA with language acquisition and performance is one of the most accepted phenomena in psychology and education (Horwitz, 2000). For this reason, it is unsurprising that research on FLA has been predominantly concerned with examining the potential effect of anxiety on learners' L2 achievement, measured through course grades, language tests, and GPA[1].

Early research on anxiety revealed mixed findings with respect to FLA effects, indicating it could potentially be either detrimental (Gardner et al., 1976) or facilitative (Chastain, 1975) for language achievement. However, with the introduction of the FLCAS in 1986, more consistent findings emerged, revealing a negative relationship between FLA and L2 achievement across different target languages, including Japanese, French, English and Spanish (e.g., Aida, 1994; Coulombe, 2000; Phillips, 1992; Saito & Samimy, 1996; Saito et al., 1999; Young, 1986). Hence, experiencing FLA usually leads to lowered grades and poor academic achievement (MacIntyre, 2017). Notably, Coulombe (2000) found that this negative relationship is observable at various L2 proficiency levels (i.e., beginning, intermediate and advanced).

Moreover, two meta-analyses conducted by Teimouri et al. (2019) and Botes et al. (2020) yielded firm evidence of the negative role of FLA in course grades, language tests, and academic achievement in FL courses. Teimouri et al.'s (2019) analyses included 97 studies between 1985 and 2017 across 23 countries, and its findings indicated a moderate negative correlation (r = -.36) between FLA and achievement scores. This means that the participants with higher levels of FLA exhibited poorer performance on course grades, language tests, and GPA. In fact, FLA was found to account for 13% of the variance in L2 achievement scores.

A year later, Botes et al.'s (2020) meta-analysis focused on the relationship between FLA and general academic achievement plus the four competence-specific scores (reading, writing, listening, and speaking). A total of 67 studies from 1986 to 2018 were incorporated, and findings corroborated the negative relationship between FLA and overall academic achievement (r = -.39). Furthermore, significant moderate negative correlations were found between FLA and language-specific competency scores, with listening exhibiting the strongest correlation (r = -.52), followed by writing (r = -.43), reading (r = -.34), and finally speaking (r = -.25). Consequently, it can be concluded that

[1] Grade Point Average: A common scale used at colleges and high schools.

experiencing FLA is detrimental, and the more anxious learners are, the more likely they are to obtain lower achievement scores in language learning courses.

In addition to overall L2 academic achievement, previous research has explored the effects of FLA on L2 performance, as well as its psychological and cognitive impact (MacIntyre, 2017; Rubio-Alcalá, 2017; Steinberg & Horwitz, 1986; Tóth, 2011, 2017). Findings have revealed that when learners experience anxiety, they tend to describe visual stimuli less interpretively and are less willing to communicate (MacIntyre, 2017; Steinberg & Horwitz, 1986; Tóth, 2011) They also make more errors during oral performances (Gregersen, 2003) and experience other negative emotions in L2 classes (i.e., frustration) (Tóth, 2011) as well as mental blocks (MacIntyre, 2017) which can lead to difficulties in understanding interlocutors (Tóth, 2017) and retrieving vocabulary (MacIntyre & Gardner, 1994b).

FLA can also encompass physical and psychological effects, such as bodily reactions (e.g., palpitations and sweating), accompanied with feelings of insecurity, discomfort, and passive learning attitudes (Rubio-Alcalá, 2017). Finally, FLA has cognitive effects, including increased self-related cognition (i.e., thoughts of failure and performance worry) (MacIntyre, 2017) and difficulties in processing information (MacIntyre, 2017; MacIntyre & Gardner, 1994b). For example, anxious learners have more difficulties processing information quickly and accurately due to anxiety acting as a filter. Hence, they may require more time to recognize words, hold them in short-term memory, and recall vocabulary and grammar rules from long-term memory (MacIntyre, 2017; MacIntyre & Gardner, 1994b). As a result, students with high levels of anxiety may perform poorly compared to less anxious learners.

Given the negative impact of FLA on language acquisition, it is crucial to investigate strategies that both teachers and learners can employ to prevent and manage these adverse effects. By identifying and implementing effective techniques, educators can create a more supportive learning environment that reduces anxiety, while learners can develop coping mechanisms that enhance their confidence when performing in the target language.

To being with, anxious learners feel uneasy in classroom settings due to the anticipation of scrutiny by others, which is closely tied to self-esteem. In other words, highly anxious learners tend to focus on their perceived limitations and fear of failure, leading them to worry about being worried (Oxford, 2017). Oxford (2017) explored possible strategies in relation to positive psychology (i.e., optimism and emotional intelligence) which could potentially alter students' perceptions to minimize FLA. One of the main recommendations put forth by Oxford (2017) involves fostering positive self-talk, such as repeating "I want to learn this language because it is interesting and challenging". This practice is believed not only to boost motivation, but also to engage a constructive mindset that will support effective language learning. Moreover, by fostering positive self-talk, learners can maintain enthusiasm and resilience, which are crucial for overcoming challenges and achieving proficiency in the L2.

On the teachers' side, teachers should adopt a learner-centered approach to encourage cooperative work and prevent limiting the chances for oral interaction (Rubio-Alcalá, 2017; Horwitz, 2017). Changing from a teacher- to a learner-centered approach would lessen the atmosphere of "intimidation" in class and students would be more involved in their language learning process. Thus, it would benefit anxious learners who set excessively high standards for themselves and have unrealistic expectations about the level of proficiency they wish to achieve (Oxford, 2017).

Moreover, teachers also play a pivotal role in instructing learning strategies to identify what Oxford (2017) calls learners' "signature strengths", which can help mitigate FLA. Other important suggestions include helping students develop realistic expectations for language learning (Price, 1991), correcting mistakes gently, fostering small-group and pair activities rather than whole-class activities, and enhancing learners' emotional security. The latter can be achieved through "active-constructive responding", characterized by enthusiastic responses, together with praises, and assertive communication when required (Oxford, 2017; Rubio-Alcalá, 2017). This provides students with reassurance and support. Horwitz (2017) further advises acknowledging students' discomfort and openly discussing their experiences in class to reduce the sense of isolation often associated with anxiety. This way students would feel reassured that their feelings are normal and (possibly) shared by others, creating a supportive and empathetic classroom environment. Moreover, as Sellers (2000) points out, once students identify their fears, they can gradually learn to view anxiety-inducing situations more realistically, which may encourage them to approach rather than avoid these challenging experiences.

In terms of assessment, Rubio-Alcalá (2017) suggests providing guidelines prior to evaluation, conducting mock exams and presenting the evaluation criteria together with a rubric to mitigate anxiety. These actions promote clarity, predictability, and preparedness among students. Thus, students understand what is expected of them, can get familiarized with the assessment criteria, and practice their skills. In addition, the researcher recommends initiating oral exams with interactive activities, such as short personal questions, which contribute significantly to reducing students' anxiety levels. While these strategies can be effectively applied within the classroom, another promising approach is the use of virtual reality (VR) for anxiety training. This technology is not widely implemented in schools due to its challenging technical configuration, high initial costs, and the need to integrate VR into the curriculum, which requires time and resources for both setup and training. However, VR has proven to be an efficient tool for anxiety management by simulating real-world scenarios, allowing students to experience and practice high-anxiety triggering tasks, such as public speaking, in a controlled environment (Dhimolea et al., 2022; Harris et al., 2002; Kaplan-Rakowski & Gruber, 2023; Parmaxi et al., 2023; Quiu et al., 2023). All in all, this technology would enable students to gradually build their confidence in speaking situations, providing them with an opportunity to manage anxiety in a low stake setting before facing real classroom assessments.

These strategies represent only a fraction of the myriad approaches available for mitigating the impact of FLA within the classroom setting. By implementing these practices and exploring additional interventions, teachers can promote a supportive learning environment which will minimize FLA and foster students' confidence during their L2 acquisition journey.

1.2.3. FLA and socio-biographical variables

This section aims to briefly review some results of the main research on FLA and socio-biographical variables: age, gender, proficiency and multilingualism. FLA is a complex and often intense experience with both internal and social dimensions. It arises from different factors, including internal physiological processes, cognitive and emotional states, as well as situational demands and the presence of others, among other factors (MacIntyre, 2017). Dewaele (2007) aligns with this view, adding that FLA is partly influenced by the individual's personality, partly by situational factors (e.g., formality of situation, number, and identity of interlocutors), as well as by socio-educational and political contexts.

Hence, research on IDs in FLA has two main strands of investigation, *intra-individual* variability or situational variables and *inter-individual* variability or learner variables. Studies that analyze intra-individual variability examine the type of tasks and situations that are likely anxiety-provoking; meanwhile, inter-individual variation studies focus on socio-biographical variables via questionnaires, observations, reflective journals and/or interviews, with the aim to identify learner characteristics that would induce greater FLA. Both strands of research interact in multiple ways since FLA is subject to both internal and external factors. This dissertation targets intra-individual variability and situational variables; however, understanding socio-biographical variables can help achieve a better understanding of individuals' susceptibility to suffer from FLA (Jin & Dewaele, 2018).

Research findings on the role of age in FLA are mixed. Many studies have found older learners (e.g., university students) to report higher levels of FLA than younger learners (e.g., high school students) (Bailey et al., 2000; Dewaele, 2007; Donovan & MacIntyre, 2005; Onwuegbuzie et al., 1999). These findings may be attributed to the greater self-awareness older learners have regarding how they are perceived by others, leading them to put more pressure on themselves and have higher expectations of their performance. However, other studies have found the opposite (Arnaiz & Guillén, 2012) or have failed to find a relationship between age and FLA in classroom settings (Dewaele et al., 2008) and immersion (MacIntyre et al., 2002).

Similarly, research has yielded inconclusive findings regarding gender and FLA. While some studies suggest that female learners exhibit higher levels of FLA compared to their male counterparts (Arnaiz & Guillén, 2012; Dewaele, 2007; Dewaele & MacIntyre, 2014; Dewaele et al., 2016; Park & French, 2013), and interestingly, female learners can experience more enjoyment than male students in class (Dewaele et al., 2016), other studies indicate that male learners experience higher levels of FLA than female learners (MacIntyre et al., 2002; Kitano, 2001) or have found no gender differences (Dewaele & Al-Saraj, 2015; Dewaele, 2017; Donovan & MacIntyre, 2005; Park & French, 2013).

These mixed findings regarding age and gender in FLA studies may be attributed to a variety of personality traits, such as perfectionism (Gregersen & Horwitz, 2002; Dewaele, 2017), and experiential factors (e.g., first year vs. last year of high school, previous negative learning experiences). All these factors can significantly shape attitudes towards mistakes and influence students' feelings of pressure to excel in class, fear of not meeting expectations, and doubt in their own L2 competence (Ewald, 2007; Tóth, 2011).

Concerning proficiency, research indicates that low proficiency learners often experience high levels of FLA due to linguistic limitations (e.g., gaps in lexical knowledge) (Arnaiz & Guillén, 2012; Frantzen & Magnan, 2005; Sevinç, 2018), especially when they have limited opportunities to use the language regularly (Sevinç, 2018). However, findings also show that as learners become more proficient, their FLA may either decrease (Dewaele & Al Saraj, 2015; Dewaele & Meftah, 2023; Gardner et al., 1977; Liu, 2006) or, conversely, increase (Ewald, 2007; Jee, 2014; Kitano, 2001; Marcos-Llinás & Garau, 2009; Saito & Samimy, 1996; Tóth, 2009, 2011, 2017). Lower anxiety in more proficient learners can be explained by increased familiarity with the tasks (e.g., at school) and enhanced confidence in the speaking environment. However, higher anxiety among advanced students can be due to the increasingly challenging linguistic demands of the learning environment (Tóth, 2009).

Finally, multilingualism appears to yield straightforward outcomes regarding FLA. Research by Dewaele (2007) and Dewaele et al. (2008) demonstrated that knowing more languages was associated with lower reports of anxiety in different situations, such as speaking with friends, interacting with strangers, and speaking in public. Moreover, multilingualism enhanced learners' self-confidence and self-perceived competence (Dewaele, 2007), particularly when learners socialized more frequently in the new language (Dewaele et al., 2008).

In conclusion, the relationship between factors such as age, gender and proficiency with anxiety is complex, which justifies the need for further research. Socio-biographical variables are important because they may affect the measurement of FLA and should be taken into account in when planning FLA research and participant recruitment.

1.2.4. FLA and cognitive factors

Cognitive factors, such as aptitude, working memory (WM), inhibition and attention, are an important source of individual differences in foreign language acquisition (Carroll & Sapon, 1959; Juffs & Harrington, 2011; Kormos, 2015; Mora, 2022). As Kormos (2015) highlights, these factors are key predictors of success in language learning, and in terms of L2 speech production, they play a crucial role in facilitating effective communication. However, anxiety disrupts these cognitive functions, making it more difficult to focus, process information, and communicate in the foreign language. Previous research has linked WM and attention to anxiety. Therefore, before examining the interplay between FLA and these cognitive factors, it is important to provide a comprehensive overview of WM and attention.

As explained in *Section 1.2.1*, the most widely accepted conceptualization of short-term memory today is the working memory model developed by Baddeley & Hitch (1974) (for a review, see Baddeley, 2003). WM involves the temporary storage and manipulation of information needed for a wide range of complex cognitive activities, for example, when solving a math problem (Baddeley, 2003). The WM model proposed by Baddeley & Hitch (1974) comprises a multi-component memory system consisting of the central executive (responsible for coordinating the phonological loop), the visuo-spatial sketchpad, and the episodic buffer. While the phonological loop is in charge of the manipulation and retention of speech, the visuo-spatial sketchpad works with visual and spatial information. The central executive performs several functions, including attentional control, directing the flow of information through the system and planning. Thus, WM plays an important role in cognitive activities such as comprehension, reasoning and learning (Kormos & Sáfár, 2008).

Attention is the capacity to focus on relevant stimuli while ignoring competing information (Segalowitz & Frenkiel-Fishman, 2005). Traditionally, theories in SLA have viewed attention as the executive control mechanism that directs the sequential transfer of information from short-term memory to long-term memory. There are three conceptualizations of attention: firstly, depicting the selection of information (where attention is directed towards items for further processing); secondly, describing the capacity of attentional resources (varying levels of attention can be available depending on the task); and thirdly, describing the effort needed to sustain attention towards task objectives (attention levels can be maintained or can decline over time). While distinct, these uses of attention are interrelated, each one related to separate functions, but frequently operating concurrently (see Robinson, 2003 for a review).

Cognitive factors such as WM and attention are inextricably linked to emotions (Dewaele, 2002b; Kormos, 2015; Simard et al., 2023; Zuniga & Simard, 2022). For example, FLA is known to interfere with the attentional resource pools available for speaking, hindering learners' ability to focus on their own output and, in turn, notice gaps in their L2 knowledge (Kormos, 2015). In terms of WM, the negative effect of anxiety on WM was first raised by Eysenck (1979) with his Processing Efficiency Theory (later elaborated as Attentional Control Theory) which remarks that anxiety has a greater effect on processing efficiency than on performance effectiveness. That is, anxiety may not detract from task performance accuracy, but efficiency decreases as more resources have to be invested to attain a given performance level. Thus, experiencing anxiety imposes a substantial demand on the processing and storage capacity of WM, and according to Eysenck et al. (2007) this has one main effect: anxiety induces disruptive thinking filled with worrisome and irrelevant thoughts, overloading the functioning of cognitive process, and consuming the limited attentional resources of WM. These attentional resources, in turn, are less available for concurrent task performance resulting in more errors. Plenty of research on cognition, psychology and neuroscience have related lack of attention control to anxiety demonstrating how it impairs one's ability to inhibit the processing of task-irrelevant information (see Shi et al., 2019 for a review). However, less research has focused on cognitive factors and anxiety in relation to foreign language acquisition.

According to Tobias (1979, 1986), anxiety not only affects WM and attention when producing output, but also at the input and speech processing stages. For this reason, MacIntyre & Gardner (1994b) developed the Input, Processing, and Output Anxiety Scale (IPOAS), an 18-item questionnaire that tackled L2 French anxiety related to these three stages. For example, the input anxiety scale includes statements such as "I get upset when I read in French because I must read things again and again". The processing anxiety scale includes statements like, "I am anxious with French because, no matter how hard I try, I have trouble understanding it", and the output anxiety scale includes statements such as "I may know the proper expression but when I am nervous it just won't come out" (MacIntyre & Gardner, 1994b, p. 304-305).

Spoken interactions offer learners the possibility to expand their knowledge through reflecting on their output and feedback (see Swain's Output Hypothesis, 2005). Nevertheless, usually L2 learners experience limitations on WM and attentional capacity that makes it challenging for them to attend to their own speech production processes, manage the interaction, and process feedback into their language system. If we add FLA to this picture, these limitations might be exacerbated. To illustrate, MacIntyre and Gardner (1994a, 1994b) found that when experiencing anxiety, learners had difficulties retaining new words or maintaining focus on sentence structure. These difficulties were attributed to anxiety hindering learners' attentional resources. Another example is Sheen's (2008) study which showed how high-anxiety learners modified their utterances significantly less than the low-anxiety learners after a recast had been provided. The researcher argued that this finding lends support to Krashen's Affective Filter Hypothesis

(Monitor Model Theory, 1982), suggesting that negative feelings such as anxiety would create a filter or barrier that prevents FL input from being processed cognitively.

Focusing on output, speech production is traditionally explained to consist of four important processes (see Levelt, 1989 for an L1 production model; De Bot, 1992 for a bilingual production model; Kormos, 1999 for an L1-L2 production model): the conceptualizer, the formulator, the articulator, the acoustic-phonetic processor, and the parser (see Figure 2). Briefly summarized, the conceptualizer is in charge of generating and monitoring the communicative intention and the message. The formulator is responsible for giving grammatical and phonological shape to the conceived message. The articulator turns the internal speech (or phonetic plan) into overt speech through the speech organs (lungs, vocal tract, larynx, etc.). The acoustic-phonetic processor transforms the acoustic signal into phonetic representations and, finally, the parser monitors the correctness and appropriateness of the overt speech output. However, spontaneous self-repairs show that we can monitor our internal speech (Levelt et al., 1999). All these processes require certain automaticity on the part of the speaker, since each stage must be accomplished in a limited time during which speakers need simultaneously to plan their message and map their thoughts into appropriate language. As Kormos (2006) reviews, there is an agreement that conceptualization, formulation, and articulation of an intended message follow each other, and that in L1 production planning the message requires attention. Formulating and articulating are automatic; therefore, this is a fast process in our L1.

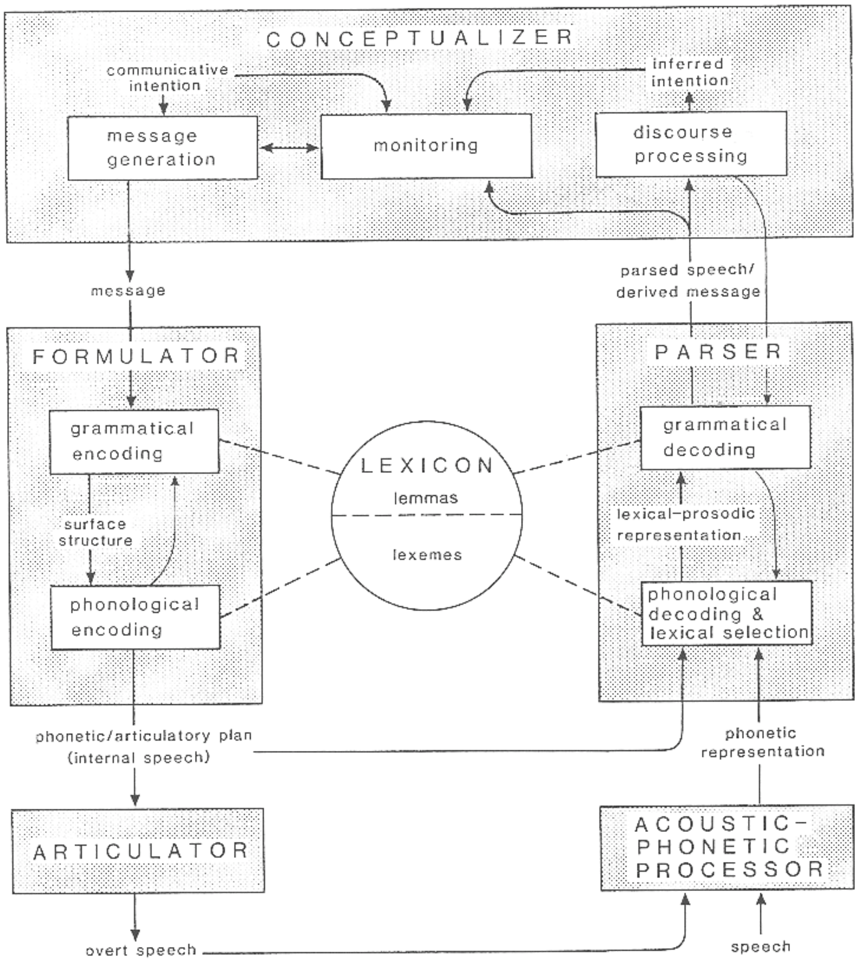


Figure 2: Levelt's (1989, p. 9) model of speech production.

On the other hand, L2 learners generally suffer from lack of automatization and exhibit resource deficits (e.g., they may not have the lexical resources they need to speak fluently). Hence, L2 production requires more attentional control, as the L2 resource pool may be insufficient or may need to be consciously recalled (Kormos, 2006, 2015). Since speech production processes require conscious control that rely on WM and attention, and FLA interferes with those, anxious learners find it more effortful to retrieve words from their mental lexicon and encode syntactic structures (MacIntyre and Gardner, 1991b; MacIntyre & Gardner, 1994b; MacIntyre, 2017; Tóth, 2017), or to inhibit their L1. Furthermore, Robinson (2007) found that IPOAS questions related to output anxiety and processing anxiety correlated with the complexity and accuracy of oral production, measured as clauses per C-unit and error free C-units. However, input anxiety showed no such correlation.

FLA interference may also result in increased time-gaining mechanisms such as pausing (i.e., Mora et al, 2023; Pérez Castillejo, 2019) or a detriment in the overall quality of L2 speech by producing more self-repairs (Zuniga & Simard, 2019, 2022), making the speech production process more difficult (Kormos, 2015). In turn, experiencing this difficulty getting the message across may increase learners' FLA, creating a self-reinforcing circle. To exemplify, Zuniga & Simard's (2022) study investigated the relationship between attention shift capacity, FLA and the production of self-repairs. After creating two separate groups of participants based on the results obtained on the attention shift task, findings revealed that learners in the high attention group had lower mean anxiety scores. However, only for anxious learners with low attentional control, FLA was positively related to the production of self-repairs. These results reflect the findings from Zuniga & Simard (2019), who also observed fewer instances of disfluencies in the speech of participants with faster attention shift capacity, supporting their assertion that quicker attention shift facilitates more fluid pre-articulatory processes, consequently reducing the need for self-repairs. However, to the best of the author's knowledge, it remains an under-researched empirical question whether both high WM capacity and more efficient attention control provide an advantage to anxious learners when performing speaking tasks, resulting in more fluent speaking performance.

1.3. The construct of L2 speaking anxiety

L2 speaking anxiety is the feeling of apprehension, usually accompanied by physiological responses, when using the L2 in a communicative situation, for instance, giving an oral presentation or interacting with native speakers (Woodrow, 2006). L2 speaking anxiety[2] may be conceptualized as either a skill-specific or a situation-specific anxiety. Nonetheless, it remains intrinsically linked to FLA owing to its manifestation within the context of FL use and acquisition. In light of this association with FLA, there is also a dynamic interplay between L2 speaking anxiety and other related anxieties such as fear of negative evaluation and social anxiety (Gkonou, 2014), which directly affects individuals' willingness to communicate (MacIntyre & Charos, 1996).

L2 speaking anxiety is related to low self-perceptions of speaking ability (Gkonou, 2014). Previous research has unveiled that learners consistently report

[2] Throughout this dissertation, the terms *L2 speaking anxiety* or *anxiety* will be used to refer to the specific conceptualization of the construct by Woodrow (2006).

experiencing higher levels of L2 speaking anxiety when they perceive their performance in oral tasks to be less competent than that of their peers (Beatty, 1988; Gkonou, 2014; Gregersen, 2003; Kitano, 2001; Yan & Horwitz, 2008). This is in line with other survey-based studies which have also found significant negative correlations between anxiety and self-perceived speaking ability (Kitano, 2001) or oral performance (Woodrow, 2006).

Given the importance of English as an international lingua franca nowadays, most English language teachers prioritize speaking tasks in their classrooms. While speaking tasks in an EFL classroom tend to be rather enjoyable activities, it is known that learners may experience anxiety during speaking tasks, and this anxiety can be detrimental for their performance and, ultimately, their learning progress (Ewald, 2007; Horwitz et al., 1986; Gkonou, 2014, 2017; Phillips, 1992; Piechurska-Kuciel, 2008; Tóth, 2011, 2012, 2017; Woodrow, 2006).

Previous research has consistently shown that speaking is the most anxiety-provoking language skill (MacIntyre & Gardner,1991c; Young, 1990, 1992). This is because speaking is a skill characterized by being under time pressure, requiring leaners to formulate responses when asked a question in class or during other speaking circumstances. Given that learners are typically less proficient and fluent in the L2 than in their L1 (Segalowitz, 2010), not being able to find the appropriate words and structures when attempting to speak, may cause them to missed their turn. Also, anxious learners are more likely to subjectively appraise speaking activities negatively in the L2 classroom because they are critical of their communicative competence (Beatty, 1988; Gkonou, 2014; Gregersen, 2003; Kitano, 2001; Yan & Horwitz, 2008) and fear negative evaluation from others. Certainly, speaking is the skill in which learners' egos are most vulnerable due to the higher level of self-exposure that it imposes on them (Gkonou, 2014; Price, 1991). Consequently, students might not want to participate in class and will miss opportunities to develop their communicative skills during lessons.

Despite the growing interest in L2 speaking anxiety, researchers have addressed FLA primarily from broad perspective. More research is needed to further explore how L2 speaking anxiety affects L2 speech, anxiety sources, and the relationship between different measures of anxiety. The intricate interplay between psychological factors and linguistic development holds an immense potential to shape learners' language learning processes. For this reason, tackling this gap in L2 speaking anxiety research is key to uncovering invaluable insights into effective pedagogical approaches, learner strategies, and the overall improvement of language learning methodologies.

1.4. Non-verbal anxiety cues

Anxiety is one reliable predictor of language learning achievement, and the literature suggests various strategies to mitigate its effects (see *Section 1.2.1*). However, addressing this issue depends on first identifying anxiety in our students through non-verbal behavior. In fact, according to Gregersen et al. (2017), non-verbal behavior is the main mechanism through which individuals communicate emotions while learning a language

Anxiety has predominantly been assessed through introspection. These subjective measures can also be used to investigate others' perceptions of this emotion, such as identifying the most prominent nonverbal cues and examining the extent to which external perceptions align with self-perceptions. The latter is particularly interesting since according to King and Smith (2017), anxious learners are hypersensitive to their surroundings and tend to overestimate the extent to which their anxiety symptoms are visible to others fueled by negatively biased thoughts. In fact, Savitsky and Gilovich (2003) called this idea the "illusion of transparency" and unveiled that when learners are explicitly informed of their negatively biased thoughts, they self-evaluate their speeches more positively. Moreover, once informed, learners were also seen as more composed by observers, resulting in more favorable evaluations of their performance.

Despite the existent studies in psychology regarding the illusion of transparency, to the best of the author's knowledge, only Gregersen et al. (2017) have investigated the degree of alignment between external perceptions by a peer and a teacher and self-perceptions of nervousness in SLA. However, this study only included a cohort of six participants and correlations were not tested for significance. Instead, Gregersen et al. (2017) focused on examining *when* the alignment of anxiety ratings between participants and observers took place using the idiodynamic task. Interestingly, their findings revealed that the patterns of self-ratings were more variable than the peer or expert ratings, implying that observers had difficulties in identifying the fluctuations of distress experienced by the presenter (see Figure 3). These challenges in recognizing distress, as identified by Gregersen et al. (2017) can be due to individual variations in the extent to which individuals display and perceive an emotion. In other words, some learners may be hypersensitive to their own perceived anxiety and overestimate it, while others may be lenient in assessing anxiety when the signs are not easily perceptible.

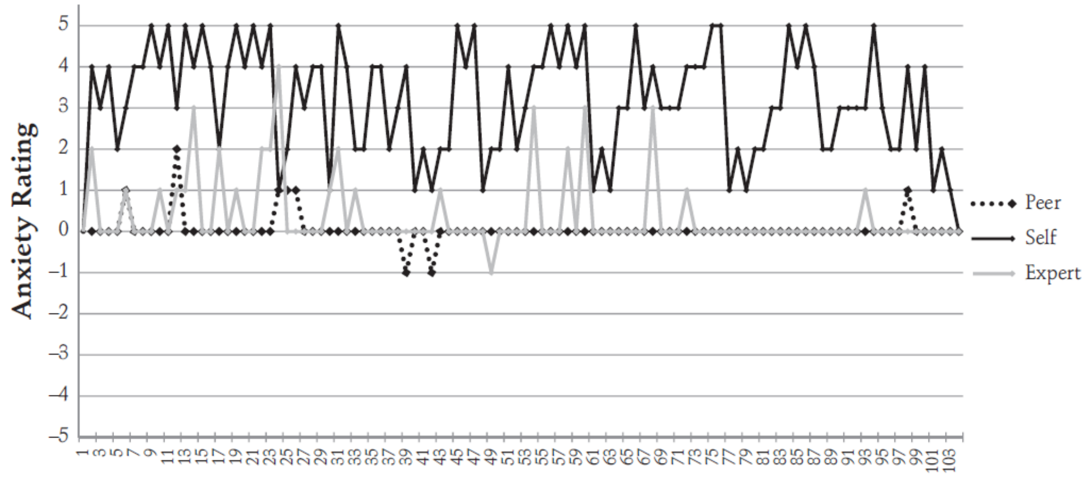


Figure 3: Self- and observers' idiodynamic anxiety ratings (from Gregersen et al., 2017, p. 120)

When experiencing anxiety, learners often report a myriad of unpleasant physiological symptoms (King & Smith, 2017; Simsek & Dörnyei, 2017), in addition to cognitive, communicative and behavioral symptoms (Gregersen, 2005, 2007b, 2009; Gregersen et al., 2017; Lee & Kleinsmith, 2019; Lee et al., 2020b; Lindberg et al., 2021; Rubio-Alcalá, 2017; see Table 1 for a summary). In extreme cases, anxiety may result in severe symptoms such as unexplained pain, muscle stiffness, or even fainting. The fact of being aware of these symptoms or cues can be employed to train educators in recognizing anxious language students in the classroom and in responding to the classroom's emotional atmosphere. Hence, teachers could provide effective support by planning future interventions.

Although not all these anxiety cues are always perceptible to external observers, behavioral symptoms of anxiety are more visible by just observing students than any other symptoms such as physiological ones, which are challenging to discern unless verbally reported or manifested as nervous tics (Rubio-Alcalá, 2017). The most common examples of anxiety cues encompass avoiding eye contact with people, limited facial expressions in the brow and mouth regions, fidgeting, restlessness and the adoption of closed or rigid postures, such as crossing their legs (Gregersen 2007a, 2009; Gregersen et al., 2017). On the other hand, facial expressions involving smiling and brow movement, and head nods are found to be representative of low arousal (Lindberg et al., 2021).

Table 1: Summary of anxiety symptoms from Gregersen's studies (2005, 2007a, 2009), Lindberg et al. (2021) and Rubio-Alcalá (2017)

Physical symptoms	- High heart rate - Stomach pain - Trembling voice - Blushing - Shaking of hands or legs - Clammy hands - Sweating - Higher pitch
Cognitive symptoms	- Feelings of confusion - Mental blanks - Distorted thoughts - Failing to recall learnt material
Communicative symptoms	- Freezing and blocking - Stuttering - Hesitations and pausing - Sensitivity to error correction - Avoiding participating - Adopting passive language learning attitudes - Rhythm: either speaking more slowly or more rapidly
Behavioral symptoms	- Touching one's hair continually - Fidgeting - Facial gestures such as tensing the jaw and frowning - Avoiding eye contact - Limited facial expressions - Reduced use of gestures - Rigid postures - Shifting weight continually

To further illustrate, Gregersen (2005) gathered evidence through observing the behavior of anxious and non-anxious language learners when taking an oral language exam, a high-anxiety-evoking situation. She discovered that anxious language learners exhibited diminished control over their eye movements and were less inclined to establish eye contact. This finding is in line with Lindberg et al. (2021) who also reported increased blinking behavior in anxious learners. Indeed, language teachers may find this anxiety cue familiar, as it is often evident in students who actively avoid making eye contact by looking elsewhere when confronted with the prospect of speaking (Gregersen, 2007b; Lee & Kleinsmith, 2019; Lee et al., 2020b). With respect to gestures, Gregersen (2005) noticed that anxious learners gesticulated more with their hands, but for purposes unrelated to communication, such as self-touching (e.g., hair-twisting) and object manipulation. Their counterparts used gestures with the purpose of enhancing communication.

Teachers and teacher trainees are generally effective at recognizing signs of anxiety. However, as unveiled by Gregersen (2007a, 2009), observers might struggle when interpreting gestures and body language, particularly closed body positions, such as crossed legs and arms. Another question raised in Gregersen (2009) was whether observers relied more on auditory (e.g., raising pitch, more filled and silent pauses) or visual cues (gestures, fidgeting) in the identification of anxiety, and her findings revealed that observers were more accurate decoding anxiety cues in visual-only (73.9%), followed closely by the combination of audio-visual modality (73.5%). In contrast, accuracy significantly decreased under the audio-only condition (69.6%). Hence, visual signals are more salient than auditory cues alone.

In conclusion, understanding how external observers, such as teachers, can detect and interpret anxiety cues can lead to more tailored and responsive strategies that address the specific needs of anxious learners. However, research to date is limited and several questions remain unanswered, such as: What are the *actual correlations* between self- vs. other's impressions of anxiety? In a large sample of participants, how well do students' self-perceptions accurately reflect others' impressions? In addition, what are the most frequently mentioned non-verbal cues by a *group* of observers?

1.5. Causes of FLA

Since FLA poses a significant obstacle in the language acquisition process, numerous researchers have been motivated to explore the underlying causes of anxiety, especially during the *Specialized Approach* period. Numerous factors contribute to the emergence of FLA, including, but not limited to, learner characteristics (e.g., being a perfectionist), instructional methods (e.g., classroom dynamics, teaching styles and evaluation practices), and prior language learning experiences.

Previous research has categorized FLA sources as either internal or external factors (Tóth, 2020; Szyszka, 2017; Young, 1990, 1991). However, MacIntyre (2017) proposed an alternative classification that goes beyond this internal-external dichotomy: academic, social, and cognitive causes. Nonethless, several of the reported causes by MacIntyre (2017) closely resemble, if not align directly with, those identified by Young (1990). This section will be based on Young's (1990, 1991) findings, and organized following the *internal* versus *external* dichotomy (see Table 2 for a summary).

Table 2: Summary of FLA causes

Internal causes	• Personal and interpersonal anxieties (e.g., low self-esteem, perfectionism) • Learner beliefs • Fear of negative evaluation
External causes	• Teacher beliefs • Instructor-learner interactions • Classroom procedures • Language testing • Speaking-oriented activities: oral presentations

Internal factors refer to the sources that originate within the learner. For example, Young (1991) identified personal and interpersonal anxieties, as well as learner beliefs about language learning, as the most frequent internal sources of FLA. Personal and interpersonal anxieties are fears that learners experience that arise from low self-esteem and competitiveness (Bailey, 1983; MacIntyre, 2017; Piechurska-Kuciel, 2008; Tóth, 2011) or fear of failure and perfectionism (Gregersen & Horwitz, 2002), and are very present in social interactions (e.g., learners have to speak in front of peers) (Young, 1991). Both personal and interpersonal anxieties can contribute significantly to causing FLA. To illustrate, numerous studies have shown low self-perceptions of L2 academic competence (e.g., Krashen, 1982; Onwuegbuzie et al., 1999; Price, 1991) or learning abilities (e.g., Baran-Łucarz, 2011; Gkonou, 2013; Kitano, 2001; Szyszka, 2011) lead to experiencing high levels of FLA.

Learner beliefs about language learning can also shape their experience and feelings towards the foreign language. As Young (1991) explains, sometimes L2 learners may set unrealistic learning goals which can lead to anxiety and frustration, especially if learners present personality traits such as perfectionism (Gregersen & Horwitz, 2002). To further illustrate, advanced learners in Tóth (2011) exhibited perfectionist tendencies and revealed that their major source of anxiety was perceiving differences between the language they used in the classroom and the language used outside of it. Moreover, these learners feared appearing less competent than others and mainly aimed at avoiding mistakes.

Another important internal source of FLA is the fear of negative evaluation, especially when speaking in the L2 (e.g., Aida, 1994; Horwitz et al., 1986; Kitano, 2001; Mak, 2011; Tóth, 2011). To illustrate, L2 learners consistently report being anxious and concerned about being incorrect in front of their peers and making a fool of themselves (MacIntyre, 2017; Young, 1991). This phenomenon is particularly present during oral activities in the classroom, where the learners experience heightened self-exposure, and they are afraid of being spotlighted in front of others (Young, 1990). Fear of negative evaluation as a source of anxiety is closely linked to an external factor, namely conflicts within peer relationships (Rubio-Alcalá, 2017). In other words, if learners find themselves in a classroom environment where they do not feel accepted or connected to their peers, they will be prone to developing a diminished sense of belonging, lower self-concept, and heightened FLA.

External FLA sources are those that arise from the learning context, such as the type of activities, the classroom atmosphere, the teacher, etc. For instance, Young (1991) identified as external causes of FLA: teacher beliefs about language learning, instructor-learner interactions, classroom procedures, and language testing. FLA may arise due to disparities between learners' and teachers' perceptions of effective learning. However, anxiety is most likely to emerge in situations where the teacher is perceived as intimidating and unsupportive, especially during error correction (Gkonou, 2013; Kitano, 2001; Tóth, 2011; Rubio-Alcalá, 2017; Szyszka, 2017). Consequently, a teacher who leans more towards criticism may contribute to heightened levels of language anxiety among learners. Other teacher practices that significantly contribute to the emergence of language anxiety in the classroom are overcorrection, failure to encourage collaborative work, and an overuse of teacher-talk (Young, 1991). Regarding language testing, it is very common that learners experience nervousness when being evaluated (Gkonou, 2013;

Rubio-Alcalá, 2017; Pérez Castillejo, 2019; Szyszka, 2017; Young, 1990). However, higher levels of anxiety are present when learners are unfamiliar with the evaluation criteria, the test items, test formats or question types (Szyszka, 2017).

Another important external source of FLA are class activities, especially speaking-oriented tasks. As proficiency in speaking a FL is highly valued in our society, teachers emphasize its importance in their classrooms and incorporate activities such as oral presentations and role-plays to create communicative opportunities resembling real-life interactions. However, as Tóth (2019) reflects, teachers find themselves in a dilemma, wherein certain classroom activities seem ideal for maximizing communicative opportunities in the TL but may not be ideal for addressing learners' affective needs, particularly, minimizing their anxiety.

Speaking-oriented activities are consistently regarded as the most anxiety-inducing for L2 language learners, in comparison to any other in-class activity (Arnaiz & Pérez-Luzardo, 2014; Gkonou, 2011, 2013; Koch & Terrell, 1991; Liu & Jackson, 2008; Mak, 2011; MacIntyre & Gardner, 1991c; Price, 1991; Tian, 2019; Woodrow 2003, 2006; Young, 1990, 1991, 1992), regardless of their proficiency level (Tóth, 2009, 2011) or the instructional setting (Young, 1990). In fact, MacIntyre and Gardner (1991c) found that 87% of their participants regarded speaking as the most anxiety-evoking situation when reflecting on their L2 experiences.

Although being called on in class or participating in class discussions can cause nervousness (Tóth, 2010; Woodrow, 2006), speaking on the spot and speaking in front of the class are often highlighted as the main sources of L2 speaking anxiety (Young 1990, 1992; Woodrow, 2006). Oral presentations are great learner-centered activities that provide learners with the opportunity to speak for an extended period (Koch & Terrell, 1991). However, this class or assessment activity is often associated with apprehension, even when performed in the native language (L1; Croft et al., 2004; McCroskey, 1970, 1978; McCroskey et al. 1985; Westenberg et al., 2009). Despite this, to the best of the author's knowledge, comparisons of whether performing oral presentations in the L1 and L2 induce similar anxiety levels are not found in the literatura.

Besides the fear of negative evaluation, the question remains as to why oral presentations are so anxiety-evoking. To the best of the author's knowledge, only Tóth (2019) has conducted a comprehensive investigation to explore the specific sources of anxiety undergraduate EFL learners experience after performing a 5 - 10-minute oral presentation. Firstly, Tóth (2019) observed that most participants (79.5%) experienced at least moderate to high levels of anxiety (*M* = 6.51 on a 10-point Likert scale). Based on interview responses, Tóth (2019) distinguished between four interconnected causes of anxiety: 1) the communication situation itself, (2) audience related fears, (3) self-related concerns, and (4) talk-related concerns. Many participants attributed their anxiety levels to the nature of the communicative situation, highlighting salient factors such as "moving to the front", "being in the spotlight" or "standing in front of the others" as especially anxiety-inducing. Concerning the audience, Tóth (2019) noted that even though learners had roles as both presenters and audience in the class, they tended to see each other as strict judges rather than fellow presenters. Consequently, learners were concerned about their peers' perceptions of them and their language abilities, which reflects personal and interpersonal anxieties (Young, 1991) and the fear of negative evaluation as key determinants in shaping learners' vulnerability to FLA (Horwitz et al., 1986).

With regards to self-related concerns, participants reflected on their awareness of bodily reactions such as shaking hands, a pounding heart, blushing, and shaky voice, among others. As reviewed in *Section 1.4*, when nervous, students are hypersensitive to their surroundings and tend to overestimate the visibility and severity of their anxiety symptoms; in fact, Tóth's (2019) participants highlighted that experiencing these physical symptoms caused them discomfort and heightened their feelings of vulnerability and self-consciousness, particularly in front of their peers.

Lastly, in terms of task-related concerns, participants directed their attention to two aspects when delivering their speeches: language concerns and cognitive demands. For instance, participants mentioned experiencing issues during their presentations, such as "forgetting what I want to say", "losing track of my thoughts", "leaving out parts of my talk", "not finding the right words", "making grammar mistakes", "mispronouncing words", etc. This finding aligns with Aubrey (2022), who found that most participants reported struggles to remember words, thereby triggering anxiety.

Another aspect is whether there are specific elements affecting anxiety when performing speaking tasks sin the L2. According to Young (1990), there are three key aspects that can affect learners' levels of anxiety: preparedness, speaking errors, and the type of task. However, other factors found to determine L2 speaking anxiety include the degree of familiarity with the interlocutor (Dewaele, 2007; Baran-Łucarz, 2014; Liu, 2006; MacIntyre & Thivierge, 1995), the nationality of the interlocutor (native vs. non-native of the L2; Çagatay, 2015; Tóth, 2017; Woodrow, 2003, 2006), and the interlocutor's collaborativeness (i.e. whether interlocutors interact positively or negatively with FL learners; Bassett et al., 1973; Hartanto et al., 2014; Sosa-López, 2019; Steinberg & Horwitz, 1986).

Regarding the level of preparedness, previous research has reported that students reported feeling more relaxed when they had the opportunity to prepare in advance (Aubrey 2022; Young, 1990; Choi, 2016). However, learners in Young's (1990) study reported being moderately anxious regardless of whether they prepare the role play in advance or performed spontaneously. This suggests that, despite prior task preparation, performing in front of the class presents a greater "threat" and implies a significant degree of self-exposure for the learners. However, this "threat" and self-exposure can be intensified by the size of the audience (Beatty & Behnke, 1991). For example, Beatty and Behnke's (1991) findings indicate that students experienced more anxiety when performing in front of 15-20 peers compared to a single member audience.

In a similar vein, but under laboratory conditions, Aubrey's (2022) participants identified lack of planning time and time limitations as sources of anxiety during task performance. While planning time fostered self-confidence, anxiety grew as the time limit to complete the task approached. Moreover, learners reflected that their emotions changed depending on whether they used previously planned words and structures or perceived themselves as having failed at producing an appropriate word or structure that was important for expressing a specific idea. Overall, these insights suggest that the lack of planning time and strict time constraints can evoke anxiety.

As discussed throughout this thesis, L2 learners are very sensitive to others' evaluations, whether real or imagined (Horwitz et al., 1986); thus, they fear being laughed at or seen as less competent (Beatty, 1988; Gregersen, 2003; Kitano, 2001; Yan & Horwitz, 2008). For example, most learners report they would volunteer in class if they were not afraid of making mistakes (Young, 1990). Although making language mistakes is part of the learning process, the anxiety levels mistakes generate depend greatly on the L2 learning environment. If L2 learners are in a context where overt error correction or harsh criticism is present, they will be more prone to experience FLA (Gkonou, 2013). Only by engaging in active-constructive feedback can teachers foster learners' emotional security, thereby diminishing FLA (Gkonou, 2017; Oxford, 2017; Rubio-Alcalá, 2017).

Regarding types of activity, previous studies have found that learners rank different types of speaking activities differently (Young, 1990; Woodrow, 2006). For instance, students tend to prefer engaging in communicative activities in small groups rather than in front of the class (Young, 1990; Woodrow, 2006). Despite this, little is known about whether specific task types may elicit varying levels of speaking anxiety. To the best of the author's knowledge, only Kormos and Préfontaine (2017), Sosa-López and Mora (2022a) and Tóth (2017) have found that certain task types (and not others) can trigger higher anxiety levels. In the case of Kormos and Préfontaine (2017), learners reported being more anxious when performing a picture narration task with unrelated pictures than when retelling a story. Retrospective interviews revealed that participants' anxiety perceptions were strongly associated with the lack of task structure (i.e., the unmatching pictures), together with the lack of prior planning and lack of linguistic resources.

On the other hand, Sosa-López and Mora (2022a) and Tóth (2017) found argumentative tasks to be more anxiety-evoking than descriptive tasks. In the case of Tóth's (2017) study, the researcher sought to compare FLA sources in high- versus low-anxiety students when performing three different tasks: exchanging information about themselves and the interlocutor (named "lead-in phase" by the researcher), expressing their opinions on a controversial issue (argumentative task), and describing an ambiguous picture (descriptive task). Tóth's (2017) findings unveiled that high-anxiety learners found the "lead-in phase" to be the easiest, while the argumentative task posed the greatest challenge. Interestingly, in the argumentative task, anxious learners consistently expressed dissatisfaction with their performances and exhibited more concerns over accuracy compared to low-anxiety learners, who more frequently commented that they enjoyed the task. In the descriptive task, high-anxiety participants also worried about the linguistic demands and reported difficulties such as recalling the appropriate words.

All in all, FLA can arise due to several causes, but previous findings underscore the significance of internal factors such as fear of negative evaluation and self-consciousness while speaking-oriented activities is noted as the most frequent external factor. In this context, an important question arises: what happens when we combine some of the aforementioned factors? Specifically, what would occur when Spanish EFL learners are tasked with delivering oral presentations across various formats, such as descriptive and argumentative tasks, with minimal preparation, and in the presence of unfamiliar individuals?

1.6. Measuring anxiety

Research on anxiety has predominantly adopted a retrospective approach, primarily focusing on learners' self-reports. While "psychometric instruments" such as questionnaires and personality tests have played a central role in obtaining FLA data (Gkonou et al., 2017), the integration of objective measures, including physiological ones, offers a means to capture bodily responses of FLA. This section aims at reviewing the main tools employed in FLA research, categorizing them into subjective and objective measures. Subjective measures will be further divided into the FLCAS and dynamic measurements of FLA, while objective measures will be subdivided into heart rate monitoring and galvanic skin response analysis. Lastly, a subsection will be dedicated to examining the relationship between subjective and objective measurements.

1.6.1. Subjective measures: Questionnaires

Language anxiety has traditionally been investigated through subjective measures obtained from diaries (e.g., Bailey, 1983; Gkonou, 2013, 2017) and questionnaires (e.g., FLCAS; Horwitz et al., 1986). These instruments hold significant importance in assessing FLA as they shed light on the complexity of affective factors (Bailey, 1983). Indeed, self-reports provide valuable insights into learners' subjective experiences, perceptions, and emotional responses during language learning. For example, Bailey (1983) showed, through diary entries, that whereas some learners feel stimulated when experiencing anxiety, others feel debilitated by it.

Questionnaires offer a structured path, in which participants are presented with predetermined factors. Hence, learners only have to write a response or select the option that most accurately reflects their feelings, agreement level, or other relevant responses. From these, researchers can obtain both qualitative and quantitative data. Questionnaires addressing FLA commonly incorporate Likert-type scales that quantitatively self-assess levels of distress and are typically administered pre- and/or post-performance. Likert-type scales with a broad range of points, such as 7 or 9-point scales are frequently favored in research due to their validity, statistical reliability, and discriminatory power.

Although questionnaire data depend largely on the participant's willingness and ability to accurately report their experiences, questionnaires remain very popular owing to several advantages such as their low cost, efficient administration in large participant samples, and the ease to gather multiple responses in a single instrument (see Dörnyei &

Taguchi, 2009 for a review). However, as Schwerdtfeger (2004) points out, questionnaires present a significant limitation concerning laboratory settings: they may fail to elicit sufficient emotional reactivity. Consequently, the data obtained from questionnaires in such settings may be different from those obtained in other contexts (e.g., in a classroom setting).

This section reviews some of the main tools that have been used to study FLA. The FLCAS is one of the most popular instruments used to evaluate anxiety levels in language learners. However, alongside the FLCAS, several other questionnaires have been devised to measure anxiety, such as the French Class Anxiety Scale (Gardner & Smythe, 1975); the French Use Anxiety Scale (Gardner et al., 1997); the English Use Anxiety Scale (Clément et al., 1977) and the Input Processing Output Anxiety Scale (IPOAS; MacIntyre & Gardner, 1994b). Furthermore, numerous tools have been created to assess anxiety related to specific language skills, such as the Foreign Language Listening Anxiety Scale (Elkhafaifi, 2005), the Foreign Language Reading Anxiety Scale (Saito et al., 1999), the Second Language Writing Anxiety Inventory (Cheng, 2002, 2004), the Second Language Speaking Anxiety Scale (Woodrow, 2006), the Foreign Language Pronunciation Anxiety scale (Kralova et al., 2018) and the Phonetics Learning Anxiety Scale (Baran-Łucarz, 2013). These tools not only contribute to a comprehensive understanding of FLA, but also address a myriad of facets of anxiety in a variety of language learning environments and skills.

1.6.1.1. Foreign Language Classroom Anxiety Scale

Prior to the development of the Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986), several other questionnaires addressing anxiety had been developed even though most of them held little relevance to second language learning (see Maclntyre & Gardner, 1988, for a review of anxiety scales available in the 80s). Among these, only the *French Classroom Anxiety* and the *French Class Anxiety Scale* (Gardner, 1985) included items to capture typical experiences encountered in language-learning environments. However, it is noteworthy that these tools were part of a larger battery of scales focused on attitudes and motivation, namely the Attitude Motivation Test Battery or AMTB. As observed by MacIntyre (2017), the focus was neither on anxiety itself, nor was it consistently integrated into the socio-educational model of motivation.

Hence, the development of the FLCAS was a necessary step towards a more comprehensive exploration of FLA, particularly within the context of foreign language learning. As explained by Horwitz et al. (1986), the FLCAS originated from group meetings with students at the beginning of language classes at the University of Texas. These group meetings centered on students' concerns and difficulties in language learning, effective learning strategies, and anxiety management exercises. All these insights contributed to the development of the FLCAS, a 33 item self-report questionnaire that addresses specific situations that might elicit FLA. The questions employ 5-point Likert-scales, ranging from "strongly disagree" (= 1) to "strongly agree" (= 5) to specifically assess aspects focusing on speaking and listening in the target language. To illustrate, the FLCAS includes questions on comparisons with peers (question 7: 'I keep thinking that the other students are better at languages than I am'), fear of negative performance (question 9: 'I start to panic when I have to speak without preparation in language class') or physiological symptoms (question 20: 'I can feel my heart pounding when I'm going to be called on in language class'), among others.

Once scores on negatively worded items are reversed and recoded, the final FLCAS scores range from 33 to 165, with higher scores indicating greater levels of FLA. Since its introduction, this scale has gained popularity, and it continues to be one of the main tools for the analysis of FLA owing to its high internal reliability (Cronbach alpha = .93) and test-retest reliability (r^{tt} = .83) (Aida, 1994; Horwitz, 1986; Elkhafaifi, 2005). Notably, the FLCAS has been adapted in several studies (e.g. Aida, 1994) as well as translated into numerous languages (e.g., Arabic in Dewaele & Al-Saraj, 2015; Spanish in Pérez-Paredes & Martínez-Sánchez, 2000; Hungarian in Tóth, 2008).

Returning to the students' accounts of anxiety-provoking aspects, these included experiences speaking the L2, testing and fear of being negatively evaluated by both teachers and peers (Horwitz et al., 1986), all of which were taken as theoretical inspiration when developing the questionnaire items (Botes et al., 2022). As posited by Horwitz et al. (1986), given that FLA involves performance evaluation within both academic and social contexts, it was helpful to draw parallels between FLA and three other related performance anxieties: communication apprehension (CA) (McCroskey, 1970), fear of negative evaluation (Watson & Friend, 1969), and test anxiety (Sarason & Sarason, 1990).

The first, CA is described as a type of shyness characterized by fear or anxiety about interacting with others, whether in real or anticipated public communication with an individual or a group of people (Horwitz et al., 1986; McCroskey, 1978). McCroskey (1978) further noted that communicatively apprehensive people tend to avoid and withdraw from communication with others. Early research on CA, which used McCroskey's scale Personal Report of Communication Apprehension (PRCA, McCroskey 1970, 1978; McCroskey et al. 1985), revealed that CA in the L1 served as a predictor of anxiety levels in the L2 (McCroskey et al., 1985). This notion was also echoed by Horwitz et al. (1986) who suggested that individuals who struggle to speak in groups are likely to face even more challenges in a foreign language class. In addition, in classroom settings, individuals not only feel less in control of the communicative situation, but also may perceive that their oral contributions are constantly being scrutinized.

Second, test anxiety refers to a type of performance anxiety which arises from fear of failure (Horwitz et al., 1986). According to Sarason and Sarason (1990), test anxiety is a situation-specific personality trait, which varies according to the learners' predisposition to experience worry and emotional arousal in academic settings. In particular, test anxiety is influenced by how individuals approach evaluative situations with their distinctive personal traits, and their assumptions, concerns, and expectations (Sarason & Sarason, 1990). Therefore, this form of performance anxiety is triggered by the demands and pressures individuals impose upon themselves.

Lastly, the fear of negative evaluation encompasses apprehensions about others' judgments, the tendency to avoid evaluative situations, and the expectation that others (e.g., the teacher or peers) will evaluate oneself negatively (Watson & Friend, 1969). Additionally, Young (1990) suggests that the way L2 learners respond to error correction and their fear of negative evaluation indicates that social anxiety might be an important element in the theoretical model of language anxiety. However, as Watson and Friend (1969) highlight, this apprehension over negative evaluations does not necessarily imply that individuals hold negative self-perceptions. In the context of instructed language learning, learners have to perform in front of teachers and peers very often; hence, constant comparison of their performance relative to that of other students is inevitable.

It is essential to clarify that Horwitz et al. (1986) did not suggest that these performance anxieties equaled FLA, as has been the case in several subsequent interpretations (Horwitz, 2017). Rather, Horwitz et al. (1986) identified them as situation-specific anxieties ***related*** to FLA. As warned by Horwitz (2017), this is a common misreading of the original paper since the emotions evoked by FLA are much deeper than a simple combination of anxieties. In fact, in order to establish validity, Horwitz (1986) correlated FLCAS with CA, test anxiety, fear of negative evaluation, and even trait anxiety, finding non-significant correlations between the FLCAS and McCroskey's PRCA and the Fear of Negative Evaluation Scale developed by Watson and Friend (1969). Only the Test Anxiety Scale by Sarason and Sarason (1990) was found to be significant. These findings suggest that CA and fear of negative evaluation are different constructs, and that people who experience FLA are more prone to experience test anxiety and vice versa. Furthermore, Horwitz (1986) found a significant correlation between the FLCAS and the Trait scale of the State-Trait Anxiety Inventory (Spielberger, 1983), suggesting that people who are generally anxious in their daily lives may be more likely to be anxious in language learning contexts.

The only downside of the FLCAS is the concern about its construct validity when assessed by exploratory factor analysis. According to Botes et al. (2020), researchers have analyzed the component/s or factor/s of FLCA differently because the original publication of the measure by Horwtiz et al. (1986) did not explore its underlying factor structure, nor was this aspect addressed in Horwitz's (1986) work (see Table 3 for a summary). While Dewaele and MacIntyre (2014) believe the FLCAS to be unidimensional, other researchers such as Aida (1994) among others (see Table 3) propose it is a multidimensional questionnaire structure. To illustrate, after conducting a factor analysis study, Aida (1994) revealed four important factors accounting for variance in FLCAS scores: speech anxiety and fear of negative evaluation, fear of failing, comfortableness in speaking with native Japanese, and negative attitudes toward Japanese class. According to Botes et al. (2022), the use of different statistical analyses on the data sets may have contributed to identifying various factor structures. Moreover, Park and French (2013) noted that these differences may arise depending on the participant's proficiency levels and learning contexts.

All in all, these differentiating factors make the FLCAS construct validity complicated due to the absence of a well-defined theoretical foundation regarding the number of factors underlying FLA (Botes et al., 2020). To address this issue, Botes et al. (2020) developed a unidimensional scale known as the short-form FLCAS (S-FLCAS). The validity of the S-FLCAS was thoroughly assessed (see Botes et al., 2020, for its validity assessment process), including tests for measurement invariance, which are crucial when comparing questionnaire (or scale) scores across key participant characteristics (e.g., target language, language-learning context, age, gender, country) (see Sudina, 2023 for detailed information on measurement invariance and the FLCAS)[3]. Nonetheless, given the FLCAS reliability, internal consistency, and consistent pattern of validity, Horwitz et al.'s (1986) tool has been widely accepted by the research community, leading to its predominant use in the majority of self-reported language anxiety studies.

Table 3: Examples of FLCAS factor analysis conducted in SLA research.

Publication	L1	Target language	Proposed factors	Methods and reliability
Aida (1994)	English	Japanese	1. Speech anxiety and fear of negative evaluation 2. Fear of failing the class 3. Comfortableness in speaking with Japanese people 4. Negative attitudes towards Japanese class	Principal component analysis with varimax rotation Cronbach alpha = .94
Cheng et al. (1999)	Chinese	English	1. Low self-confidence in speaking English 2. General English classroom performance anxiety	Principal component analysis with varimax rotation Cronbach alpha = .95
Matsuda & Gobel (2004)	Japanese	English	1. General English classroom performance anxiety 2. Low self-confidence in speaking English	Principal component analysis with varimax rotation Cronbach alpha = .78
Liu & Jackson (2008)	Chinese	English	1. Fear of negative evaluation 2. Communication apprehension 3. Test anxiety	Factor analysis with varimax rotation Cronbach alpha = .92

[3] The concept of *measurement invariance* refers to the situation in which a latent variable representing a theoretical construct and comprised of one or more observed variables, such as questionnaire items, is interpreted consistently by respondents across different groups or by the respondents in the same group over time (Sudina, 2023).

Mak (2011)	Chinese	English	1. Speech anxiety and fear of negative evaluation 2. Uncomfortableness when speaking with native speakers 3. Negative attitudes towards the English class 4. Negative self-evaluation 5. Fear of failing the class/consequences of personal failure.	Factor analysis with varimax rotation Cronbach alpha = .91
Park (2014)	Korean	English	1. Communication apprehension and understanding 2. Communication apprehension and confidence	Maximum likelihood exploratory factor analysis with direct oblimin rotation Cronbach alpha = .86
Tóth (2008)	Hungarian	English	1. Communicative apprehension 2. Fear of inadequate performance in English classes	Principal component analysis with direct oblimin rotation Cronbach alpha = .93

Following the framework set by FLA studies, L2 speaking research has also relied on the FLCAS, which is dominated by speaking-related items (Gkonou, 2011). However, recognizing that existing tools mainly addressed in-classroom anxiety, Woodrow (2003, 2006) developed the Second Language Speaking Anxiety Scale to capture both formal learning environments and everyday situations. While the questionnaire assesses speaking anxiety, self-efficacy, motivation, and learning strategies, only the speaking anxiety subscale will be reported.

The Second Language Speaking Anxiety Scale's design is based on both the FLCAS and McCroskey's PRCA (1985). It is comprised of 12 items rated on a 5-point Likert scale, ranging from "not anxious at all" (= 1) to "extremely anxious" (= 5). The items presented were tailored to address communicative situations participants were likely to encounter, considering both in-class and out-of-class communicative settings as well as variables such as the interlocutor's relation to the participant (e.g., teacher, classmate, native speaker) (see Figure 4). Finally, this tool was tested with English for Academic Purposes (EAP) learners studying in Australia. The reliability coefficient for in-class items was found to be .89, for out-of-class items it was .87, and when combined the reliability increased to .94, making it a highly reliable tool (Woodrow, 2003).

Situation	*Anxiety*
1. The teacher asks me a question in English in class.	○ ○ ○ ○ ○ 1 2 3 4 5
2. Speaking informally to my English teacher out of class.	○ ○ ○ ○ ○ 1 2 3 4 5
3. Taking part in a group discussion in class.	○ ○ ○ ○ ○ 1 2 3 4 5
4. Taking part in a role-play or dialogue in front of my class.	○ ○ ○ ○ ○ 1 2 3 4 5
5. Giving an oral presentation to the rest of the class.	○ ○ ○ ○ ○ 1 2 3 4 5
6. When asked to contribute to a formal discussion in class.	○ ○ ○ ○ ○ 1 2 3 4 5
7. Talking to administrative staff of my language school in English.	○ ○ ○ ○ ○ 1 2 3 4 5
8. Taking part in a conversation out of class with more than one ***native speaker*** of English.	○ ○ ○ ○ ○ 1 2 3 4 5
9. Starting a conversation out of class with a friend or colleague who is a ***native speaker*** of English.	○ ○ ○ ○ ○ 1 2 3 4 5
10. A lecturer/supervisor in my intended university faculty of study asks me a question in English.	○ ○ ○ ○ ○ 1 2 3 4 5
11. Asking for advice in English from a lecturer/supervisor in my intended university faculty of study.	○ ○ ○ ○ ○ 1 2 3 4 5
12 A ***native speaker*** I do not know asks me questions.	○ ○ ○ ○ ○ 1 2 3 4 5

Figure 4: Woodrow's (2003, 2006) Second Language Speaking Anxiety Scale.

1.6.2. Subjective dynamic measures

As stated in *Section 1.2*, the *Dynamic Approach* is a relatively recent line of research which considers that affective factors such as anxiety are not static constructs and that they fluctuate over time rising and falling while an activity such as communicating unfolds (MacIntyre, 2012, 2017). In fact, as Gregersen et al. (2014) highlighted, no two speeches are ever truly identical. Even if the same person speaks about the same topic with the same audience a second time, the reactions would be expected to show different patterns because each event unfolds differently each time. In other words, variations in the intensity of affective factors are subject to the continuous interaction between internal physiological processes and the speaker's ability to regulate these emotions; yet every presentation will unfold differently.

With these considerations in mind, MacIntyre and Gregersen (2012) advocate that studies based on questionnaires or interviews can only give partial answers; they are not well suited to analyze anxiety fluctuations. For this reason, other subjective measures such as stimulated recalls (Gass & Mackey, 2013) and the idiodynamic task (MacIntyre, 2012) are valuable tools to capture how the speaker's perception unfolds and how they express their own understanding of the impulses that changed their emotional state. Nevertheless, it is important to acknowledge that researchers may view recall methods as susceptible to the detrimental effects of time on memory. As time passes, performance details may fade, thus posing an inherent "risk" of memory reconstruction.

1.6.2.1. Stimulated Recalls

Verbal reports or process tracing are subjective measures that allow researchers to qualitatively analyze affective or cognitive changes that unfold before, during and after a task such as communicating in the L2. Verbal reports cover a wide range of elicitation types, including those which occur concurrently with an event (think aloud), thinking about a previously performed task (retrospective protocols) or prompted interviews after completing a task (stimulated recall). On the other hand, process tracing is a qualitative research method, or family of methods with the aim to trace how a potential cause (or causes) leads to a known outcome (or set of outcomes) by identifying the underlying causal process.A stimulated recall is a type of introspective research methodology which uses a stimulus, either a video recording or an audio of the performance, which is stopped at strategic points (e.g., when there is a communication breakdown) in order to stimulate a comment upon the learner's thought processes at a given time (Gass & Mackey, 2013). This technique was first used in the 50s by Bloom to investigate students' ability to recall classroom events. Video recordings are meant to provide a richer source of stimulation for recall than audio recordings, as they enable participants to re-live an original situation with vividness and accuracy (Bloom, 1953). This is due in part to the fact that visual cues tend to dominate perception, while audio cues play a secondary role (Denley & Bishop, 2010).

As outlined by Gass and Mackey (2013), stimulated recalls require meticulous realization, or one can easily find oneself with contaminated data. This is because participants' recall comments may reflect what they are thinking about when seeing the video or listening to the audio rather than their thoughts during the original task production. Moreover, researchers must bear in mind some limitations of stimulated recalls, such as the differences in individuals' verbal expressiveness (Lyle, 2003) and memory biases (Gregersen et al., 2014). Gass and Mackey (2013) offer practical guidance and support mechanisms, which will be further discussed in the Instruments section (see *Section 4.4.2.3*).

Regarding FLA, numerous studies have used semi-structured interviews after task performance to, for instance, unravel the sources of FLA (Baran-Łucarz, 2011; Tóth, 2011, 2017; Gregersen et al., 2014, Tian, 2019; among many others). However, there has been comparatively little research employing stimulated recalls while analyzing the participants' video or audio recordings (Gregersen, 2003; Gregersen & Horwitz, 2002; MacIntyre & Legatto, 2011), and to the best of the author's knowledge, no study has specifically addressed the sources of FLA as the oral performances unfolded.

While MacIntyre and Legatto, (2011) explored willingness to communicate more specifically through stimulated recalls, an instance of this approach concerning FLA is found in Gregersen (2003) and Gregersen and Horwitz's (2002) exploratory studies. Thanks to the stimulated recall comments, the researchers revealed that anxious and non-anxious FL learners (previously categorized through the FLCAS) reacted differently to their own oral performances. To illustrate, anxious participants identified fewer errors during the stimulated recall process (Gregersen, 2003), reported higher standards for their English performance, and were more concerned about their mistakes and others' opinions compared to their non-anxious counterparts (Gregersen & Horwitz, 2002)

1.6.2.2. The idiodynamic task

According to MacIntyre (2012), affective patterns of change within an individual cannot be fully addressed with cross-sectional or longitudinal research designs because the process of change is not studied directly as it unfolds in time. For this reason, recent research on affective factors such as FLA has adopted more innovative methodologies such as the idiodynamic method using a novel data-gathering software (MacIntyre, 2020; see MacIntyre & Ducker, 2022 for a quick practical guide). Given the relative novelty of this approach, there are few studies using the idiodynamic method, especially for FLA. However, to date, this method has been used to examine willingness to communicate

(MacIntyre & Legatto, 2011; MacIntyre, 2012; MacIntyre & Gregersen, 2022), enjoyment (Aubrey, 2022; Boudreau et al., 2018), FLA (Gregersen et al. 2014; Gregersen et al., 2017) comprehensibility (Nagle et al., 2019) and fluency (Wood, 2016).

The idiodynamic method requires researchers to first record, either in video or audio, a sample of communication from a participant. Then, using the idiodynamic software, also called Anion Variable tester, participants are shown their communicative performance and are asked to rate their own or another person's moment-to-moment changes on an affective or cognitive variable that is relevant to the research (MacInyre, 2012). The software allows participants to indicate their responses on a continuous scale ranging from +5 to -5, and it includes a feature that gradually decreases the rating by one point towards zero if the participant does not respond. This feature ensures active engagement (Gregersen et al., 2014), with ratings being recorded approximately once per second (MacIntyre, 2012). Additionally, after the assessments, the software provides a graph with the changes in the variable under study. Hence, the idiodynamic task enables researchers to quantitatively track affective states that have accompanied communication over time. MacIntyre (2012) and MacIntyre and Ducker (2022) recommend that these assessments are complemented by follow-up interviews to further explore the fluctuations in participants' reactions. For example, researchers can print the graph of the participant's ratings and conduct a stimulated recall.

According to MacIntyre and Gregersen (2022), the idiodynamic method offers multiple strengths. It evaluates real-time changes continuously, operates systematically by gathering the data on a per-second timescale, provides flexibility by allowing assessment of various target variables, and adopts a mixed-methods approach by collecting quantitative data that can be complemented with stimulated recalls. However, despite these strengths, this method also presents certain limitations. Firstly, it requires the presence of distractors, such as video cameras, which can make participants more self-conscious and affect their natural behavior, prior to its implementation. In addition, there is the potential for error variance due to the "auto-zero" feature, which gradually decreases ratings to zero if participants are reluctant to click. As in stimulated recalls, it participants may have difficulties in articulating their internal states or the reasons for their fluctuations. Another limitation is that previous studies have examined relatively small samples, despite conducting highly detailed analyses.

Despite these limitations, research on FLA using the idiodynamic task has yielded interesting findings. For instance, Gregersen et al. (2014) triangulated physiological heart rate measurements with the idiodynamic self-assessment, unveiling that the more anxious learners exhibited higher heart rates and reported higher average idiodynamic anxiety scores than less anxious learners did (see Figure 5). Gregersen et al. (2017) used the idiodynamic task to triangulate self-assessments with assessments from peers and language practitioners. Their findings revealed that self-assessment patterns displayed significantly greater variability, while peer and language practitioner ratings were more consistent and homogeneous. Hence, unlike studies that typically provide a correlation coefficient, studies using the idiodynamic task allows for the identification of specific moments where the data aligns, offering more nuanced insights into participants' experiences.

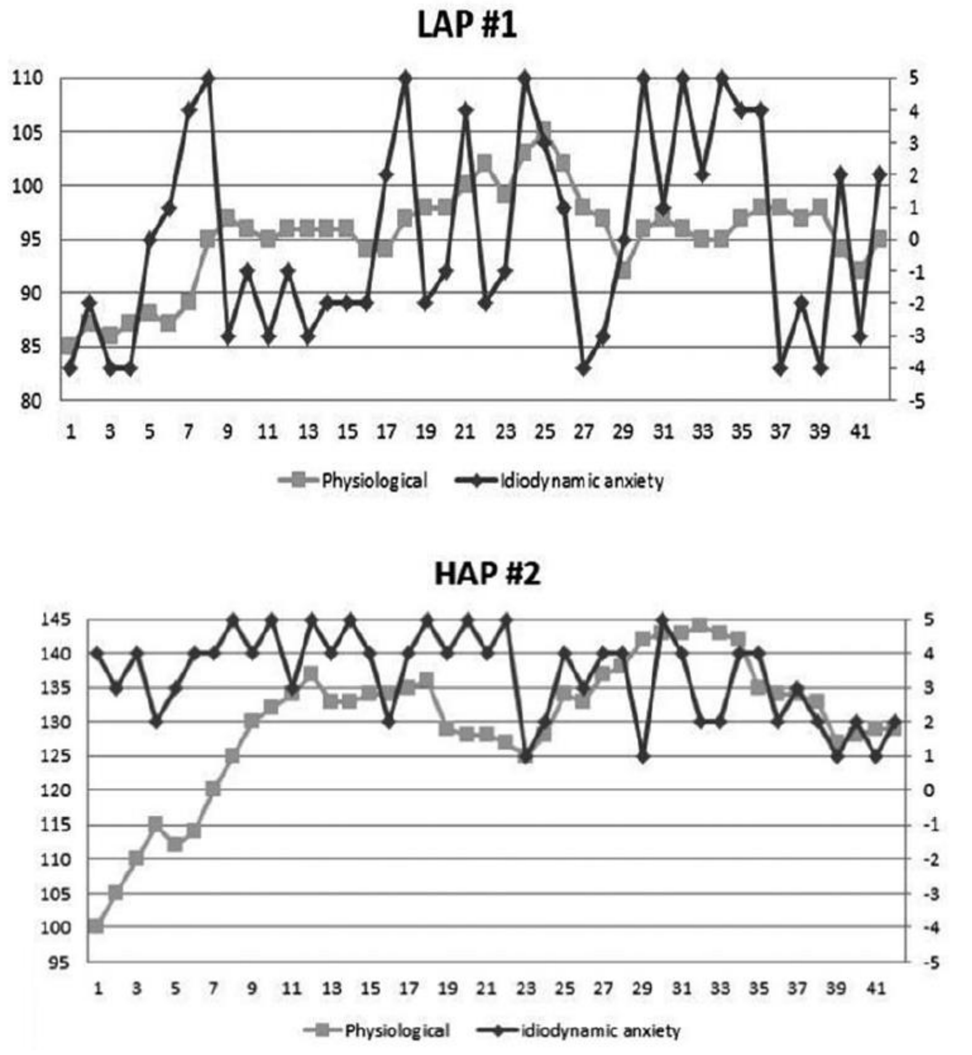


Figure 5: Heart rate (in grey) and idiodynamic ratings (in black) from Gregersen et al. (2014, p. 579)

Note: LAP= low anxiety participant; HAP= high anxiety participant. Heart rates are on the left y-axis and idiodynamic anxiety scores on the right y-axis.

1.6.3. Objective measures

While much research on FLA is based on self-reported measures and qualitative assessments, objective measurement methods, such as heart rate (HR) and Galvanic Skin Response (GSR), also known as electrodermal activity (EDA), have gained interest and importance in recent years. Advances in wearable sensor technology have made it possible to obtain physiological responses associated with anxiety during language learning tasks, which may be complementary to self-reported measures.

The use of sensors offers numerous benefits for studying FLA. Firstly, they provide measures of anxiety-related physiological responses. Physiological processes can affect emotions without conscious awareness, raising the question of whether emotions can be experienced or expressed when consciousness is diminished or absent (Pace-Schott et al., 2019). Thus, by combining physiological data with self-report measures, researchers can attain a more comprehensive understanding of FLA, including its triggers and impacts on language acquisition. Moreover, compared to electroencephalograms (EEG), wearable sensors are typically unobtrusive, allowing learners to engage in tasks without disruption (Rivers, 2022; Garbarino et al., 2014). Finally, bearing in mind the Dynamic Systems Theory, wearable sensors facilitate obtaining continuous, real-time physiological data while learners perform a task by capturing fluctuations in FLA over time through learners' physiological responses.

Despite the increased availability of these tools and their advantages, most of the research on FLA has primarily focused on subjective reports. Reasons for this include their higher cost compared to traditional self-report measures, and specialized knowledge that may be required for accurate data interpretation. In addition, wearing sensors for extended periods of time may cause slight discomfort and potentially affect learners' natural behavior. Furthermore, technical issues such as signal interference or data loss may arise, compromising the reliability and validity of the collected data (Roos & Slavich, 2023). However, according to Roos et al. (2021), physiological measurements contain relatively few artifacts from participants' movements or technical issues when studies are carried out in lab settings. Finally, another disadvantage when studying FLA is the limited contextual information provided by physiological data alone. Therefore, relying solely on physiological data may limit the depth of analysis and interpretation (Barrett, 2006; Roos et al., 2021; Sevinç, 2008).

HR and GSR responses arise as a result of a close interaction between the subsystems of the central nervous system: specifically, the sympathetic nervous system and the parasympathetic nervous system (see Figure 6). While the sympathetic nervous system is responsible for high arousal, including the fight or flight response (preparation for and response to stressors), the parasympathetic nervous system oversees the resting and self-regulation phase (Pace-Schott et al., 2019). As Pace-Schott et al. (2019) explain, under safe conditions, the sympathetic and parasympathetic systems work together to maintain homeostasis, which is the balance or stability maintained by our body's internal processes despite changes in the external environment. However, this dynamic interaction between the systems adapts in response to environmental changes altering their balance (Val-Calvo et al., 2020).

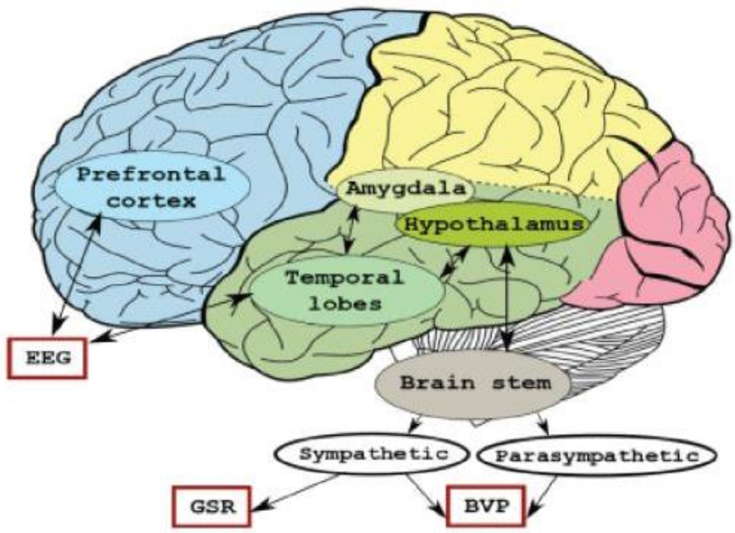


Figure 6: Brain areas involved in emotion processing (From Val-Calvo et al., 2020, p. 134052)
Note: EEG corresponds to electroencephalogram and BVP to blood volume pulse.

Therefore, as exemplified in Pace-Schott et al., (2019), when sympathetic efferent nerves, which carry signals away from the central nervous system, are activated, it leads to an increase in heart rate, blood pressure and sweat gland activity (see also Roos et al., 2021). This is often associated with situations requiring quick action or responses, such as being stressed or exercising. On the other hand, parasympathetic efferent nerves carry signals to body organs, including the heart, and prioritize different functions. They focus on conserving energy and maintaining the body's internal environment. For instance, they tend to lower heart rate and increase heart rate variability, which can be associated with a more relaxed state (Rivers, 2022). As further explained in Pace-Schott et al. (2019), when feeling danger, the ANS activates and modulates regions in the insular cortex, medial prefrontal/anterior cingulate cortices, hippocampal formation and amygdala which, on the subjective level of experience, individuals can experience as tension, heat, trembling, throbbing heart, sweating (i.e., sympathetic activation), nausea, faintness, and even cold (i.e., a defensive parasympathetic activation to the gut and heart respectively). These physiological responses align with the emotional experiences encountered by individuals exhibiting high levels of anxiety (Rubio-Alcalá, 2017; Gregersen, 2005, 2007b, 2009; Gregersen et al., 2017).

The following subsections will now provide an overview of physiological responses (HR and EDA) in relation FLA, highlighting the potential benefits and challenges of these measurements.

1.6.3.1. Heart Rate (HR)

From a research perspective, heart rate (HR) is a key physiological response because of the reactivity of the heart to emotional changes (Gregersen et al., 2014). Although some people might be able to regulate their own emotions deliberately through conscious processes, the brain is constantly regulating our physiological processes unconsciously. To assess cardiac activity, two main methodologies are commonly employed: the electrocardiography (ECG) and the photoplethysmography (PPG). Both modalities serve as bio-signals or physiological indicators, enabling the measurement of the heart's electrical activity using electrodes in direct contact with the skin (Soehn, 2017; Electrocardiography iMotions guide, 2020) to obtain the HR measures, that is, the number of heart beats per unit of time (usually measured per minute).

Although both ECG and PPG offer HR measurements, these methodologies diverge in their approaches to obtaining this data. On one hand, the ECG measures the electrical activity of the heart which is recorded by the electrodes, usually placed on the chest and limbs. As Soehn (2017) explains, this series of coordinated electrical signals are sent by the ANS to the heart and can be quantified with the QRS complex, which is the typical zigzag patterns displayed on hospital monitors (see Figure 7). The QRS complex comprises three waves that mark the location of the electrical signal: the Q wave, the R wave, and the S wave. Initially, the Q wave represents the initial dip in the ECG signal before it spikes, and this occurs when the ventricles fill with blood from the atria. Subsequently, during the R wave, the signal increases from its lowered position, reaching its peak of electrical activity, a phenomenon accentuated by the considerable size of the ventricles. Additionally, this phase coincides with the closure of valves separating the ventricles and atria, thus, generating the first of the two audible thumps characteristic of a heartbeat. Notably, the R peak is frequently used in ECG HR measurements. Lastly, the S wave denotes the pronounced signal descent from its peak to below the baseline. Then, the blood is pumped out of the ventricles, and newly arriving blood enters the atria to repeat the process (Electrocardiography iMotions guide, 2020).

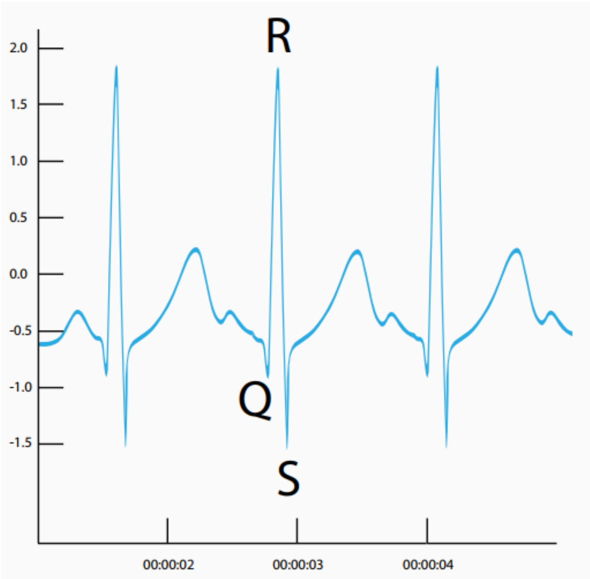


Figure 7: Example of HR data (QRS complex; from Electrocardiography iMotions guide, p. 24)

On the other hand, the PPG method measures the HR by sensing changes in the volume of blood flow that flows through the vessels in a localized area using light transmission or reflection during a cardiac cycle (Soehn, 2017). These changes are recorded in photoplethysmographic waveforms, also known as pulse oximeter waveform (see Figure 8, part a). Wearable sensors (e.g., Empatica E4) make it possible to measure PPG thanks to their photoplethysmogram sensor which records the blood volume pulse (BVP). PPG sensors are composed of a LED light source and a photodetector. The light source illuminates the tissue, while the photo-detector measures changes in the intensity of absorbed or reflected light, which varies depending on the blood volume, blood vessel wall movement, and the orientation of red blood cells (Blasco & Peris-Lopez, 2018; Milstein & Gordon, 2020; Roos & Slavich, 2023). As PPG is an indirect method, meaning that it does not directly record the heart's activity at its origin, Soehn (2017) remarks that it may not always provide precise measurements down to the millisecond level as the

ECG does. Nevertheless, if the LED light is placed in an area where arteries are close to the skin, such as the wrist or a fingertip, HR data will be reliable (Soehn, 2017).

Several measures can be derived from the PPG data, such as the HR, the Inter-Beat Interval (IBI) and the Heart Rate Variability (HRV). Firstly, HR represents the frequency of a full heartbeat cycle, from its beginning to the onset of the next beat, within a specific time frame. This measurement is commonly expressed as beats per minute (bpm), and the normal resting HR is usually between 60 and 100 bpm (Roos et al., 2021). However, HR can be influenced not only by affective arousal, but also by gesturing and changes in respiration (Gregersen et al., 2014; Kreibig, 2010; Milstein & Gordon, 2020; MacIntyre et al, 2010), and even increased mental workload (Beatty and Behnke, 1991; Roos et al., 2021). According to Beatty and Behnke (1991), when a task poses a significant challenge to an individual's cognitive abilities, it activates the cortical system, leading to elevated heart rates in participants. Thus, even non-anxious learners would present increased HR if tasks were *extremely* demanding.

Although heartbeats are quite regular, there is a surprising amount of variability between beats (MacIntyre et al., 2010). For this reason, it is important to record this variability with the IBI. As the name indicates, the IBI represents the time interval between individual beats of the heart (Electrocardiography iMotions guide, 2020) or the R waves. The IBI is also known R-R interval (when abnormal heartbeats can be found) or N-N interval to highlight that the heartbeats are normal. This data is obtained from the BVP signal by extracting the variations in the intensity of the refracted light caused by the fluctuations in blood flow (see Figure 8, part b). In a normal functioning heart, each IBI value varies from beat to beat, and with these variations HRV measurements are computed.

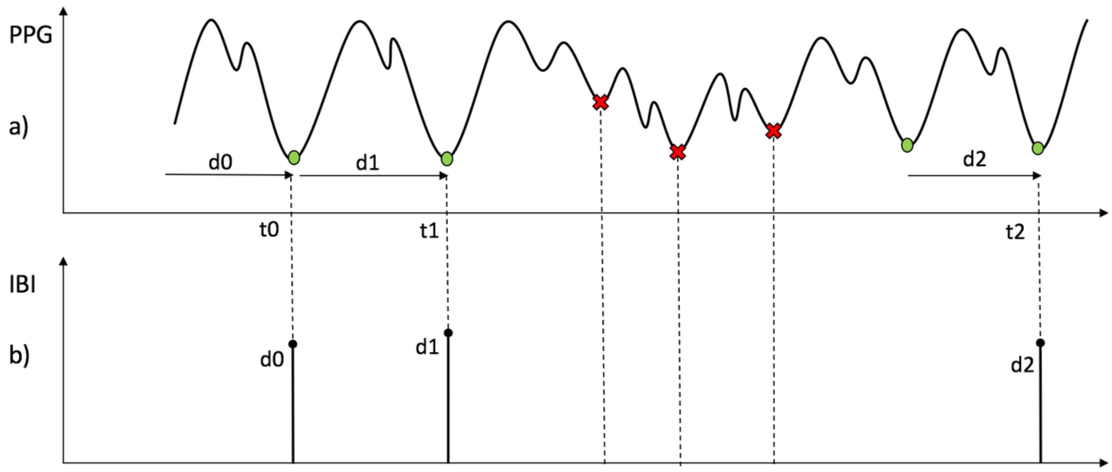


Figure 8: How PPG data is transformed into IBI data (from Empatica support website on IBI, n.d.)

Note: In graph (a), green dots correspond to good heartbeats, while red crosses depict heartbeats corrupted with motion artifacts

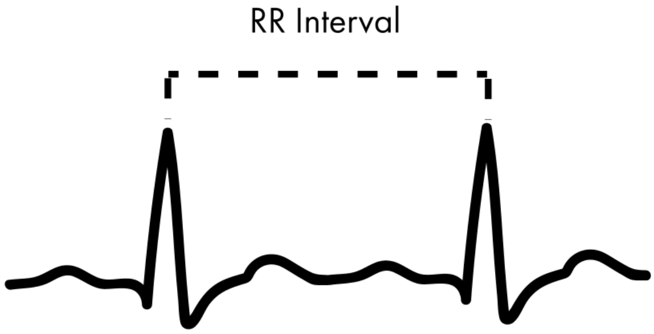


Figure 9: RR interval (from Electrocardiography iMotions guide, 2020, p.21)

HRV refers to the variance in duration between heart beats over a period time (Rivers, 2022; see Figure 9 above). Several HRV metrics have been used in research and their classification distinguishes the calculation techniques used: time-domain, frequency-domain, and non-linear measurements. While time-domain indices of HRV focus on quantifying the amount of variability in IBI measurements, frequency-domain indices estimate the distribution of absolute or relative power into frequency bands categorized from ultra-low to high (see Shaffer & Ginsberg, 2017 for a detailed account). Non-linear measurements are used to quantify the unpredictability of a time series with metrics such as the Poincaré plot.

As reviewed in Shaffer and Ginsberg (2017), the choice of these measurements depends on the research area (clinical vs. non-clinical), the device used (ECG vs. PPG-based), and the duration of the data collection (24h; short-term, ~ 5 minutes; or ultra-short term, < 5 minutes). For example, frequency-domain and non-linear measurements are commonly employed in clinical research together with respiratory data (Kreibig, 2010). Furthermore, reliable HRV measures should be above 250 Hz for frequency domain measures and above 100 Hz for time-domain measures (Milstein & Gordon, 2020; Kwon et al., 2018), and long data recordings are preferred over short ones.

It is evident that HR and HRV analyses offer valuable insights into the functioning of the ANS and cardiovascular health (Held et al., 2021; Roos et al., 2021; Shaffer et al. 2014). For example, in clinical anxiety disorders research, HRV has been associated with parasympathetic influences of the ANS. Research findings suggest that people who suffer clinical anxiety disorders exhibit reduced HRV (Held et al., 2021), and this low variation between heart beats is an indicative of the fight-or-flight response (Rivers, 2022). This happens because, in a fight-or-flight situation, the body reacts by producing rapid changes in heartbeats to adapt to environmental demands (Held et al., 2021). Thus, HRV is associated with the ability to regulate emotions to better cope with stressors (MacIntyre et al., 2010). However, contrary to HRV, HR is interpreted as directly indicative of anxious arousal, meaning that an increase in HR is associated with increased anxiety (Roos et al., 2021).

In the field of linguistics, research has mainly focused on HR (Booth-Butterfield, 1987; Croft et al., 2004; Gregersen et al., 2014; Gregersen et al., 2017; Westenberg et al., 2009) rather than HRV, except for MacIntyre et al. (2010) and Gallego et al. (2021). Both MacIntyre et al. (2010) and Gallego et al. (2021) compared self-reported data with physiological measures. While Gallego et al. (2021) included RMSSD (i.e., root mean square of successive N-N intervals) and galvanic skin response, MacIntyre et al. (2010) focused on HR and HRV across different conditions (e.g., speaking while sitting vs. standing). However, MacIntyre et al. (2010) did not specify which HRV metrics were used in their study. Interestingly, MacIntyre et al.'s study (2010) revealed that while HR was influenced by body position (sitting vs. standing) due to muscle tension, HRV was unaffected. These results suggest that HR might be more sensitive to physical factors such as posture. Moreover, the researchers found that HR and HRV correlated negatively, a finding that is consistent with clinical studies (Kazmi et al., 2016; Kreibig, 2010).

Research exclusively analyzing HR has yielded noteworthy findings concerning its association with public speaking anxiety (read Bodie, 2010 for a detailed review), and L1 anxiety. Previous studies on L1 anxiety revealed that individuals exhibited increased HR during speech performances at university, compared to measurements obtained before or after the task's completion (Behnke & Beatty, 1981; Booth-Butterfield, 1987; Croft et al., 2004; Felnhofer et al., 2014; Westenberg et al., 2009). This indicates that speaking in front of others triggers significant physiological stress responses.

Despite the studies conducted on L1 anxiety, research in SLA and FLA has not commonly employed HR measurements. To the best of the author's knowledge, very few studies have monitored HR during L2 production (Gregersen et al., 2014; Mora et al., 2023; Sosa-López, 2019; Sosa-López & Mora, 2022a, 2022b). Previous studies have shown that highly anxious learners obtain higher HRs than less anxious learners during an L2 oral production task (Gregersen et al., 2014). Although different task types (descriptive vs. argumentative) have not been found to affect HR (Sosa-López & Mora, 2022a), performing complex tasks has been linked to greater HR variability (Mora et al., 2023). Further research using physiological measures is needed in SLA to obtain valuable insights into FLA experiences, especially when combined with self-report measures.

1.6.3.2. Galvanic Skin Response (GSR)

Since the discovery of Galvanic skin response (GSR), recording electrodermal activity has become a widely used method in psychophysiology. GSR, also known as Electrodermal Activity (EDA) or Skin Conductance (SC), originates from the activation of the sweat glands in our skin[4]. Our bodies have millions of sweat glands and, when facing emotional stimulation, these glands are triggered and become more active by secreting moisture through pores towards the skin surface (GSR iMotions guide, 2020). This provokes a change in the balance of positive and negative ions in the secreted fluid, enhancing the flow of electrical current flows and resulting in measurable changes in skin conductance (i.e., increased skin conductance corresponds to decreased skin resistance) (GSR iMotions guide, 2020). In other words, GSR or EDA refers to the electrical changes that arise when the skin receives stimulation signals from the sympathetic nervous system (Boucsein, 1992).

As skin conductivity is solely regulated by the sympathetic system of the ANS, this physiological marker offers highly valuable insights into individuals' emotional and cognitive states by reflecting unconscious behaviors that are not subject to cognitive control (GSR iMotions guide, 2020). That is, the sympathetic system regulates bodily

[4] In this dissertation, the mentioned terms will be used interchangeably, although Boucsein (1992) advises using EDA rather than GSR because the skin is a galvanic element, which fails to capture the multiplicity and complexity of the electrodermal phenomena.

processes, as well as cognitive and emotional states at a subconscious level (Blasco & Peris-Lopez, 2018; GSR iMotions guide, 2020; Milstein & Gordon, 2020; Rivers, 2022). In fact, researchers such as Barrett (2006) believe that a change in HR is not always necessarily the same as experiencing anxiety, as heart rate monitors can sometimes be influenced by parasympathetic activity (Rivers, 2022). For this reason, some researchers consider that an arousal is more clearly expressed through increases in sweating than through cardiac acceleration (Barry and Sokolov, 1993; Lindberg et al., 2021; Rivers, 2022).

EDA consists of two components: the Skin Conductance Level (SCL; or tonic phase) and the Skin Conductance Response (SCR; or phasic phase) (see Figure 10; Braithwaite et al., 2013). SCL corresponds to stimuli-free levels of skin conductance, which would represent the "baseline" level of skin conductance. According to the GSR iMotions guide (2020), the SCL slowly varies over periods ranging from tens of seconds to minutes, and these fluctuations can be influenced by hydration, skin dryness, and autonomic regulation. Moreover, the tonic level of skin conductance can differ significantly across individuals (Sevinç, 2018), making it challenging to interpret SCL in terms of "high" or "low" for an individual. This variability suggests that SCL alone may not be that informative (GSR iMotions guide, 2020; Sevinç, 2018).

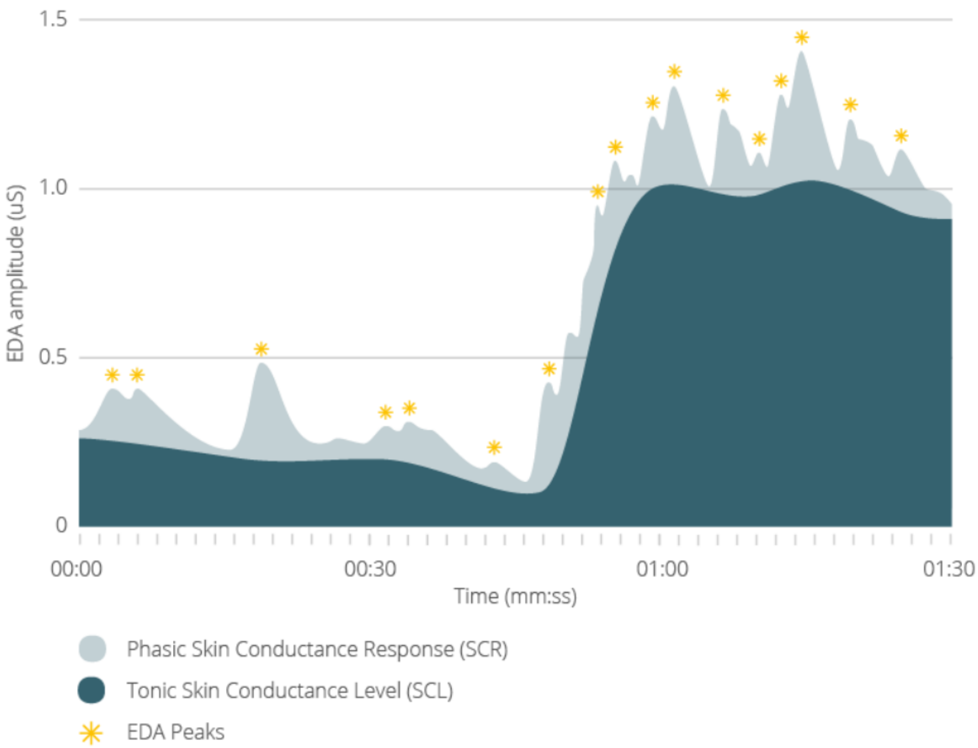


Figure 10: EDA's tonic and phasic phases (from GSR iMotions guide, 2020, p. 22)

On the other hand, SCR captures the faster changing elements of the signal in reaction to an external stimulus (Empatica EDA support page, n.d.). These faster fluctuations are discernable in the data and are called "GSR peaks" or "GSR bursts" (see Figure 11), which occur within 1-5 seconds after the onset of a stimulus (GSR iMotions guide, 2020; Ross et al., 2021). Hence, SCR serves as a valuable tool for assessing the strength and timing of physiological reactions to stimuli, providing insights into emotional responses.

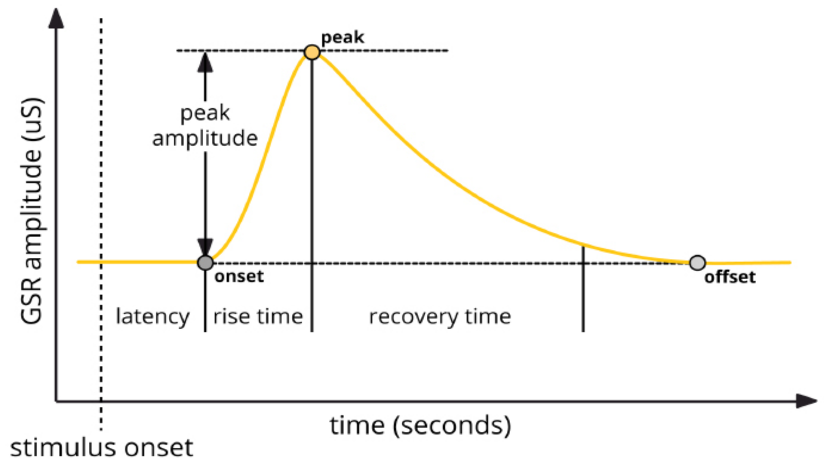


Figure 11: EDA peak (from GSR iMotions guide, 2020, p. 25)

EDA can be measured using the endosomatic and the exosomatic methods. The former refers to the electrodermal recordings which do not use an external current, while the latter uses current applied to the skin (Boucsein, 1992). For instance, wearable devices such as Empatica E4 apply a minuscule amount of current between two electrodes in contact with the skin (reported in Empatica support page on EDA, n.d.) and later measure the skin conductance in response to the sweat secretion made by the eccrine sweat glands. It is worth noting that depending on whether the current is direct or alternating, different measurements can be obtained including skin potential, resistance, conductance, admittance, and impedance (see Boucsein, 1992 for a detailed review in biophysics). All these measurements are obtained in Siemens (S), which measure the conductance ($G = \frac{1}{R}$). As explained by Boucsein (1992), conductance, denoted by the symbol G, is a fundamental electrical property quantifying the ease with which electric current flows. Conductance is defined as the inverse of resistance (R), and measures how easily a material allows electrical charge to flow when a voltage is applied. In essence, conductance reflects the permissiveness of a material to electrical conduction, with higher conductance indicating lower resistance. In biological studies, it is frequent to use the kiloohm (kΩ) and correspondingly the microSiemens (μS) as the units of resistance and conductance.

In the process of collecting EDA data, several crucial considerations need attention. To start with, according to Kurniawan et al. (2013) skin conductance is subject to inter-person variability, with differences in age, gender, ethnicity, and hormonal cycles. For this reason, it is important that participants belong to a homogenous group, to avoid strong variability effects. Furthermore, EDA is especially sensitive to room temperature and humidity (reported in Empatica support on EDA, n.d.; Roos et al., 2021). For instance, Barry & Sokolov (1993) highlighted in their study that the room was air-conditioned, and the temperature was maintained between 22 and 24°C.

Another important factor is the electrode placement. Early research in the 70s found palms and fingers to show greater skin current compared to other body parts such as the wrists or elbows (Boucsein, 1992). Even today, EDA studies continue to be conducted by attaching sensors or electrodes to the fingertips or the palm of the non-dominant hand (Roos et al., 2021; Roos & Slavich, 2023; van Dooren & Janssen, 2012). In fact, van Dooren & Janssen (2012) found that the aforementioned areas contain a high density of eccrine sweat glands of approximately 600 to 700 glands/cm^2. Nevertheless, the magnitude of sweating varies between individuals depending on the density of sweat glands, their relative size, and the output of individual glands (van Dooren & Janssen, 2012).

A concluding recommendation by Sevinç (2008) and Boucsein (1999) is that while elevated SCLs and SCRs are indicative of higher levels of anxiety (Kreibig, 2010), it is advisable to obtain additional measures such as self-reports because there is no clear evidence that distinct physiological patterns correspond to a specific emotion. In fact, research on the relationship between self-reported anxiety and EDA has yielded inconclusive findings. While Gallego et al., (2021) found no significant associations between self-reports and EDA, Sevinç (2018) did observe a moderate association. Since different emotions can elicit similar levels of general arousal (Boucsein, 1999), relying solely on physiological recordings is considered insufficient in psychophysiological studies.

In the field of linguistics, as in studies involving HR, EDA has been used as a metric for assessing L1 public speaking anxiety (Croft et al., 2004; Westenberg et al., 2009). However, previous research on the relationship between EDA and anxiety has yielded mixed results. Westenberg et al. (2009) reported increased sweat levels during speech performances compared to measurements obtained before or after task's completion. In contrast, Croft et al. (2004) observed an increase in HR during speech but found no corresponding rise in EDA. These findings reflect the complexity of physiological responses to anxiety, suggesting that HR and EDA may capture distinct aspects of the emotional experience during stressful tasks. Hence, relying on a single physiological measure may not fully capture how anxiety manifests during performance situations.

Once more, despite the studies conducted on L1 anxiety, investigations using physiological measures such as EDA within the context SLA are not frequent. To the best of the author's knowledge, previous research has focused on examining the relationship between GSR peaks and non-verbal behavior during oral presentations (Lee & Kleinsmith, 2019; Lee, 2020; Lee et al., 2020a; Lee et al., 2020b; Lindberg et al., 2021) (see *Section 1.4.*), as well as the association between L2 speaking anxiety and physiological measures (Sosa-López, 2019; Sosa-López & Mora, 2022a, 2022b). Moreover, very few studies have monitored EDA during L2 production (Sosa-López, 2019; Sosa-López & Mora, 2022a, 2022b). All in all, given the clear lack of studies, further exploration of EDA in SLA could enhance our understanding of the physiological arousal caused by language anxiety.

1.6.4. The relation between subjective and objective measures

Understanding the intricate relationship between physiological arousal and subjective experiences remains a critical endeavour, even in clinical research, because of the relatively little empirical evidence suggesting that emotions involve closely coordinated changes in experiential, behavioural, and physiological response systems (Mauss et al., 2004; Trotman et al., 2019). Clinical anxiety studies have found a positive relationship between self-reported anxiety and perceived physiological responses (e.g., racing heart, blushing, shortness of breath and sweaty palms), suggesting that when individuals report feeling anxious, they also perceive corresponding physiological reactions of anxiety (Mauss et al., 2004; Trotman et al., 2019). However, research has not found a significant association between self-reports and physiological measures (Mauss et al., 2004), illustrating an intricate relationship between the data. To further illustrate, Trotman et al. (2019) unveiled that perceived HR, rather than actual HR, was more closely linked to anxiety ratings during anxiety-inducing tasks, such as competitions and mental arithmetic tasks.

In the study of anxieties related to learning (i.e., CA and FLA), research findings are mixed and sometimes contradictory, with very few studies attempting to triangulate anxiety self-reports with physiological measures, especially EDA. While most studies have not found significant associations between CA self-reports and HRV (MacIntyre et al., 2010; Gallego et al., 2021) or between FLA self-reports and both HR (Shachter et al., 2022; Schwerdtfeger, 2004; Sosa-López, 2019) and EDA (Gallego et al., 2021; Sosa-López, 2019), other studies have reported moderate to strong associations between self-reports of CA and either HR (Behnke & Beatty, 1981) or EDA (Sevinç, 2018). To the best of the author's knowledge, only Gregersen et al. (2014) have shown a relationship between self-reported FLA and HR. These mixed findings underscore the complexity of interpreting physiological responses in relation to subjective experiences.

In fact, Sosa-López (2019) observed no significant association between HR and EDA. The lack of a relationship between the HR and EDA measures observed in the study reflect the intricate nature of physiological responses and the myriad factors that can influence these measures. For instance, the absence of controlled room temperature, individual variability such as age and gender, the lack of control over other emotions which may cause arousal too (i.e., enjoyment), caffeine consumption, as well as methodological considerations such as the choice of equipment, the experimental set up, etc.

Schwerdtfeger (2004) highlighted two decades ago that self-reported measures of emotions often fail to align with autonomic reactivity assessments because autonomic arousal does not always correspond to one's subjective experience, especially when it is a retrospective judgment. To illustrate, participants can be influenced by memory biases and the changes in emotional state between the time of the physiological measurement and the reporting (Lindberg et al., 2021). In addition, learners may initially experience high anxiety levels in a specific situation (e.g., performing an oral presentation) but become habituated over time. This familiarity, combined with memory biases, can lead to the mismatch between their physiological responses and their subjective perception of anxiety.

Moreover, individuals might vary in their sensitivity to physiological arousal and their ability to accurately report their anxiety. Some learners might overestimate their anxiety despite experiencing minimal physiological responses, while others may experience significant physiological changes but underestimate them when self-reporting anxiety. This discrepancy arises because individuals' interpretations of their emotional state are influenced by how they perceive a situation and their previous experiences (Gallego et al., 2021). For example, Gallego et al. (2021) suggests that learners can interpret their physical reactions as excitement or facilitative energy, which could lead them to not report feeling very anxious about public speaking.

Another factor explaining these contradictory findings are methodological differences in collecting and analyzing physiological data. For example, Gregersen et al. (2014) found a strong degree of convergence in the rising and falling trends when comparing HR in relation to self-ratings with the idiodynamic task by segmenting their data into approximately 30 to 45 second intervals. Nevertheless, the researchers did not provide specific correlation coefficients but adopted a moving window to identify where data aligned (see Figure 12). According to Gregersen et al. (2014), longer time periods generated smaller correlations as the emotional dips and spikes essentially cancelled each other out. All in all, despite the efforts to understand the complex interplay between autonomic arousal and subjective experiences, studies have unveiled different outcomes, indicating that further research is needed.

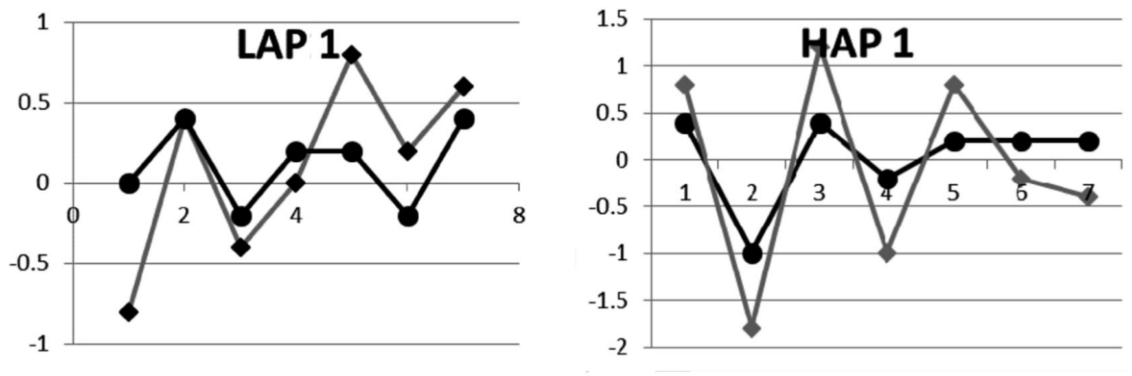


Figure 12: Slopes of heart rate (grey - rhombus) and the idiodynamic ratings (black - circles) from the low anxiety group and a high anxiety group (Gregersen et al., 2014, p. 579)

Note: LAP corresponds to low anxiety participant and HAP to high anxiety participant.

CHAPTER 2: L2 speaking anxiety and L2 speech

"Anxiety is both a cause and
a consequence of performance"
(MacIntyre, 2017, p.27).

This chapter examines the relationship between L2 speaking anxiety and L2 speech, starting with an overview of the effects of anxiety in Section 2.1. The chapter is organized into three subsections focusing on fluency (Section 2.2), accentedness and comprehensibility (Section 2.3), and lexical and structural accuracy and complexity (Section 2.4). Each section begins with a brief theoretical background on the respective construct, followed by a discussion of the documented effects of L2 speaking anxiety on that aspect.

2.1. Anxiety effects on L2 speech

In the last decades, much research has focused on examining the impact of FLA on testing situations and overall course grades, which serve as indicators of language achievement (See *Section 1.2.2*). These studies have consistently demonstrated a negative correlation between FLA and general L2 achievement. Despite this body of research, the question of whether anxiety causes poor performance or poor performance causes anxiety remains unanswered since the relationship between FLA and poor language performance can be bidirectional, with learners feeling increasingly anxious as they realize their limitations in the L2 while performing (Bailey, 1983; Pérez Castillejo, 2019). Despite the amount of research available on FLA, and research acknowledging speaking as the most anxiety-inducing skill, there remains a noticeable gap in the literature concerning the specific effects of anxiety on L2 speech. This section aims to address this gap by reviewing studies that focus on general oral test scores and learners' self-ratings of anxiety.

Although Teimouri et al. (2019) and Botes et al. (2020) meta-analyses did not find a strong relationship between FLA and language-specific competency scores (i.e., speaking), research has demonstrated a negative correlation between anxiety and general oral test scores (Gallego et al., 2021; Hewitt & Stephenson, 2011; MacIntyre & Gardner, 1994b; Pérez Castillejo, 2019; Phillips, 1992; Sutarsyah, 2017), official language exams such as IELTS (Woodrow, 2003) and university entrance exams (Young, 1986).

Moreover, anxious students tend to exhibit a predominantly negative attitude toward oral exams and their performances (Phillips, 1992; Tóth, 2011, 2017). In fact, Phillips (1992) described learners experiencing "going blank", frustration from their inability to express what they knew, being distracted, and a sense of panic. These observations are also present in research conducted by Tóth (2011, 2017) which further highlighted that experiencing anxiety led learners to have negative emotions towards their English classes and prompted reticent behavior in the classroom. Tóth (2017) offered additional examples of the challenges encountered by advanced EFL learners. For instance, highly anxious students complained about their inability to articulate their ideas effectively, leading to discomfort in front of their interlocutors (e.g., "I couldn't express my thoughts and feelings as well, as precisely as I would have liked to", p. 168). Additionally, they experienced difficulties in understanding the interlocutor and the content of the speaking prompts (e.g., "when I'm very anxious I don't understand what I'm reading, the letters are blurred before my eyes, I can't see words or sentences, just letters", p. 167), and anxiety hindered their ability to retrieve vocabulary (e.g., "in such an anxious state, words simply don't come; actually, I think they don't come because I'm afraid they won't", p. 167), and generate ideas (e.g., "it was not merely a language problem: that I didn't remember how to say something in English, but that I had no idea whatsoever, nothing came to my mind", p.167).

Besides the association of higher levels of anxiety with poorer performance in the use of complex grammatical features in an oral exam (Hewitt & Stephenson, 2012; Phillips, 1992), L2 speaking anxiety also interferes with vocabulary retrieval (MacIntyre & Gardner, 1994b; Tóth, 2017) and with accuracy in terms of grammatical complexity (Dörnyei & Kormos, 2000; MacIntyre & Gardner, 1994b; Gregersen, 2003; Robinson, 2007; Steinberg & Horwitz, 1986). Anxious learners' oral productions tend to be shorter and exhibit less lexical richness and language accuracy. Additionally, Gregersen (2003) observed that anxious learners not only made more errors, but also resorted to their native language more frequently, and recognized making fewer errors during the stimulated recalls.

Nonetheless, there remains a noticeable gap regarding the specific effects of L2 speaking anxiety on aspects of L2 speech production such as fluency, comprehensibility and accentedness. The following subsections report on the literature available. The current thesis will address this gap with the aim of contributing to our understanding of the complex interplay between affective factors and language learning outcomes.

2.2. Speaking fluency

"Does anxiety impair students' oral fluency, or do they become anxious in oral production tasks because their speech skills are low?" was a question posed by Bailey (1983, p. 68) concerning the relationship between FLA and oral fluency. While it is plausible that low proficiency levels trigger anxiety, leading to worse performance, previous research has indicated that FLA impacts proficient learners as well (See *Section 1.2.3*). Consequently, Bailey's question seems to reflect the classic chicken-and-egg scenario in language learning[5]. Knowing that FLA leads to worse performance, it would be logical to assume a causal relationship between heightened anxiety levels and poorer L2 speaking fluency. However, prior to exploring the possible impact of L2 speaking anxiety on speaking fluency, it is essential to first establish a clear definition of the concept and its measurement.

In teaching and assessing L2 speech production, educators and researchers consider the concept of speech fluency as mirroring the ability of L2 learners to formulate messages under time constraints. Thus, speaking fluency indicates ease of production without (unnatural) hesitations (De Jong, 2016) and efficiency of language processing (Kormos, 2006). However, Lennon (1990) distinguished between two interpretations of fluency: a broad definition, where fluency encompasses a learner's overall oral proficiency in the L2, and a narrow one, where fluency constitutes a component of oral proficiency. The present thesis adopts Lennon's narrow definition of fluency (fluency as one, isolatable, component of oral proficiency[6]), together with Segalowitz's (2010) framework for operationalizing L2 fluency quantitatively and qualitatively. According to Segalowitz (2010), fluency is an observable characteristic of real-time speech behavior, which can be broken down into three distinct subdomains: cognitive fluency, perceived fluency, and utterance fluency.

Briefly summarized, cognitive fluency refers to the fluency that characterizes a speaker and has to do with the speaker's abilities to efficiently plan and execute speech. In other words, it refers to the fluid operation (speed, efficiency) of the cognitive processes responsible for performing L2 speech acts. Perceived fluency, on the other hand, refers to the inferences listeners make about a speaker's cognitive fluency based on their perception of utterance fluency. Hence, it is a judgement made about speakers based

[5] It is impossible to decide which of two things caused the other one.
[6] This contrasts with the broad definition that may include accuracy and complexity.

on impressions drawn from speech samples. Finally, utterance fluency has to do with the actual properties of the utterances. As Segalowitz (2010) explains, utterance fluency refers to the temporal, pausing, hesitation, and repair characteristics which can be measured in speech. That is, they can be operationalized from speech samples (e.g., speech rate). Additionally, utterance fluency is closely related to perceived fluency (Derwing et al., 2004; Kormos & Dénes, 2004; Préfontaine, 2013; Rossiter, 2009) and cognitive fluency (de Jong et al., 2013; Kahng, 2014, 2020).

Regarding utterance fluency, Skehan (2003) thought of finer-grained distinctions, drawing attention to three subconstructs: breakdown fluency, speed fluency and repair fluency. Table 4 provides examples of widely used measures for utterance fluency across the three subconstructs, recognized for their robustness and reliability (see De Jong & Wempe, 2009; Kormos & Dénes, 2004). Nevertheless, despite their reliability, there is variability in the analysis of silent pauses due to pauses being defined with varying minimum durations by different researchers (Segalowitz, 2010). Although De Jong and Bosker (2013) recommend a lower cut-off point for silent pauses of 250–300 milliseconds (ms), other researchers have used a wide variety of thresholds, from 100 to 1000 milliseconds (ms) (Segalowitz, 2016).

Breakdown fluency has to do with the hesitation phenomena and the flow of speech. Measures of breakdown fluency involve quantifying the number and length of pauses, whether filled (e.g., *uhm, err*) and/or silent. Additionally, Kormos (2006) highlighted the significance of pause location, as it reflects the different cognitive processes involved in speech. For example, pauses occurring at the end of clauses signal the speaker's conceptualization of intended content (i.e., *what* to convey), while pauses within clauses denote the speaker's ongoing linguistic encoding to articulate the message (i.e., *how* to say the message).

On the other hand, speed fluency is related to the speech rate, that is, the speed with which a speech is delivered. An example of a speed fluency measure would be computing the number of syllables uttered per unit of time (e.g., seconds). As for repair fluency, it involves false starts, reformulations, self-corrections, or repetitions made by the speaker during speech production. Similarly to speed fluency, measures of repair fluency can be obtained by computing the frequency of these dysfluencies per unit of time (e.g., a minute). These three subconstructs can vary depending on the speech elicitation tasks used and the task difficulty (see Tavakoli & Skehan, 2005).

Table 4: Examples of utterance fluency measures (from Segalowitz, 2010, p. 6)

Measure (units)	**Definition**
Speech rate[1]	The total number of syllables is divided by total time (including pauses) in seconds.
Articulation rate[1]	The total number of syllables is divided by total time (excluding pauses) in seconds.
Phonation-time ratio[1]	Percentage ratio of time speaking to time to time taken from the whole speech sample.
Mean length of run[1]	Average number of syllables between pauses (period of silence ≥ 250 ms).
Silent pauses per minute[2]	The total number of pauses (periods of silence ≥ 250 ms) is divided by the total time speaking in seconds.
Mean length of pauses (seconds)[2]	Mean length of all pauses (periods of silence ≥ 200 ms).
Filled pauses per minute[2]	The total number of filled pauses (*uhm, mm, er*) divided by the total time speaking in seconds.
Dysfluencies per minute[3]	The total number of dysfluencies (repetitions, restarts, repairs) divided by the total time speaking in seconds.

Note: [1] = speed, [2] = breakdown, [3] = repair

When investigating fluency, two main issues emerge. The first aspect concerns the relationship between L1 and L2 performance (De Jong, 2015; Derwing et al., 2009; Kahng, 2020). Compared to their knowledge and fluency in the L1, learners are typically not only less proficient in their L2, but also considerably less fluent (Segalowitz, 2010). However, there are IDs in temporal aspects of speech (Kahng, 2020). That is, individuals vary in their fluency levels when speaking and this distinction is evident not only among non-native speakers but also among native speakers of any language. To illustrate, individuals may exhibit differences in speaking speed, frequency of filled pauses (*uhm, ehm*), or the duration of pauses (De Jong, 2013). Thus, if a speaker frequently uses filled pauses in their native language, it may be futile to expect this individual not to use significantly more filled pauses in their L2 or any other FL in which they are less proficient. In fact, utterance fluency measures such as speed, pause, and repair phenomena exhibit moderate to strong correlations between the L1 and the L2 (De Jong et al. (2015), and L2 fluency measures can be predicted to a certain extent based on L1 fluency (e.g., 57% for breakdown fluency) (De Jong et al., 2015; Derwing et al., 2009;

Kahng, 2020). Therefore, as Segalowitz (2010) highlights, obtaining fluency data in the L1 to use as baseline measures is key. Thus, when assessing aspects of L2 fluency, researchers should collect data in both the L2 and the speaker's L1 to determine whether fluency-related phenomena (e.g., hesitations, speech rate) are characteristic of the individual's general speaking patterns rather than just characteristic of their L2 speech (Segalowitz, 2010).

Another key aspect in L2 fluency research is not the number and length of disfluencies, but rather their distribution (De Jong, 2016; Kahng, 2014, 2020). Previous research has found that while pause distribution is an indicator of L2 proficiency, pause duration is not (De Jong, 2013; De Jong 2016; De Jong et al., 2015). For example, fluent speech tends to have pauses at grammatical junctures, whereas non-fluent L2 speech often presents pauses within clauses or utterances (De Jong, 2016; Kahng, 2014, 2018). As De Jong (2016) explains, proficient L2 learners and NSs can produce multi-word chunks, allowing them to focus their attention on planning the upcoming message without needing to concentrate on the linguistic formulation. Consequently, pauses in their speech tend to occur predominantly at clause boundaries, that is, at syntactic boundaries. In contrast, less proficient L2 speakers, who may not yet have a large repository of prefabricated chunks, might need to pause within clauses to concentrate on lexical retrieval. In fact, previous research has demonstrated that one of the biggest differences between L1 and L2 utterance fluency was in the number of pauses *within* a clause (De Jong, 2016), which reflects difficulties in L2 planning or speech encoding (Kahng, 2014). Moreover, research have found that when speaking in the L2, learners tended to pause more frequently and for longer within clauses. However, at ASU boundaries, speakers tended to pause for similar durations both in the L2 and in the L1. In fact, Kahng (2014) found that silent pause time within clauses in L2 speech was almost double compared to that in L1 speech (32% vs. 17%). Finally, findings showed that the higher the L2 proficiency, the less likely speakers were to pause within clauses, and the pauses within clauses became shorter (De Jong, 2016).

2.2.1. Anxiety and L1 speaking fluency

The relationship between fluency and L1 anxiety has predominantly focused on scrutinizing pause duration and hesitations, with findings consistently indicating that experiencing anxiety leads to dysfluencies (Buchanan et al., 2014; Christenfeld &

Creager, 1996; Goberman et al., 2011). According to Buchanan et al. (2014), behavioral indicators of stress, such as speech and language measures, may more accurately reflect physiological stress responses than self-reports because they happen simultaneously with the physiological reactions and are less influenced by conscious control compared to self-reported measures. In fact, Buchanan et al. (2014) identified low but significant correlations between the physiological measures (HR and cortisol) and pausing time, especially during the first and last minutes of public speaking. Christenfeld and Creager (1996) and Goberman et al. (2014) also found significant relationships between L1 anxiety and pause duration, with the key distinction being that these studies relied on self-reports to measure anxiety. These results suggest the existence of a relationship between speech fluency and both subjective measures and physiological stress responses.

2.2.2. L2 anxiety and L2 speaking fluency

Factors related to L2 fluency include general L2 proficiency (e.g., Kormos & Dénes, 2004), and L1 fluency (e.g., Derwing et al., 2009), as well as other user-external and user-internal variables (see Lintunen et al., 2020 for a review). Among all these factors, the affective characteristics of L2 learners, such as FLA, have received comparatively less attention regarding their potential impact on L2 fluency. In fact, much of the initial investigation into the relationship between anxiety and fluency centered around participants' native languages and L1 communicative apprehension (see *Section 2.2.1*). Fortunately, with the growing interest in FLA and its implications for L2 acquisition, more studies have begun investigating the effects of FLA on L2 fluency.

Given the findings of L1 studies and the relationship between cognition (WM and attention) and anxiety, L2 speaking anxiety is hypothesized to negatively affect L2 speaking fluency. One of the first studies to explore the relationship between L2 fluency and anxiety was Sanaei et al. (2015), which examined L2 speed fluency measures including speech rate (SR), articulation rate (AR), phonation-time ratio (PTR) and mean length of runs (MLoR), along with FLA, measured with the FLCAS. Findings revealed significant strong correlations between FLCAS scores and PTR as well as MLoR, while no significant correlations were found with AR and SR. This suggests that anxiety influences certain aspects of speed fluency, such as the amount of time spent speaking and the average length of continuous speech without pauses, but it does not impact the rate of speech production. Anxious learners spend less time speaking and more time pausing, and have shorter runs of uninterrupted speech, meaning their speech is more frequently interrupted by pauses. FLA seems to influence the overall flow of speech, but it does not necessarily slow down the rate at which words are uttered when the learner is speaking.

However, Pérez Castillejo (2019) acknowledged various constraints that limit the generalizability of Sanaei et al.'s (2015) results: the oral proficiency level of the learners was not controlled for, it was a small sample size, and their FLCAS scores were very low, classifying them as "fairly anxious" rather than highly anxious learners. Furthermore, Pérez Castillejo (2019) argued that Sanaei et al. (2015) provided participants with too much planning time, raising the question of whether this extra time was used to allocate more attentional resources for encoding, rather than for planning the message.

Similar results were found by Pérez Castillejo (2019, 2023) who also examined the relationship between L2 speed fluency and FLA, measured with the FLCAS. However, the researcher controlled global oral proficiency through an elicited imitation task (Wu et al., 2021) and did not provide planning time prior to task performance. Additionally, she computed breakdown fluency measures, including pause length and frequency (joining both filled and unfilled pauses). These pauses were differentiated according to their position: pauses between Analysis of Speech Unit (AS-Unit or ASU), which would indicate to what extent the speaker needs to stop the performance to think of what to say, or within AS-Units, which would indicate the time the speaker needs to stop to encode the message. As defined by Foster et al. (2000), an ASU is "a single speaker's utterance consisting of an independent clause, or a subclausal unit, together with any subordinate clause(s) associated with either" (p. 365). Interestingly, only two breakdown fluency indices were found to be significantly related to anxiety: ratio of mid-AS pauses and mean length of pauses between AS-Units. Hence, anxious learners made longer pauses at AS-Unit boundaries and more pauses within AS-Units than their non-anxious counterparts, the latter probably due to difficulties in encoding the message lexically and grammatically (Kormos, 2006). As for the longer pauses observed at the end of AS-Units, Pérez Castillejo (2019) posited that they may be due to the lack of planning time prior to the speaking performance.

Finally, to further explore the relative contributions of FLA and oral proficiency to the variance in PTR, MLR, and pause duration at AS-Unit boundaries, the researcher ran regression analyses which revealed that FLA and proficiency significantly accounted for a part of the variance in each of these three fluency measures: 26.2% of the variance in pause duration at AS-Unit boundaries, 36.2% in PTR, and 40.1% in MLoR. Furthermore, the Beta coefficients showed that anxiety contributed considerably more than proficiency to these three fluency measures. A later study by Bielak (2022) also revealed that proficiency and FLA accounted for variance in speed and breakdown fluency measures. However, in contrast to Pérez Castillejo's (2019) findings, oral proficiency (as measured by Elicited Imitation Task as well) emerged as the strongest predictor. This disparity may be attributed to the advanced proficiency level of Bielak's (2022) participants. Accordingly, the researcher inferred that at higher proficiency levels, emotions might have a comparatively lower impact on fluency, presumably due to more automatized L2 use.

In a similar vein, Aubrey (2022), Simard et al. (2023) and Bielak (2022) found FLA to be negatively related to speed and breakdown fluency measures. Aubrey's (2022) case study reported that FLA was weakly to moderately related with pause duration in the speech of three out of four participants. Simard et al. (2023) found a negative relationship between SR and FLA. In turn, Bielak (2022) compared FLA in the performance of a group task and a monologue task, and findings revealed that FLA experienced in the group task was negatively correlated with MLR and AR, and positively with end AS-Unit pause ratio. In the monologue task, FLA was only negatively associated with end AS-Unit pause ratio. Hence, unlike Pérez Castillejo (2019), Bielak's (2022) study indicated that FLA correlated with an index of breakdown fluency representative of message conceptualization. In contrast to previous studies, Kormos and Préfontaine (2017) did not find any significant relationship between anxiety and either pause frequency or pause duration measures across three monologue task types, suggesting that anxiety did not affect performance.

However, none of the previous studies mentioned differentiated between silent and filled pause behavior as done by Lindberg et al. (2021), Mora et al. (2023), Rood and De Jong (2023) and Szyszka and Lintunen (2023). Lindberg et al. (2021), Rood and De Jong (2023) and Szyszka and Lintunen (2023) found significant differences between anxious learners and their non-anxious counterparts regarding filled pause frequency. Findings suggest that anxious learners produce much more filled pauses than their non-anxious counterparts. Additionally, Szyszka and Lintunen (2023) observed that the length and frequency of silent pauses did not differentiate between highly anxious learners and those with lower anxiety levels.

It must be highlighted that while previous studies controlled for L2 proficiency levels, none of them controlled for WM, except Mora et al. (2023) who used the Digit Span test. Moreover, to the best of the author's knowledge, Mora et al. (2023) is the only study examining FLA both subjectively, via 9-point Likert scale self-reports, and objectively, through HR measurements, in relation to L2 utterance fluency, comprehensibility, accentedness and lexico-grammatical accuracy. Focusing on L2 utterance fluency, findings aligned with previous studies even though weak to moderate correlations were found between learners' self-reports, repair fluency measures and pause frequency only when combining silent and filled pauses per 100 syllables (Kahng, 2020) in mid-clause position and when not differentiating their distribution. Also, in line with previous studies, no significant correlations were found regarding pause duration. These findings imply that learners who exhibited a higher frequency of pauses and repairs in their speech also reported experiencing greater levels of anxiety during task performance. On the other hand, the physiological measure of anxiety (HR) was only weakly associated with mid-clause pause frequency, suggesting that those learners who produced more mid-clause pauses experienced higher levels of anxiety.

Finally, like Pérez Castillejo (2019) and Bielak (2022), Mora et al. (2023) conducted further examinations through hierarchical regressions to unveil the relative contributions of L2 proficiency (Elicited Imitation Task scores), WM, and speaking anxiety in explaining the variance in L2 oral production. Findings revealed that L2 proficiency accounted for a significant amount of variance in breakdown fluency measures, while WM made only an additional significant contribution in the case of mid-clause pause frequency. However, when controlling for L2 proficiency and WM, speaking anxiety significantly explained an additional variance in pause frequency measures (13-15%), but not in pause duration measures. Furthermore, Mora et al. (2023) found that self-reported speaking anxiety contributed more strongly than heart rate. These results suggest that, independently of proficiency and WM, speaking anxiety in L2 learners is primarily reflected in the number of pauses they make rather than the length of these pauses, and that self-reports were more sensitive than HR when capturing individual variability in pause frequency. Although Mora et al. (2023) controlled for proficiency and WM, the role of attention alongside these individual factors remains unknown.

Overall, the previous research reviewed above suggests a clear relationship between L2 speaking anxiety and L2 utterance fluency. However, several questions remain unanswered, such as the role of attention in addition to WM and proficiency. Moreover, further research is needed to understand how L2 learners' anxiety interacts with L2 speaking fluency since only Mora et al. (2023) have combined physiological and self-reported measures of L2 speaking anxiety and have scrutinized its association with L2 speech. So far, to the best of the author's knowledge, no study has examined HR together with EDA and L2 speech effects.

2.3. Accentedness and comprehensibility

Concerning L2 production assessment, specific constructs like breakdown fluency can be operationalized through objective measures; however, L2 speech can also be evaluated subjectively through listener ratings such as comprehensibility, fluency and accentedness (Munro & Derwing, 1995a). Considering the relationship between L2 speaking anxiety and L2 utterance fluency, it can be hypothesized that anxiety may also impact other aspects of L2 speech, such as accentedness and comprehensibility. However, before examining previous research on the impact of L2 speaking anxiety on accentedness and comprehensibility, it is key to establish clear definitions of these concepts.

On one hand, accentedness can be defined as "non-pathological speech produced by L2 learners that differs in partially systematic ways from the speech characteristics of native speakers" (Munro 1998, p. 139). In other words, it refers to how much of a non-native accent can be perceived in someone's speech, which might be attributed to both segmental and suprasegmental errors (Pinget et al., 2014). To exemplify, segmental errors occur when one phoneme is substituted for another (e.g., producing /ʌ/ instead of /æ/ for the word "fan") . In contrast, suprasegmental errors involve assigning lexical stress to the wrong syllable (e.g., saying 'computer instead of com'puter) and intonation (i.e., pitch variations).

On the other hand, comprehensibility refers to "the listener's perception of how easy or difficult it is to understand a given speech sample" (Derwing & Munro 2009, p. 478). As Derwing and Munro (2009) clarify, this dimension implies a judgment of difficulty, not the degree of a listener's actual comprehension of an utterance, which would be intelligibility. Therefore, these ratings capture the effort required by the rater to process utterances, even if they are ultimately perfectly understood. Comprehensibility is considered more important in successful L2 communication than accentedness owing to the fact that a lack of perceived fluency and comprehensibility may lead to the loss of listeners' attention (Derwing et al., 2004; Suzuki & Kormos, 2020). In fact, according to Trofimovich and Isaacs (2012), non-native accent does not necessarily impede successful communication. Indeed, learners whose L2 productions are not native-like but who are able to communicate effectively are successful L2 users. However, a strong foreign accent may lead to partial or complete communicative misunderstandings due to listeners' inability to decode a phonetic segment, a word, or larger units. Thus, accentedness is considered one of the most salient aspects of speech and a highly sensitive dimension for listeners (Derwing & Munro, 2009; Trofimovich & Isaacs, 2012), encompassing both dialectal differences and phonological variations resulting from L1 influence on the L2.

Comprehensibility has been found to be closely related to, but independent from accentedness (Derwing & Munro, 2009; Shintani et al., 2019; Trofimovich & Isaacs, 2012). Although heavily accented speech can be highly comprehensible (Derwing & Munro, 2009), Munro and Derwing (1995b) found that a strong foreign accent led to lower perceived comprehensibility ratings and entailed a greater response latency. In fact, previous research has shown that raters tend to judge accentedness more harshly than comprehensibility (Derwing et al., 2004; Derwing & Munro, 2009).

Several studies have investigated which aspects impact listeners' understanding through subjective judgments of comprehensibility (i.e., pronunciation, fluency, lexis, and grammar). According to Kennedy and Trofimovich (2019), the linguistic elements linked to L2 speakers' English comprehensibility can be separated into two dimensions: pronunciation and lexicogrammar, the latter encompassing the variety and richness of vocabulary as well as the accuracy or complexity of grammar. Indeed, previous research by Trofimovich and Isaacs (2012) and Saito et al. (2017) have found that while raters' comments on speech samples primarily focused on segmental errors (vowels and consonants) and word stress for accentedness, they relied on pronunciation (i.e., intonation and rhythm) as well as lexicogrammar aspects for comprehensibility (see Saito, 2021 for a meta-analysis). However, other studies such as Suzuki and Kormos (2020) have also found that besides morphological errors, listeners' ratings of comprehensibility can also be related to speed and breakdown fluency measures (i.e., articulation rate, mid-clause pause duration and ratio). According to Saito et al. (2017), differences between raters' assessments of comprehensibility and accentedness reflect the specific demands of each task, with comprehensibility judgments considering all the linguistic information available in L2 speech to process meaning, and accentedness evaluations focusing specifically on pronunciation to determine inter-learner differences in L2 pronunciation accuracy.

Nevertheless, Strachan et al. (2019) suggested that the speakers' and listeners' L1 background and the demands of the speaking task would determine which linguistic features are important, especially for comprehensibility (Crowther et al., 2015; Suzuki & Kormos, 2020). Regarding the listeners' background, most research has employed native speakers of the learners' target language; however, research also advocates that linguists and language teachers are more accurate perceivers than inexperienced listeners when assessing FL speech accentedness, precisely because of their special training (Flege, 1984; Gallardo del Puerto et al., 2015).

Previous research has found mixed results on how listeners' background profiles affect their L2 speech ratings. Some studies have unveiled that native listeners rate accentedness or comprehensibility more leniently when they are familiar with the non-native accent in the speech that they are judging, especially if they have teaching experience (see Bergeron & Trofimovich, 2017; Crowther et al., 2015; Gass & Varonis, 1984; Rossiter, 2009; Shintani et al., 2019). This finding echoes Bradlow and Bent's (2008) notion of 'perceptual adaptation' which suggests that native speakers can adjust to foreign-accented L2 English that deviates substantially from pronunciation norms. Conversely, other studies have not found L2 familiarity to be a significant predictor in judgements of accentedness or comprehensibility (Isaacs & Thomson, 2013; Kennedy & Trofimovich, 2008; Saito et al., 2019). Thus, given these results, it is important to carefully consider the choice of raters when assessing accentedness and comprehensibility.

2.3.1. L2 anxiety, accentedness and comprehensibility

Previous research investigating the relationship between FLA and accentedness and comprehensibility is scarce, compared to the studies that have examined the impact of FLA on speaking fluency. However, some studies have revealed negative relationships between FLA and self-perceptions of pronunciation accuracy (e.g., Baran-Łucarz, 2011, 2022; Szyszka, 2017). Additionally, there is empirical evidence supporting the construct of pronunciation anxiety (Baran-Łucarz, 2014), and a specific 44-item questionnaire to measure this dimension (Baran-Łucarz, 2013, 2016, 2017, 2021).

Emotional states have been reported to affect learners' English pronunciation (Derwing & Rossiter, 2002; Price, 1991). For example, Price (1991) found that highly anxious students believed their pronunciation was not accurate, and Derwing and Rossiter (2002) reported that 60% of their participants perceived a change in their accent, particularly when they were excited or nervous. Baran-Łucarz (2011) revealed that high anxiety learners rated their pronunciation as poorer than their low anxiety counterparts. Similar findings were also observed in studies by Szyszka (2011) and Kralova et al.'s (2018), suggesting that anxiety shapes learners' impression of their competence and accuracy in pronunciation.

Nevertheless, despite the existing body of research on pronunciation anxiety, the possible effects of L2 speaking anxiety on accentedness and comprehensibility are under-researched. To the best of the author's knowledge, there are very few studies which have examined the relationship between L2 speaking anxiety and the degree of perceived accentedness and comprehensibility. One of the first studies was MacIntyre and Gardner's (1994b), in which judges analyzed L2 French speeches in terms of fluency, accentedness and sentence complexity. Findings unveiled that anxious learners were perceived to have less of a French accent, exhibited lower levels of fluency, and produced less complex sentences.

It is only recently that comprehensibility (Saito et al., 2018; Nagle et al., 2022; Nagle et al., 2023), or comprehensibility together with accentedness (Mora et al., 2023; Suzukida & Saito, 2023; Zhou, 2024) have been included in studies on FLA. For instance, Saito et al. (2018) investigated comprehensibility gains in Japanese high-school students learning English depending on their different motivations, emotions (anxiety vs. enjoyment), and EFL experience profiles (e.g., L2 use). The findings revealed that the participants' comprehensibility scores were mainly explained by anxiety (10.6%) and to a lesser extent by motivation (4.8%) at the onset of the 3-month academic term. In addition, anxiety correlated negatively with Japanese EFL high-school students' speech comprehensibility both at the onset of the 3-month period and with their comprehensibility gains over that period. This suggests that anxiety can interfere with L2 comprehensibility and its development. Furthermore, research by Nagle et al. (2022) revealed that comprehensibility could be predicted by the speakers' perceived levels of anxiety, suggesting greater comprehensibility was associated with less perceived anxiety levels.

However, Nagle et al. (2023) and Suzukida and Saito (2023) found no evidence that anxiety contributed to comprehensibility scores. Instead, they observed that experiential factors, such as English learning outside the classroom, were more influential. In contrast, for accentedness, both Suzukida and Saito (2023) and Zhou (2024) showed that anxiety played a significant role. Specifically, Suzukida and Saito (2023) demonstrated that anxiety, alongside phonemic coding ability, contributed to determining accentedness score. For this reason, the researchers concluded that L2 learners with higher phonemic coding ability and lower anxiety levels were more successful at minimizing the influence of their L1 sound system on their English pronunciation.

Despite the previous findings, Mora et al. (2023) found that while anxiety was related to some breakdown fluency measures, neither physiological measures (HR) nor self-reported speaking anxiety levels were related to accentedness or comprehensibility. Hence, the researchers suggest that the native listeners did not detect the extent to which L2 learners' levels of anxiety may have manifested in their speech when assessing L2 speech. All in all, given the limited number of studies and the mixed results found when examining anxiety in relation to accentedness and comprehensibility, further research is needed to explore L2 speaking anxiety effects.

2.4. Lexical and structural accuracy and complexity

Anxiety has been shown to have a detrimental effect on the attention-related processes governing speech production, such as lexical activation and grammatical encoding (see *Section 1.2.4*). As a result, higher levels of anxiety while performing an oral task in an L2 may be expected to lead to decreased lexical and grammatical accuracy. However, to the best of author's knowledge, not many studies have examined L2 speaking anxiety in relation to lexical and structural accuracy and complexity.

Phillips (1992) sought to examine associations between anxiety and various oral test performance variables related to accuracy and comprehensible output (e.g., total words in communication units, average length of communication units, percent of error-free of communication units, and number of dependent clauses, among others). Findings revealed that students with higher levels of anxiety performed worse on their oral tests since they used noticeably fewer dependent clauses and produced shorter communication units (a measure similar to T-units for evaluating syntactic maturity).

In a similar vein, the replication by Hewitt and Stephenson (2012) also revealed that students with higher levels of anxiety performed worse on their oral exams compared to their more relaxed peers. Their study also aligned with Phillips (1992) in showing a negative correlation between FLCAS scores and the total words in communication units[7], indicating that highly anxious students produced shorter communication units. Nevertheless, unlike Phillips's (1992), Hewitt and Stephenson (2012) found a significant relationship between FLCAS scores and the percentage of total words in mazes (i.e., stuttering and repetitions), suggesting that anxious students produced more repetitions during their oral exams.

Other studies have observed that anxious learners construct less complex sentences (MacIntyre & Gardner, 1994b), make more errors and resort to their native language more often (Gregersen, 2003), use less varied vocabulary (Kormos & Dörnyei, 2004), have less ability to self-repair production errors (Sheen, 2008), and produce more repetitions as well as grammatical and lexical errors in their speech (Mora et al. 2023). Despite the general consensus on the detrimental effects of anxiety, it is worth noting that Révész (2011) found in her study that anxiety was unrelated to measures of grammatical and lexical accuracy (i.e., errors per AS-Unit, error-free AS-Units, and self-repairs). While substantial evidence highlights the negative impact of L2 speaking anxiety on language performance, more research is always needed to fully understand the complexities of this relationship.

[7] A communication unit (C-unit) is a measure used in discourse analysis, especially in spoken language research, to segment speech into meaningful units. A C-unit consists of an independent clause and any of its dependent (subordinate) clauses. It is similar to a T-unit but can also include sentence fragments used in spoken language.

Part II

THE STUDY

CHAPTER 3: Objectives and research questions

The aim of this chapter is to summarize the main contributions of the present study with respect to the literature reviewed in Chapters 1 and 2 (Section 3.1). Then, the research questions that have motivated this study are stated (Section 3.2) and, subsequently, the corresponding hypotheses are formulated (Section 3.3).

3.1. Goals

The present study is concerned with five key areas related to L2 speaking anxiety under various communicative situations that differ in the extent to which they can potentially lead to emotional arousal: (1) the effects of task type, context and language; (2) the alignment between self-reported anxiety and others' perceptions and the most prominent non-verbal anxiety cues; (3) the alignment between subjective and physiological measures; (4) anxiety effects on L2 speech focusing specifically on fluency, accentedness and comprehensibility; and (5) sources of L2 speaking anxiety. With these goals, this study aims at advancing our knowledge in the field of FLA and SLA, contributing to research on L2 speaking anxiety effects on second language acquisition in Spanish EFL learners studying an English degree at university. Thus, there are several important areas to which this study aims to contribute to:

1. **Effects of L2 speaking anxiety**: This study seeks to contribute to the literature on the effects of L2 speaking anxiety on L2 speaking fluency, accentedness and comprehensibility. While a considerable amount of research in the last decade has measured the effects of FLA on overall language achievement, such as grades, the impact of anxiety on L2 speech has been researched to a lesser extent, especially with regards to accentedness and comprehensibility (Mora et al., 2023; Suzukida & Saito, 2023).

2. **L2 speaking anxiety, audience and task type**: This study evaluates the potential detrimental effects of anxiety on L2 speech elicited in front of different types of audience (one-to-one vs. an audience of seven), a contextual condition that has not been previously examined while performing different kinds of oral tasks (Kormos & Préfontaine, 2017; Tóth, 2017). Based on previous research, the level

of preparedness before performing the oral task and the degree of familiarity with interlocutors were considered when designing the study (Baran-Łucarz, 2014).

3. **Dynamic approach of FLA**: This study aims to contribute to a *Dynamic Approach* in measuring FLA (MacIntyre, 2017). L2 speaking anxiety was examined from a dynamic perspective through the idiodynamic task and physiological sensors (as in Croft et al., 2004; Gregersen et al., 2014; Gregersen et al., 2017; Westenberg et al., 2009).

4. **Subjective and objective measures of L2 speaking anxiety:** This study intends to clarify the intricate relationship between physiological arousal caused by anxiety and its subjective evaluations, for which research has found mixed results (see *Section 1.6.4*). Therefore, this study includes not only self-reports but also the idiodynamic task along with both HR and EDA as physiological measures. To date, and to the best of the author's knowledge, no study has used all of these measurements together to examine the relationship between subjective and objective measures of anxiety.

5. **Self- and others' impressions**: This study intends to determine the extent to which self- and others' impressions of anxiety during an oral presentation performance are related to one another (Gregersen, 2007a, 2009; Gregersen et al., 2017). Furthermore, this study aims to expand Gregersen (2005) and Gregersen et al.'s (2017) research by examining FLA non-verbal cues with a larger sample size.

6. **Sources of L2 speaking anxiety**: This study aims to contribute to the extensive research on the sources of L2 speaking anxiety. Similar to Tóth (2019), this study was conducted with undergraduate EFL learners; however, the participants in the present study are representative of EFL Spanish university students, most of whom are L1 Spanish/Catalan speakers. Moreover, while Tóth (2019) examined the sources of anxiety in oral presentations in general, this study will expand on that by examining whether the causes are different depending on the types of audience (one-to-one vs. small group of seven) and the types of tasks (descriptive vs. argumentative).

7. **IDs and L2 speech performance**: This study considers individual differences as predictors of L2 speaking performance. Previous studies have not simultaneously considered proficiency along with WM and attention. Hence, this study is designed to offer insights into the role of anxiety bearing in mind experiential and cognitive factors (Mora, 2022; Suzukida & Saito, 2023).

Overall, this research aims to enhance our understanding of the interplay between emotional and linguistic factors in L2 learning, which will contribute to more effective language teaching methodologies. For this purpose, eighty-six first-year university EFL students, primarily L1-Catalan/Spanish speakers, performed a total of eight speaking tasks in counterbalanced order. Four of the presentations were performed in their L2 (English) and the remaining four in their L1 to obtain baseline data. This study included two presentation modes (one-to-one vs. group, *N* = 7) and two task types (descriptive vs. argumentative). After task performance, the "audience" assessed the perceived level of nervousness, while the EFL learners self-assessed their distress with post-task questionnaires and the idiodynamic task. Additionally, stimulated recalls were carried out for the L2 oral presentations, in which the EFL learners explained the reasons of their distress during the oral presentations. Control tests for proficiency, vocabulary size, WM and attention were administered. Finally, accentedness and comprehensibility were assessed through native speakers' ratings.

3.2. Research Questions

This dissertation poses five main research questions (RQ):

Research question 1 (RQ1): Effects of Language, Task Type and Context on Speaking Anxiety

- RQ1.1: Is there a relationship between the L1 and L2 speaking anxiety scores (obtained from the LBQ) and physiological measures of anxiety (i.e., HR, HRV and EDA) as well as subjective measures of anxiety (i.e., self-reported anxiety levels and the idiodynamic task measures)?
  - RQ1.1.1: Does the relationship between L2 speaking anxiety scores, physiological measures of anxiety and subjective measures of anxiety

differ as a function of whether participants are highly or non- anxious learners?

- RQ1.2: Are learners' speaking anxiety levels, as measured through physiological and subjective measurements, affected by task type (descriptive vs. argumentative), context (one-to-one vs. group audience), and language (L2 vs. L1)?
  - RQ1.2.1: Does task type affect speaking anxiety levels?
  - RQ1.2.2: Does context affect speaking anxiety levels?
  - RQ1.2.3: Does language affect speaking anxiety levels?
  - RQ1.2.4: Are language, task type and context independent of one another?

Research question 2 (RQ2): Self-reports vs. others' perceptions

- RQ2.1: To what extent do self-reports of anxiety correlate with others' perceptions during oral presentations?
  - RQ2.1.1: Are others' perceptions of speaking anxiety levels affected by task type (descriptive vs. argumentative), context (one-to-one vs. group audience), and language (L1 vs. L2)?
- RQ2.2: What non-verbal cues are most indicative of L2 speaking anxiety for the audience?

Research question 3 (RQ3): Subjective and physiological measures of anxiety

- RQ3: Are subjective measures of anxiety (i.e., self-reported anxiety levels and the idiodynamic measures) related to physiological measures of anxiety (i.e., HR, HRV and EDA)?
  - RQ3.1: Do relationships between subjective measures of anxiety and physiological measures differ between languages (L1 vs. L2)?
  - RQ3.2: Do relationships between subjective and physiological measures of anxiety differ between individuals who reported high and low anxiety levels?

Research question 4 (RQ4): L2 Speaking Anxiety and L2 speech

- RQ4.1: Is there an association between L2 learners' anxiety levels and L2 linguistic features of speaking performance?

- RQ4.2: To what extent do objective and subjective measures of L2 speaking anxiety contribute to explaining variance in L2 speaking performance, while controlling for IDs in L2 proficiency and cognitive factors (WM and attention)?

Research question 5 (RQ5): Sources of L2 speaking anxiety

- RQ5: What are the main sources of L2 speaking anxiety among EFL Spanish university students?
  - RQ5.1: Do highly and non-anxious learners differ in the sources of L2 speaking anxiety?

3.3. Hypotheses

Several hypotheses are formulated supported by previous research in the field of FLA and L2 speaking anxiety. Concerning RQ1.1, it is hypothesized that there will be a positive relationship between L1 and L2 speaking anxiety (as measured from the LBQ) and both physiological measures of anxiety and subjective measures of anxiety. While the questions in the LBQ target trait anxiety, and the study measures (i.e., physiological and subjective anxiety measures) assess state anxiety, it is hypothesized that these two types of anxiety will be related. Previous studies have demonstrated a positive relationship between CA and HR (Behnke & Beatty, 1991; Booth-Butterfield, 1987; Croft et al., 2004; Felnhofer et al., 2014; Westenberg et al., 2009). Therefore, similar findings are expected with EDA and subjective measures of anxiety (i.e., self-reports and idiodynamic scores). Moreover, since CA in the L1 can be used as a predictor of FLA (McCroskey et al., 1985), it is anticipated that students who show L1 trait anxiety will also experience L2 anxiety.

RQ1.2 aims to examine how L2 speaking anxiety levels are affected by task type, context and language. Main effects of these three variables are expected to be found. First, it is hypothesized that different communicative contexts will generate different levels of distress even though both contexts involve giving an oral presentation, a highly-anxiety provoking activity (Woodrow, 2006), in front of an unfamiliar audience (MacIntyre & Thivierge, 1995) with limited time to prepare (Young, 1990).

Regarding task type, it is expected that the argumentative tasks will lead to higher levels of distress as observed in Sosa-López & Mora (2022a) and Tóth (2017). Although the descriptive tasks involve reporting on a personal experience, which could cause discomfort due to the personal nature of the task requiring participants to expose themselves by sharing personal details, perceptions of anxiety depend on the cognitive demands and the design features of the task (Kormos & Préfontaine, 2017). In this case, argumentative tasks require a higher level of complexity and accuracy, needing more linguistic command to justify choices.

To the best of the author's knowledge, only Beatty & Behnke (1991) has explored how audience size affects L1 anxiety, and previous research on L2 speaking anxiety has not explored its effects. Previous studies have found that L2 students are more comfortable engaging in communicative activities in small groups rather than in front of the class (Young, 1990). However, even in more "comfortable" situations such as pair work (Woodrow, 2006; Young, 1990) or performing in the L1 (Croft et al., 2004; Westenberg et al., 2009), learners report experiencing speaking anxiety. Hence, in line with previous findings, it is predicted that participants will experience more heightened L2 speaking anxiety during an oral presentation before an audience of seven compared to the one-to-one scenario because performing in front of a larger group of people can be more intimidating.

Similarly, to the best of the author's knowledge, previous research has not compared L1 and L2 speaking anxiety during oral presentations. It is highly likely that L1 anxiety levels are not comparable to those experienced when presenting in the L2. Although oral presentations are a typical classroom activity, learners encounter more difficulties when presenting in a foreign language class due to feeling less in control of the communicative situation (Horwitz et al., 1986).

With respect to RQ2 on the relationship between self-assessments of anxiety and others' perceptions, previous studies have claimed the existence of the "illusion of transparency", meaning that when anxious, learners are hypersensitive to their environment and tend to overestimate the extent to which their anxiety symptoms are visible (King & Smith, 2017; Gregersen et al., 2017; Savitsky & Gilovich, 2003). It is likely that participants in this study will experience heightened sensitivity when performing in front of the group audience, particularly in the L2. The context and fear of negative evaluation is expected to exacerbate learners' self-focused attention, making them feel subtle physiological changes more pronounced. As a result, it is expected that this hypersensitivity will likely cause participants to overestimate their anxiety levels.

Although many anxiety cues are not often perceptible to external observers, such as faster HR, other behavioral symptoms of anxiety (i.e., restlessness) are highly visible (Gregersen et al., 2017). It is hypothesized that the conditions of this study will trigger enough anxiety so that symptoms of anxiety are highly perceptible. As a result, it is predicted to find a significant relationship between self-assessment of anxiety and others' perceptions.

Within RQ2, RQ2.2 aims to explore the most salient non-verbal cues of L2 speaking anxiety. Examples of recurring non-verbal indicators encompass, but are not limited to, avoidance of making eye contact, limited facial expressions, restlessness and the adoption of closed or rigid postures (Gregersen, 2005; Gregersen, 2007b; Gregersen et al., 2017). Based on previous research, similar non-verbal indicators are predicted to be perceived by the audience (in both the one-to-one and group contexts), particularly the avoidance of making eye contact and restlessness.

With respect to RQ3 regarding the relationship between subjective measures of anxiety (i.e., self-reported anxiety) and physiological measures of anxiety (i.e., HR and EDA), previous research has yielded mixed results regarding the association between these measurements. While many studies have found physiological measures and subjective measures to be unrelated (MacIntyre et al., 2010; Gallego et al., 2021; Schwerdtfeger, 2004; Shachter et al., 2022; Sosa-López, 2019), fewer studies have found a relationship between EDA and self-reports (Sevinç; 2018) and HR and self-reports (Behnke & Beatty, 1981; Gregersen et al., 2014). Although the author acknowledges that physiological measures and subjective measures provide different insights into anxiety, this study is predicted to generate enough variability in anxiety responses to find a relationship between both measures of anxiety. Furthermore, it is hypothesized that HR will show a stronger positive association with self-reports of anxiety compared to EDA. This expectation is based on the premise that HR is a more salient physiological response and easily perceptible to participants. In contrast, EDA may be less noticeable to individuals and less intuitive for individuals to link to anxiety.

Sub-questions RQ3.1 and RQ3.2 investigate potential differences in the relationship between subjective and physiological measures of anxiety across languages (L1 vs. L2) and anxiety levels (high vs. low). L2 presentations are expected to elicit heightened anxiety, which should be reflected in stronger associations between subjective and physiological measures. Similarly, individuals who report higher anxiety levels are anticipated to show a more pronounced alignment between these measures compared to those with low anxiety.

RQ4 explores the association between L2 learners' anxiety levels and L2 speech measures of fluency, complexity, accuracy, and comprehensibility and accentedness.

Additionally, it seeks to determine the extent to which L2 speaking anxiety accounts for variation in these speech measures while controlling for IDs in L2 proficiency and cognitive factors. Previous research has revealed detrimental effects of anxiety on comprehensibility gains (Nagle et al., 2022; Saito et al., 2018), accentedness (Suzukida & Saito, 2023; Zhou, 2024), accuracy (Hewitt & Stephenson, 2012), the overall flow of speech (speed fluency; Pérez Castillejo, 2019), the production of self-repairs (Zuniga & Simard, 2019, 2022), increased pausing time (Pérez Castillejo, 2019) and pause frequency (Aubrey, 2022; Bielak, 2022; Lindberg et al., 2021; Mora et al. 2023; Pérez Castillejo, 2019; Rood & De Jong, 2023; Simard et al., 2023; Szyszka & Lintunen, 2023). In this study, L2 speaking anxiety is expected to be negatively associated with L2 speech measures and explain some of the variation in these measures. This is expected to be particularly evident in the breakdown fluency measures within AS-units. When L2 learners perform their oral presentations and get anxious, it is predicted that anxiety will negatively affect the allocation of resources to the message formulation (either lexical, syntactic, or phonological encoding); hence, participants are expected to produce significantly more pauses within AS-Units, as found in Pérez Castillejo (2019), Mora et al. (2023) and Szyszka & Lintunen (2023).

Also, previous research has revealed that experiencing L2 speaking anxiety is not associated with pause duration (Mora et al., 2023; Szyszka & Lintunen, 2023), even though Pérez Castillejo (2019) found anxious learners made longer pauses at AS-Unit boundaries. However, as the researcher suggested, longer pauses at the end of AS-Units might result from insufficient planning time prior to the oral task. Based on the findings in Mora et al. (2023) and Szyszka and Lintunen (2023), it is anticipated that pause frequency will be much more related to L2 speaking anxiety than pause duration.

Regarding accentedness and comprehensibility, research is scarce, and results are mixed. However, anxiety has been found to negatively impact both accentedness and comprehensibility since it contributes to more accented speech (MacIntyre & Gardner, 1994b; Suzukida & Saito, 2023) and interferes with L2 comprehensibility (Saito et al., 2018). Nevertheless, Mora et al. (2023) did not find physiological measures (HR) or self-reported speaking anxiety levels to be related to accentedness or comprehensibility. In this study, it is predicted that L2 speaking anxiety will be negatively associated with accentedness and comprehensibility. Thus, experiencing L2 speaking anxiety is expected to be reflected in speech through breakdown fluency and consequently detected by native listeners when evaluating L2 speech overall.

Within RQ4, RQ4.2 aims to explore the contribution of objective and subjective measures of L2 speaking anxiety in explaining differences in L2 speaking performance once proficiency and cognitive factors are controlled for. It is hypothesized that L2 proficiency and cognitive factors will significantly contribute to L2 speech accuracy and fluency, as supported by previous findings from Bielak (2022), Mora et al. (2023) and Simard et al. (2023). However, after controlling for the effects of L2 proficiency and WM, it is expected that both objective and subjective measures of L2 speaking anxiety will account for a significant amount of variance, especially fluency measures (Bielak, 2022; Mora et al., 2023; Pérez Castillejo, 2019; Simard et al., 2023).

Concerning accentedness and comprehensibility, based on previous research, proficiency and L2 speaking anxiety are expected to account for variability in their respective scores. Mora et al. (2023) found that higher proficiency learners were found to speak more fluently and accurately and to be more comprehensible and less foreign accented than lower proficiency learners. However, such relationships were not found for speaking anxiety. On the other hand, Suzukida & Saito (2023) found that FLA accounted for accentedness scores but not comprehensibility. This study expects that the context and the task type will induce enough L2 speaking anxiety to make its effects noticeable on both measures.

Lastly, RQ5 aims to identify the sources of L2 speaking anxiety. Following Tóth's (2019) findings, it is anticipated that different sources of L2 speaking anxiety will be reported by the participants, allowing the identification of different categories. Some of the anticipated sources of L2 speaking anxiety are: being in the spotlight or standing in front of others, fear of negative evaluation and experiencing mental blocks, which will likely be connected to other anxiety sources such as forgetting words and making mistakes. Tóth's (2019) study distinguished four main causes of L2 speaking anxiety: the communication situation itself audience-related fears, self-related concerns, and talk-related concerns. Similar categories are expected to be present in this study, especially the communication situation and audience-related fears. However, it is hypothesized that sources of L2 speaking anxiety will differ in terms of frequency in the self-reports depending on the group that learners belong to, either high or low anxiety. High-anxiety learners are expected to be more concerned with being in the spotlight (Young, 1990, 1991) and to comment much more frequently on issues such as mental blocks than low-anxiety learners.

CHAPTER 4: Methodology

The present chapter provides an overview of the experimental design of the study (Section 4.1), describes the demographic and linguistic characteristics of the participants (Section 4.2) and the raters involved in this study (Section 4.3.1 for audience members and Section 4.3.2 for native listeners), describes the instruments for the data collection and the measures derived (Section 4.4 and 4.5 for control tests), explains the overall study procedure (Section 4.6), and presents the specific L2 speech measures computed from the oral presentations (Section 4.7).

4.1. Research Design

One of the main questions addressed in the current study is whether L2 speaking anxiety affects L2 speaking fluency, comprehensibility and accentedness. Therefore, in order to answer the research questions and test the hypotheses formulated in Section 3.3, we opted to elicit speech from the participants through oral presentations.

Prior to initiating data collection, a pilot study was run in which the effects of L2 speaking anxiety in two different task types (descriptive vs. argumentative) were compared (Sosa-López & Mora, 2022). This pilot study helped me re-design the descriptive task to encourage participants to provide longer and more detailed responses. In the pilot, twenty-one (N = 21) 3rd-year university learners performed a descriptive and argumentative task in front of their peers, and results showed that the argumentative task generated significantly higher levels of distress.

While the pilot study did not use mixed methods, this dissertation does. The use of a mixed methods framework in research on FLA is scarce, and as noted by Gkonou (2014), there is limited attention given to the quantitative data of FLA. Instead, the emphasis often lies on the follow-up interviews. Nonetheless, in the present study, both methods are given equal importance, and the quantitative and qualitative findings will be thoroughly presented.

Moreover, for the present study, it was important that students performed the speaking tasks not only in their L2 (English), but also in their L1 (i.e., Spanish/Catalan) to obtain baseline data on speaking anxiety and fluency measures (see Figure 13). These mixed methods within-subject design had fluency, accentedness and comprehensibility as dependent variables, and the different communicative contexts (one-to-one vs. audience) and tasks (descriptive vs. argumentative) as the independent variables.

Speaking anxiety was also considered a dependent variable with respect to the communicative contexts and an independent variable in relation to the L2 speech output measures. The study included two different speaking tasks for each communicative context (one-to-one oral and before an audience of seven people) as outlined in Table 5 below:

Table 5: Simplified summary of the task

		Context	
		one-to-one oral presentation	oral presentation before an audience
Task type	Descriptive task	task 1	task 3
	Argumentative task	task 2	task 4

Task type and context order were counterbalanced, with participants assigned to one of four groups (1A, 1B, 2A, 2B), each differing in the initial task type or context they began with. This approach was chosen based on the assumption that the first task might elicit higher levels of anxiety compared to the subsequent task; thus, counterbalancing was implemented to equalize the level of nervousness experienced across participants (see Table 5 and 6). A detailed table of the design and procedure of the data collection can be found in Appendix A1. Moreover, the study protocol adhered to the good practices of data collection, data anonymization, data processing, and data storage of the Institutional Review Board of the University of Barcelona (IRB0003099) and in accordance with the Declaration of Helsinki (see Appendix A2).

Table 6: Summary of the design including counterbalance

GROUP	SESSION	CONTEXT	TASK TYPE	ORDER
1A	1	One-to-one	Descriptive (task 1)	descriptive > argumentative
1B			Argumentative (task 2)	argumentative > descriptive
1A	2	Group	Descriptive (task 3)	descriptive > argumentative
1B			Argumentative (task 4)	argumentative > descriptive

GROUP	SESSION	CONTEXT	TASK TYPE	ORDER
2A	1	Group	Descriptive (task 3)	descriptive > argumentative
2B			Argumentative (task 4)	argumentative > descriptive
2A	2	One-to-one	Descriptive (task 1)	descriptive > argumentative
2B			Argumentative (task 2)	argumentative > descriptive

Further details of the tasks will be provided in the section *Instruments (4.3)*. In fact, the tasks were piloted twice. As aforementioned, the first time some of the tasks were done by 21 3rd -year English students from the University of Barcelona who were enrolled in the subject "Spoken English". In this first pilot it was seen that the descriptive task did not generate more than 1- to 2-minute-long speeches. Afterwards, the tasks were changed and the eight tasks were piloted again with seven of the author's friends and colleagues.

It was thought that multiple data-elicitation measures of learners' thoughts would help understand L2 speaking anxiety. For this reason, several tasks were included in the design that prompted participants to reflect on their performance and feelings (i.e., anxiety). First of all, there was a post-task questionnaire with 9-point Likert scale questions on anxiety (see *Section 4.4.3.3*) followed by two tasks on self-reflection: the idiodynamic task and a stimulated recall (see *Section 4.4.2.3* and *4.4.2.4* for further details). Contrary to the post-task questionnaire, which is an overall rating, the idiodynamic task and stimulated recall capture the dynamicity of L2 speaking anxiety, as explained in the literature (Gass & Mackey, 2013; MacIntyre, 2012), and is expected to give an insightful view of the "ups and downs" as well as the sources of L2 speaking anxiety.

With the objective of controlling IDs that may interfere in the observation of the variables, L2 proficiency, vocabulary size and working memory were controlled for (see *Section 4.4* for further details). To avoid long data collection sessions, it was decided not to include the stimulated recall task when participants performed the oral tasks in their L1. Instead, that time was allocated to control tests. L2 proficiency was measured with an EIT (Ortega et al., 2002; Wu et al., 2021), vocabulary size with the Yes/No test from *Lognostics* (Meara & Miralpeix, 2015a, 2017), WM measured with a digit span task

(DST) (Conway et al., 2005; Lee et al., 2019), and general attention with a Trail making test (TMT) (Isaacs & Trofimovich, 2011).

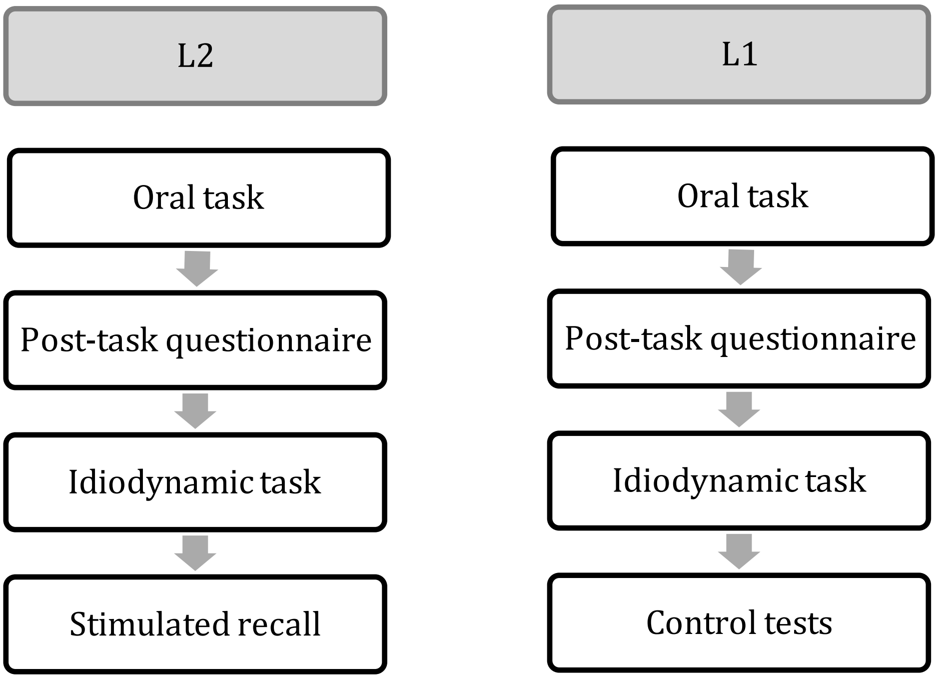


Figure 13: Data collection procedure.

4.2. EFL learners

Data collection took place in two years' time between 2021-2022 and 2022-2023. During the first year 68 participants were recruited from the subjects *Llengua Anglesa II* and *Fonètica i Fonologia I* from the English studies degree at the UB. The remaining 18 participants were recruited only from *Llengua Anglesa II*. In both years, the participants had similar backgrounds; thus, they were a homogeneous group.

Eighty-six (N = 86) L2 learners of English (74 females, 12 males; see Table 7 for a brief summary of the demographics and Appendix B for detailed information) took part in this study for course credit. Participants were young adult university learners (M_{age} = 20.16, SD= 4.09, *Range* = 18-47) who had started learning English as a foreign language mainly in a formal setting at the M_{age} of 6.57 (SD = 3.24). They were mostly Spanish-Catalan bilinguals (N= 75; 87%); however, 13% of the participants had other mother tongues such as Arabic (N = 2), French (N = 2), Italian (N = 2), Russian (N = 1), Dutch (N = 2), Chinese (N = 1) and Armenian (N = 1). Although their AOL varied (*Range* = 2-22), all of them had English as their L2. In addition, those who had different mother tongues indicated in the language background questionnaire (LBQ) that they spoke Spanish daily (*Range*: 5-60%).

In terms of the learners' mother tongue, as aforementioned, the LBQ reported a dominance of L1 Spanish-Catalan bilinguals (87%); 50% were Catalan-dominant and 20.93% were Spanish-dominant. The remaining 29.07% stated that they used Catalan and Spanish indifferently. However, Catalan-Spanish bilinguals differed in the degree of L1 daily use. Learners reported using Spanish approximately 32.84% of the time, Catalan 44.53% of the time and English around 20.21% of the time.

As for their FL experience, most of the participants (N = 56, 65.11%) stated they had received English input from school (*Range* = 3-4 hours/week) and around 2-4 hours/week of extra-curricular lessons from private language academies (M = 2.69, SD = 0.96). The remaining 34.89% only received English input from school. While 31.39% (N = 27) of the participants did not hold any certificate of English, 84.88% (N = 73) had obtained a certificate of English. While most of them had taken the Cambridge First Certificate of English (FCE, N = 29, 33.72%), some participants had taken both the FCE and the Certificate in Advanced English (CAE, N = 11, 12.79%). Very few learners had obtained solely CAE (N = 8, 9.30%), TOEFL (N= 2, 2.32%), IELTS (N = 3, 3.48%), Certificate of Proficiency in English (CPE, N = 1, 1.16%) and B1-level certificates (N= 5, 5.81%).

As for their FL use, the learners reported being exposed quite often to L2 spoken input (M = 4.33, SD = 0.73) as well as L2 written input (M = 3.24, SD= 1.06) by doing activities such as watching movies or videos or reading books in English. Similarly, they produced frequently spoken (M = 3.37, SD = 0.95) and written (M = 3.18, SD = 1.02) output (e.g., writing emails).

Moreover, participants considered themselves to have more of a General American accent (N = 56, 65.11%) than British English (N = 24, 27.90%) hen speaking. Only five participants couldn't decide on their accent (N = 5, 5.81%). Interestingly, participants mentioned being slightly more exposed to General American (53.48%) than Received Pronunciation (29.06%) accents of English; the remaining 17.44% considered being equally exposed to both accents. Regarding their proficiency level, participants self-rated their overall L2 proficiency in the different skills (i.e., reading, listening, writing, speaking) on a 9-point Likert scale. They judged themselves to be upper-intermediate to advanced level in the skills of reading, listening, writing, speaking, and considered their pronunciation to be quite accurate. Results from their self-reported proficiency are consistent with the scores obtained on the EIT (*M* = 97.42, *SD*= 13.21, *Range* = 0-120) and with their vocabulary size (*M* = 6546.92, *SD*= 1211.43; test from Yes-No vocabulary test by Meara and Miralpeix, 2015a, 2017). Finally, most of the participants (*N* = 54, 62.79%) had not lived in an English-speaking country abroad for more than 3 weeks, whereas 16.27% (*N* = 14) had (*M* = 19.40, *SD*= 26.11, *Range*: 3-96 weeks).

Last but not least, two learners reported having either a speech or hearing impairment: one of them reported being half deaf in one ear while the other stated having problems pronouncing the /r/ sound as a child. Both learners stated that nowadays this difficulty only happens from time to time. The remaining 97% of the participants did not report any impairment.

Participants were recruited in the classroom in the following way: during the presentation of the course syllabus, students were informed that they could volunteer to do a research task which would grant them an additional 5% of course credit. Then, they were informed in class and via Moodle that they could voluntarily sign up in a Google Drive Excel sheet. Afterwards, they were contacted by email to confirm their scheduled time and day.

Table 7: Participants' demographics summary

Variables	*M*	*SD*	95% CI	
Age	20.16	4.09	19.28	21.04
AOL English (years)	6.57	3.24	5.88	7.26
Spoken L2 Input*[8]	4.33	0.73	4.17	4.48
Spoken L2 Output*	3.37	0.95	3.17	3.57
Written L2 Input*	3.24	1.06	3.01	3.46
Written L2 Output*	3.18	1.02	2.96	3.40
Self-estimated Overall Proficiency**[9]	6.86	1.19	6.61	7.12
Elicited Imitation (EI, 0-120)	97.42	13.21	96.43	98.41
Vocabulary size (0-10.000 words)	6546.92	1211.43	6287.19	6806.65

[8] * 1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily
[9] **(Reading, Listening, Speaking, Writing): Likert scale 1=VERY POOR and 9=NATIVE-LIKE

Forward digit-span (TE_ML)	5.86	1.24	5.59	6.12
Backward digit-span (TE_ML)	5.53	1.56	5.20	5.86

Note: For a more complete Table check Appendix B.

4.3. Raters

4.3.1. Audience members

Seven people (*N* = 7) were recruited as an audience for one of the study's contexts (i.e., group audience). From this group of people, one audience member was also present in the one-to-one sessions. This person was a PhD colleague who from now one will be designated as "the helper". Given that the data collection took place in two different academic years, the author tried to contact the same audience members from the first round of data collection for the second round the following year. However, they were unavailable to participate again, except for the helper mentioned above. Thus, people whose profiles were similar to the first group of audience members were searched for and chosen based on their availability. All the information on the audience members' demographics was extracted from a Google Form questionnaire (see Appendix C).

In the first year of data collection, the seven audience members (5 females, 2 males; see Table 8 for a summary and Appendix D1 for detailed information) were young adults (M_{age} = 20.71, *SD* = 4.78, *Range* = 20-32). They were mostly Spanish-Catalan bilinguals (*N* = 5, 71.42%); however, there was a Spanish-English bilingual (*N*=1, 14.28%) and the helper who was Hindu-English bilingual (*N* = 1, 14.28%) (see Table D2.1 in Appendix D1 for detailed information). The helper reported being familiar with Spanish-accented English as well as Spanish and Catalan. She reported having lived in Spain for 9 years and having received instruction of Spanish for 5 years. Although she had not received any formal instruction of Catalan, she self-rated herself as having an intermediate level. Except the English bilinguals, the rest of the raters had started learning English as a FL mainly in a formal setting (M_{age} of 5.86, *SD* = 4.48), and four out of five of them had taken extracurricular English classes in a private language academy (*Range*: 1.5- 3h/ week) and also hold a certificate of English level (FCE, *N* = 4, 57.14%). It is worth adding that one of the bilingual raters had the Certificate in Advanced English (CAE) (*N* = 1, 14.28%). As for their EFL use, the raters self-rated their overall proficiency on a 9-point Likert scale, and they judged themselves to be at an advanced level (M = 7.74, SD = 1.26) (see Table D1.3 in Appendix D1 for a complete Table).

Table 8: Summary of audience members' demographics in 2021-2022

Variables	*M*	*SD*	95% CI	
Age	26.71	4.78	22.29	31.14
AOL English (years)	5.86	4.48	1.71	10.01
Spoken L2 Input*[10]	4.19	0.50	2.77	4.78
Spoken L2 Output*	2.57	0.93	1.64	3.13
Written L2 Input*	3.71	1.29	2.38	4.50
Written L2 Output*	3.00	1.53	1.75	3.92
Self-estimated Overall Proficiency**[11]	7.74	1.26	5.16	8.88

In the second year of data collection, the seven audience members (4 females, 3 males; see Table 9 for a summary of the demographics and Appendix D.2 for detailed information) were also young adults (M_{age} = 33.71, SD = 9.93, *Range* = 25-43). In this group, three were Spanish-Catalan bilinguals (N = 3, 42.85%), one was a native-speaker of Polish (N = 1, 14.28%) and two native speakers of English (N = 2, 28.57%) from the United States. The two NSs of English completed a questionnaire (see Table D2.2 in Appendix D2for detailed information) in order to assess their level of Spanish and Catalan due to the fact that most of the participants were expected to their L1 oral presentation in Spanish or Catalan. Except the NSs of English and the Hindi-English bilingual, the raters had taken extracurricular English classes in private language academies (*Range*= 1-4h/ week), and three of them also hold a certificate of English (FCE and IELTS). As for living abroad, disregarding the NSs who spent their childhood and teen ages in the USA, three of the audience members had lived in an English-speaking country for more than 2 weeks. Likewise, audience members additionally self-rated their overall English proficiency on a 9-point Likert scale and judged themselves to be at an advanced level (M = 8.26, SD = 0.94).

[10] * 1 = never; 2 = a few times a year; 3 = monthly; 4 = weekly; 5 = daily
[11] **(Reading, Listening, Speaking, Writing): Likert scale 1 = VERY POOR and 9 = NATIVE-LIKE

Table 9: Summary of the audience members' demographics in 2022-2023

Variables	*M*	*SD*	95% CI	
Age	33.71	9.93	24.53	42.90
AOL English (years)	3.57	3.69	0.16	6.98
Spoken L2 Input*[12]	4.67	0.47	4.23	5.10
Spoken L2 Output*	4.21	0.95	3.33	5.09
Written L2 Input*	4.29	0.70	3.64	4.93
Written L2 Output*	4.43	1.13	3.38	5.48
Self-estimated Overall Proficiency**[13]	8.26	0.94	7.38	9.13

4.3.2. Native raters

Seven expert native raters (*N* = 7; 2 males, 5 females; M_{age} = 31.86, *SD* = 3.80, *Range* = 28-39) were recruited to assess the speech samples of L2 learners for both comprehensibility and accentedness. Appendix E shows the questionnaire used to collect their demographic information. They spoke either British English (*N* = 3, 42.85%) or American English (*N* = 4, 57.14%) and had had a predominantly monolingual upbringing but had lived in a Spanish-speaking country for the last 5.43 years (*SD* = 2.29, *Range* = 2-8 years). Most of these native speaker raters were experienced EFL teachers (*N* = 6, 85.71%), except one who was a content writer. While all of them reported learning Spanish (*M* = 9.29, *SD* = 10.51), only 4 of them declared having learnt Catalan for 2-3 years (*M* = 2.87, *SD* = 0.62). As for their Spanish and Catalan proficiency, the raters self-rated their overall proficiency on a 5-point Likert scale, and they judged themselves to be more proficient in Spanish (*M* = 3.76, *SD* = 1.17) than in Catalan (*M* = 2.19, *SD* = 0.99). Finally, on the 9-point Likert scale, the raters reported being familiar with foreign-accented English (*M* = 8.28, *SD* = 1.11) as well as British (*M* = 7.28, *SD* = 1.97) and American accents (*M* = 8.85, *SD* = 0.37) of English (see Table 10 for a summary, and Appendix F for detailed information).

[12] * 1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily
[13] **(Reading, Listening, Speaking, Writing): Likert scale 1=VERY POOR and 9=NATIVE-LIKE

Table 10: Native raters' demographics summary

Variables	*M*	*SD*	95% CI	
Age	31.86	3.80	28.34	35.38
SCORE Spanish	3.76	1.16	2.68	4.84
SCORE Catalan	2.19	0.99	1.26	3.11
Years learning Spain	9.29	10.51	-.44	19.01
Years learning Catalan	1.64	1.59	0.16	3.12
Years living in Spanish	5.43	2.29	3.30	7.55
Familiarity with Spanish -accented English	8.29	1.11	7.26	9.31
Knowledge of Spanish	5.71	2.28	3.60	7.83
Everyday use of Spanish	3.57	2.37	1.38	5.76
Everyday exposure to Spanish-accented English	6.43	2.37	4.24	8.62

4.4. Instruments and measures

In this section, the instruments used in this study and the measures derived are described. First, the speaking tasks are presented. Second, the L2 speaking anxiety instruments are described, including the Empatica E4 wristband, the idiodynamic task and the stimulated recall. Third, the questionnaires are explained in detail: the language background questionnaire, the L2 speaking anxiety post-task questionnaire, and the questionnaire completed by the audience ("anxiety raters"). Finally, the control tests are described: the EIT and yes/no vocabulary size test for L2 proficiency, the digit span test for WM, and the trail making test for attention.

4.4.1. Speaking tasks

To elicit learners' oral productions in both the L2 and the L1, eight speech elicitation tasks were used (see Figure 14 and Appendix G for further details). They comprised an equal number of descriptive or argumentative tasks, with a total of four tasks per language (four in the L2 and four in the L1) to mitigate potential task effects. The ecological validity of these tasks lies in the essential and frequent nature of oral narratives in real-life situations (e.g., describing an event) and tasks used in EFL school or university classes (e.g., discussing the pros and cons between different options). Furthermore, they are widely employed in foreign language classrooms for both learning and assessment purposes. Consequently, such tasks are familiar to L2 learners. Two speaking tasks were adapted from other studies (De Jong et al., 2012, 2013), and the remaining tasks were created by the author, ensuring that they followed the same design as the adapted ones. In the argumentative tasks, participants were presented with three pictures and required to provide an explanation of the pros and cons related to the contexts presented in the pictures. To develop the argumentative tasks, the author referred to the *Ready for Advanced* coursebook (Norris & French, 2014) for topic ideas which were familiar to students. This book is an EFL coursebook typically used at the university where participants studied.

Figure 14 presents a summary of the tasks included in the present study. One of the argumentative tasks was selected from De Jong et al. (2012, 2013), accessible through the IRIS repository (https://www.iris-database.org/iris/app/home/index). In this task, participants were required to compare three means of transport (i.e., private transport like cars, public transport, and bicycles) in terms of sustainability and comfort for moving around a city. In another argumentative task, participants were presented with three options designed to potentially stop global warming. Their task was to engage in a discussion about which option would have a more substantial impact and would be more likely to be embraced by society. The argumentative tasks in the L1 centered on the comparison of three learning contexts (i.e., online, face-to-face, autonomous learning contexts) based on the student's role, their popularity, and flexibility. Finally, participants were required to compare living contexts (i.e., living in a city, in the countryside and an island) in terms of mobility, working or studying options and life quality.

In the L2 descriptive tasks, one was adapted from Crowther et al. (2015) and participants had to explain the toughest challenge they had experienced and what they learned from it. The other task in the L2 involved describing a memorable childhood experience. For the L1 tasks, one task prompted participants to explain a moment in which they had experienced frustration and how it was overcome, and the other required students to describe the best trip they had ever taken.

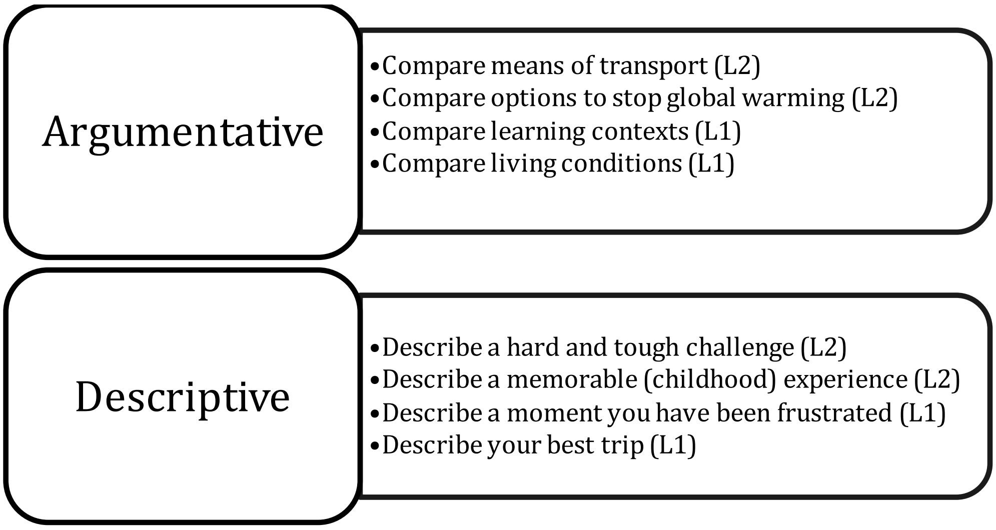


Figure 14: Summary of the speaking tasks

All the tasks required semi-formal language rather than informal speech but varied in content (personal vs. academic) as well as discourse type (descriptive vs. argumentative). Differences were anticipated across tasks in terms of speaking fluency, linguistic accuracy and complexity, and the anxiety levels they would elicit. The argumentative tasks were deemed to be slightly more difficult than the descriptive task due to the linguistic resources learners had to recruit in order to justify their arguments. Moreover, tasks overall were considered appropriate since they provided the participants with the opportunity to generate sufficiently long samples of uninterrupted semi-unprepared speech (3-5 minutes approximately), from which measures of fluency, accentedness and comprehensibility could be obtained.

The tasks mentioned above were presented on plasticized paper, and participants had 1 minute and a half to review the tasks and plan their productions, but they were not allowed to write down their ideas. Each task included 4–5 speaking prompts to generate ideas for their presentations; however, students were not confined to speaking solely about those prompts. While some support (either visual or written) is allowed in L2 classrooms during presentations, this research aimed to elicit more spontaneous speech, which would not have been possible if participants had had more time to prepare or to have notes with them. Furthermore, the presentations were recorded by the helper rather than the author. This decision was influenced by the participants' familiarity with the author who was their teacher in *Llengua Anglesa II*. The helper was in charge of delivering the tasks and recording. Additionally, she was instructed to intervene as minimally as possible to avoid influencing the participants' productions. However, in cases where participants did not speak at all, she could help by posing a question.

4.4.2. Instruments for L2 speaking anxiety

In order to comprehensively assess L2 speaking anxiety, my objective was to gather data through a combination of subjective and objective measures. To achieve this, various instruments were employed (see Table 14 in *Section 4.4.4* for a summary of L2 speaking anxiety measures derived from the different instruments).

4.4.2.1. Empatica E4 Wristband and physiological measures

The Empatica E4 wristband (https://www.empatica.com/en-eu/research/e4/) served as a crucial instrument for the objective measurement of participants' heart rate and electrodermal activity during the oral presentations. Commonly used in research studies related to physiological responses and health monitoring, the Empatica E4 is a non-invasive wireless wristband sensor which is strategically placed on the participants' non-dominant wrist (Figure 15) following the sensor manual guidelines to minimize motion artifacts. For instance, a right-handed individual would wear the wristband on their left wrist. Although this instrument has not been widely applied in a classroom setting yet, it has received strong endorsement for its use (Lee et al., 2020; Rivers, 2022; Schwerdtfeger, 2004) and has been piloted in various studies (Lee & Kleinsmith, 2019; Lee, 2020; Lee et al., 2020a, 2020b). Other investigations have employed similar devices such as Biopac (Sevinç, 2018) or Neulog (Mora et al., 2023). These studies collectively demonstrate the potential of utilizing physiological measures to discern students' emotional states, particularly in gauging the intensity of anxiety during oral presentations in classroom settings.

The sensor provides wireless connectivity for data transmission and real-time monitoring and is equipped with four sensors, including two electrodes for EDA, a 3-axis accelerometer, a skin temperature sensor, and PPG for BVP, from which HR and the IBI are derived. The IBI raw data is processed from the PPG and BVP signal and uses an algorithm designed to eliminate erroneous peaks resulting from noise in the BVP signal (Empatica support page on IBI, n.d.). For this reason, data obtained is highly reliable. For the research scope of this thesis, a detailed exploration on HR, HRV and EDA was conducted. The remaining data, including accelerometer and temperature sensor readings, proved valuable for controlling factors such as movement and body temperature.

Figure 15: Emaptica E4 technical specifications and how to wear it (E4 manual, n.d, p. 22[14])

Furthermore, following the user manual, the device was positioned with the case on the top of the wrist and the EDA electrodes facing the inner part of the wrist. Moreover, it was crucial to secure the sensor snugly to prevent any movement. This precaution was emphasized because inadequate tightness or improper placement could lead to potential data loss or inaccuracies as noted in studies by Milstein & Gordon (2020) and the Empatica EDA support page (n.d.).

It is noteworthy that various studies (Borrego et al., 2019; Milstein & Gordon, 2020; Menghini et al., 2019; McCarthy et al., 2016; Schuurmans et al., 2020) have scrutinized the E4 sensor's reliability and validity when comparing it to other high-cost equipment. The results suggest that this device is straightforward and a promising tool

[14] Also available at https://support.empatica.com/hc/en-us/articles/202581999-E4-wristband-technical-specifications

for research, especially regarding HR data. However, it is important to mention that EDA data was not found to correlate with other standard devices in Milstein & Gordon (2020) and Menghini et al. (2019). Menghini et al. (2019) suggested that this discrepancy may be attributed to variations in the sweat gland density sensor, as the E4 electrodes are placed on the wrist while the fingers and palmar sites were not measured, although they are known to be influenced by psychophysiological processes. Moreover, Menghini et al. (2019) reported low percentages of detected IBIs and frequent time gaps within the recordings that were found, potentially introducing a biased estimation of the IBI variability.

All the researchers unanimously caution that significant amount of movement, such as gesturing, and fluctuations in temperature can impact the accuracy of the recordings and subsequently, the reliability of the data (McCarthy et al., 2016; Schuurmans et al., 2020) (see Figure 16). Finally, another remark by the Empatica EDA support page (n.d.) notes that the EDA reading can be influenced not only by the material and the position of the electrodes, but also by the environmental conditions during the recording (temperature and humidity of the room) and the length of the recording, meaning that the coupling of the electrodes with the skin could take some time. Bearing all that in mind, the data collection for this study took place in ventilated rooms and from April to May, so the room temperature would be stable. It is important to note that studies such as Shachter et al. (2022) collected HR data outside the classroom context as baseline data. However, since it was not feasible to collect participants' resting HR data outside the classroom in this study, the planning time is considered the baseline data. Finally, concerning the measures derived from the Empatica E4 wristband, Section 4.6.1.1 will provide a detailed account.

Figure 16: Part of Empatica E4 sensor's data (EDA, BVP and accelerometer respectively) from a trial
Note: Before the red line, the subject was moving continuously.

With respect to the measures derived from the Empatica E4 wristband, individual physiological indices of anxiety were obtained by computing the descriptors[15] of HR and EDA for each speaking performance based on the raw output data files (i.e., one per oral task). By averaging the physiological response of each participant across the duration of their speaking performance (HRm, HRVavnn, and EDAm), which typically lasted between 180-240 seconds, and by calculating an individual standard deviation measure (HRsd and EDAsd), a representative of the magnitude and frequency of anxiety peaks was obtained. This methodology aligns with previous studies, such as those by Croft et al. (2004) and Mora et al. (2023).

The HR measure was obtained in beats per minute (bpm; min.= 60, max.= 200 [16]) with a sampling rate of 1 Hz (i.e., every second). This raw data was computed by the E4 algorithm which derived from the IBI (HR from IBI [bpm] = 60/IBI) and BVP data collected at 64 Hz. On the other hand, the EDA was obtained in microSiemens (mS; min.= 0, max.= 100) with a sampling rate of 4 Hz (every 0.25 seconds). Despite the EDA extensive range, according to Garbarino et al. (2014), in laboratory studies EDA does not usually exceed the bottom quartile of this range, and it is only in real-world medical events that the upper half of the range has been observed. Thus, normal human EDA typically ranges from 1 to 20 mS. Furthermore, according to Empatica's support page (n.d.), there are no algorithms available to compute the tonic or the phasic features of EDA separately[17], and to measure a person's truly lowest tonic SCL, long-term measurements would be needed to look for the lowest smooth period of skin conductance where there was no physical movement.

Raw data from the planning time was discarded to eliminate potential initial noise. Although Empatica E4's algorithm cancels noise and removes "wrong beats when the difference between two successive beats is not in line with other intervals in the time-series" (Milstein & Gordon, 2020, p. 4), the first seconds of the raw data can be misleading due to "variation in the contact between the skin and the recording electrode caused by pressure, excessive movement, or adjustment of the device" (Taylor et al., 2015, p. 1934).

[15] *M*, *SD*, minimum, maximum and distance, computed as the maximum score minus the minimum score.
[16] The HR range may vary slightly since it is related to age.
[17] https://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment

Further steps were taken to check any source of error variance through manual inspections using EDA Explorer[18], considering the accelerometer data, which helps to identify motion affecting the data and visual inspection to detect unrealistic values (e.g., heartbeat measurements below 40 beats and over 220 per minute). Kocielnik et al. (2013) suggest taking additional precautions by discarding both the first and last minute of data, as the sweat level in the skin needs to adjust to the new situation when the device is put on, and movement occurs in the wrist when the device is removed.

Concerning HRV, due to the metrics' technical complexity and the device used in this study, primarily time-domain indices were computed for each oral task and participant. The most common measures of time-domain HRV are the average and standard variation of the IBI (Shaffer & Ginsberg, 2017), either including only normal beats (SDNN) by removing the abnormal ones or including both normal and abnormal beats (SDRR), and the RMSSD (see Table 11). According to the Empatica manual on IBI expected signal[19] (n.d.), "incorrect" beats caused for example due to strong movement (and more than 30% of the time) are not included in the IBI data; thus, when values were unavailable, the data had to be interpolated to connect the last available point with the first point after the lapse in data. Steps were taken to check any source of error variance through manual inspections using the NN50 and Pnn50 metrics.

Table 11: HRV time-domain metrics (following Shaffer & Ginsberg, 2017)

Metric	Description
AVNN	Average of NN intervals
AVRR	Average of R-R intervals
SDNN	Standard deviation of N-N intervals
SDRR	Standard deviation of R-R intervals
pNN50	Percentage of successive R-R intervals that differ by more than 50 milliseconds (relative to the total time)
NN50	Number of N-N intervals which differ by more than 50 milliseconds from the previous interval
RMSSD	Root mean square of successive R-R interval differences

4.4.2.2. Idiodynamic task

The idiodynamic task served as one vital instrument for the subjective measurement of L2 speaking anxiety after task performance. The Anion Variable tester is a free software

[18] Available at https://eda-explorer.media.mit.edu/
[19] https://support.empatica.com/hc/en-us/articles/360030058011-E4-data-IBI-expected-signal

developed by Dr. Peter MacIntyre (https://petermacintyre.weebly.com/idiodynamic-software.html) in which the participants watch or just listen to their productions and self-assess themselves in 1-second increments by clicking to raise or lower on a continuous scale. Although the software allows a scale from -5 to +5 regarding the level of the rated construct, in this study, participants were encouraged to use the scale from 0 to +5, as any negative value would indicate not being nervous. The fluctuations in the affective state, specifically speaking anxiety, were recorded when the participant clicks the button labelled "increase" or "decrease" (see Figure 17). Each successive click of the mouse corresponded to an additional increase or decrease in the rating, appearing as an upward or downward block on a color bar graph. In the absence of any rating activity, the software records a built-in auto-zero function, returning the rating to the baseline at the rate of one point (click) per second (MacIntyre, 2012). Once finished, a visual output (i.e., a graph) and a CSV file with the data were generated and saved (see Figure 18).

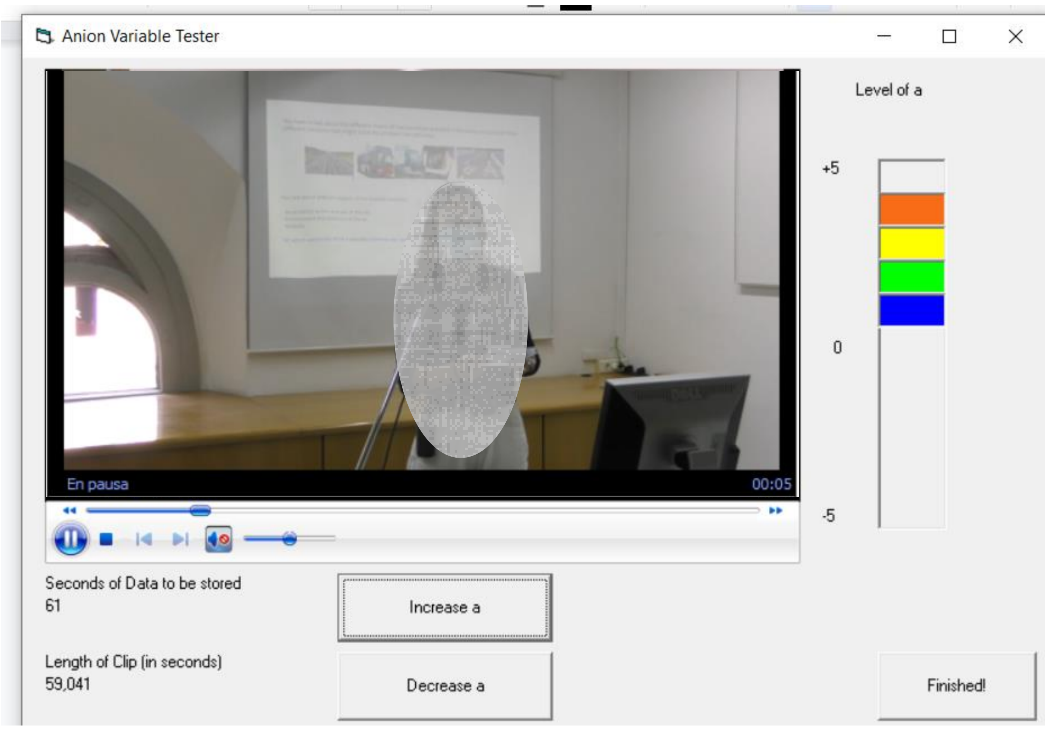


Figure 17: Anion Variable Tester.

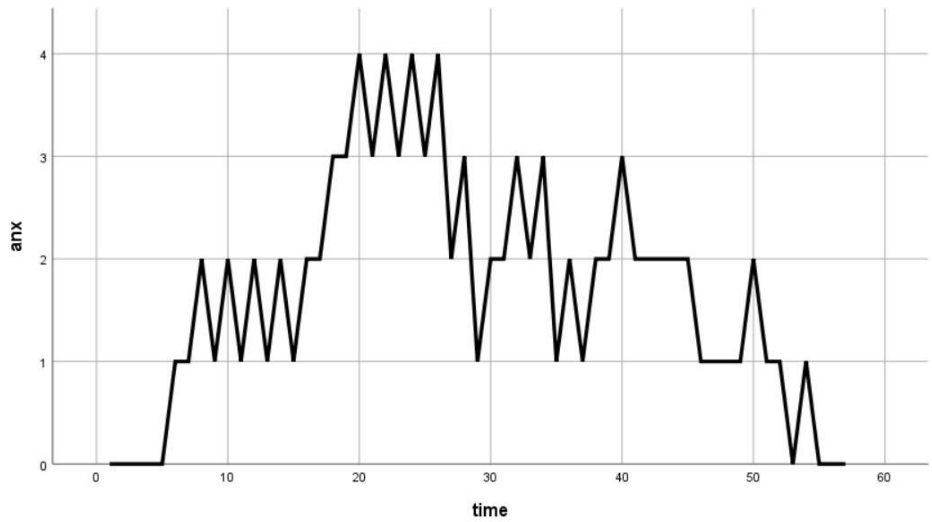


Figure 18: Post self-assessment idiodynamic anxiety level graph
Note: The peaks and falls correspond to the speaking anxiety levels (y axis) rated by the participant throughout the duration of the oral performance (x axis, time in seconds).

Concerning the measures derived from the idiodynamic task, following Gregersen et al. (2014), general descriptors and the ratio peaks per minute were calculated. The latter was computed by dividing the total number of peaks in each task by the task's duration in minutes.

4.4.2.3. Stimulated recalls

In L2 performances, participants conducted a stimulated recall after the idiodynamic task. Stimulated recalls are another valuable instrument for obtaining a subjective measurement of L2 speaking anxiety, aiming to extract participants' thoughts on why they experienced this emotion. Participants had to watch their videos again and were instructed to stop the video at any moment they recalled experiencing anxiety or nervousness during their performances (not their impressions while watching) and also at any point they encountered a communication problem (e.g., difficulty finding a word or realizing making a mistake).

According to Gass & Mackey (2013), when performing a stimulated recall, there are validity concerns, such as whether participants are recalling memories or responding directly to the stimulus, and whether introduction of the stimulus alters the original memory. For this reason, the delivered instructions strongly emphasized that participants had to explain what they recalled from the experience rather than respond to the stimulus. After stopping the video, they specified the minute and second of the recording and provided an explanation using a Tascam DR-05x recorder.

Following Gass and Mackey's (2013) recommendations to avoid critical issues in stimulated recalls, decisions were taken regarding the focus (specific vs. non-specific), structure (high vs. low), support (strong support vs. unsupported) and who initiates the task (learner-initiated vs. researcher-initiated). In this case the focus of the task is specific since it focuses solely on L2 speaking anxiety, with low structure allowing the learners to specify what, when and how much they verbalize. In addition, to no constraints with time or over questions, with strong support (video) and learner initiated. Regarding the structure, participants were given clear instructions to avoid the likelihood of collecting data that is not fruitful for analysis.

Another critical issue considered in the stimulated recall was the time elapsed between the performance and the task, as the greater the delay, the greater the potential for memory decay. Gass and Mackey (2013) exemplify that stimulated recalls can be immediately consecutive, delayed or nonrecent; and depending on the time elapsed, the support provided should be stronger. For example, in a very delayed stimulated recall, besides the video, it would be encouraged to provide a transcript. Although in this study it was not immediately consecutive, participants watched themselves twice, first in the idiodynamic task and then for the stimulated recall. Thus, there wasn't much transition time from the L2 performance to setting up the stimulated recall.

Moving to the concern of participant training, Gass and Mackey (2013) advocate for minimal training. In other words, participants should be able to carry out the procedure without being informed of the experimental objectives or receiving any unnecessary information. Participants in this study were not subject to formal training. However, the researchers assisting the author were trained before data collection because instructions needed to be standardized since any change (i.e., word or tense) can affect the nature of the participant's recall (Gass & Mackey 2013). Consequently, clear instructions were made following Gass and Mackey's guidelines (p. 59-60, 153-155) and these were printed to ensure consistent delivery by both the assisting researchers and the author.

A final procedural consideration, as outlined by Gass & Mackey (2013), revolves around the language employed during the recall session. Conducting the stimulated recall in the L1 may be more comfortable for the participants and help prevent limitations related to their ability to express themselves in the L2 (e.g., their limited language resources can affect how they express certain details). Despite these considerations, participants were encouraged to carry out the task in English because they had enough language competence (given that they belong to the degree "English Studies"). However, recognizing the potential limitations in expressing certain nuances in English, participants were also granted the option to perform the task in or switch to Spanish or Catalan, should they feel more adept at articulating specific aspects in their L1.

Stimulated recall data can later be analyzed quantitatively and/or qualitatively. Quantifying what is typically regarded as subjective coding of verbal content can be challenging, as there is often no straightforward correspondence between predefined categories and the verbalized data without some inferencing or judgment. For this reason, Gass & Mackey (2013) recommend combining analytical techniques and following their suggestion: qualitative data (i.e., extracting quotes) is used to shed light on findings of quantitative analysis (i.e., quantifying categories in percentages). In this dissertation percentages were computed to represent the frequency of a specific answer reported by the participants, and quotes will be included to illustrate the coding.

The protocol of coding the items was done by the author of this study and an additional researcher (the helper) to avoid any rater subjectivity effect. It followed a bottom-up structure, and the items were processed by means of systematic content analysis (Dörnyei & Taguchi, 2009) in ATLAS.ti. This was achieved through a stepwise process that involved 5 phases:

1. *Data familiarization*: The first step was deciding how much or how many samples were going to be analyzed. Following Gass and Mackey (2013), it is common in L2 research to analyze a subset of an interview or a selection of participants as a representative sample. In this case, the whole stimulated recall audio was transcribed into a Word document, but a selection was made among participants. Instead of analyzing the stimulated recalls of the 86 participants, thirty people were selected among those who scored the highest (N = 10), around the mean (N = 10) and the lowest (N = 10) in the FLA questionnaire (as in Szyszka & Lintunen, 2023; Tóth, 2017).
2. *Systematic data coding and generating initial themes*: Each text unit was assigned a code depending on the theme it belonged to. Coding was necessary to categorize and quantify how many instances the different sources of anxiety appeared. Codes were assigned inductively.
3. *Developing and reviewing themes*: Themes were redefined when needed, and the transcriptions and the initial codes were given to the helper (following Gass & Mackey, 2013, p. 72). The helper assigned the codes to the text units and then it

was agreed if any discrepancy arose. Inter-code agreement coefficients were calculated to observe the extent to which the helper and the author agreed.

4. *Data analysis and writing of the report*: The themes obtained were analyzed in ATLAS.ti with the Code-Document analyzer. This led to the generation of quantitative data expressed in percentages. Following the ATLAS.ti manual[20] regarding data analysis, by using the Code-Document researchers can obtain frequency counts of the codes as well as a word count of the quotations coded by the selected code or code group within and across documents or group comparisons. Furthermore, some of the most relevant responses were quoted in the discussion for the purpose of exemplification. Finally, the author placed the codes into an Excel document to observe and group them into bigger categories such as internal or external sources of L2 speaking anxiety (e.g., experiencing a "mental block" would be considered internal, while "being recorded" would be external).

4.4.3. Questionnaires

In this section we provide a detailed account of the questionnaires used in this study to gather information about (1) the participants' demographics and linguistic characteristics and (2) their impressions of their feelings and performance safter each task. The online demographic and linguistic background questionnaire was created to obtain the EFL learners' personal and linguistic information as well as gain insights into their trait anxiety levels both in the L2 and in the L1. Lastly, in the online post-task questionnaire, the participants expressed their perceptions and evaluation regarding the oral tasks previously performed.

4.4.3.1. Demographic and linguistic background questionnaire

An extensive body of research exists on the impact of individual differences on the acquisition of FL (see Mora, 2022 for a review). Factors including age of acquisition, the extent of L2 experience and usage, as well as the quality and quantity of L2 input among others have been extensively studied. Consequently, a Language Background Questionnaire (LBQ) was created using Google Forms (see Appendix H) and distributed to the participants via email prior to data collection[21]. The participants' responses were

[20] Available at https://atlasti.com/manuals-and-documents

[21] The LBQ was adapted from the questionnaire available in the GRAL website: https://ubgral.wordpress.com/gral-materials/

collected, and the data was scored in a separate file on the Google server. Subsequently, the author downloaded the responses into an Excel spreadsheet for further analysis.

The LBQ consisted of four sections: The first section gathered information on general demographic information; the second section focused on the learners' English experience (i.e., their English onset age, whether they had English certificates, etc.); the third section, divided into two sub-sections, addressed questions related to FLA, first only about the L2 and after about the L1; and the final section explored the learners' language profile and their use of both their native language (L1) and the target language (L2).

In the initial part of the questionnaire, participants provided basic demographic data such as age and sex. The subsequent section inquired about L2 learning experience, covering questions such as the age of onset of L2 learning (Q1), weekly hours dedicated to English (Q2-2.1), possession of English certificates (Q3-4; i.e., FCE, CAE, etc.), and the percentage of exposure and production (Q4-6) of English. Participants also self-evaluated their command of English across the five competences (Q7), reported their time spent in English-speaking countries (Q8 – Q8.1), and indicated if they had been diagnosed with any speech/ hearing problem as well as clinical anxiety.

The third section focused on FLA (Q12-14) and L1 anxiety (Q15-16). Adapted statements from the FLCAS (Horwitz et al., 1986), the Second Language Speaking Anxiety Scale (Woodrow, 2006) and the Pronunciation Anxiety Questionnaire (Baran-Łucarz, 2016) were included. From Q12 to Q14, participants rated their distress levels on a 9-point Likert-scale (1 = extremely relaxed, 9 = extremely anxious) for various contexts such as "Talking in English to their teachers", "Talking in English to non-native speakers of English", "Talking in English to native speakers of English", "Doing an oral presentation in English", "Doing a task with a peer (e.g., a dialogue) in English" and "Taking part in a group discussion in class in English". Subsequently (Q13-Q14), participants used a 9-point Likert-scale to express the extent to which they agreed or disagreed (1= strongly disagree, definitely not true about you; 9= strongly agree, definitely true about you) with statements about speaking English such as: "I feel stressed knowing that others are listening to me", "I never feel quite sure of myself when I am speaking in English in class", "I tremble when I know that I'm going to be called on in class", and "I would not be nervous speaking English with native speakers" among others.

The subsection on L1 speaking anxiety mirrored the structure of Q12-14. Q15 required participants to rate their distress levels on a 9-point Likert-scale for situations such as "Talking to their teachers", "Doing an oral presentation", "Doing a task with a peer (e.g., a dialogue)", "Taking part in a group discussion in class" and "Talking to their classmates". Meanwhile, in Q16 participants had to rate their agreement or disagreement on a 9-point Likert-scale with statements adapted from McCroskey's Communication Apprehension (McCroskey, 1970).

The fourth section aimed at obtaining detailed information about the participants' language profile, covering aspects such as the participant's mother tongue (Q17), the languages they spoke daily (Q18) and their daily use of Catalan, Spanish and English (Q18.1).

The data obtained from the questionnaire was categorical (sex), ordinal (self-estimated language proficiency, frequency of exposure, etc.) and scale (age, number of years studied L2, FLA statements, etc.). Demographic summaries are presented in *Section 4.2*. The calculation of spoken and written input/ output was derived from the sum of responses to different statements. First, *Spoken L2 input* comprised the statements "Watching English language television", "Listening to songs in English" and "Watching videos or movies in English". *Written L2 input* was obtained from the sum of "Reading newspapers/ magazines" and "Reading books in English". Then, *Spoken L2 output* included the statements "Speaking English with native or fluent speakers" and "Speaking English with non-native speakers". Finally, Written L2 output came from "Writing emails/letter in English". Learners' self-perceived L2 proficiency was computed with the average responses regarding the aforementioned competences.

Similarly, a general FLA score and L1 anxiety score were computed by summing the responses to statements Q12-14 for FLA and Q15-16 for L1 anxiety. Nonetheless, specific statements were removed from the FLA section (Q13-14) due to the author's decision not to address pronunciation anxiety, and some questions were deemed irrelevant. The statements removed from the LBQ scores are detailed in Table 12. The data yielded a 0.89 Cronbach alpha coefficient for L2 speaking anxiety with the removed statements, and a 0.98 Cronbach alpha coefficient for L1 speaking anxiety.

Table 12: Statements removed from the LBQ FLA score

Rate the extent to which you agree/disagree with the following statements when speaking ENGLISH
I fear others might find my pronunciation of English foreign-accented.
English pronunciation is difficult for me.
My pronunciation is at a lower level than that of people around me.
I'd rather people did not hear me making a pronunciation mistake.
I (would) feel more embarrassed making a pronunciation mistake than any other type of mistake when I speak in English.
I remember the pronunciation of new words easily.
Whenever I think of my future career, I imagine myself using English with native-like pronunciation.
I think I have a talent to acquire English pronunciation.
Being understandable depends on my pronunciation.
I keep thinking that the other students are better at languages than I amI am usually at ease during tests.
I am usually at ease during tests.
I worry about the consequences of failing my foreign language classes.
I am satisfied with my current level of English speaking.
I don't worry about making mistakes in English.
It frightens me when I don't understand what the teacher is saying in the foreign language.
During class, I find myself thinking about things that have nothing to do with the course.
I get upset when I don't understand what the teacher is correcting.

Moreover, statements such as "I am calm when I have to speak in English without preparation time", "I enjoy expressing my ideas in English", "I would not be nervous speaking English with native speakers" and "I like speaking to myself in English" were subjected to reversal when computing the final score. Similarly, within the subsection of L1 anxiety, the positive worded statements detailed in Table 13 were reversed and recoded during the score computation process.

Table 13: Statements removed from the L1 anxiety score

Rate the extent to which you agree/disagree with the following statements when speaking your L1 (e.g., Spanish)
I have no fear of facing an audience.
I look forward to expressing my opinions in class.
I find the prospect of speaking mildly pleasant.
I look forward to an opportunity to speak in public.
I would enjoy presenting a speech on a local television show.
I face the prospect of making a speech with complete confidence.
I feel relaxed and comfortable while speaking in any context.
Although I am nervous before a presentation, I soon forget my fears and I enjoy the experience.
I feel that I am more fluent when talking to people than most other people.
I have no fear of expressing myself in a group.

The L2 speaking anxiety scores from the LBQ were subsequently used to categorize learners into three distinct groups: “high”, “mid” and “low”. This categorization was done to select the ten most representative learners belonging to those groups to transcribe the stimulated recalls. The individuals in the high and low group were those who scored the highest and the lowest scores, respectively (M = 123.08, SD = 34.72; min = 55, max = 192). Participants placed in the mid group were those scoring around the mean. Furthermore, the L1 speaking anxiety scores were employed to merely identify highly anxious learners. Despite participants exhibiting slightly higher than the mean scores (M = 135.96, SD = 47.59; min = 34, max = 228), none of them scored extremely high scores.

4.4.3.2. Audience and native raters’ demographic and linguistic background questionnaire

Following the LBQ, a similar questionnaire was devised using Google Forms and distributed to the audience members and the helper prior to data collection (see Appendix I). The participants’ responses were collected, and the data was scored in a separate file on the Google server. Subsequently, the author downloaded the responses into an Excel spreadsheet for further analysis. This questionnaire included two sections: the first section gathered information on general demographic information and the second section focused on their English experience (i.e., their English onset age, whether they had English certificates, etc.).

In the first part of the questionnaire, the audience members provided basic demographic data such as age and sex. The following section inquired about L2 learning experience, covering questions such as the age of onset learning (Q1), English courses taken outside school (Q2-2.1) and possession of English certificates (Q3, i.e., FCE, CAE, etc.). Participants also self-assessed their command of English across the five competences on a Likert-scale (Q4); reported weekly hours dedicated to activities done in English (Q5) and the time spent in English-speaking countries (Q6 – Q6.1); and indicated if they had been diagnosed with any speech/ hearing problem (Q7).

Concerning the native raters, a questionnaire used in previous studies such as Ortega et al. (2022) was distributed prior to data assessment (see Appendix E). The participants' responses were collected, and the data was saved in an Excel spreadsheet for further analysis. This questionnaire gathered information on general demographic information and their Spanish-Catalan learning experience and proficiency (i.e., their language onset age, whether they had any certificate, etc.).

In the first part of the questionnaire, the native raters provided basic demographic data such as age, sex, place of birth and their parents' place of birth. The second section focused on L2 learning experiences, specifically Spanish and Catalan. In this part of the questionnaire, native raters were asked to self-assess their proficiency in Spanish and Catalan, provide the age of onset of these languages, and indicate their average use of these languages. Additional questions covered their English accent and profession, required them to rate on 9-point Likert-scales statements regarding their familiarity with Spanish-accented English and other accents of English (i.e., British and American), and their overall knowledge and use of Spanish.

As reported in the previous section (4.4.3.1), data obtained from the questionnaire was categorical (sex), ordinal (self-estimated language proficiency, frequency of exposure, etc.) and scale (age, number of years studied L2). Demographic summaries are presented in *Section 4.3.1* and *4.3.2*, respectively.

4.4.3.3. L2 speaking anxiety post-task questionnaire

In addition to the idiodynamic task and the qualitative part of this study involving the stimulated recalls, an online post-task questionnaire was administered immediately after the participants' performances through Google Forms. The questionnaire included 9-point Likert scale questions and sought information regarding participants' impressions of the tasks related to (1) anxiety, (2) task difficulty, (3) mental effort, (4) task performance and (5) the potential distress caused by specific elements in the performance (see Appendix J). The post-task questionnaire in this study was based on the questions used in Préfontaine & Kormos (2015) and Mora et al. (2023).

The inclusion of questions on perceived cognitive demands posed by the tasks (i.e., mental effort and task difficulty) was deliberate (Révész & Gurzynski-Weiss, 2016; Révész et al. 2016; Robinson, 2007), considering that these cognitive demands can affect speech (i.e., fluency as in Préfontaine & Kormos, 2015) and anxiety may influence or be influenced by these constructs (Kormos, 2015; Tobias, 1979). Mental effort denotes the cognitive resources participants invest in completing a specific task, influenced by factors such as the complexity of the task, the learners' prior knowledge and their cognitive capacity. On the other hand, task difficulty refers to the inherent complexity or challenge posed by the task itself because of the complexity of the linguistic features, the unfamiliarity of the topic, or the presence of distractions.

In practical terms, a task can be difficult in the sense that it poses a significant challenge to learners, but learners may invest varying degrees of mental effort depending on their skills, motivation, and strategies. In this instance, tasks were not regarded as inherently challenging by the author, as they resembled tasks commonly undertaken by EFL learners in a classroom setting. Nonetheless, the perceived mental effort was anticipated to vary, given that the argumentative tasks needed more sophisticated linguistic resources, more specifically the ability to articulate justifications for their arguments, compared to the descriptive tasks.

To start with, participants were required to self-assess their perceived task performance (Q1; 1 = very poorly, 9 = very well), the perceived difficulty of the task (Q2, 1 = not difficult at all, 9 = extremely difficult) and the level of anxiety experienced during task performance (Q3; 1 = I did not feel anxious at all, 9 = I felt very anxious).

Next (Q4), participants were presented with a list of factors and were asked to rate, using a 9-point Likert-scale (1 = totally relaxed, 9 = the highest anxiety you have ever felt), the extent to which these elements might have caused them distress during their performances. Statements were included such as “the reactions of the person/ people in front of them”, “speaking in English” and “being audio recorded” among others. Participants had the opportunity to share additional factors causing distress in Q5, an open-ended question. Open-ended questions serve as a valuable resource, allowing for greater freedom of expression and providing insights that may not have been anticipated by the researcher (Dörnyei & Taguchi, 2009). Question 6 (Q6) inquired about the amount of mental effort that the task entailed for the participant (1 = no mental effort at all, 9 = extreme mental effort). Finally, participants evaluated their overall impressions of task performance (Q7) responding to statements regarding the ease of task planning and their perception of overall comprehensibility and fluency of their speeches among other factors.

The responses to the post-task questionnaire were analyzed quantitively and qualitatively for RQ5 on causes of distress following the same procedure as in the stimulated recalls (see *Section 4.4.2.3*). Thus, with ATLAS.ti codes were created and later percentages with the frequency counts of the codes were computed. Overall, the questionnaire data (i.e., without separating each individual task response) yielded a Cronbach alpha coefficient of .91.

When computed for each individual task, the Cronbach’s alpha for the descriptive tasks in the L2 regarding a challenging moment and a memorable childhood experience were .68 and .72, respectively. The argumentative tasks in the L2 (about means of transport and global warming) produced Cronbach’s alpha coefficients of .77 and .81, respectively. Regarding the L1, the descriptive tasks about a frustrating experience and a memorable trip obtained a .82 and .87 Cronbach’s alpha coefficient, respectively. Similarly, the argumentative tasks in the L1 about learning contexts and living contexts yielded coefficients of .76 and .87, respectively.

4.4.3.4. Audience L2 speaking anxiety questionnaire

As previously mentioned, seven people (N = 7) were recruited as an audience for the group audience context. During both data collection dates, the audience members were personal contacts of the author such as current and former work colleagues. In order to be recruited, they needed to have a high B2 or C1 level of English and be available for all the sessions. Prior to data collection, audience members met via Zoom to receive all the information necessary and to make sure they understood and knew how to use the materials. Moreover, they were sent a questionnaire that asked about their personal information and L2 experience (*Section 4.4.3.2*).

During data collection, the seven audience members were seated in a distributed way (i.e., two or three people per row) in a quiet meeting room at the University of Barcelona. They were asked to complete a questionnaire, preferably after the participants' task performance. However, they were also informed that they could take notes during the presentation if needed. The time between presentations, combined with the planning time before each participant's performance, provided enough time to complete the questionnaire.

In the questionnaire, the audience members were required to note the participant's code, which was provided by the participant before their task performance, and the task code provided by the helper. The questionnaire included four 9-point Likert-scale questions and one open question addressing the following aspects: the perceived level of the participant's anxiety (Q1), the anxiety indicators (Q2), the audience members' confidence in their ratings (Q3), the difficulty they experienced in rating the participant (Q4), and the overall comprehensibility of the participant's speech (see Appendix K). The questionnaire data yielded a 0.81 Cronbach's alpha coefficient. For this study, the perceived level of the participant's anxiety was particularly important, and data yielded a 0.84 Cronbach's alpha coefficient for the 2021-2022 academic year ratings, and a 0.85 Cronbach's alpha coefficient for 2022-2023. Finally, an average of the different rating scores were computed across audience members for each participant.

4.4.4. Summary of L2 speaking anxiety measures:

Table 14: Summary of L2 speaking anxiety measures

Type	*Measure*	*Label*	*Description*	*Calculation*
Objective	Heart rate mean	HRm	Mean beats per minute	Average computed with the HR data provided by the E4
	Heart rate standard deviation	HRsd	Standard deviation of mean beats per minute	SD computed with the HR data provided by the E4
	EDA mean	EDAm	Mean of skin conductance in response to sweat secretion	Average computed with the EDA data

				provided by the E4
	EDA standard deviation	EDAsd	Standard deviation of skin conductance	SD computed with the EDA data provided by the E4
	HRV mean 1	HRV avnn	Average of heartbeat intervals (excluding abnormal beats)	Mean computed with the IBI data provided by the E4
	HRV mean 2	HRV avrr	Average of heartbeat intervals (including abnormal beats)	Mean computed with the IBI data provided by the E4
	HRV standard deviation 1	SDNN	Standard deviation of heartbeat intervals (excluding abnormal beats)	SD computed with the IBI data provided by the E4
	HRV standard deviation 2	SDRR	Standard deviation of heartbeat intervals (including abnormal beats)	SD computed with the IBI data provided by the E4
	Root mean square of successive NN intervals	RMSSD	Beat-to-beat variability	√AVRR
Subjective	L2 speaking anxiety score	FLA Score	L2 speaking anxiety	Sum of 1-9 x 26 statements (26-234)
	L1 communication apprehension score	L1 CA Score	Communication apprehension	Sum of 1-9 x 29 sentences (29-261)
	Self-reported anxiety	Self-reports	Self-reported anxiety levels	Rating of anxiety (1 = I did not feel anxious at all; 9 = I felt very anxious)
	Others' perceptions	Others' perceptions	Audience assessments of perceived anxiety levels	Rating of anxiety (1 = extremely relaxed, 9 = extremely anxious)
	Idiodynamic score	IDm	Mean beats per minute	Average computed with the data provided by the idiodynamic task (1-5)
	Peaks per minute	Peaks per minute	Ratio of idiodynamic peaks per minute	Number of peaks/ (total duration/60)
	Stimulated recalls	-	Frequency of counts of the codes (%)	Provided by ATLAS.ti

4.5. Control tests

A battery of tests was administered at the end of data collection to ensure the participants belonged to a homogenous group (see Table 15 in *Section 4.5.4* for a summary of control measures derived from the different instruments). Having both a high working memory capacity and high proficiency may facilitate L2 oral production, as it may help the learners to control or diminish anxiety levels while speaking (Mora et al., 2023).

4.5.1. L2 Proficiency tests: EIT & Yes/No vocabulary size test

Proficiency measures were obtained through an EI task (originally designed by Ortega et al., 2002; Wu et al., 2021) and a receptive vocabulary test (Yes/No test; Meara and Miralpeix, 2015a, 2017) that measures up to 10k words.

Firstly, in the EIT participants were required to listen to 30 English sentences (See Appendix L1), ranging from 7 to 19 syllables, and repeat them as accurately as possible. These sentences progressively increased in both grammatical and lexical complexity, providing the flexibility to align with diverse proficiency levels. Indeed, the EIT has been demonstrated to have the potential to distinguish different proficiency levels effectively and reliably (McManus & Liu, 2020; Kostromitina & Plonsky, 2022; Yan et al., 2016 for meta-analyses).

This task includes high frequency vocabulary items, a range of syntactic complexity and grammatical features designed to pose challenges for L2 learners. Importantly, when the stimulus exceeds the working memory capacity of the test-takers, mere reliance on repetition becomes insufficient. Decoding and understanding the overall meaning are necessary to reconstruct and repeat the stimulus (Wu et al., 2021). Thus, the rationale behind employing the EIT as a measure of language proficiency relies on test-takers' linguistic ability to ensure correct repetition, considering that the sentences are intentionally designed to exceed the typical working memory capacity. In addition, it is essential that a temporal delay between the stimulus presentation and subsequent response exists.

The EIT was presented in a PowerPoint presentation and had a duration of 7-8 minutes. The sentences were read by an American native speaker of English (Speech rate: $M = 4.58$, $SD = .58$) and were presented auditorily over headphones for delayed repetition. Participants were instructed to repeat the sentences as accurately as possible after a 250-millisecond *beep* signal to preempt rote repetition. Participants' productions were recorded with a Shure SM58 voice microphone and a Marantz PMD-661 mkii recorder and later assessed for accuracy following Ortega et al.'s (2002) rubric (see Appendix L2). The sentences' accuracy score ranged from 0 to 4 as a function of how much of each sentence was repeated, the type of inaccuracies present and unrepeated elements. The lowest score (= 0) indicated either an inability to repeat the sentence or a minimal repetition, whereas the highest score (= 4) was given when the sentence was repeated exactly as presented in the stimulus. Scores between 1 to 3 were determined based on specific criteria provided by Ortega et al. (2002) depending on the parts of the sentences uttered or the mistakes made during the repetition. Inter-rater reliability analyses were high (α=.85) when scoring the EIT, and the author of this dissertation and the helper discussed any differences in the scoring until a strong consensus was reached.

Secondly, the Yes/No vocabulary size test is a tool for assessing language proficiency, specifically in terms of the breadth of vocabulary knowledge. That is, it does not test the depth of vocabulary knowledge (how well learners know a word in terms of spelling, collocation, etc.) but rather the size of vocabulary or the number of words a learner has (Masrai, 2022). This test assesses L2 receptive vocabulary size by presenting participants with two hundred words, encompassing both words in the target language (at varying frequency levels) and pseudowords. The latter are words that are designed to be phonologically and orthographically plausible in the language of the test which aim to control for guessing if practiced by the test-takers.

Participants then are required to indicate whether they are familiar with each word by responding with "Yes" or "No", bearing in mind that "knowing" a word means being able to understand and use it in a sentence. Individual scores could range from 0 to 10.000 words (Meara & Miralpeix, 2015b, 2017) and participants had four potential responses to the test items: (1) a "hit" when responding "yes" to a real word, (2) a "false alarm" when responding "yes" to a pseudoword, (3) a "miss" when responding "no" to a real word, and (4) a "correct rejection" when responding "no" to a pseudoword. The total test score is provided by using a correction for guessing formula, where the ratio of "yes" responses to word items (hits) was adjusted by subtracting the ratio of "yes" responses to pseudoword items (false alarms).

The results of the test can be analyzed to estimate an individual's vocabulary size, thereby providing an indirect measure of their overall language proficiency (Montero-Pérez, 2020). In accordance with the manual (Meara & Miralpeix, 2015b, 2017), the maximum achievable score is 10,000 words. Scores falling within the range of 4.500 to 10.000 words are considered good for non-native speakers, while scores ranging from

2.500 to 4.500 words are typical of intermediate-level learners. Scores below 2.500 words are likely to be unreliable and should be approached with extreme caution. Past research has demonstrated the test's correlation with general EFL proficiency (Miralpeix & Muñoz, 2018) and L2 speaking proficiency (Uchihara & Clenton, 2020). These findings underscore the utility of the Yes/No vocabulary size test as a reliable tool for assessing language competence.

4.5.2. Working Memory: Digit span test

Individual measures of short-term memory or working memory capacity were obtained through a forward and backward digit span test, which are the most widely used neuropsychological tests of working memory (Woods et al., 2011). These tests were implemented using Inquisit 6 computer software (Draine, 1998). The reason to administer both DSTs lies in the fact that while the forward DST provides a measure of the learners' short-term memory, which assesses its storage capacity, the backward DST requires both storage and manipulation of information (Montero-Pérez, 2020). In fact, the backward digit span test is recognized as a measure of complex working memory capacity involving both the central executive and the phonological loop (Kormos & Trebits, 2011). Moreover, complex WM has been found to be predictive of a greater ability to focus, divide, and switch attention among various task demands (Montero-Pérez, 2020).

In the DST, a sequence of numbers is shown on a computer screen one at a time for one second. For the forward DST, participants were instructed to recall the digits in the order the numbers were presented (see Figure 19), while for the backward DST, they had to click the digits in the reverse order. The digit string's length increased with each trial until the participant was unable to correctly recall two consecutive sequences with the same length. Therefore, the learner could only go to the next span if they had typed the correct digit sequence. In the event of an incorrect response, the same span was presented for a second attempt. If a consecutive error occurred, the participant would go back to a lower level. Each test lasted approximately five minutes and comprised 17 trials, including 2 practice trials. In the forward version, the test started at level 3, whereas in the backwards version, the test started at level 2.

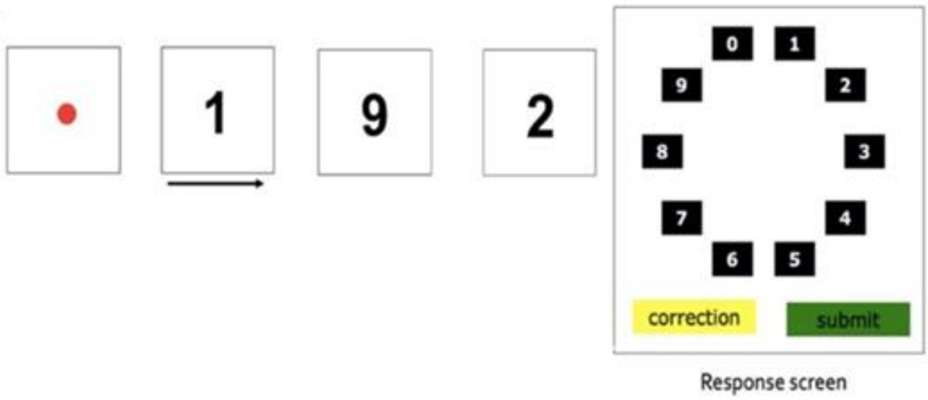


Figure 19: (Forward) Digit Span test from Inquisit 6.

Inquisit 6 data includes the two-error maximum length (TE-ML), which represents the last span correctly recalled before two consecutive errors, the number of trials before reaching TE_ML, the maximal span recalled, and the mean span providing an estimate of the score a participant would obtain 50% of the time on the basis of overall performance during the test trials (Woods et al., 2011; Inquisit Manual[22]).

TE_ML, the traditional measure of a participant's DST (as in Kormos & Sáfár, 2008; Kormos & Trebits, 2011), and the mean span (Woods et al., 2011) were extracted for both forward and backward tests. The choice of the mean span derived from Woods et al.'s (2011) findings, which showed reduced variance, improved test-retest reliability and obtained higher correlations with the results of other neuropsychological results relative to two-error maximum-length measures. This metric also heightened the sensitivity of forward versus backwards span comparisons, explained changes in digit span performance with age and education level, and offered useful insights into performance variability around maximal span. Moreover, following Mora et al. (2023), a global working memory measure was computed from the sum of both forward and backward digit span mean scores.

4.5.3. Attention: Trail making test

The trail making test[23] (TMT) is a neuropsychological test (Reitan, 1958) for evaluating visual attention and task-switching. Recognized for providing a language-independent

[22] Available at: https://www.millisecond.com/download/library/v6/digitspan/digitspan_visual/digitspanvisual/digitspanvisual.manual
[23] Available at https://www.center-tbi.eu/files/approved-translations/English/ENGLISH_TMT.pdf

assessment of an individual's ability to shift attention between two sets of stimuli, the TMT comprises two parts. Part A requires test-takers to connect consecutive digits from 1 to 25 (1-2-3-4-5-6, etc.), whereas Part B involves connecting alternating digits and letters (1-A-2-B-3-C, etc.) in ascending sequence (See Appendix M).

Part A establishes the baseline for the individual's motor and visual acuity, while part B provides a measure of the individual's executive control, assessing the ability to switch attention between two stimulus sequences. The time difference in completing the two parts serves as an indicator of attention-switching between different stimuli (Arbuthnott & Frank, 2000; Corrigan & Hinkeldey, 1987).

Participants completed first Part A followed by Part B, and completion times for both tests were measured using a stopwatch. As in Isaacs & Trofimovich (2011), the values were rounded to seconds. The attention control measure for each participant was determined by the difference in completion times between Part B and Part A (B-A), with a smaller difference indicating more efficient attention control. Furthermore, two TMT ratios were computed following established procedures: first by dividing part B by Part A (B/A) (Arbuthnott & Frank, 2000), and the second, in line with previous studies on applied linguistics, by dividing this difference between Part B and Part A (B-A) by its base (Part A) to ensure comparability of the results obtained from this subtraction (Simard et al., 2023).

4.5.4. Summary of control measures:

Table 15: Summary of control measures

Measure	*Label*	*Description*	*Calculation*
Elicited imitation	EI	Proficiency score	Sum of 0-4 x 30 sentences (0-120)
Yes/No vocabulary size	Y/N	Vocabulary size score	Score provided by the test
	fWM	Forward WM score	Forward digit span test score provided by Inquisit
WM	bWM	Backward WM score	Backward digit span test score provided by Inquisit
	WM	Overall WM score	fWM + bWM
Attention	Att	Difference in completion times between Part B and Part A	Part B time - Part A time

TMT ratio 1	Ratio of time to complete the TMT (Arbuthnott & Frank, 2000)	Part B time/Part A time
TMT ratio 2	Ratio of time to complete the TMT (Simard et al., 2023)	(Part B time - Part A time)/Part A time

4.6. Procedure

Data collection took place in two academic years' 2021-2022 and 2022-2023 once the study was approved by the University of Barcelona Bioethics Board (Institutional Review Board – IRB00003099). The first year, 68 participants were recruited from the subjects *Llengua Anglesa II* and *Fonètica I Fonologia I* in English studies undergraduate degree. The remaining 18 participants were recruited exclusively from *Llengua Anglesa II* in the subsequent year. Both data collection phases followed identical procedures in recruitment and data gathering. Participants were compensated with an extra 5% of course credit for their time following ethical standards.

First, enrollment involved students signing up on a Google Drive Excel sheet, selecting suitable dates and times to accommodate different schedules. They signed up in groups of 5 even though they did not perform in front of their peers. Following sign-ups, students were contacted via email, confirming their appointment with the author, who subsequently sought permission and obtained it through a signed informed consent form (See Appendix N) the first day of data collection. Confidentiality was maintained through coding. One-to-one sessions took place in the GRAL lab[24], while group or audience sessions took place in the Sala de Professors[25], a bigger room. In both contexts, participants were briefed on the procedure outside the rooms, and upon entering, they were guided to the recording set, where the helper put on the sensor and handed the speaking task to the participant. Participants had 1 minute and a half to prepare their performance and were asked to talk for about 3-5 minutes while standing up. During preparation time, the helper made sure the sensor was recording and that all the equipment was ready. Then, participants were audio-recorded with a Shure-SM58 voice microphone on a Marantz PMD-661 (mkii solid-state digital recorder, 44.1kHz, 16-bit) and video recorded with a Small Black Dynabook laptop's webcam (Intel® Core (TM) i7-10510U

[24] The GRAL Lab at the University of Barcelona is a space where applied linguistics researchers can collect data using a recording booth, an eye-tracker among other recording materials.
[25] The Sala de Professors at the University of Barcelona is a space used for meetings, thesis defenses, colloquiums, etc.

CPU). Participants' physiological responses (HR and EDA) were taken using an Empatica E4 wristband (https://www.empatica.com/en-eu/research/e4/).

In total, participants were required to attend 4 sessions of 45-60 minutes each which were distributed over two days per week, one week apart (see Participant Layout/ Set up in Appendix A1). The first week was devoted only to the L2 speaking tasks, while the following was for the L1 speaking tasks and control tests. At the end of their participation, every participant had done a total of 8 tasks, 4 in each language (L2 – L1). In order to prevent task effects, all the tasks differed regarding the topics but always followed a similar structure and included prompts. In the descriptive tasks, personal experiences were explained, and in the argumentative tasks, three different options were compared. During the first academic year, COVID-19 safety measures were followed using facemasks, in accordance with Spanish government guidelines.

The people selected as audience members were chosen through convenience of time availability as well as interests. Most of the members were or had been English teachers and had some interest in L2 acquisition. Prior to data collection days, these audience members met with the author via Zoom and were explained the whole procedure as well as their role. Their task involved firstly rating L2 speaking anxiety on a 9-point Likert-scale and then writing down the elements (if any) that made them think the participant experienced distress. They were informed to act as naturally as possible (i.e., nodding or smiling to show understanding if they felt so called to do) but they were asked to never intervene during the presentations (i.e., provide a word, a recast…). In addition, they were encouraged to complete the questionnaire once the presentations were over to avoid creating a more tense context for the participants. One of the members of the audience was another PhD colleague who is designated as "the helper". Besides assessing the participants as part of the audience, this person helped the author with the recordings and conducted the one-to-one sessions. This choice was made because some of the participants were familiar with the author since she was her teacher in *Llengua Anglesa II*; thus, in order to make sure everyone was unfamiliar with the people they performed in front of, the helper took charge of the presentations while the author directed the study and was in charge of the subsequent tasks.

During performances, interventions by the audience and the helper were deliberately kept to a minimum, adhering to instructions not to interrupt participants and to act as naturally as possible (e.g., nodding, frowning, or smiling to show understanding only if they deemed it appropriate). Furthermore, a questionnaire was distributed to both the audience and the helper in both contexts to gather information about each participant's perceived level of nervousness, potential indicators, comprehensibility, etc. (See Appendix K).

After the task performance, the helper saved the video recording to a digital file (.wmv), and the participant moved to another room in which they completed a questionnaire via Google Forms. Meanwhile, the author prepared the participants' performance video in the Anion Variable tester program (MacIntyre, 2020). Participants then watched their video while rating their levels of L2 speaking anxiety by clicking the "Increase anxiety" button when they recalled having experienced anxiety. Participants were instructed that consecutive clicks of the mouse would represent intensity in their anxiety rating, and they were encouraged to continue clicking if they experienced high levels of speaking anxiety. In the event of no rating activity from the user, the software would register a zero. Concurrently, the author proceeded to prepare the next station with the stimulated recall in a separate room (see Figure 20 for a data collection procedure diagram).

Following suggestions from MacIntyre (2012) and MacIntyre and Ducker (2022), the participants underwent a stimulated recall immediately after the ratings, providing comments on the anxiety-triggering factors. However, unlike what the researchers suggested, the resultant bitmap graph of the participant's ratings was not printed and handed to the participant to comment on. Instead, they were shown their video recording again. Stimulated recalls were recorded with a Tascam DR-05x and assistance was needed from a PhD colleague. This task was exclusively conducted for the L2 task performances, requiring participants to watch their video again. Participants were instructed to stop the recording and specify the time when they experienced L2 speaking anxiety and provide the reasons for it. Additionally, participants were encouraged to reflect on any miscommunication issues.

For L1 tasks, the time allocated for stimulated recall was instead used for control tests, including an EIT, DST, Yes/No vocabulary test, and the TMT. All tests were conducted using closed headphones for concentration. In the case of the EI, participants were audio-recorded with a Shure-SM58 voice microphone on a Marantz PMD-661 (mkii solid-state digital recorder, 44.1kHz, 16-bit).

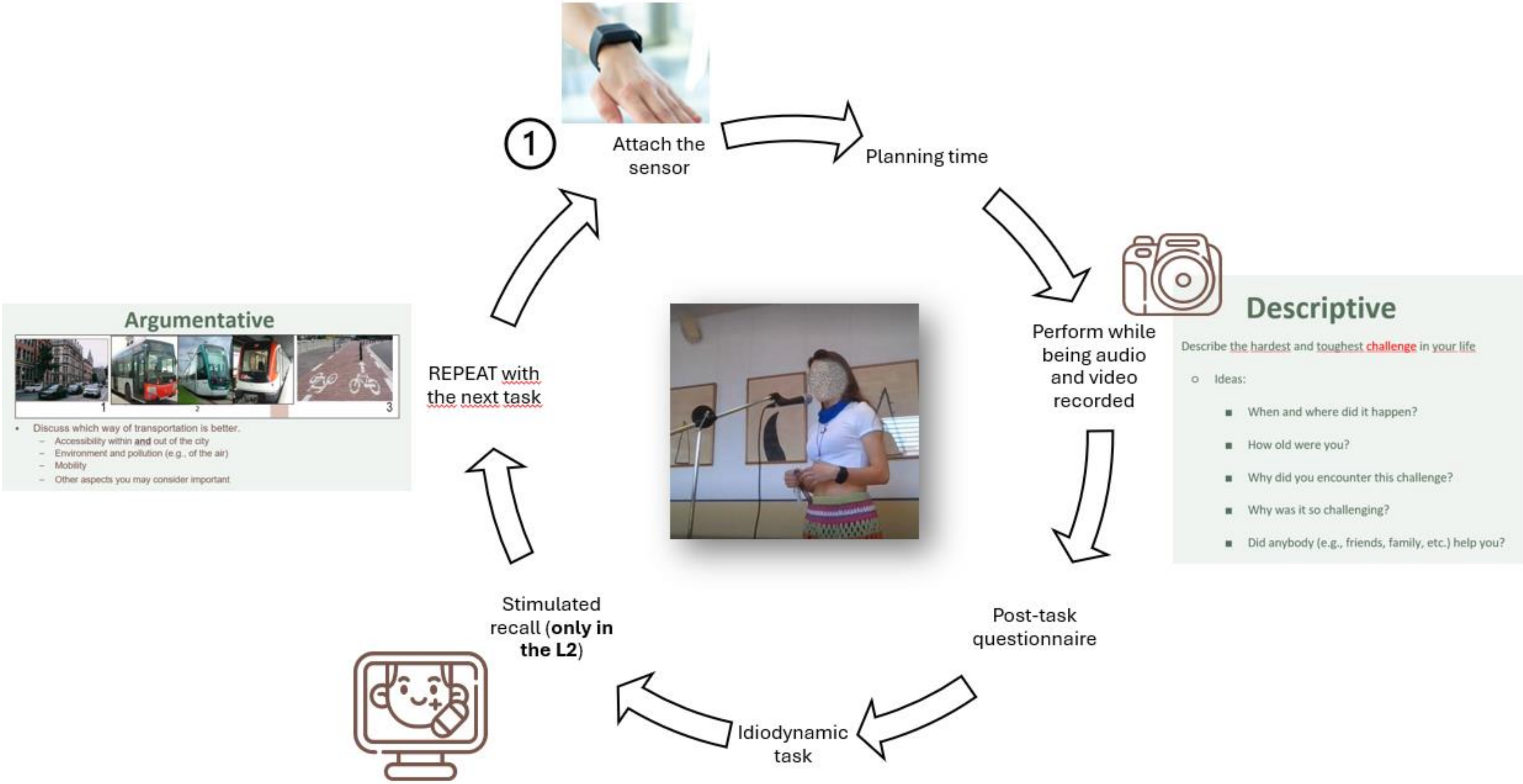


Figure 20: Data collection procedure diagram.

4.7. L2 speech measures

In this section the measures derived from the oral tasks used in this study are described.

4.7.1. Speaking fluency, accuracy, and complexity

L2 speaking fluency is a multidimensional construct consisting of utterance fluency, cognitive fluency, and perceived fluency (Segalowitz, 2010). In the current study, measurable aspects of utterance fluency were chosen as they may be indicative of the lack of automaticity in the formulation (lexical, grammatical, morphophonological, and phonetic encoding), articulation and monitoring processes characteristic of L2 speech production (Kormos, 2006; Segalowitz, 2010). All such processes may be affected by L2 speaking anxiety, leading to less fluent speech characterized by slower speed, increased pause frequency and duration, and more frequent hesitations and repetitions (as in Mora et al., 2023; Szyszka & Lintunen, 2023). Following Skehan's (2003) taxonomy, L2 utterance fluency was operationalized into three components: speed, breakdown, and repair (also in Tavakoli & Skehan, 2005; Tavakoli & Wright, 2020).

Fluency measures were obtained by running a Praat script (De Jong & Wempe, 2009) that generated two TextGrid tiers: one for counting the syllables and another segmented into silent or sound intervals. In the latter, the lower boundary for silent pauses was set at 250 milliseconds (De Jong & Bosker, 2013). The interval boundaries were then manually adjusted and re-labelled as appropriate to avoid oral closures (e.g., /k/), word-final devoiced vowels and low-intensity fricatives being labelled as silent intervals.

This TextGrid was duplicated to orthographically transcribe and manually split the transcriptions using a semi-colon (;) or a single slash (/) into AS-Units (or ASUs, Foster et al. 2000), which Foster et al. (2000) advocate as the optimal way of dividing orally-transcribed data into analyzable units. As defined by Foster et al. (2000), an ASU is "a single speaker's utterance consisting of an independent clause, or a subclausal unit, together with any subordinate clause(s) associated with either" (p. 365).

In this study, ASUs are used to mark major boundaries in speech. Pause distribution was annotated and distinguished by categorizing them as either internal pauses, occurring within ASUs, or external pauses, occurring at ASU boundaries. Silent pauses were categorized by adding *e* or *i* depending on their placement: *pe* for silent external pauses (between ASU) and *pi* for silent internal pauses (within ASU). Sound intervals consisting of hesitations were annotated in the transcripts as non-lexical filled pauses (such as *uh, uhm, er, mm*) and re-labelled as well as external filled pauses (*fpe*; between ASU) or internal filled pauses (*fpi*; within ASU). It is worth noting that when identifying filled pauses, the corresponding syllable was removed from the syllable TextGrid.

Furthermore, in a duplicate of the transcription TextGrid tier, repairs such as repetitions and self-corrections were annotated with a code between hyphens (i.e., -R-), as well as grammatical, lexical and pronunciation mistakes (phonemic substitutions and mispronunciations) (i.e., -G-, -L-, -P-, respectively) for accuracy and complexity analyses (Hewitt & Stephenson, 2012). Figure 21 provides an example of a transcript with information about ASU boundaries, repairs, and the silent and filled pauses.

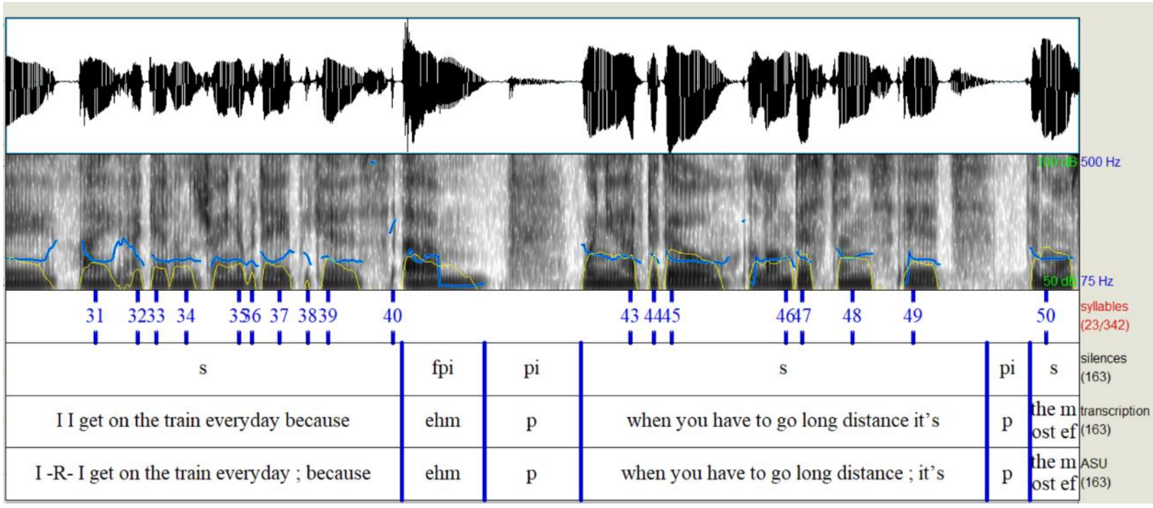


Figure 21: Example of transcription and coding.

Once the transcriptions and coding were completed, a Praat script named “labelsandduration” was run[26] (Arnhold, 2014). The script converted the TextGrids’ information into a table that included the duration of each interval, the tier and the text for each filename in a folder. From this information, speed measures obtained from the Praat script were recalculated, and repair and breakdown fluency measures were computed with respect to time (i.e., duration of speech), as used in previous research (i.e., Mora et al., 2023; Pérez-Castillejo, 2019; Szyska & Lintunen, 2023), and per 100 syllables (Kahng, 2020) (see Table 16). Speaking accuracy measures were obtained by computing error rates (number of errors per 100 syllables and per total ASUs) based on the accuracy-annotated orthographic transcriptions in Praat.

[26] Available at https://sites.ualberta.ca/~arnhold/scripts/TextgridsOnly/Print_all_intervals.praat

With respect to lexical complexity, measures were computed using CLAN (CHILDES; MacWhinney, 2000). Using the orthographic transcriptions, CLAN generated a word list containing all the tokens and types from each of the participants' oral narratives. These lists were then revised to exclude proper names, incomplete words, and loose syllables.

Table 16: L2 speaking fluency, accuracy and complexity measures

Type	*Measure*	*Label*	*Description*	*Calculation*
Speed	Speech rate	SR	Syllables per minute including pause time	nsyll[27]/duration
	Mean length of run	MLoR	Average number of syllables between pauses	nsyll/speech runs
	Average syllable duration	ASD	Average syllable duration	phonation time[28]/nsyll
	Articulation rate	AR	Syllables per minute excluding pause time	nsyll/Phonation time
Breakdown	Phonation time ratio	PhonTimeR	Proportion of time spent speaking out of total duration of speech sample	(Phonation Time*100)/Total Duration
	Duration of silent external pauses	Durpse	Average duration of silent external pauses	silent external pausing time/ npse
	Duration of silent internal pauses	Durpsi	Average duration of silent internal pauses	silent internal pausing time/ npsi
	Duration of filled external pauses	Durpfe	Average duration of filled external pauses	filled external pausing time/ npfe
	Duration of filled internal pauses	Durpfi	Average duration of filled internal pauses	filled internal pausing time/npfi
	Pause duration	pDur	Average duration of pauses	pausing time / n of pauses (silent and filled)
	Pausing time ratio	pTimeR	% of pausing time out of total duration of speech sample	(Pausing Time*100)/Total Duration
	External pausing time ratio	ExtpTimeR	% of external pausing time out of total duration of speech sample	(External pausing Time*100)/Total Duration
	Internal pausing time ratio	IntpTimeR	% of internal pausing time out of total duration of speech sample	(Internal pausing Time*100)/Total Duration

[27] Codes in the table: n= number; syll= syllables; pse = silent external pause; pfe = filled external pause; psi = silent internal pause; pfi = filled internal pause; npsfe = silent and filled external pauses; npsfi = silent and filled internal pauses.
[28] Phonation time is the total speaking time minus silent time (De Jong et al., 2013).

	External silent pause frequency	pseFreq	number of external silent pauses per minute of speaking time	npse/Phonation time
	Internal silent pause frequency	psiFreq	number of internal silent pauses per minute of speaking time	npsi/Phonation time
	External filled pause frequency	pfeFreq	number of external filled pauses per minute of speaking time	npfe/Phonation time
	Internal filled pause frequency	pfiFreq	number of internal filled pauses per minute of speaking time	npfi/Phonation time
	Pause frequency	pFreq	number of pauses per minute of speaking time	n of pauses/Phonation Time
	Pause frequency per syllables	P100Syll	number of pauses per 100 syllables	(n silent and filled pauses*100)/nsyll
	External silent pause frequency 2	Ep100Syll	number of silent external pauses per 100 syllables	(npse pauses*100)/nsyll
	Internal silent pause frequency 2	Ip100Syll	number of silent internal pauses per 100 syllables	(npsi pauses*100)/nsyll
	External filled pause frequency 2	FPE100Syll	number of filled external pauses per 100 syllables	(npfe pauses*100)/nsyll
	Internal filled pause frequency 2	FPI100Syll	number of filled internal pauses per 100 syllables	(npfi pauses*100)/nsyll
Repair	Dysfluency ratio	DysR	Number of repetitions, restarts, rephrasings per 100 syllables	Number of dysfluencies/(nsyll/100)
Accuracy	Errors per ASU	ErrASU	Errors per ASU	(nErr*100)/nASUs
	Lexical errors ratio	LexErr	% of ASUs with lexical errors	(nLexErr*100)/nASUs
	Grammatical errors ratio	GramErr	% of ASUs with grammatical errors	(nGramErr*100)/nASUs
	Pronunciation errors ratio	PronErr	% of ASUs with pronunciation errors	(nPronErr*100)/nASUs
	Error rate per 100 syllables	ErrASUsyll	number of pauses per 100 syllables	(nErr*100)/nsyll

	Lexical errors ratio 2	LexErrSyll	number of lexical errors per 100 syllables	(nLexErr*100)/nsyll
	Grammatical errors ratio 2	GramErrSyll	number of grammatical errors per 100 syllables	(nGramErr*100)/nsyll
	Pronunciation errors ratio 2	PronErrSyll	number of pronunciation errors per 100 syllables	(nPronErr*100)/nsyll
Complexity	Guiraud Index	Guiraud	Types/square root of tokens	Types/√ tokens
	Type-Token Ratio	TTR	Types/Tokens	ntypes/ntokens
	Sentence node per ASU	SNxASU	Number of sentence nodes per ASUs	
	Mean length of ASU	WDxASU	Words per AS unit	nwords/nASUs

4.7.2. Comprehensibility and accentedness

Comprehensibility and accentedness of L2 speech were measured globally through native speakers' perceptual judgements (Saito & Plonsky, 2019). Seven native speakers of English (N = 7) who had experience with comprehensibility and accentedness were recruited as expert raters. Moreover, most of them were either graduate students in applied linguistics or experienced EFL teachers familiar with Spanish-accented speech. They assessed a total of three hundred forty-four (N = 344) 40-second speech excerpts on 9-point scales for comprehensibility (1 = very difficult to understand to 9 = very easy to understand) and for accentedness (1 = not accented at all to 9 = very strongly accented) (as in Saito et al., 2016; Mora et al., 2023) presented in a randomized order through Praat. According to Isaac and Thomson (2013) who compared 5-point to 9-point Likert-scales for measuring comprehensibility and accentedness, 9-point Likert scales are preferable. The researchers argue that "smaller scales might cause a ceiling effect", especially for accentedness (Isaac & Thomson, 2013, p. 135-136) and that, in general, 9-point Likert-scales allow for finer-grained distinctions and improved reliability. In addition, the speech samples had been previously normalized for peak and mean amplitude, and band stop-filtered at 50Hz using GSU Praat tools (Owren, 2008).

Despite their expertise, NS raters were trained prior to the rating task on the rating procedures as well as on the notions of comprehensibility and accentedness in a Zoom meeting. Based on prior research, comprehensibility was defined as the degree of ease or difficulty in raters' understanding of L2 speech (Derwing & Munro, 2009). On the other hand, accentedness was defined as raters' perception of the degree to which L2 speech is influenced by nonnative features (Trofimovich & Isaacs, 2012).

A shared folder containing the task included the consent form (See Appendix O1), specific instructions and a brief explanation of the tasks performed by the participants (see Appendix O2) to eliminate familiarity bias (Derwing et al., 2004; Rossiter, 2009). Moreover, the rating task contained three practice trials which were not included in the study. After familiarizing themselves with the procedure and the tasks, the speech samples were rated using Praat (Boersma & Weenink, 2012), and the raters could hear each audio excerpt up to two times before making a scalar judgment for comprehensibility and accentedness, in that order. The 344 speech samples were randomly distributed across four sessions, with each session containing a total of 86 speech excerpts. These excerpts were presented in blocks of 26, separated by short breaks. Each block took approximately two and a half hours to complete, and raters were compensated for their time. To reduce fatigue, the raters could assess each block on separate days. Inter-rater reliability (Cronbach's alpha intra-class correlation coefficients) of native raters' ratings was high for comprehensibility (α = 0.77) and accentedness (α = 0.72), so an average comprehensibility and accentedness score were computed across raters for each learner.

CHAPTER 5: Results

In this chapter, the analyses and results of the five main research questions are presented. Section 5.1 presents the preliminary analyses, while Section 5.2 details the statistical analyses, which are further divided into subsections, each corresponding to one of the research questions: the effects of task type, context and language (Section 5.2.1); the alignment between self-reported anxiety and others' perceptions, and the most prominent anxiety non-verbal cues (Section 5.2.2); the alignment between subjective and physiological measures (Section 5.2.3); anxiety effects on L2 speech (Section 5.2.4); and L2 speaking anxiety sources (Section 5.2.5).

5.1. Preliminary analysis

Some preliminary steps were performed before conducting the primary analyses which involved screening the HR and EDA data for extreme values (2.5 standard deviations below or above). These data points were removed since it would represent an artifact or noise that would not reflect participants' physiological responses. Descriptive statistics were then extracted for each participant's physiological responses across different oral tasks, contexts, and languages.

Data were tested for normal distribution and homogeneity of variance in SPSS 27 using the Kolmogorov-Smirnov and Shapiro-Wilk tests before the statistical analyses were carried out. The Shapiro-Wilk test indicated that normality could be assumed for many variables ($W \geq .95$). However, for certain measures, including physiological and breakdown fluency measures, a logarithmic transformation was necessary to achieve normality, allowing the use of parametric statistics.

Then, we explored whether participants in this study tended to be nervous when speaking in their L1 and L2 (i.e., trait anxiety scores). To do so, descriptives from the participants' L2 speaking anxiety and L1 communicative apprehension LBQ were obtained. These revealed that participants exhibited relatively high means in L1 communication apprehension (*M* = 135.96, *Mdn* = 142.50, *SD* = 47.59; min. = 34, max. = 228)[29] and L2 speaking anxiety (*M* = 123.08, *Mdn* = 129, *SD* = 34.72; min. = 55, max. = 192). This implied that participants in this study were likely to exhibit nervous tendencies when performing both in the L1 and the L2 (See Appendix P for detailed information on the LBQ section regarding L1 and L2 anxiety descriptives).

[29] M= mean, Mdn = median, SD= standard deviation, min. = minimum, max. = maximum

Then, two-tailed Pearson correlations were conducted between L2 speaking anxiety and L1 communicative apprehension trait scores. This step aimed at exploring the relationship between L1 and L2 trait anxiety and determining whether individuals who experience anxiety in their L1 would be more likely to experience anxiety when speaking in their L2. Results showed a significant moderate-to-strong correlation between the scores (r = .738, p < .001), indicating that L1 anxiety is a reliable indicator of experiencing L2 anxiety. Hence, it can be inferred that participants who experience nervousness when performing in their L1 will tend to be nervous when performing in their L2. This further suggests that anxiety experienced can transfer from one language to the other due to similar underlying causes, which might be but not limited to fear of negative evaluation, lack of confidence, or previous negative experiences in communicative situations.

Additionally, Pearson correlation coefficients were run between physiological measures, subjective anxiety measures and mental effort and task difficulty. Significant associations were found between mental effort and task difficulty with both physiological and subjective anxiety measures, indicating that when participants found the tasks more challenging, their nervousness increased (see Appendix Q).We also ran Pearson correlation coefficients among the participants' control tests to determine whether they measured a related construct. In the case of the EIT and vocabulary size test, a significant moderate correlation was found (r = .604, p <.001), suggesting both tests measured a related construct, namely, proficiency. By contrast, a weak correlation between the WM and attention control scores ($r \geq .130$), suggesting that while there is a slight relationship between these cognitive measures, they represent distinct constructs.

Furthermore, it was found that L2 speaking anxiety scores were related weakly and negatively to proficiency scores (EI: r = -.259, p < .001) and vocabulary size (r = -.213, p < .001), suggesting that proficiency is only minimally associated with L2 speaking anxiety. Despite the weakness, this negative association suggests that more proficient learners experienced less L2 speaking anxiety. Hence, in this case, higher proficiency might lead to increased confidence in their language abilities and a better command of the language, making L2 communication less stressful and more manageable.

5.2. Statistical analyses

5.2.1. Effects of language, task type and context on speaking anxiety (RQ1)

This section examines first the relationship between the L1 and L2 speaking anxiety scores (trait anxiety) obtained from the LBQ and physiological measures of anxiety as well as subjective measures of anxiety (state anxiety) (RQ1.1), followed by the effect of task type (descriptive vs. argumentative), context (one-to-one vs. group audience), and language (L2 vs. L1) on learners' speaking anxiety levels, as measured through physiological and subjective measurements (RQ1.2).

Before reporting statistical tests analyses, Table 17 first presents the descriptive statistics of L1 and L2 physiological measures and self-rated anxiety levels experienced during their presentations with respect to the different context and task types. At first glance, in Table 15 we can observe patterns between the two languages, with both physiological and subjective measures being higher in the L2 compared to the L1. Furthermore, from the descriptives, participants' self-assessments indicated that the group audience context and the argumentative task tended to cause more distress than the one-to-one context and the descriptive task, respectively. Indeed, as Figure 22 shows, the frequencies (%) of the participants' anxiety self-rating scores tend to be higher in the L2 than in the L1. However, this pattern was not consistently reflected in physiological measures. To illustrate, lower means were present in the group audience context and the argumentative task compared to the one-to-one context and the descriptive task.

Table 17: Anxiety descriptives by language, task type and context

	L2			
	Descriptive		*Argumentative*	
	One-to-one	*Group Audience*	*One-to-one*	*Group Audience*
HRm	90.37 (15.52)	91.29 (13.88)	90.56 (12.21)	88.99 (13.11)
HRsd	11.07 (5.18)	11.16 (5.44)	11.94 (6.88)	10.20 (5.42)
HRV avnn	.57 (.13)	.54 (.12)	.59 (.14)	.56 (.13)
HRV avrr	.58 (.14)	.54 (.12)	.58 (.13)	.57 (.13)
EDAm	1.96 (2.10)	2.11 (2.89)	1.94 (2.61)	1.63 (1.94)
EDAsd	.83 (.98)	.84 (1.37)	.97 (1.04)	.72 (.95)
Self-reports	5.08 (2.32)	5.38 (2.33)	5.19 (2.22)	6.12 (2.40)

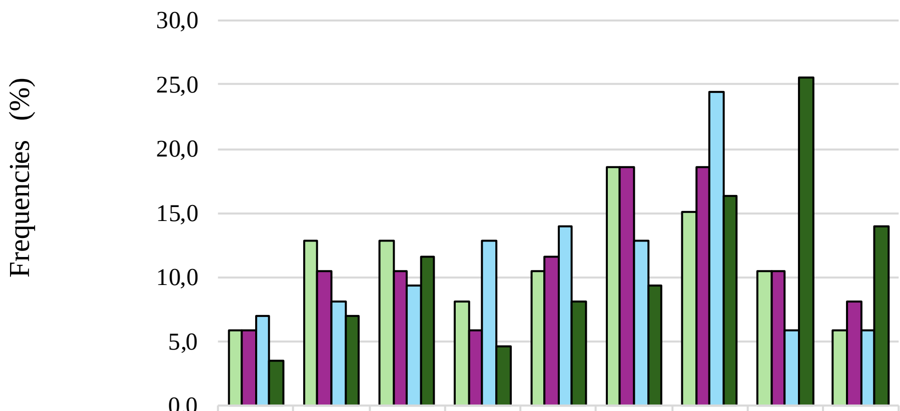
Idiodynamic score	.46 (.77)	.54 (.74)	.57 (.86)	.87 (1.09)
Peaks per minute	3.85 (2.80)	4.23 (3.00)	4.29 (3.15)	4.78 (3.36)
Audience assessment	3.38 (.84)	2.88 (.74)	3.28 (.84)	3.20 (.79)

L1

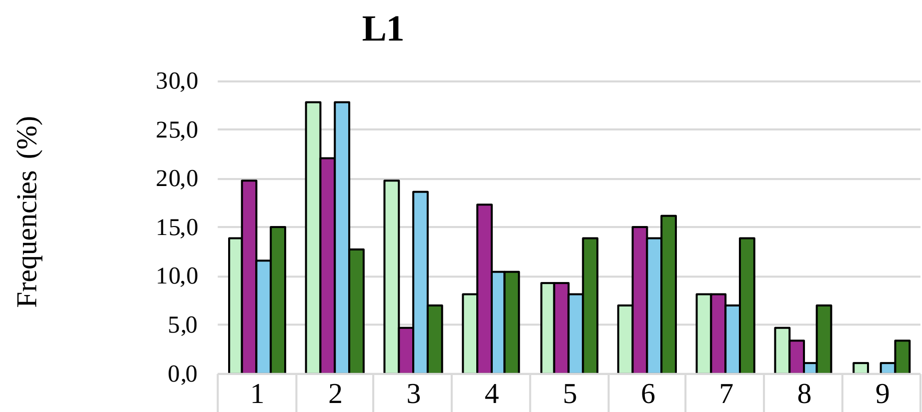
	Descriptive		*Argumentative*	
	One-to-one	*Group Audience*	*One-to-one*	*Group Audience*
HRm	87.14 (12.37)	85.74 (10.27)	87.74 (10.86)	87.11 (11.56)
HRsd	10.27 (6.92)	9.67 (5.26)	10.13 (5.40)	9.89 (5.25)
HRV avnn	.57 (.14)	.54 (.15)	.57 (.13)	.57 (.13)
HRV avrr	.58 (.14)	.54 (.15)	.58 (.14)	.57 (.13)
EDAm	1.64 (2.01)	1.34 (1.39)	2.17 (2.96)	1.73 (1.97)
EDAsd	.71 (.95)	.59 (.76)	.84 (1.32)	.80 (1.03)
Self-reports	3.55 (2.13)	3.70 (2.15)	3.58 (1.99)	4.56 (2.39)
Idiodynamic score	.23 (.63)	.26 (.62)	.25 (.51)	.36 (.82)
Peaks per minute	4.78 (3.36)	2.88 (3.00)	2.99 (2.87)	3.51 (3.43)
Audience assessment	3.22 (.82)	2.62 (.72)	3.19 (.77)	2.71 (.71)

Note: Mean (SD).

L2



	1	2	3	4	5	6	7	8	9
Descriptive one-to-one	5,8	12,8	12,8	8,1	10,5	18,6	15,1	10,5	5,8
Descriptive group	5,8	10,5	10,5	5,8	11,6	18,6	18,6	10,5	8,1
Argumentative one-to-one	7,0	8,1	9,3	12,8	14,0	12,8	24,4	5,8	5,8
Argumentative group	3,5	7,0	11,6	4,7	8,1	9,3	16,3	25,6	14,0



	1	2	3	4	5	6	7	8	9
Descriptive one-to-one	14,0	27,9	19,8	8,1	9,3	7,0	8,1	4,7	1,2
Descriptive group	19,8	22,1	4,7	17,4	9,3	15,1	8,1	3,5	0,0
Argumentative one-to-one	11,6	27,9	18,6	10,5	8,1	14,0	7,0	1,2	1,2
Argumentative group	15,1	12,8	7,0	10,5	14,0	16,3	14,0	7,0	3,5

Figure 22: The distribution of participants' anxiety self-rating scores
Note: on 9-point Likert-scales

In order to answer RQ1.1, it was examined whether the L1 and L2 anxiety scores obtained from the LBQ were related to physiological and subjective measures of anxiety obtained in this study (See Table 18). Results showed that for *L1 anxiety score* and HR, HRV (avnn) and EDA were weak and not statistically significant (HR: r = .084, p = .119; HRV: r = -.087, p = .109; EDA: r = .038, p = .484). However, moderate correlations were found between the L1 anxiety score and self-reports (r = .436, p < .001) and the idiodynamic task measures (r = .330, p < .001), indicating a moderate alignment between the subjective anxiety measures and L1 anxiety score.

Similarly, when examining the correlations between *L2 speaking anxiety score* and the physiological and subjective measures, the L2 speaking anxiety score was weak and not statistically significantly related to physiological measures (HR: r = .049, p = .364; HRV: r = -.066, p = .222; EDA: r = .013, p = .804). On the other hand, moderate significant correlations were found between the L2 speaking anxiety score and self-reports (r = .396, p < .001) and the idiodynamic task measure (r = .349, p < .001), indicating a meaningful relationship between anxiety-specific subjective measures and trait L2 speaking anxiety.

Table 18: Two-tailed Pearson correlations between L1 and L2 speaking anxiety

	HRm	HRV avnn	EDAm	IDm	Self-reports
L1 speaking anxiety score	.084	-.087	.038	.436**	.330**
L2 speaking anxiety score	.049	-.066	.013	.349**	.396**

Note: N = 344; ***p* < .001 (2-tailed)

We then questioned whether the same results would be consistent when distinguishing high- from low-anxiety learners. Then, we conducted a median split to examine whether the relationship between anxiety scores and both physiological measures of anxiety and subjective measures of anxiety differed between participants with high and low anxiety levels (RQ1.1.1) (See Table 19). However, the results revealed no differences in the relationship between L1 and L2 speaking anxiety scores and the physiological and subjective measures, except for a weak but significant association between HR and L2 speaking anxiety score (r = .167, p = .028), and between EDA and L1 speaking anxiety score (r = .192, p = .011) in the high-anxiety learners.

Table 19: Two-tailed Pearson correlations between L1 and L2 high and low speaking anxiety

		HRm	HRV avnn	EDAm	IDm	Peaks per minute	Self-reports
High	L2 speaking anxiety score	.167*	.127	.059	.384**	.282**	.246**
	L1 speaking anxiety score	.035	.013	.192*	.313**	.320**	.216**
Low	L2 speaking anxiety score	.017	.136	.041	.304**	.301**	.316**
	L1 speaking anxiety score	.008	.025	.205	.193**	.150**	.308**

Note: N = 172; **p* < .05 (2-tailed); ***p* < .001 (2-tailed)

These findings suggest that students who reported a tendency to experience L1 or L2 anxiety afterwards provided consistent reflections on their anxiety experienced in the oral tasks. Thus, despite possible individual biases (such as the tendency to under-report or over-report anxiety), participants in this study were able to recognize and report their anxiety levels. The lack of alignment between physiological arousal and speaking anxiety scores in either language, except for the weak association for the high-anxious learners, suggests that there is a dissociation between perception and physiological responses.

RQ1.2 asked whether learners' speaking anxiety levels were affected by *Task Type* (descriptive vs. argumentative), *Context* (one-to-one vs. group audience) and *Language* (L2 vs. L1). A series of linear mixed-effects models were run in SPSS 27 on the anxiety measures with *Task Type*, *Context*, *Language* and their interactions as fixed effects and a random intercept for *Subject* (see Appendix R for all parameter estimates).

For HRm, the model yielded non-significant main effects of *Task Type* (F [1, 680] = .018, p = .893) and *Context* (F [1, 680] = .573, p = .450) and a significant main effect of *Language* (F [1,680] = 9.08, p < .001). None of the interactions reached significance. A main effect of *Language* was observed because regardless of *Task Type* and the *Context*, participants' beats per minute were higher in the L2 than in the L1 (t [1, 680] = 3.64, p = .000) (See Table R1 in Appendix R for parameter estimates).

For HRsd, the model showed a significant main effect of *Language* (F [1, 680] = 9.24, p = .002), but neither the main effect of *Task Type* (F [1, 680] = .000, p = .983) or *Context* (F [1, 680] = 1.66, p = .197) reached significance. Again, none of the interactions were significant (See Table R2 in Appendix R for parameter estimates). As with HRm, HRsd was larger in the L2 (M = .991, SE = .013, 95% CI [.967, 1.01]) than in the L1 (M = .941, SE = .013, *95% CI* [.916, .966]) in both task types and contexts (See Table 17).

In addition, we explored heartbeat intervals through HRV avnn. While HR provides a direct measure of physiological arousal, HRV, specifically average normal-to-normal intervals, offers additional insight into autonomic regulation. Given that L2 speaking anxiety involves both sympathetic activation (reflected in HR) and changes in autonomic balance (captured by HRV), including both measures allows for a more comprehensive analysis.The model, whose parameter estimates are detailed in Table R3 in Appendix R, revealed significant main effects of *Task Type* (F [1, 680] = 4.04, p = .045) and *Context* (F [1, 680] = 7.79, p = .005) and no significant effects of *Language* (F [1, 680] = .037, p = .847). None of the interactions were significant. A main effect of *Task Type* was observed because participants experienced more beat-to-beat variability in the descriptive task (M = .577, SE = .010, *95% CI* [.557, .597]) than in the argumentative task (M = .559, SE = .010, *95% CI* [.539, 597]). Regarding *Context*, greater HRV was observed in the one-to-one context (M = .580, SE = .010, *95% CI* [.560, .600]) than in the group audience (M = .577, SE = .010, *95% CI* [.536, .576]).

For EDAm, the model showed non-significant main effects of *Task Type* (*F* [1, 680] = .032, *p* = .858) and *Context* (*F* [1, 680] = .117, *p* = .732) and close to significant main effects of *Language* (*F* [1, 680] = 3.74, *p* = .053). None of the interactions were significant (See Table R4 in Appendix R for parameter estimates). The main effect of *Language* arose because regardless of *Task Type* and the *Context*, participants experienced more activation of the sweat glands in the L1 (*M* = -.090, *SE* = .044, *95% CI* [-.176, -.005]) than in the L2 (*M* = -.022, *SE* = .044, *95% CI* [-.108, .064]).

The model showed for EDAsd non-significant main effects of *Task Type* (*F* [1, 680] = .777, *p* = .378), *Context* (*F* [1, 680] = 1.33, *p* = .248), and a main effect of *Language* (*F* [1, 680] = 6.23, *p* = .013). None of the interactions were significant (See Table R5 in Appendix R for parameter estimates). As with EDAm, participants experienced more variability in the sweat measures in the L1 than in the L2 (*t* [1, 680] = -2.49, *p* = .013).

In summary, physiological measures showed that task type, context and language were independent of one another, except for HRV avnn. Moreover, for physiological measures, the model showed non-significant main effects for *Task Type* and *Context*. On the other hand, speaking anxiety levels were affected by *Language* for HRm, HRsd, EDAm and EDAsd. The findings indicate that participants experienced faster beats per minute and more variability of HR values in the L2 but showed more sweating in the L1. In the case of HRV avnn, findings suggest more beat-to-beat variability, which is associated with a more relaxed state, which was present in the descriptive task and in the one-to-one context.

The outcome of the mixed-effects model on the Idiodynamic data (IDm) revealed significant effects of *Task Type* (*F* [1, 680] = 10.78, *p* = .001), *Context* (*F* [1, 680] = 15.10, *p* < .001), and *Language* (*F* [1, 680] = 126.96, *p* = .000), as well as significant *Task Type x Context* (*F* [1, 680] = 6.44, *p* = .011) and *Task Type x Language* (*F* [1, 680] = 4.04, *p* = .045) interactions (See Table R6 in Appendix R for parameter estimates). Bonferroni-adjusted pairwise contrasts revealed that in the argumentative task, the IDm was significantly higher (indicating higher anxiety levels) in the one-to-one sessions (*M* = -.897, *SE* = .076, *95% CI* [-1.04, -.747]) than in the group context (*M* = -.693, *SE* = .076, *95% CI* [-.843, -.543]), but not in the descriptive task (*t* [1, 680] = -.945, p = .345) (See Table 17). This suggests that *Context* significantly affects anxiety levels when performing argumentative tasks but not for descriptive tasks. Regarding the *Task Type x Language* interaction, pairwise contrasts (Bonferroni-adjusted) showed that the argumentative task in the L2 generated more anxiety (*M* = -.747, *SE* = .076, *95% CI* [-.896, -.598]) than the descriptive task (*M* = -.580, *SE* = .076, *95% CI* [-.728, -.431]) but not for the L1 (*t* [1, 680] = -.862, p = .389). Hence, the influence of task type on L2 speaking anxiety varied depending on the language used.

For Peaks per minute ratio responses, the model revealed no significant main effects of *Task Type* (*F* [1, 680] = 2.26, *p* = .133) and *Context* (*F* [1, 680] = 2.78, *p* = .095) and a main effect of *Language* (*F* [1, 680] = 52.65, *p* < .001). None of the interactions were significant (See Table R7 in Appendix R for parameter estimates). The main effect of *Language* arose because regardless of the context and task type, there was always a difference between the languages. In this case, participants reported more anxiety peaks in the L2 (*M* = .502, *SE* = .037, *95% CI* [.430, .573]) than in the L1 (*M* = .346, *SE* = .037, *95% CI* [.273, .419]).

Finally, for self-reports, the model unfolded significant effects of *Task Type* (*F* [1, 680] = 11.35, *p* < .001), *Context* (*F* [1, 680] = 20.98, *p* < .001), *Language* (*F* [1, 680] = 153.47, *p* = .000) as well as the *Task Type x Context* (*F* [1, 680] = 7.95, *p* = .005) interaction (See Table R8 in Appendix R for parameter estimates). Bonferroni-adjusted pairwise contrasts revealed that the argumentative task in the group audience (*M* = 5.33, *SE* = .206, *95% CI* [4.93, 5.74]) generated more anxiety than the one-to-one context (*M* = 4.38, *SE* = .206, *95% CI* [3.98, 4.78]), but this was not the case for the descriptive task (*t* [1, 680] = -1.24, *p* = .214).

In contrast to physiological measures, subjective measures showed that task type, context and language were not independent of one another, as evidenced by the significant interactions. Main effects of *Language* were found only for peaks per minute, indicating that participants reported more peaks of anxiety in the L2 than in the L1. For the IDm, in the argumentative task, participants experienced higher anxiety in one-to-one sessions than in the group context. Moreover, the influence of task type on L2 speaking anxiety depended on the language used: the L2 argumentative task induced higher levels of anxiety than the L2 descriptive task, a pattern which was not present in the L1. However, for self-reports, the model revealed a contrasting trend, with higher anxiety experienced in the group setting during the argumentative task compared to one-on-one sessions.

5.2.2. Self-reports vs. others' perceptions (RQ2)

This section examines the illusion of transparency of anxiety (as described by in Savitsky and Gilovich, 2003)[30] by analyzing the extent to which self-assessments of anxiety align with others' perceptions of anxiety in oral presentations. First, the concept "illusion of transparency" (i.e., learners overestimating the extent to which their anxiety is visible to others) was explored by computing a "disagreement score", which was derived by subtracting the participants' scores from the audience scores (both had evaluated anxiety on a 9-point Likert scale). A positive result indicated that the participant perceived their anxiety to be higher than the audience did, while a negative result suggested that the audience perceived the participant's anxiety to be higher than the participant did. See Appendix S1 and S2 for detailed information on the post-task questionnaire and the audience questionnaire descriptives.

Table 20 shows that the descriptive statistics for disagreement scores varied across different contexts and task types. The descriptors show that disagreement scores were generally higher and negative in the one-to-one context than the group audience, except for the argumentative task in the L1. This suggests that, in one-to-one contexts, the helper perceived participants' anxiety to be higher than the participants themselves reported. Conversely, in the group audience context the disagreement scores were generally lower and positive, except for the descriptive task in L1. This indicates that the seven audience members did not perceive participants' anxiety to be as high as the participants reported.

Table 20: Comparison of the disagreement scores by language, task type and context

	L2			
	Descriptive		*Argumentative*	
	One-to-one	*Group Audience*	*One-to-one*	*Group Audience*
Disagreement score	-1.37 (2.37)	.57 (2.48)	-1.11 (2.20)	.71 (2.32)
	L1			
	Descriptive		*Argumentative*	
	One-to-one	*Group Audience*	*One-to-one*	*Group Audience*
Disagreement score	-1.95 (2.53)	-.29 (1.97)	1.84 (2.22)	.30 (2.40)

Note: Mean (SD).

[30] As explained in Section 1.4, the illusion of transparency refers to the tendency for anxious individuals to overestimate how apparent their anxiety symptoms are to others.

First, we conducted two-tailed Pearson correlation analyses to explore the relationship between participants' self-ratings and the observers' ratings (i.e., those of the helper in the one-to-one sessions, and the average ratings from the seven audience members) (RQ2.1)[31]. The statistical analysis revealed weak but significant correlations between self-reported anxiety and others' perceptions (r = .235, p < .001), and weaker between the idiodynamic means (IDm) and others' perceptions (r = .158, p < .001). This finding suggests that anxiety, when self-reported by participants, generally correlated with the perceptions of others.

Next, a GLMM was conducted to assess the effects of *Task Type* (descriptive vs. argumentative), *Context* (one-to-one vs. group), and *Language* (L1 vs. L2) as well as their interactions on the level of disagreement (RQ2.1.1). Random intercepts by *Subject* were included in the random-effects structure. The model revealed a close to significant main effect of *Task Type* (F [1, 680] = 3.83, p = .051), and significant main effects of *Context* (F [1, 680] = 179.12, p = .000) and *Language* (F [1,680] = 21.12, p < .001). None of the interactions were found to be statistically significant (See Appendix S3 for parameter estimates), indicating that the disagreement scores were consistent. The disagreement score was greater in the descriptive task (M = -.762, SE = .181, *95% CI* [-1.11, -.408]) than in the argumentative task (M = -.485, SE = .181, *95% CI* [-.840, -.130]), suggesting that participants perceived themselves as less anxious than observers did. On the other hand, for the argumentative task, the smaller disagreement score suggests that participants' self-assessments of anxiety were more aligned with audience perceptions compared to the descriptive task.

Moreover, the helper (who acted as an audience member in the one-to-one context) showed greater disagreement with participants' ratings (M = -1.57, SE = .181, *95% CI* [-1.92, -.1.21]) compared to the group audience (M = .324, SE = .181, *95% CI* [-.031, .678]). In other words, the helper perceived the participants' anxiety to be higher than the participants did. Regarding *Language*, this behavior was more present in the L1 (M = -.949, SE = .181, *95% CI* [-1.30, -.594]) than in the L2 (M = -.298, SE = .181, *95% CI* [-.653, .056]) (See Table 20.

[31] As specified in Section 5.1, the Shapiro-Wilk test indicated that normality could be assumed for many variables ($W \geq .95$). However, for te physiological measures of anxiety and breakdown fluency measures, a logarithmic transformation was necessary to achieve normality, enabling the use of parametric statistics.

Within RQ2, RQ2.2 aimed to explore which non-verbal cues were the most indicative of L2 speaking anxiety for the audience. A total of 26 non-verbal and verbal cues were identified in the 785 comments for the 86 participants during their oral presentations as indices of L2 speaking anxiety by the seven audience members and the helper. The collected reports encompass various behaviors, as detailed in Figure 23 and Table 21 Some of these non-verbal and verbal cues were grouped under broader categories in ATLAS.ti as part of "movement", "touching" and "voice".

Figure 23: Word cloud of audience reports created using ATLAS.ti

The most prevalent cue was avoiding eye contact, accounting for 19.11% of the observations. This suggests that judges were particularly sensitive to whether participants directed their gaze towards them, interpreting the lack of eye contact as a sign of experiencing anxiety (i.e., "*the participant only stared at one person in the audience*" or *"not quite looking at the audience and shook the mic at one point"*)[32]. This was followed by participants being "out of words" or struggling to maintain the verbal flow which accounted for 10.06% (e.g., *"loss of words", "she got complete stuck, some things she said didn't make much sense"*).

Among the movement-related cues, hand movements were the third most common cue, comprising 9.17% of the instances. For instance, audience members considered the participants as anxious because they moved their hands constantly: *"She felt insecure about what she was saying, didn't know what to say and she was in blank, lots of fillers, nervous smiling and laugh, looking around and avoided eye-contact,*

[32] Quotes in this section may contain grammatical errors; they were copied *sic*.

nervous hand moves". Although they did not generally specify what they meant by nervous hand movements, some described these movements as "*constant*", occurring "*at the moment she did not know what to say*", or "*erratic*". It is noteworthy that audience members recognized other physical behaviors or movement-related cues less frequently, such as rocking (2.80%), leg movement (1.40%), feet movement (1.91%) and finally, shaking (0.89%). To quote some audience members:

> *"Sometimes looking around, constant moving, moving legs and feet, hands moving nervously"* (#Judge 1)
>
> *"Quite a lot of hand movements, rocking legs and feet"* (#Judge 7)
>
> *"He was crossing and de-crossing legs"* (#Judge 3)
>
> *"The participant was moving hands quite a bit, crossing arms, rocking legs, some hesitations over certain words, had to prompt her to speak a bit more but didn't say much"* (#Judge 5)

Additionally, nervous smiles or laughter (e.g., "*giggles*" or "*for a moment doubted a little bit about the answer and laughed sometimes but in a nervous way*") and silent pauses each accounted for 6.11% of the observations, reflecting vocal manifestations of anxiety. As opposed to the silent pauses, the use of fillers such as "you know", "like" or "well" (3.95%) or filled pauses (1.91%) such as "uhm" was observed as a sign of anxiety less frequently. Other vocal indicators of anxiety were also evident for the judges, with stuttering observed in 4.46% of the instances and trembling voice in 3.57%; however, low voice volume was accounted for 1.53%.

Touching behaviors also played a significant role in their assessments. For example, touching materials (e.g., the table participants had nearby, the post-it with their subject code), clothes, or accessories (e.g., "*constantly touching her scarf*") was noted in 5.10% of the instances, while touching body parts or hair was reported around 4.46% of the time. To illustrate "*the participant didn't stare the audience, stopped talking and didn't know what else to say, playing with a small paper*". These actions suggest that participants were engaging in self-soothing behaviors to cope with their anxiety.

Similarly, position and tension in the body were observed in 4.46% of the cases. For example, audience members thought the participants appeared nervous when they had "*tense shoulders, strange position of feet and legs*". On the other hand, limited movements or engaging in small repetitive actions also were notable cues, although to a lesser extent, including hands in pockets (1.91%), being static (0.89%), and fidgeting or fiddling (0.76%).

Finally, other subtle yet notable cues observed were the speech content and mistakes (4.08%), the need for prompting (1.78%) and facial expressions (0.64%). To quote some audience members:

> *"Complex rather than clear sentences, I feel he was less confident with this topic, repetition of words and ideas at the beginning"*
>
> *"Started off nervous, looked towards the paper, pauses, changing what she wanted to say, forgetting words and had to prompt"*
>
> *"Structure a little bit chaotic, at some points, little eye contact, a bit slow rhythm due to pauses"*
>
> *"Her faces, and her relief when finished"*
>
> *"She had some trouble understanding the instructions, slight tremble in the voice, frowning expression and looking upwards at times, some pauses and hesitations, moving feet and nervous laughter had to prompt"*

Table 21: Anxiety indicators by the audience members

Cues	**Instances**	**%**
Avoiding eye contact	150	19.11%
Out of words	79	10.06%
Movement: hand movement	72	9.17%
Nervous smile/ laughter	48	6.11%
Silent pausing	48	6.11%
Touching: material, clothes or accessories	40	5.10%
Position and body tension	35	4.46%
Touching: Body parts or hair	35	4.46%
Voice: stutter	35	4.46%
Speech content and mistakes	32	4.08%
Use of fillers	31	3.95%
Voice: tremble	28	3.57%
Movement: rocking	22	2.80%
Hands in pockets	15	1.91%
Movement: feet	15	1.91%
Repetitive	15	1.91%
Need of prompting	14	1.78%
Filled pauses	12	1.53%
Voice: low voice	12	1.53%
Movement: Leg Movement	11	1.40%

Speed	11	1.40%
Movement: shaking	7	0.89%
Being static	7	0.89%
Touching: fidgeting or fiddling	6	0.76%
Facial expressions	5	0.64%
TOTAL	785	

5.2.3. Subjective and physiological measures of anxiety (RQ3)

This section focuses on answering RQ3 by assessing the relationships between different L2 speaking anxiety measures obtained in this study. To determine whether subjective measures (i.e., self-assessments) are related to physiological measures of anxiety (i.e., HRm, EDAm), parametric correlation analyses were employed.

Prior to conducting the analyses, we explored the relationship between HR and HRV to ensure the measures correlated negatively, as suggested in previous research. The results confirmed this assumption with weak but significant negative correlations between HR and HRV measures (HRV avnn and HRm: $r = -.130$, $p < .001$; HRV avrr and HRm: $r = -.137$, $p < .001$). Moreover, both physiological measures, HR and EDA, were positively correlated ($r = .153$, $p < .001$), as were subjective measures of anxiety, that is, self-reports and the IDm ($r = .527$, $p < .001$). These correlations indicated consistency within these categories. Although the significant correlation between HR and EDA was weak, it suggests a consistent relationship between physiological arousal measures. In contrast, self-assessments and the idiodynamic measurement showed stronger alignment in capturing individuals' self-perceptions of anxiety.

Findings also revealed non-significant relationships ($p > .05$) between HRV measures (avnn – avrr, sdnn – sdrr, and RMSSD) and EDA. This finding can be explained by the fact that HRV is primarily influenced by the ANS, particularly parasympathetic activity, while EDA is more directly related to sympathetic nervous system activity. Hence, these two measures capture different physiological responses to anxiety.

Two-tailed Pearson correlation analyses revealed significant but weak ($r \geq .130$) correlations between self-assessment measures and both HR and EDA (see Table 22). Despite being weak, these significant positive associations between HR, EDA and self-assessment measures reflect a certain alignment between bodily physiological responses and how learners perceived and reported their own anxiety levels. This is particularly evident in the association between HR and self-reports ($r = .131$, $p < .001$), as HR was a salient physiological symptom for the learners. Very weak but statistically significant associations were found between HR and IDm (r = .087, p < .05), and peaks per minute (r = .080, p < .05). EDA, a less salient physiological response, was also found to be related, but very weakly, with self-reports (r = .086, p < .05), and IDm (r = .113, p < .05). However, these correlations were found to be non-significant with peaks per minute (p > .05).

These findings suggest that when there was an increase of physiological manifestations of anxiety, that is, more beats per minute and sweat, learners reported higher levels of anxiety. It further indicates that learners could recognize and faithfully report their physical symptoms of anxiety, particularly increases in HR, without underestimating these internal states. However, the correlations observed are very weak, indicating that physiological responses may have reflected other emotional states, not just anxiety, while self-reports might have focused specifically on anxiety.

Additionally, neither of the HRV measures (See Appendix T) were significantly associated with self-assessments (p > .05), suggesting that variability in the beat-to-beat is not reflected in self-assessments. This lack of relationship can be explained by the fact that HRV measures capture subtle, moment-to-moment variations in HR which are not consciously perceptible to individuals. Finally, audience assessments did not show a significant association with physiological measures (p > .05). This lack of relationship is logical since audience assessments are often based on specific manifestations of anxiety which tend to be highly visually perceptible, for example, behavioral symptoms. Thus, physiological symptoms are almost impossible to discern unless they are highly severe (e.g., extreme sweating) or verbally reported.

Given these results, we next sought to determine whether statistically significant would emerge and whether these associations would be stronger when considering distinctions between languages (L2 vs. L1) and anxiety levels (high vs. low). Consequently, two-tailed Pearson correlation analyses were run again differentiating languages (See Table 23a) and anxiety levels (See Table 23b). When language distinctions (L1 vs. L2) were considered, correlations between subjective and physiological measures were found to be non-significant in the L1 (see Table 23a). On the other hand, in the L2, physiological measures of anxiety were weakly associated with subjective measures of anxiety. In the case of HRm, weak but statistically significant correlations were found between HRm and subjective measures: IDm (r = .160, p < .05), Peaks per minute (r = .155, p < .05), and Self-reports of anxiety (r = .114, p < .05). These findings suggest that as HR increased (i.e., an indicator of anxiety), participants reported higher IDm values, experienced more frequent anxiety peaks, and reported greater levels of anxiety. In contrast, for EDAm, a significant but weak correlation was observed only with IDm ($r = .174$, $p < .05$). However, no significant correlations were observed between EDAm and peaks per minute or self-reported anxiety.

The differences in results between languages suggest that the emotional experience of using the L2 may differ from that of the L1. Performing in the L1 involves a lower cognitive load, as the participants are proficient and comfortable expressing themselves in their native language. In contrast, speaking in the L2 could result in varying levels of emotional engagement, stress, and cognitive load, which in turn may influence physiological responses. Overall, this ease in using the L1 may help keep physiological responses minimal and stable. Conversely, using the L2 introduces additional challenges, as participants must not only communicate ideas but also manage linguistic aspects that are less familiar or automatic.

Regarding anxiety levels, the data used to select the cases were obtained from the post-task performance questionnaire, as it reflects state anxiety the same way that physiological measures[33]. Low-anxious students were defined as those reporting ≤ 4 in the 9-point Likert-scale, whereas high-anxious students were those reporting ≥ 5. Results from the two-tailed Pearson correlation analyses revealed significant but weak correlations between subjective and physiological measures in participants with high reported anxiety levels. On the other hand, no significant correlations were observed between physiological measures and self-reported anxiety. Despite the weak alignment found in high-anxious learners, this finding reflects the fact that self-reported anxiety and physiological measures are related. Hence, it is plausible that in more stressful situations, stronger alignments between subjective and physiological measures of anxiety would emerge. That is, high-stress scenarios could lead to a clearer and more consistent relationship between how anxious participants feel and how their bodies respond. Meanwhile, the absence of significant correlations when focusing on low-anxious students might imply that students were generally relaxed during the task. Thus, if they experienced low anxiety overall, it is possible that their physiological responses were so minimal that they were undetected. Again, this could imply that a certain level of anxiety is necessary for physiological and subjective measures to align consistently.

[33] Prior to this analysis, it was observed that high levels of reported anxiety from the post-task questionnaire correlated significantly ($r = .285$, $p = .008$) with L2 speaking anxiety scores from the LBQ.

Table 22: Two-tailed Pearson correlation between subjective and physiological measures of anxiety

Note: $N = 688$; $^{*}p < .05$ (2-tailed); $^{**}p < .001$ (2-tailed)

	1	2	3	4	5	6	7
1. HRm		-.130**	.153**	.087*	.080*	.062	.131**
2. HRV avnn			-.035	.051	.044	-.013	.028
3. EDAm				.113*	.046	.029	.086*
4. Idiodynamic					.877**	.167**	.527**
5. Peaks per minute						.130**	435**
6. Audience assessments							.235**
7. Self-reports							

Table 23a: Two-tailed Pearson correlations differentiating between L1 and L2.

Note: N = 344 per language; **p* < .05 (2-tailed); ***p* < .001 (2-tailed)

	1	2	3	4	5	6	7
L1							
1. HRm		-.178*	.173**	.013	.028	.021	.078
2. HRV avnn			-.013	.066	.059	.005	.015
3. EDAm				.046	.026	.008	.038
4. Idiodynamic					.838**	.134*	.490**
5. Peaks per minute						.086*	.396**
6. Audience assessments							.158*
7. Self-reports							
L2							
1. HRm		-.082*	.118*	.160*	.155*	.069	.114*
2. HRV avnn			-.056	.039	.031	-.033	.041
3. EDAm				.174*	.105	.021	.104
4. Idiodynamic					.915**	.082*	.523**
5. Peaks per minute						.106*	.436**
6. Audience assessments							.139*
7. Self-reports							

Table 23b: Two-tailed Pearson correlations differentiating between L2 anxiety levels.

Note: N = 358 for high levels of reported anxiety and *N* = 330 for low levels of reported anxiety; $*p < .05$ (2-tailed); $**p < .001$ (2-tailed)

			High levels of reported anxiety				
	1	2	3	4	5	6	7
1. HRm		-.100*	.144*	.191**	.152**	.049	.144*
2. HRV avnn			.118	.067	.031	.005	.035
3. EDAm				.256**	.186**	.001	.172*
4. Idiodynamic					.840**	.152*	.506*
5. Peaks per minute						.176*	.415*
6. Audience assessments							.192*
7. Self-reports							
			Low levels of reported anxiety				
1. HRm		-.139*	.168**	.036	.026	.040	.032
2. HRV avnn			-.162*	.116	.138	.074	-.047
3. EDAm				.033	.013	.067	.054
4. Idiodynamic					.783**	.135*	.461**
5. Peaks per minute						.123*	.377**
6. Audience assessments							.187**
7. Self-reports							

5.2.4. L2 speaking anxiety and L2 speech (RQ4)

This section focuses on answering RQ4 on the extent to which objective and subjective measures of L2 speaking anxiety contribute to explaining variance in L2 speaking performance, while controlling for IDs in L2 proficiency and cognitive factors (i.e., WM and attention). Descriptives for L2 speech measures are available in Appendix U1. Additionally, descriptives for L1 speech measures are also available in Appendix U2. Appendix U3 presents the descriptive statistics of NS ratings for comprehensibility and accentedness across each task and context.

Before assessing the relative contribution of L2 proficiency, cognitive factors (i.e., WM and attention) and anxiety (i.e., physiological and subjective measures) in explaining variance in L2 speech (RQ4.2), correlation analyses were conducted to understand the associations between variables (RQ4.1). As illustrated in Table 24 (See complete Table in Appendix U4), proficiency, as measured via EIT, was weakly associated with subjective anxiety measures and the dependent L2 speech measures, that is, comprehensibility, accentedness and breakdown fluency measures (i.e., pause frequency and pause duration). Accuracy and complexity measures were found to be moderately associated with L2 proficiency. Thus, more proficient learners reported experiencing less anxiety after task performance, produced shorter and less frequent pauses and made fewer errors. As expected, they were also more comprehensible and less strongly accented.

Moreover, WM was found to be weakly related to L2 proficiency ($r \geq .300$), suggesting that learners with larger WM had achieved higher proficiency levels in English. Cognitive factors were found to be unrelated to accuracy measures and weakly associated with lexical complexity measures (i.e., Guiraud). On the other hand, cognitive factors, especially attention, were negatively associated with pause duration and pause frequency measures, indicating that learners with better cognitive abilities, such as attention capacity, needed less time to formulate clauses when speaking.

Table 24: Two-tailed Pearson correlation between EIT, WM, attention, L2 speaking anxiety measures and L2 speech measures

	EIT	WM	Att.	HRm	EDAm	IDm	Self-reports
EIT	-	.284**	.053	-.041	-.047	-.165**	-.144*
WM	.284**	-	.127*	-.038	-.034	-.113*	-.070*
Attention	.053	.127*	-	-.014	-.031	-.085*	-.115**
Comp.	.303*	.022	.031	.020	-.040	-.052	-.275**

	Acc.	-.227*	-.051	-.050	.012	.013	.027	.234**
Breakdown Fluency	Durpse	-.070	-.054	-.102	.010	.072	.062	.021
	Durpsi	-.052	-.100	-.078	.058	.013	.192**	.116*
	Durpfe	-.089	-.058	-.146*	.060	.001	.231**	.134*
	Durpfi	-.127*	-.049	-.114*	.020	.001	.145**	.091
	pDur	-.094	-.050	-.122*	.002	.025	.062	.017
	pTimeR	-.147**	-.081	-.124*	.043	.102	.095	.170**
	P100Syll	-.250**	-.107*	-.123*	.054	.095	.192*	.219**
	Ep100Syll	-.203**	-.065	-.230**	.053	.124**	.045	.123*
	Ip100Syll	-.305**	-.107*	-.155*	.075	.150**	.230*	.268**
	Fpe100Syll	-.022	-.056	-.226*	.073	.160*	.178**	.167**
	Fpi100Syll	-.201**	-.126	-.124*	.003	.042	.214**	.149**
Accuracy and Complexity	ErrASU	-.495**	-.023	-.033	.007	.024	.141*	.173*
	LexErr	-.391**	-.037	-.049	.021	.015	.074	.012
	GramErr	-.453**	-.064	-.047	.082	.068	.175*	.171**
	PronErr	-.347**	-.010	-.066	.064	.123	.144	.118
	Guiraud	.385	.012	.159*	-.065	-.015	-.260**	-.354**

Note: N = 344; *p < .05 (2-tailed); **p < .001 (2-tailed); Comp.= Comprehensibility, Acc. = Accentedness

Regarding the correlations between physiological measures of anxiety (i.e., HRm and EDAm) and L2 speech measures (See Table 24, and complete Table in Appendix U4), there was no significant relationship (p > .05) between HR and any of the L2 speech measures (i.e., comprehensibility, accentedness, breakdown fluency measures, and accuracy and complexity measures). This indicates that higher levels of anxiety as measured in beats per minute (HR) were not associated with the speakers' fluency, accuracy, or complexity during their oral presentations. On the other hand, EDA was found to be weakly related to pause frequency measures (r = .12 – .16). Although the relationship was found to be too weak to draw definitive conclusions, this finding suggests that learners who experienced higher sweat secretion (EDA) paused more frequently.

Moving to subjective measures of L2 speaking anxiety (i.e., IDm and self-reports), self-reports were weakly associated with comprehensibility (r = - .275) and accentedness (r = .234). Hence, when learners reported feeling anxious, their speech was less comprehensible and more strongly accented. Moreover, weak associations emerged between anxiety, pause duration and frequency, and grammatical accuracy. Again, these findings were too weak (r = .11 – .27) to make any conclusive judgments, but they suggest that when learners reported feeling anxious, they produced longer and more frequent pauses and made more grammatical errors. Additionally, negative weak-to-moderate associations were found between subjective measures of L2 speaking anxiety and lexical complexity (i.e., Guiraud index), indicating that when anxious, learners used simpler language in their presentations.

The relative contribution of L2 proficiency (EIT scores), WM (*fWM* + *bWM*), attention (TMT ratio 1) and anxiety (HRm and EDAm; IDm and self-reports) in explaining variance in L2 speech was examined through a series of hierarchical multiple regressions. In these regression models, fluency (i.e., speed and breakdown), accuracy (i.e., ErrASU, LexErr, GramErr and PronErr) and complexity (i.e., TTR, Guiraud) measures were used as dependent measures and L2 proficiency and cognitive factors (i.e., WM and attention) were entered at steps 1 and 2 respectively as predictors. Step 3 included anxiety as a predictor. However, after observing the differing behaviors in the correlations between objective and subjective measures of anxiety, it was decided to differentiate between the two set of measures and enter them in step 3 separately. In other words, two separate sets of hierarchical multiple regressions were run: the first regression included only objective measures (HRm and EDAm) in step 3, whereas the second one included subjective measures (IDm and self-reports) in step 3. In this way, we could evaluate the contribution that L2 proficiency, WM and attention make in explaining variance in L2 speech measures. In addition, we could also assess the amount of variance that could be uniquely attributed to anxiety, with IDs in L2 proficiency, WM and attention being controlled for.

As illustrated in Table 25, (see Appendix V1 for a complete table and additional information), the first regression, which included physiological measures as predictors of L2 speech measures in step 3, revealed that L2 proficiency accounted for a significant amount of variance in comprehensibility and accentedness ratings (9.2% and 5.2%, respectively), and accuracy and complexity measures (between 11% - 25%), whereas cognitive factors only made an additional significant contribution in the case of lexical complexity, as measured by Guiraud (2%), errors per ASU (1.7%) and grammatical mistakes (1.7%). This suggests that L2 proficiency is a much stronger predictor of complexity and accuracy than cognitive factors are. In this case, higher proficiency learners were found to be more comprehensible, less accented and grammatically more accurate. Once the effects of L2 proficiency and cognitive factors were controlled for, anxiety did not explain an additional significant amount of variance in complexity or accuracy measures except for pronunciation errors every 100 syllables (5.1%). Hence, this indicates that anxiety specifically influenced pronunciation errors in language performance, but it did not have a significant impact either on comprehensibility and accentedness or on other aspects of language complexity or accuracy.

Table 25: Hierarchical multiple regression models on fluency, accuracy and complexity measures with L2 proficiency, cognitive factors, and physiological measures of anxiety as predictor variables.

		EIT		**WM and attention**		**HRm and EDAm**	
		ΔR²	*p*	*ΔR²*	*p*	*ΔR²*	*p*
	Comp.	0.92	< .001	.008	.220	.000	.993
	Acc.	.052	< .001	.010	.153	.003	.556
Speed fluency	SR	.112	< .001	.056	< .001	.002	.645
	ASD	.020	.009	.014	.081	.003	.580
	MLoR	.016	.018	.060	< .001	.017	.041
	PhonTimeR	.027	.002	.017	.051	.011	.134
Breakdown fluency	Durpse	.005	.193	.014	.085	.005	.387
	Durpsi	.003	.332	.018	.044	.004	.496
	Durpfe	.008	.117	.026	.017	.003	.603
	Durpfi	.016	.019	.017	.056	.000	.918
	pDur	.009	.083	.024	.015	.001	.857
	pTimeR	.022	.006	.020	.029	.011	.130
	ExtpTimeR	.000	.682	.059	< .001	.005	.399
	IntpTimeR	.042	< .001	.021	.022	.018	.055
	pFreq	.012	.039	.005	.406	.011	.147
	pseFreq	.000	.927	.044	< .001	.015	.073
	psiFreq	.044	< .001	.007	.287	.024	.014
	pfeFreq	.010	.076	.049	< .001	.032	.005
	pfiFreq	.005	.212	.020	.030	.004	.486
	P100Syll	.062	< .001	.027	.008	.011	.122
	Ep100Syll	.041	.061	.050	< .001	.016	.045
	Ip100Syll	.093	< .001	.017	.043	.025	.008
	Fpe100Syll	.000	.699	.051	< .001	.031	.006
	Fpi100Syll	.040	< .001	.045	< .001	.002	.705
	DysR	.010	.073	.057	< .001	.001	.804
Accuracy and Complexity	ErrASU	.245	< .001	.017	.027	.001	.855
	LexErr	.153	< .001	.019	.113	.003	.682
	GramErr	.205	< .001	.017	.043	.012	.108
	PronErr	.120	< .001	.023	.134	.032	.057

ErrASUsyll	.252	< .001	.010	.109	.001	.831
LexErrSyll	.129	< .001	.009	.383	.006	.495
GramErrSyll	.203	< .001	.009	.176	.010	.146
PronErrSyll	.114	< .001	.013	.321	.051	.011
TTR	.006	.139	.006	.346	.016	.061
Guiraud	.148	< .001	.020	.018	.003	.588
WDxASU	.001	.555	.016	.067	.004	.473

Note: Comp. = Comprehensibility, Acc. = Accentedness. The highlighted values represent the significant unique variance explained by a given predictor after controlling for all other predictors.

Regarding speed fluency measures (i.e., SR, ASD and MLoR), hierarchical regressions revealed that L2 proficiency predicted speed fluency significantly, accounting for 1.6% - 11.2% of the variance. On the other hand, cognitive factors only made an additional significant contribution in the case of SR (5.6%) and MLoR (6%). Once the effects of L2 proficiency and cognitive factors were controlled for, anxiety did not explain an additional significant amount of variance in speed fluency except for MLoR (1.7%). This suggests that L2 proficiency is the strongest predictor of speed fluency, but cognitive factors provide an additional contribution. Hence, cognitive factors are also relevant for sustaining speech flow. On the other hand, anxiety did not impact speed fluency measures, except for MLoR, which suggests that anxiety may only influence the average number of syllables a speaker produces between pauses during speech. In this case, higher distress corresponded to shorter MLoR, which indicates less fluent and more fragmented speech.

Moving to breakdown fluency measures (i.e., pause frequency and pause duration), hierarchical regressions indicated that L2 proficiency significantly influenced pause duration and pause frequency, particularly in the case of internal pauses, accounting for 1.6% to 9% of the variance. In this case, higher proficiency learners spoke more fluently, that is, paused less, and produced shorter pauses. However, L2 proficiency did not significantly contribute to the variance in external pause duration and frequency, and dysfluency ratio. Instead, cognitive factors emerged as significant predictors, explaining between 2% - 6% of the variance. This finding indicates that both L2 proficiency and cognitive factors (i.e., WM and attention) are predictors of breakdown fluency, but proficiency is particularly influential for internal pause behavior, while cognitive factors are influential in determining external pause duration and frequency. Hence, L2 proficiency primarily impacts internal pausing behavior, which is more closely related to the learners' command of the language (i.e., difficulties in planning or encoding speech). On the other hand, external pauses, which might occur due to difficulties in planning the message, are influenced by cognitive factors such as WM and attention.

After controlling for the effects of L2 proficiency and cognitive factors, anxiety accounted for an additional significant amount of variance in pause frequency measures. Specifically, anxiety explained 2.4% of the variance in silent internal pause frequency (psiFreq), 3.2% in filled external pause frequency (pfeFreq), 1.6% in external silent pauses per 100 syllables (Ep100Syll), 3.1% in filled external pauses per 100 syllables (Fpe100Syll), and 2.5% in silent internal pauses per 100 syllables (Ip100Syll). This finding indicates that anxiety is reflected in the speech of L2 learners mainly in the amount of pauses they make, both internal and external, rather than how long they are. As a result, anxiety can significantly impact the performance of some learners by disrupting the smoothness and flow of their speech, as higher levels of anxiety may lead to more frequent pauses.

Following the first regression analysis, a second hierarchical regression was conducted, but this time with subjective measures (IDm and self-reports) in step 3 (See Table 26, see also Appendix V2 for a complete table and additional information). These hierarchical regressions revealed that L2 proficiency accounted for a significant amount of variance in comprehensibility and accentedness ratings (9.1% and 5.7%, respectively), and accuracy and complexity measures (between 11% - 25%), whereas WM and attention only made an additional significant contribution in the case of lexical complexity (as measured by Guiraud, 2.2%). This suggests that L2 proficiency is a much stronger predictor of comprehensibility, accentedness, linguistic complexity and accuracy compared to cognitive factors. However, neither proficiency nor cognitive factors accounted for a significant amount of variance in Words per ASU and Token-Type ratio (TTR). Instead, anxiety emerged as significant predictors, explaining 2.1% and 2.2% of the variance, respectively. This suggests that anxiety affects lexical and structural complexity as high-anxious participants perform less complex oral presentations, with shorter ASUs and reduced lexical variety.

Furthermore, after controlling for the effects of L2 proficiency and cognitive factors, anxiety was found to account for a significant amount of variance in comprehensibility (4.1%) and accentedness ratings (3.4%), as well as lexical complexity (i.e., Guiraud, 3.3%). This finding highlights that anxiety is a significant predictor of language performance, influencing how well learners are understood, their perceived accent, and the complexity of their vocabulary. In this case, higher levels of reported anxiety were associated with delivering oral presentations that were less comprehensible, more accented, and presented a poorer vocabulary diversity.

Table 26: Hierarchical multiple regression models on fluency, accuracy and complexity measures with L2 proficiency, working memory, and subjective measures of anxiety as predictor variables.

Note: Comp. = Comprehensibility, Acc. = Accentedness. The highlighted values represent the significant unique variance explained by a given predictor after controlling for all other predictors.

		EIT		WM and attention		IDm and self-reports	
		ΔR²	*p*	*ΔR²*	*p*	*ΔR²*	*p*
Speed fluency	Comp.	.091	< .001	.006	.339	.041	< .001
	Acc.	.057	< .001	.009	.219	.034	.003
	SR	.120	< .001	.051	< .001	.036	< .001
	ASD	.015	.027	.013	.108	.022	.023
	MLoR	.017	.019	.064	< .001	.013	.108
Breakdown fluency	PhonTimeR	.040	< .001	.016	.063	.005	.407
	Durpfe	.009	.094	.027	.018	.039	.052
	Durpfi	.021	.008	.016	.071	.013	.118
	pTimeR	.032	.001	.020	.034	.003	.555
	ExtpTimeR	.000	.884	.067	< .001	.001	.772
	IntpTimeR	.049	< .001	.016	.060	.050	.051
	pFreq	.025	.004	.005	.419	.018	.049
	pseFreq	.000	.977	.045	< .001	.024	.016
	psiFreq	.055	< .001	.006	.342	.036	.002
	pfeFreq	.015	.057	.054	< .001	.020	.038
	pfiFreq	.006	.172	.021	.029	.024	.019
	P100Syll	.086	< .001	.025	.011	.021	.020
	Ep100Syll	.043	.071	.053	< .001	.008	.256
	Ip100Syll	.112	< .001	.010	.157	.040	< .001
	Fpe100Syll	.001	.513	.057	< .001	.032	.006
	Fpi100Syll	.045	< .001	.045	< .001	.022	.022
	DysR	.012	.053	.050	< .001	.001	.889
Accuracy and Complexity	ErrASU	.263	< .001	.012	.071	.003	.519
	LexErrSyll	.124	< .001	.009	.382	.002	.827
	GramErrSyll	.222	< .001	.007	.261	.009	.192
	PronErrSyll	.109	< .001	.014	.319	.002	.860
	Guiraud	.149	< .001	.022	.014	.033	.001
	WDxASU	.001	.641	.016	.074	.021	.030

In examining speed fluency measures (i.e., SR, ASD and MLoR), hierarchical models revealed that L2 proficiency accounted for a significant amount of variance in speed fluency measures, ranging from 1.5% to 12%. Cognitive factors also made significant contributions, particularly in the case of SR (5.1%) and MLoR (6.4%). Once the effect of L2 proficiency and cognitive factors were controlled for, anxiety emerged as a significant predictor, explaining an additional 3.6% of the variance in SR and 2.2% in ASD. This finding indicates that while L2 proficiency is a key factor in determining speed fluency, cognitive factors also play a role in managing the pace and flow of speech. The significant effect of anxiety on SR and ASD (after controlling proficiency and cognitive factors) indicates that anxiety can disrupt the smoothness and pace of speech. Hence, experiencing anxiety leads to less fluent speech production caused by speech disruptions, which in turn makes it difficult to maintain smooth communication.

Finally, for breakdown fluency measures (i.e., pause duration and pause frequency), hierarchical models showed that L2 proficiency also accounted for a significant amount of variance (between 1.7% - 11%) in almost all pause duration measures. Additionally, L2 proficiency emerged as a particularly strong predictor of pause frequency, especially for internal pauses, both silent and filled. However, this was not the case for external pause frequency (filled or silent). Instead, cognitive factors emerged as significant predictors for external pause frequency, explaining between 2.1% - 5.4% of the variance. This finding suggests that proficiency affects overall fluency, indicating that higher proficiency learners spoke more fluently, paused less, and produced shorter pauses. Meanwhile, cognitive factors are stronger predictors of external pause behavior. A possible explanation is that external pauses occur due to difficulties in planning the message, which is influenced by cognitive factors such as WM and attention. In this case, learners with better cognitive abilities exhibit fewer and shorter external pauses.

Once the effect of L2 proficiency and cognitive factors were controlled for, anxiety accounted for an additional significant amount of variance in pause frequency measures. Specifically, anxiety explained 1.8% of the variance in pause frequency (pFreq), 2.4% in silent external pause frequency (pseFreq), 3.6% in silent internal pause frequency (psiFreq), 2% in filled external pause frequency (pfeFreq), 2.4% in filled internal pause frequency (pfeFreq). Similarly, anxiety accounted for an additional significant amount of variance for pause frequency per 100 syllables, between 2.1% - 4% (See Table 264). The significant effect of anxiety on pause frequency, after controlling for proficiency and cognitive factors, suggests that anxiety primarily impacts L2 learners' speech by increasing the amount of pauses they make, rather than their duration. This finding is consistent with observations from physiological measures, which also show that anxiety accounts for higher frequency of pauses.

5.2.5. Sources of L2 speaking anxiety (RQ5)

This section examines the sources of L2 speaking anxiety when performing oral presentations, as well as the differentiation between the sources of highly and non-anxious learners. The post-task performance questionnaires included Likert scales for participants to rate their anxiety levels as regards specific statements (e.g., being audio and video recorded) (see Appendix S1 for descriptive data), as well as an open-ended question (Q5) about their sources of anxiety. However, most participants did not provide any comments, and those who did provided brief explanations. Despite this lack of comments, a total of 10 codes with varying frequencies were identified, as shown in Table 27. Participants that answered Q5 revealed that their anxiety was mainly due to feeling unprepared (25.49%), the topic (21.57%) and time issues (17.65%). Although participants mentioned that there was not enough planning time to prepare their presentation and that they spoke less than expected (see Figure 24 for a word cloud created using ATLAS.ti), they did not specify in the post-task performance questionnaires why the topic was a source of L2 anxiety (e.g., whether it was difficult or involved something personal).

Figure 24: Word cloud of L2 speaking anxiety sources from the post-task questionnaires

Table 27: Coding instances and frequencies (%) of the post-task performance questionnaire

Source	Instances (%)
Feeling unprepared	13 (25.49%)

Lexical issues ("can't find the word")	4 (7.84%)
Not expecting an audience	2 (3.92%)
Pressure to do it right	2 (3.92%)
Problems to recall the ideas	3 (5.88%)
Problems to recall the instructions	4 (7.84%)
Realizing making mistakes	1 (1.96%)
Sounding repetitive	2 (3.92%)
Time issues: talking less than expected	9 (17.65%)
The topic	11 (21.57%)
Total	51 (100%)

Given the limited number of participants who commented on the sources of anxiety in the post-task questionnaire, we decided to focus on the stimulated recalls, which provided more detailed accounts on L2 anxiety sources. As explained in *Section 4.4.2.3*, the stimulated recalls of all 86 participants were not transcribed. Instead, a selection method following Szyszka & Lintunen (2023) and Tóth (2017) was used. A total of thirty (N = 30) participants' stimulated recalls, representing 34.88% of the population, were transcribed for each task performed in the L2 and context, resulting in 120 audio files. This selection included participants who scored the highest (N = 10), around the mean (N = 10), and the lowest (N=10) in the FLA section of the LBQ questionnaire. Thus, making it possible to differentiate any distinction in the sources between levels of speaking anxiety.

By analyzing the stimulated recalls through ATLAS.ti, it was seen that almost every respondent had a reference to an anxious moment, which suggests that, despite the different reasons and regardless of the task type or context, performing oral presentations generates L2 speaking anxiety. Several sources of anxiety were identified which revealed that participants struggled, for instance, remembering words, organizing their thoughts, fearing being repetitive, making grammatical mistakes, among many other sources (see Figure 25 for a word cloud created using ATLAS.ti). In total, there were 22 codes with varying frequencies, as shown in Table 28. Then, four big closely-related categories were drawn from all the codes: (1) performance-related concerns, (2) linguistic-related concerns, (3) cognitive-related issues, and (4) audience-related fears.

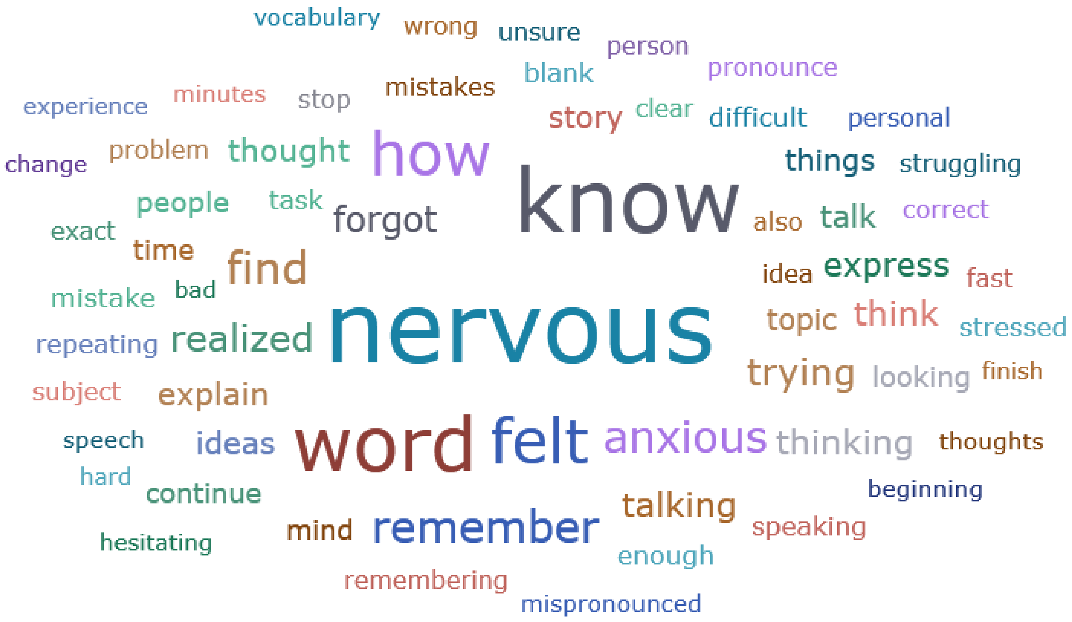


Figure 25: Word cloud of L2 speaking anxiety sources from the stimulated recalls

Participants expressed that some specific performance-related elements triggered anxiety. Performance-related issues accounted for a total of 8.86% of the L2 speaking anxiety sources, and participants pointed to different, salient features such as being recorded both with a microphone and the laptop's camera. To quote some students[34]:

> *"I was overall nervous <u>because of the microphone and the camera</u>. It didn't make it difficult the tasks but it's true… but the fact <u>it's something new</u> that I'm doing and being recorded with a microphone, which <u>I'm not used to</u>, made it a little trickier and added a little bit more of nervousness, which eventually I got rid of…"* (#S050)

> *"Also, the fact that <u>you can see how you're being recorded</u> gave me a lot of trouble because it <u>made me more aware of being recorded</u> and what I was saying"* (#S053)

[34] Quotes in this section may contain grammatical errors; they were copied *sic*.

What these comments suggest is that participants felt an increased pressure when performing due to the awareness of the recording devices, which is not something usual when performing an oral presentation. As noted by #S053, being recorded made this person more aware of their speech. Other factors related to this category include the difficulty of the topic, time issues and having to use English. To quote some comments:

> *"The topic was difficult for me because I don't know about this topic, and I didn't want to mess up and I felt nervous about messing it up or fail or say something that is wrong"* (#S004)

> *"During the first minute of my like speech, I was totally frozen because I didn't know how to approach the subject and make myself clear. The minute and a half to prepare was not enough for me in this topic so I was really, really anxious about not knowing what to say"* (#S050)

> *"I was speaking about a specific story that happened to my … in my job and I was remembering things and my nervousness got higher because usually I don't explain this kind of stories in English"* (#S063)

> *"I needed to talk more but I was running out of ideas, and this made me nervous because I had to think of something quickly"* (#S032)

> *"I didn't even know what I was saying, I was just talking for talking because I wanted to reach the minutes"* (#S012)

> *"I felt a little bit nervous because I had run out of ideas, I had already talked about the 3 pictures and I thought I had to speak more but the person controlling the machine told me it was ok, so I felt relieved"* (#S005)

> *"I was nervous because didn't know how to tackle very large subjects, there is a lot to say and yeah, just to start such a… start talking about it, it's a large topic so I was getting a little nervous and try to time for myself"* (#S053)

Unfamiliar experiences, such as being recorded, or unfamiliar topics often cause distress for EFL learners, making it more challenging for them to perform effectively. For example, as noted by #S050, planning time was not enough for this participant to prepare the given topic. Hence, the combination of insufficient knowledge about the subject and the limited preparation time exacerbated their anxiety levels. Nevertheless, anxiety can also be triggered by familiar topics. As exemplified by #S063, describing a memorable childhood memory might seem easy; however, having to use their second language to explain that specific story induced anxiety because they had never done so before. This suggests that even when learners are familiar with the topic, the fact of using the foreign language can still generate significant levels of anxiety.

Another clearly identifiable anxiety-triggering factor was linguistic-related concerns, which encompass issues with grammar, pronunciation and vocabulary, as well as L1 interference and the fear of sounding repetitive or pausing excessively. These sources accounted for a total of 24.51% of the L2 speaking anxiety episodes. Anxiety was often triggered when participants did not know the correct word or structure, and their awareness of making mistakes further heightened their nervousness. To quote some participants:

> *"I started to <u>not really know the vocabulary well the word I wanted to say</u>, I was rumbling, and <u>I realized I was making a lot of mistakes of pronunciation and vocabulary</u>... that made me nervous and making more mistakes (...) I was <u>realizing I was making grammatical mistakes</u> so again I thought that the audience would think I am stupid either for what I am saying or for how I expressed myself (...) At the end of the speech, I didn't what to say anymore and I started to <u>make more mistakes which kind of snowballed into more mistakes and more mistakes and made me more anxious which in turn made me making more mistakes</u>"* (#S050)

> *"I got nervous at that point because <u>I realized that the story was not well-told with respect the grammatical use</u>, I mean I made mistakes"* (#S063)

> *"I got nervous because <u>I didn't remember how to pronounce the word 'renewable'</u>, and I still don't know"* (#S061)

"I <u>didn't know how to say 'carril de bici' (bike lane) in English</u> so I just started to feel really anxious because <u>I didn't know the vocabulary that I needed</u> in order to say what I wanted to say" (#S042)

"I knew what I wanted to say, but <u>I didn't know the word in English</u> and that made me nervous" (#S014)

"I said something <u>in Spanish</u>, and I got nervous because <u>I wasn't sure if I could do that</u>" (#S028)

"Words <u>in Catalan kept coming to my mind</u> and that made me nervous (...) I wanted to say, well... <u>I didn't want to repeat again the word</u> 'grandmother', so I went for 'my mother's mother' but in the end that made me more nervous because I realized it was weird" (#S036)

"I started to notice that <u>I was stopping a lot</u> and <u>saying the word 'basically' too much</u>, so I just kept thinking 'don't stop' and I got more nervous" (#S032)

"I <u>didn't want to use the same expression twice</u>, but I <u>couldn't find a synonym</u>, so I was <u>hesitating a lot</u> and that made me nervous because I knew what I wanted to say but I was repeating myself" (#S077)

"I felt that my problem with this topic was that <u>I wanted to use specific terminology related but since I didn't know the words in English</u>, I couldn't continue talking after some minutes (...) I had to think really hard about another thing I wanted to say but I felt it was another larger topic to tackle and <u>I got silent for a moment, and thought: Oh no, I have to talk</u> (#S053)

"Here I was struggling, and I got nervous because <u>I was struggling to find the words</u>. I was <u>trying not to repeat all the time the same word</u>, so I was searching for synonyms. This topic was more extensive, and I needed to think fast about it. The first topic was easier because it involved personal experience, so I didn't have to think that much but this one... I was all the time I was thinking further... and I got more nervous, and <u>I felt I paused much more</u>" (#S001)

Many participants highlighted that their speaking anxiety stemmed from linguistic concerns. For example, struggling with the pronunciation of a word contributed to their nervousness as expressed by #S042 and #S061. In addition, the fear of making a mistake or realizing that they had made one significantly intensified participants' anxiety, as they worried about being judged negatively. This is exemplified in the case of #S050, who mentions feeling "dumb" and believed the audience would think poorly of them. On top of that, #S050 mentions this "snowballed into more mistakes" reinforcing the idea of a vicious circle of FLA as self-exacerbating: when feeling anxious, learners worry about how this anxiety will affect them, and ironically, this worry makes their anxiety even worse.

The third major category of participants' source of anxiety is related to cognitive issues, which accounted for a total of 56.92% of the L2 speaking anxiety episodes. Similar to findings by Aubrey (2022), participants in this study reported feeling anxious due to experiencing mental blocks, problems expressing themselves and their ideas, and struggles not only in remembering words, but also in recalling their pre-planned ideas and task instructions. Moreover, running out of ideas or the lack of them induced speaking anxiety during the oral tasks. To quote some participants:

> *"<u>I was completely blocked</u> because I <u>couldn't remember what the question was</u>, so <u>I lost track of the question</u>, despite trying to remember (...) I didn't continue talking because <u>my mind was completely blank</u> so that's all"* (#S006)

> *"I <u>panicked</u>, I don't own a bicycle, and I don't know why I said I have one, I was just so nervous"* (#S012)

> *"<u>My mind was like empty</u>. <u>I didn't remember what I wanted to say or what I was saying</u>"* (#S001)

> *"The first minute I was really comfortable with the subject and not really nervous but afterwards <u>I started not knowing what to talk about because I got blank</u>, so I started not to know what to say, and I started to rumble a little bit. This made a bit me nervous"* (#S050)

"I was totally <u>frozen</u>, and I <u>didn't know how to approach the subject and make myself clear</u>" (#S023)

"I just panicked here, I kind of realized that <u>my words came faster than my thoughts</u> (...) In here I <u>went blank</u>, I <u>couldn't remember what the task options</u> were, so I had <u>troubles to continue</u>" (#S026)

"I started to panic because again <u>words didn't come to my mind</u>, and I was <u>totally blank</u> (...) I just run out of ideas of things to say, and I stayed there saying nothing. <u>That made me more nervous than I already was because I was trying to think what to say</u> (...) at the end I thought <u>I was talking nonsense</u>, and I had the feeling I did terrible" (#S042)

"I started laughing nervously because <u>I didn't know if I was explaining the story in the right way</u>, I mean, if I was making myself clear (...) I was anxious because <u>I wanted to talk more about that, but I didn't know what else to say</u>" (#S077)

"I felt nervous because I <u>didn't know how to express myself, how to express my thoughts</u>. I <u>couldn't structure</u> it" (#S038)

"I was anxious because I was <u>desperately looking for the right words</u> and this task had more specific terminology and <u>I felt when I was for preparing for it there were some ideas</u> on there, but <u>part of my mind wasn't really "activated",</u> so in the presentation I had to <u>work really hard to focus a lot on recalling those words</u> that I needed" (#S053)

Overall, these comments reflect participants' feeling of being blocked and unable to remember the question or their planned responses (#S006). Although it is not possible to determine whether these mental blocks were caused by anxiety, in these cases, anxiety interfered with the cognitive and attentional resources available for speaking (Kormos, 2015). This interference has detrimental effects, such as diminishing the learners' confidence (#S050), making it harder to think of the right words or ideas (#S042) or causing learners to feel that their minds are not fully engaged or "activated" enough to focus (#S050), express their ideas (#S038) and retrieve necessary vocabulary (#S053).

Another clearly identifiable source of speaking anxiety that the stimulated recalls pointed to was audience-related fears, which accounted for a total of 9.71% of the L2 speaking anxiety sources. Participants described various kinds of fears related to the audience such as the mere presence of an audience, the fear of negative evaluation and judgement on their behalf, concerns about notetaking, sharing personal stories, and reactions from the audience. The selected quotes give unique insights into how these concerns intrude on the participants' minds, making it difficult for them to stay focused:

> *"I was really anxious because I had to speak in front of people that I didn't know and that were taking notes so yeah, I was really, really nervous (...) I was thinking that people thought "what is she saying? She doesn't know enough of this topic, she is dumb" so I was feeling really anxious for looking like dumb in front of them (...) I was realizing I was making grammatical mistakes so again I thought that the audience would think I am stupid either for what I am saying or for how I expressed myself"* (#S050)

> *"I saw that the person in front of me started writing something on the page and I felt a little bit anxious because I wanted to know what they were writing"* (#S026)

> *"I was looking for the 'greenhouse effect' word and I could not for the life of me remember this word and so I was hesitating and being silent because of that and that made me anxious. These people were looking at me, expecting a response, to say things and I was not talking. That was a bit nerve-wrecking"* (#S053)

> *"I was really nervous because of the reactions of the people in front of me could have and kind of felt really anxious"* (#S004)

> *"I was explaining something private and suddenly I was aware that they were listening to me, so I got nervous"* (#S012)

"In that moment I was really nervous because of <u>the topic. The story I had to tell related to something bad that happened</u> and I really struggled with, so <u>I was kind of nervous explaining it</u>" (#S042)

Anxiety poses a substantial challenge to the individual's self-concept and confidence, and these comments are clear examples of it. Overall, these worries are indicative of fear of negative evaluation and the opinion others will have of them and their abilities. As exemplified by #S050, there is a clear concern of looking "dumb" or not making a desired impression.

Table 28: Coding instances and their frequency (%) with respect to the total instances.

Factors	Sources	Total instances (%)	
Performance	1. Being recorded	4 (0.55%)	
	2. Topic difficulty/ unfamiliarity	19 (2.63%)	
	3. Time constraints	39 (5.40%)	
	4. Using English	2 (0.28%)	8.86%
Linguistic	5. Grammar issues	12 (1.66%)	
	6. Pronunciation issues	25 (3.46%)	
	7. Not knowing the word in English	29 (4.02%)	
	8. L1 interference	13 (1.80%)	
	9. Sounding repetitive	25 (3.46%)	
	10. Realizing making a mistake	55 (7.62%)	
	11. Realizing pausing excessively	18 (2.49%)	24.51%
Cognitive	12. Mental blocks	53 (7.34%)	
	13. Lack of ideas	97 (13.44%)	
	14. Problems to recall ideas	30 (4.15%)	
	15. Problems to recall task instructions	11 (1.52%)	
	16. Problems to recall words	120 (16.62%)	
	17. Problems to express themselves	100 (13.85%)	56.92%
Audience	18. Fear of negative evaluation	36 (4.99%)	
	19. Notetaking	2 (0.28%)	
	20. Having an audience	8 (1.11%)	
	21. Sharing something personal	22 (3.05%)	
	22. Audience reactions	2 (0.28%)	9.71%
	TOTAL	722 (100%)	

Finally, the codes and percentages were placed into an Excel document to categorize whether sources of anxiety were internal or external. It was observed that most of the anxiety sources in this study were due to internal reasons rather than external. Out of the 22 codes, only 8 codes (36.36%) were categorized as external (i.e., being recorded, having an audience, notetaking by the audience, audience reactions, task difficulty, topic difficulty/ unfamiliarity, time constraints, and having to perform in English). The remaining 14 (63.64%) were categorized as internal factors. Hence, most of the sources identified in this study were primarily driven by internal factors, such as the learners' concerns during their performances (e.g., mental blocks, linguistic issues), rather than external factors such as having an audience.

Sub-question RQ5.1 aimed to analyze the differences in L2 speaking anxiety sources between participants with high and low anxiety. To address this, the frequencies of the codes were calculated as percentages of their total occurrences within each group (high vs. low anxiety). Table 29 provides the number of instances and its frequency (%) for each code used by highly and non-anxious students, bearing in mind that this data is relative to the total instances for each anxiety level (high, mid, and low). Appendix W1 provides a detailed table showing both the number of instances of each code and its frequency as a percentage for each anxiety level (high, mid, and low). Additionally, Appendix W2 includes a graph comparing high and low frequency of each code. The more significant differences between these groups will be explained.

The analysis of the frequencies revealed notable differences between high-anxious and low-anxious students in some of the L2 speaking anxiety sources. Regarding performance-related factors, these sources were present for both groups but were slightly more frequently pointed out by high-anxious students. Overall, performance-related factors accounted for 10.43% for high-anxious participants, while for low-anxious participants comprised 8.58%. Specifically, high-anxious students reported topic difficulties (3.26%), and time constraints (6.51%) more frequently compared to non-anxious students (2.71%, and 4.52%, respectively).

With respect to linguistic-related concerns, low-anxious learners reported more anxiety episodes due to linguistic-related concerns (27.60%) than high-anxious learners (18.89%). Low-anxious participants reported being nervous due to pronunciation (3.62%), not knowing the words in English (4.98%), as well as L1 interference (3.62%) more frequently than high-anxious (1.63%, 2.61% and 0.65% respectively). Although realizing making a mistake was a concern for both groups, low-anxious students reported nervous arousals more frequently when realizing they had made a mistake than high-anxious learners (9.50% vs. 4.56%). Overall, this suggests that non-anxious students expressed greater concern over linguistic aspects than high-anxious counterparts.

Table 29: L2 speaking anxiety sources for high- vs. low- anxious learners: instances and their frequency (%)

Factors	Sources	Total instances (%)			
		Highly anxious		Low anxious	
Performance	Being recorded	1 (0.33%)		2 (0.90%)	
	Topic difficulty/ unfamiliarity	10 (3.26%)		6 (2.71%)	
	Time constraints	20 (6.51%)		10 (4.52%)	
	Using English	1 (0.33%)		1 (0.45%)	
			10.43%		8.58%
Linguistic	Grammar	4 (1.30%)		4 (1.81%)	
	Pronunciation	5 (1.63%)		8 (3.62%)	
	Not knowing the word in English	8 (2.61%)		11 (4.98%)	
	L1 interference	2 (0.65%)		8 (3.62%)	
	Sounding repetitive	13 (4.23%)		6 (2.71%)	
	Realizing making a mistake	14 (4.56%)		21 (9.50%)	
	Realizing pausing excessively	12 (3.91%)	18.89%	3 (1.36%)	27.60%
Cognitive	Mental blocks	26 (8.47%)		9 (4.07%)	
	Lack of ideas	44 (14.33%)		23 (10.41%)	
	Problems to recall ideas	10 (3.26%)		8 (3.62%)	
	Problems to recall task instructions	6 (1.95%)		5 (2.26%)	
	Problems to recall words	60 (19.54%)		38 (17.19%)	
	Problems to express themselves	38 (12.38%)	59.93%	32 (14.48%)	52.03%
Audience	Fear of negative evaluation	17 (5.54%)		10 (4.52%)	
	Notetaking	1 (0.33%)		1 (0.45%)	
	Having an audience	7 (2.28%)		0 (0%)	
	Sharing something personal	6 (1.95%)		15 (6.79%)	
	Audience reactions	2 (0.65%)	10.57%	0 (0%)	11.76%
	TOTAL	*307 (100%)*		*221 (100%)*	

On the other hand, cognitive-related concerns were slightly more prominent in the high-anxious group (59.93%), compared to the low-anxious group (52.03%). This difference was especially notable in anxiety arousal due to mental blocks (8.47% vs. 4.07%), lack of ideas (14.33% vs. 10.41%) and problems with recalling words (19.54% vs. 17.19%). As far as audience-related factors are concerned, low-anxious learners reported slightly more anxiety episodes (11.76%) due to audience-related factors than high-anxious learners (10.57%). These differences between the groups were not notable differences except for the mere presence of an audience and fear of negative evaluation. Specifically, high-anxious students reported these factors more frequently (2.28% and 5.54%, respectively) compared to low-anxious students (0% and 4.52%, respectively). Although the low-anxiety group did not mention being nervous because of the presence of an audience, they more frequently mentioned being nervous due to sharing something personal with the audience (6.79%), which was not as anxiety-inducing for the high-anxiety group (1.95%). This analysis suggests that while both high- and low-anxious participants faced similar challenges, the intensity and frequency of these challenges varied. High-anxious learners generally experienced more issues, especially related to cognitive factors, whereas low-anxious learners were more concerned with linguistic-related factors.

PART III

DISCUSSION AND CONCLUSIONS

CHAPTER 6: Discussion

The current study sought to examine the role of L2 speaking anxiety on L2 speech (i.e., fluency, accentedness, comprehensibility) in English learners, focusing on how different tasks and communicative situations influence anxiety. It aimed to contribute to our understanding of the causes and consequences of L2 speaking anxiety by considering both subjective and objective measures, assessing the relationship between self- and others' impressions of speaking anxiety and identifying the non-verbal cues of anxiety as well as the sources of speaking anxiety. To this end, we conducted a study where 86 EFL learners performed two types of oral presentations (descriptive and argumentative) in two different contexts (one-to-one and group audience). These presentations were performed in English (L2) and in the participants' L1.

The discussion chapter is addressed in 5 sections corresponding to each research question, and following the same order presented in the results section: the effects of language, task type and context on speaking anxiety (Section 6.1), self-reports vs. others' perceptions (Section 6.2), subjective and physiological measures of anxiety (Section 6.3), L2 speaking anxiety and L2 speech (Section 6.4) and sources of L2 speaking anxiety (Section 6.5).

6.1. The effects of language, task type and context on speaking anxiety

In addressing RQ1, firstly, this study explored the relationship between L1 and L2 speaking anxiety scores and the different measures of anxiety (i.e., physiological and subjective). Results unveiled moderate correlations between both L1 and L2 anxiety scores and subjective measures of anxiety (i.e., self-reports and the idiodynamic task). However, this relationship was not present with physiological measures (i.e., HR, HRV and EDA).

One of the challenges when implementing subjective and objective FLA measuring tools concerns the possible dissociation between perceived anxiety and physiological arousal. As mentioned in the literature, language anxiety has traditionally been investigated through subjective measures such as diary studies (e.g., Bailey, 1983; Gkonou, 2013, 2017) and questionnaires (e.g., FLCAS; Horwitz et al., 1986). However, it is undeniable that, unlike questionnaires, the use of wearable sensors facilitates obtaining continuous, real-time physiological data while learners perform a task (Rivers, 2022; Garbarino et al., 2014). Thus, this finding underscores the different insights objective and subjective data offer into learner experiences, suggesting that they should be treated independently in analysis. To this end, considering them as separate constructs when examining their effects on speech could yield a more nuanced understanding of how each dimension influences learners' performances.

We could argue that, in this case, L1 and L2 anxiety scores reflect trait-like anxiety, while physiological measures correspond to state anxiety. Hence, the lack of alignment. However, subjective measures of anxiety (i.e., self-reports and the idiodynamic task measure) are also representative of state anxiety measures. One possible explanation is due to both trait-anxiety scores and subjective measures are based on self-perceptions. In other words, learners who perceive themselves as generally anxious (trait anxiety) are likely to report being anxious in specific situations (state anxiety). All in all, the alignment between subjective measures suggests that students generally provided consistent self-assessments of their anxiety, despite the possible individual biases (such as the tendency to under-report or over-report anxiety).

Moreover, the present findings appear to be consistent with previous research which also found a lack of alignment between subjective and objective measures of anxiety (MacIntyre et al., 2010; Gallego et al., 2021; Shachter et al., 2022; Schwerdtfeger, 2004). As Schwerdtfeger (2004) already warned, this lack of alignment between self-reports of emotions and autonomic reactivity is heavily influenced by memory biases and the changes in emotional state between the time of physiological measurement and the reporting. Perhaps this dissociation between perception and physiological responses would not be present if students had experienced more severe nervousness. That is, while students in this study might have experienced nervousness having to present a topic orally, they are used to performing such activities in class and the resulting nervousness was likely not severe. Thus, in a more extreme circumstance, trait anxiety would have correlated with physiological measures.

Using linear mixed-effects models, RQ1.2 investigated the effects of language (L1 vs. L2), task type (descriptive vs. argumentative) and context (one-to-one vs. group audience) on physiological arousal anxiety measures and subjective anxiety measures when speaking the L2. Results of the current study showed physiological measures were largely unaffected by *Task Type* and *Context*, except for HRV. HRV data showed that participants experienced more beat-to-beat variability, which is generally associated with a more relaxed state (Held et al., 2021; Rivers, 2022), in the descriptive task than in the one-to-one context. In other words, these findings suggest participants were less anxious in the descriptive task than in the one-to-one context. On the other hand, physiological measures were affected by *Language*, indicating that participants experienced faster beats per minute in the L2, but more sweat activation in the L1.

The lack of significant effects of *Task Type* and *Context*, but not *Language* in the physiological measures, suggests potential methodological or data analysis issues. While participants' English proficiency was enough to perform the tasks, they may have still experienced moments of anxiety (e.g., when realizing they made a mistake or struggling to find the word in English). However, these key moments of anxiety might not have been frequent or intense enough in either task type or context to have a significant impact when averaged out in the overall data. As a result, the mean scores may have diluted these anxiety spikes, hiding any potential patterns related to task type or context. A more fine-grained analysis, such as examining specific high-stress moments might reveal more nuanced insights that the averages missed. On the other hand, *Language* emerged as a significant factor, suggesting that participants' anxiety levels were more consistently affected by whether they were using their L1 or L2. Cognitively speaking, it is more challenging to perform orally in the non-native language, even for highly proficient English speakers. The fear of making mistakes or not being understood may have been more pronounced in the L2 context, leading to more anxiety episodes captured by the sensor throughout the task, regardless of task type or context.

Regarding subjective measures, significant effects of *Task Type*, *Context*, and *Language* were found. In line with HRV results, findings for self-reports indicated that learners experienced more anxiety in the argumentative tasks in the group audience than the one-to-one context when performing in the L2. Moreover, participants reported more anxiety peaks per minute ratio in the L2. However, the idiodynamic task measure presented slightly different results. Although the IDm unveiled that the argumentative task in the L2 generated more anxiety, this was significantly higher in the one-to-one sessions than in the group context, contradicting self-reports (i.e., questionnaires).

These findings on the effects of *Language*, *Task Type*, and *Context* on speaking anxiety go partially in line with our hypothesis and previous research findings. Performing an oral presentation is known to generate anxiety both in L1 (Croft et al., 2004; McCroskey, 1970, 1978; McCroskey et al. 1985; Westenberg et al., 2009) and in L2 (Woodrow, 2006), especially in front of an unfamiliar audience (MacIntyre & Thivierge, 1995) with limited time to prepare (Young, 1990). Although it may seem logical to assume that L2 presentations would always generate more anxiety than those in the L1, to the author's best knowledge, none of the previous studies had analyzed the effects of *Language* on anxiety. This study found that learners reported higher levels of anxiety in the L2, as well as experiencing faster beats per minute. When speaking in the L2, heightened perceived anxiety is likely to arise due to the additional cognitive load and perceived communicative challenges. Moreover, these communicative challenges may exacerbate the fear of negative evaluation, one of the main sources of anxiety (Horwitz, 1986; Gkonou, 2014).

Additionally, greater sweat activation was present during L1 performance. This result should be considered with caution given that the results were close to significant, and EDA is especially sensitive to the location of the electrodes (Boucsein, 1992), as well as the room temperature and humidity (Barry & Sokolov, 1993; Empatica support on EDA, n.d.; Roos et al., 2021). Regarding the electrode placement, Empatica E4 had to be placed on the wrist in this study. However, previous research suggests that fingertips and the palm of the non-dominant hand are ideal body parts to locate the electrodes due to high density of eccrine sweat glands (Roos et al., 2021; Roos & Slavich, 2023; van Dooren & Janssen, 2012). Finally, although the room was maintained at a constant temperature, humidity was not controlled for. Hence, future research may need to consider controlling for humidity, especially if data collection takes place on different days.

Concerning *Task Type*, as expected, the argumentative tasks generated higher levels of distress. While HRV showed that participants were more relaxed in the descriptive task, subjective measures indicated the argumentative tasks as more anxiety-inducing, despite the contradiction on the *Context*. Previous research has also observed that argumentative tasks trigger more anxiety than descriptive ones (Sosa-López & Mora, 2022a; Tóth, 2017). One possible explanation is that argumentative tasks in these studies were more cognitive complex and required more linguistic demands (Tóth, 2009) than the descriptive tasks. Moreover, this study did not include describing an ambiguous picture, as in Tóth (2017) and Kormos & Préfontaine (2017), which might have contributed to perceiving the argumentative tasks as more cognitively demanding than describing a personal experience. Consequently, learners' distress could have been caused by different reasons such as perceiving the task as difficult or realizing their linguistic knowledge did not correspond to the task demands. Given that the participants were English studies university students, their linguistic knowledge was likely to be good enough to perform the tasks. However, presenting in an unfamiliar context and with limited prior planning time may have affected their perception of the task's difficulty.

Finally, mixed results were observed regarding *Context*. It was hypothesized that participants would feel heightened L2 speaking anxiety in front of an audience of seven people compared to a one-to-one scenario because it would be more intimidating. Moreover, previous research by Beatty & Behnke (1991) has shown that students experience more anxiety when performing in front of 15-20 peers compared to a single audience member. In line with our assumptions and previous research, HRV data indicated that participants were more relaxed in the one-to-one sessions. Similarly, self-reports showed that participants indicated higher anxiety levels in the argumentative task in the group audience. However, the idiodynamic task measure revealed that the argumentative task generated more anxiety in the one-to-one sessions.

This disparity between the two subjective measures can arise from how the data is gathered. In the idiodynamic task, anxiety levels are captured in detailed from moment-by-moment recollections, whereas questionnaires gathered the generalized, post-task impression. Hence, when participants continuously reflected on their anxiety, they may have re-experienced anxiety more intensely at specific moments during the one-to-one sessions, compared to the group audience context. However, their general impression was that the group audience context made them feel more nervous. One of the reasons the group audience was expected to trigger anxiety is that it consisted of seven unfamiliar people (for the participant). This also could have induced social pressure as they are being observed by multiple individuals at once. Moreover, in the group setting, learners must manage the reactions of several people simultaneously, which can be overwhelming. The variability in facial expressions, body language, and engagement levels across the group could further heighten anxiety. On the other hand, the one-to-one sessions may have been equally intimidating for participants, as the idiodynamic task measure indicated. Participants were also unfamiliar with the helper, and they might have felt under greater scrutiny, with the full attention of the helper focused solely on them.

6.2. Self-reports vs. others' perceptions

In addressing RQ2 regarding the transparency of anxiety, this study examined the alignment between self-reports of anxiety and the perceptions of others. Additionally, it explored the non-verbal cues exhibited by the participants that indicated L2 speaking anxiety.

Starting with correlation analyses, a significant alignment between participants' self-reported anxiety and others' perceptions was revealed, as hypothesized. Findings in this study contrast with those of Gregersen et al. (2017), who reported that observers had difficulties identifying distress. In this study the audience was able to detect signs of anxiety as high self-reported anxiety was shown to correspond to high anxiety in others' perceptions. Nevertheless, contrary to expectations, the strength of this alignment was found to be weak. This suggests that while the audience could identify anxiety episodes, higher levels of anxiety may be necessary for it to become more visibly noticeable to others, and for the correlation between self-reported and observed anxiety to strengthen.

Moving to the linear mixed models, results showed that the disagreement scores varied across task type, being higher in the descriptive task compared to the argumentative task. This finding suggests that the audience and the helper perceived participants as more anxious during the descriptive task than the participants self-reported. In contrast, the lower disagreement scores in the argumentative task indicate a closer alignment between participants' self-assessments of anxiety and others' perceptions. This closer alignment could indicate that in more demanding tasks, such as argumentative tasks, participants may be more aware of their anxiety, making their self-assessments more consistent with external observations. In other words, the increased cognitive and emotional load of argumentative tasks may heighten participants' sensitivity to their anxiety cues, and this is reflected in what the audience perceived. On the other hand, for descriptive tasks, participants perceived themselves as less anxious. However, observers might have noticed external cues of anxiety which led them to these differences in the scores.

Additionally, in the one-to-one context, the helper perceived participants' anxiety as higher than the participants did. Hence, compared to the group audience (N = 7), the helper seemed more receptive to anxiety signs. This finding challenges the concept of "illusion of transparency" (King & Smith, 2017; Savitsky & Gilovich, 2003) and Gregersen et al.'s (2017) findings, which revealed self-ratings to be more variable than the observers' ratings. In contrast to individuals overestimating their ratings and anxiety symptoms, this study revealed that it was the audience members, particularly the helper, who perceived participants as more anxious than the participants themselves reported.

This finding can be attributed to several factors, including participants' familiarity with performing oral presentations. It is plausible that after the first performance, participants in this study became used to the procedure and the task format. As Savitsky & Gilovich (2003) found, when learners are explicitly informed of this so-called "illusion of transparency", they tend to evaluate their own speeches more positively. Although participants in this study were not explicitly informed that their anxiety was being assessed, they likely gained insight into the objective of data collection after their initial presentations. As a result, it is plausible that their self-ratings of anxiety were influenced by this awareness combined with task familiarity, leading to a more favorable self-assessment. Similarly, the awareness of the audience regarding the study's objective may have influenced their own evaluations. That is, knowing that they were part of a research study on anxiety could have led them to scrutinize participants' performances more closely and critically.

Within RQ2, salient non-verbal cues of L2 speaking anxiety were examined. Previous studies have highlighted that learners with FLA often avoid making eye contact, exhibit restlessness, reposturing, limited facial expressions and adopt closed or rigid postures (Gregersen, 2005; Gregersen, 2007b; Gregersen et al., 2017). Consistent with these findings, this study also identified similar non-verbal indicators. The comments gathered by the audience through their observations revealed that avoiding eye contact was the most prominent anxiety cue (19.11%), followed by participants appearing at a loss for words (10.6%) and exhibiting "erratic" and "constant" hand movements (9.17%).

While physical cues such as restlessness, reposturing and rigid postures were also noted down by the audience (as in Gregersen, 2005), they were mentioned less frequently, ranging from 0.89% - 4.46%. This is also the case of cues such as facial movements and fidgeting, which were far less common in this study compared to Gregersen's (2005) findings, representing only 0.64% and 0.76% of the observations. Instead, the audience in this study focused more on behaviors such as nervous laughter (6.11%), moments of silence (6.11%) and excessive touching of the material and clothing (5.10%) over physical cues. Interestingly, while previous studies highlight behaviors like touching clothing and hair as anxiety cues (Gregersen, 2005; Gregersen, 2007b; Gregersen et al., 2017), nervous laughter or moments of silence as indicators of anxiety are not mentioned. Hence, this study found other anxiety indicators that have not been previously pointed out in the literature.

6.3. Subjective and physiological measures of anxiety

RQ3 addressed the relationship between subjective measures (i.e., self-assessments) and physiological measures of anxiety (i.e., HR, EDA). Previous research has yielded mixed findings, with most studies failing to find an alignment between self-reports of anxiety and physiological measures of arousal (MacIntyre et al., 2010; Gallego et al., 2021; Shachter et al., 2022; Schwerdtfeger, 2004; Sosa-López, 2019). Despite previous findings, it was hypothesized that a relationship between the self-reports and physiological measures of anxiety would be observed, as reported in studies by Behnke & Beatty (1981), Sevinç (2018) and Gregersen et al. (2014). This hypothesis was confirmed when data was analyzed collectively (including both L2 and L1).

Results in this study showed significant but very weak associations between self-assessment measures of anxiety (i.e., self-reports and IDm) and both HR and EDA. Additionally, in contrast to Sosa-López's (2019) findings, this study revealed a weak but significant relationship between the HR and EDA, indicating that both measures capture aspects of anxiety arousal, as an increase in heart rate aligns with heightened sweat secretion. As anticipated, HR also showed a slightly stronger correlation with self-reported anxiety than EDA, suggesting that HR is a more salient physiological response than EDA, as well as easily perceptible to L2 learners.

Although previous research has observed moderate correlations between physiological measures and self-assessments (Behnke & Beatty, 1981; Sevinç, 2018), this study found only very weak relationships. Moreover, findings indicated differences in the alignments concerning language (L1 vs. L2) and anxiety levels (high vs. low). While significant but weak alignments were found when focusing on the L2 and high reported anxiety, no association was found in the L1 performances and low reported anxiety. The lack of significant correlations when examining the L1 suggests that students were relaxed during these tasks. In other words, if participants experienced low anxiety overall, their physiological responses may have been minimal and undetected. However, this finding could also be explained by learners becoming familiar with the data collection procedure. Since L1 data collection always occurred after the L2, students may have anticipated what the tasks would involve.

Overall, these findings lead us to hypothesize that, while an association exists between physiological responses and learners' self-perceptions of anxiety, a more stressful situation might yield a stronger correlation between the measures. Consequently, it can be concluded that the oral presentations in this study, regardless of their type and context, probably did not generate enough anxiety to reflect a stronger relationship. Moreover, this could also explain why prior studies (Gallego et al., 2021; Shachter et al., 2022; Schwerdtfeger, 2004; Sosa-López, 2019) have found no relationship between subjective and physiological measures of L2 anxiety.

Finally, another plausible explanation for this weak alignment between self-assessment measures and physiological measures of anxiety may be that individuals experienced and reported anxiety differently (Gallego et al., 2021) due to memory bias (Schwerdtfeger, 2004; Lindberg et al., 2021). While physiological measures captured real-time responses during the oral presentation, self-assessments took place after task performance. As previously mentioned, a more intense situation than an oral presentation may yield stronger associations, as students would likely report higher distress, recalling the experience as more stressful.

6.4. L2 speaking anxiety and L2 speech

RQ4 addressed the possible association between L2 learners' anxiety levels and L2 speech measures of fluency, complexity, accuracy, as well as ratings of comprehensibility and accentedness. Additionally, it aimed to determine the extent to which L2 speaking anxiety accounted for variations in these speech measures, while controlling for IDs in L2 proficiency and cognitive factors.

Firstly, correlation analyses revealed that subjective measures of anxiety (i.e., IDm and self-reports) were significantly, though weakly, correlated with proficiency, cognitive factors (i.e., WM and attention), comprehensibility, and accentedness. Learners who reported higher anxiety were less proficient, had weaker cognitive abilities, and were less comprehensible and more accented in their presentations. This result should be considered with caution given the weakness of the alignment. However, this outcome highlights the detrimental effect of anxiety on comprehensibility and accentedness, an aspect that has been less studied compared to its effects on L2 fluency or accuracy. In addition, this finding goes in line with previous studies, such as Saito et al. (2018) and Nagle et al. (2022), who found a detrimental effect of anxiety on comprehensibility, and Suzukida and Saito (2023) and Zhou (2024), who reported similar effects on accentedness. A plausible explanation of the detrimental effect of anxiety on comprehensibility is that anxiety disrupted the participants' speech flow (Sanaei et al., 2015; Simard et al., 2023; Pérez Castillejo, 2019) and increased pausing frequency (Lindberg et al., 2021; Mora et al., 2023: Rood & De Jong, 2023; Szyszka & Lintunen, 2023). Hence, native listeners had more difficulties understanding learners' speeches. Concerning accentedness, anxious learners may experience cognitive overload (Eysenck et al., 2007; Kormos, 2015) due to the simultaneous demands of managing anxiety and focusing on language production. Anxiety may shift learners' focus away from pronunciation as they are occupied with other concerns (e.g., fear of making mistakes, performance anxiety). As a result, the speech in their performances might sound more strongly accented.

Additionally, subjective anxiety measures were weakly and negatively associated with certain L2 speed and breakdown fluency measures, as well as accuracy and complexity measures. Again, given the weakness of the correlations, these results should be interpreted carefully. Focusing on L2 speed fluency, participants with higher reported anxiety levels produced their oral presentations at a slower SR, with shorter stretches of fluent speech (MLoR) and they spent less time speaking (phonation time ratio). Previous research by Sanaei et al. (2015) and Pérez Castillejo (2019) also found anxiety to be related to PTR and MLoR. This observation further demonstrates the detrimental effect of anxiety as anxious learners spend less time speaking, and have shorter runs of uninterrupted speech, meaning their speech is more frequently interrupted by pauses. Although Sanaei et al. (2015) and Pérez Castillejo (2019) did not find a significant relationship between SR and anxiety, this study aligns with Simard et al.'s (2023) findings, which linked anxiety to a reduction in the overall flow of speech, suggesting that anxious learners tend to speak more slowly. Finally, consistent with the findings of Sanaei et al. (2015), this study did not find significant correlations between anxiety and AR. Thus, while anxiety appears to disrupt speech flow, it does not seem to impact how fast or slow learners articulate their speech.

Higher reported anxiety levels were also weakly linked to increased pause duration and frequency as well as more grammatical mistakes and reduced lexical variety in speakers' presentations. This outcome aligns with previous literature, which unveiled FLA to be negatively related to breakdown fluency measures (Aubrey, 2022; Bielak, 2022; Pérez Castillejo, 2019, 2023; Lindberg et al., 2021; Mora et al., 2023; Rood & De Jong, 2023; Szyszka & Lintunen, 2023), accuracy (MacIntyre & Gardner, 1994b; Mora et al., 2023; Hewitt & Stephenson, 2012; Phillips, 1992) and lexical variety (Kormos & Dörnyei, 2004). Pérez Castillejo (2019, 2023) found that when anxious, learners made longer pauses at AS-Unit boundaries. In this study, however, a weak relationship was observed not only between self-reported anxiety and filled pauses at AS-Unit boundaries, but also between anxiety and silent pauses, both at and within AS-Unit boundaries. This suggests that, when anxious, participants tended to make longer filled pauses at AS-Unit boundaries and longer silent pauses regardless of their position. Anxiety, together with cognitive overload, might have affected learners' ability to plan and execute speech effectively.

As hypothesized, anxiety also negatively influenced pause frequency. Following previous research findings, learners who self-reported higher anxiety levels paused more frequently (Bielak, 2022; Pérez Castillejo, 2019, 2023; Lindberg et al., 2021; Mora et al., 2023; Rood & De Jong, 2023; Szyszka & Lintunen, 2023). While previous research has indicated that anxiety affects pause frequency either within AS-Units (Mora et al., 2023; Pérez Castillejo, 2019) or at AS-Unit boundaries (Bielak, 2022), this study observed that, when in distress, learners paused more often both at and within AS-Unit boundaries. This evidence reveals that anxiety interferes both with conceptualization (pauses at ASU boundaries; Kormos, 2006; Pérez Castillejo, 2019) and lexical retrieval (pauses within ASUs; De Jong, 2016; Kahng, 2014, 2018). However, this alignment was found to be stronger with internal silent pauses, indicating that anxious learners may struggle more to retrieve the vocabulary needed to express their ideas.

Regarding accuracy and reduced lexical variety, higher reported anxiety levels were also weakly associated with more grammatical mistakes, fewer words per AS-Units and reduced lexical variety in their presentations. These findings go in line with previous studies, which have reported that anxious learners produce more grammatical errors (Mora et al., 2023) and use fewer complex sentences (MacIntyre & Gardner (1994b) and less varied vocabulary (Kormos & Dörnyei, 2004) in their speeches.

All in all, these weak yet significant findings suggest an overall detrimental effect of L2 speech due to L2 speaking anxiety. Considering students' detailed accounts of their anxiety during L2 tasks, the weakness of these associations can be confounding. It is possible that the task types and context of this study did not trigger enough anxiety, which may explain the weakness of the correlations. However, Kormos (2015) suggested that many studies report only weak or moderate associations between anxiety and L2 performance because global assessments of speaking performance (e.g., comprehensibility) may fail to capture the nuanced effects of anxiety. Additionally, studies tend to assess anxiety with scales that may oversimplify the complex emotional landscape learners experience. Indeed, other factors, such as motivation, willingness to communicate, or enjoyment can interact with FLA, influencing its impact. This aligns with findings from Simsek & Dörnyei (2017), who observed that highly anxious students sometimes reported enjoying the tasks despite their anxiety, highlighting the variability of affective factors.

In contrast to subjective measures, no significant correlations were found between physiological measures of anxiety (i.e., HR and EDA) and comprehensibility, accentedness, L2 speed and breakdown fluency, or accuracy and complexity measures. The exception was EDA, which was weakly associated with MLoR, specific pause frequency measures (internal and external silent pauses per minute and every 100 syllables) and pronunciation errors. Similar to self-reported anxiety, more anxious participants produced shorter stretches of fluent speech (MLoR), paused more frequently during their presentations and made more pronunciation errors. Concerning pause frequency, this observation aligns with the findings of Mora et al. (2023). While using HR instead of EDA, the researchers found a weak association between HR and mid-clause pause frequency. This finding indicates that learners who experienced higher levels of anxiety produced more pauses both at and within AS-Units. However, the lack of alignment of physiological measures with other L2 speech measures suggests that physiological responses do not always translate to observable changes in speech performance. It is possible that subtle changes in HR and EDA occur when learners realize they have made a mistake, but the arousal may not be as pronounced as when learners need to pause during their speech for planning or word retrieval. However, future analyses could focus on specific moments during speech production to observe whether there is a slight increase in HR and EDA within a 5-second window.

Further results from hierarchical regression analyses revealed that L2 proficiency, cognitive factors (i.e., WM and attention), and anxiety played distinct roles in shaping L2 speaking performance, as their contributions varied across different aspects of speech. L2 proficiency consistently emerged as the strongest predictor of L2 speech performance, explaining variance in comprehensibility, accentedness, accuracy, lexical complexity, as well as speed and breakdown fluency. Hence, more proficient learners were more comprehensible, were perceived as less accented, exhibited fewer pauses, and made fewer errors. Cognitive factors emerged as the most significant predictor of external pausing behavior, both in terms of duration and frequency. This indicated that learners with better

WM and attention control were more capable of reducing external pauses, which are potentially caused by challenges in planning their speech.

Physiological measures of anxiety did not emerge as the strongest predictor of any L2 speech measure but made additional contributions to explaining variance in pronunciation errors, MLoR and pause frequency measures (especially silent internal and external pauses). The regressions for both physiological measures and subjective measures of anxiety converged in explaining the variance observed in pause frequency measures (both internal and external pauses). Unlike Pérez Castillejo (2019), this study did not find that FLA was the strongest predictor of L2 breakdown fluency during oral presentations. Instead, as in Bielak (2022) and Mora et al. (2023), the present study found L2 proficiency to be a much stronger predictor of L2 breakdown fluency. As in Bielak (2022), participants' proficiency levels might influence the impact of emotions like anxiety on their fluency. That is, when the L2 is more automatized, negative emotions such anxiety might have a comparatively less impact on L2 speech.

Although previous studies, such as Mora et al. (2023), found that anxiety accounted for a significant 13%–15% of additional variance in pause frequency, the percentages observed in this study were not as high. Yet, after controlling for L2 proficiency and cognitive factors, speaking anxiety was found to significantly explain an additional 1.6% – 4% of the variance in pause frequency. It is highly plausible that proficiency and cognitive factors helped to mitigate the effects of anxiety. In other words, higher L2 proficiency and better cognitive factors (i.e., WM and attention) compensated for the detrimental effect anxiety had in L2 speech production. This may explain why anxiety was not a strong predictor in the regression models but added an additional amount of variance. Moreover, as in Mora et al. (2023), anxiety contributed to pause frequency rather than pause duration, indicating that independently of individual differences in L2 proficiency and cognitive factors, higher levels of anxiety were mainly reflected in pause frequency measures.

Furthermore, self-reported anxiety was a significant predictor for words per ASU and TTR, suggesting that anxiety interfered with structural complexity and lexical variety. Although the percentage of the contributions were not high, subjective measures of anxiety also contributed to comprehensibility, accentedness, and speed fluency (SR and AR). These findings from both objective and subjective measures of anxiety indicate that, while L2 proficiency and cognitive factors play crucial roles, anxiety can significantly disrupt speech fluency, as evidenced by the increased frequency of pauses among learners. With regard to comprehensibility and accentedness, this study's findings coincide with previous studies. As found, previous literature reports subjective measures of anxiety significantly contributing to determine comprehensibility (Saito et al., 2018; Nagle et al., 2022) and accentedness scores (Suzukida & Saito, 2023). It is possible that anxiety might have interfered with learners' ability to focus on their ideas and output (Eysenck et al., 2007; Kormos, 2015), which hindered their focus on pronunciation, leading to more accented speech. This interference could have caused mental blocks, resulting in more pauses, reduced comprehensibility, and the inability to produce more complex sentences.

6.5. Sources of L2 speaking anxiety

The last section of the discussion focuses on learners' sources of L2 speaking anxiety (RQ5), and the sub-question of whether the sources are different when comparing highly and non-anxious learners. The present study revealed that in the stimulated recalls all participants experienced at least one anxiety episode, which were categorized into four groups: (1) performance-related concerns (e.g., being recording devices, topic difficulty), (2) linguistic-related concerns (e.g., grammar, pronunciation issues), (3) cognitive-related issues (e.g., mental blocks), and (4) audience-related fears (e.g., audience reactions). An important finding was that most sources of anxiety were internal (63.64%) rather than external factors (36.36%). This finding implies that in this study anxiety was more self-generated than caused by external stressors such as the audience.

Among these categories, participants mentioned cognitive-related issues most frequently, accounting for 56.92% of all the sources of anxiety. Participants often reported mental blocks (or going blank) (#S006), inability to recall ideas (#S001), words (#S053), or task instructions and prompts (#S006 and #S026), running out of ideas (#S050), or even feeling that their minds were not "activated" (#S053), which caused them distress. Moreover, these cognitive-related issues were identified by the participants to interfere with the learners' performance (#S042). As found in previous research such as Tóth (2009) and Aubrey (2022), participants in this study also mentioned recurringly being anxious due to experiencing problems with recalling words and expressing their ideas (e.g., *"I felt nervous because I <u>didn't know how to express myself, how to express my thoughts</u>. I <u>couldn't structure</u> it"* #S038). Although "problems to express oneself" could fit in the category of "linguistic-related concerns", participants did not clarify that this problem arise from lacking vocabulary or struggles with grammar, but difficulties from organizing their thoughts (#S038). Moreover, it was thought convenient to separate "problems to express themselves" from "mental blocks" because participants distinguished between knowing what to say but struggling when organizing their thoughts from not being to think of what to say (i.e., mental block).

The fact that many anxiety episodes originated from cognitive-related issues reinforces previous findings on how anxiety interferes with the attentional resource pools available for speaking, hindering learners' ability to focus on their own ideas and output (Eysenck et al., 2007; Kormos, 2015). As Eysenck et al. (2007) highlighted, anxiety induces disruptive thinking filled with worrisome and irrelevant thoughts, overloading the functioning of cognitive process and consuming the limited attentional resources. This WM overload and limited attentional resources explains why participants in this study experienced distress in response to their problems with recalling ideas, task instructions and words.

Cognitive-related issues were followed by linguistic-related concerns, even though the percentage was much lower (24.51%). In this category, the main source of anxiety was realizing they were making a mistake. Previous research has stated that L2 speaking anxiety is closely related to low self-perceptions of speaking ability (Gkonou, 2014) due to the self-exposure that is imposed on the learner (Gkonou, 2014; Price, 1991). Hence, it is understandable that, while performing an oral presentation, participants got anxious when realizing they were making a mistake. In fact, anxious learners tend to exhibit more concerns over accuracy (Tóth, 2017), and they are very sensitive to others' evaluations (Beatty, 1988; Gregersen, 2003; Kitano, 2001; Yan & Horwitz, 2008).

In this study, this is exemplified with participant #S050, who reflected on how anxious she was after making one mistake which "snowballed into more", and how she feared others perceived her as "dumb". As previously found in Young (1991), many learners struggle with the idea of being perceived as less competent than they are. In this case, being in the spotlight intensified her concerns about being judged over her language skills and her self-consciousness about her speech. Such worries not only increased her distress, but also induced more disruptive thoughts, creating a vicious circle that is very difficult to stop during an oral presentation. Not knowing the word in English was also repeatedly mentioned in this category as an L2 speaking anxiety source because participants felt they did not know the vocabulary needed for the task (#S042), and this disrupted the presentation flow (#S053). Other recurring linguistic concerns focused on grammar issues, pronunciation of words, L1 interference, sounding repetitive and realizing they were pausing excessively.

To a lesser extent, participants experienced distress due to performance-related concerns (8.86%) and audience-related fears (9.71%). Although less frequent, participants mentioned feeling distress due to being recorded, topic difficulty or unfamiliarity, time constraints and using English to perform the task. As reflected by #S050 and #S053, anxiety arose from unfamiliar experiences, such as being recorded or discussing unfamiliar topics. These sources are understandable, especially being recorded, to trigger distress for learners given they are not used to performing under such conditions. However, in this study, #S063 noted that while the topic was not difficult or unfamiliar to her, the fact that she had never explained it in English before caused her anxiety. This suggests that even when learners are familiar with a topic, the novelty of explaining it in the L2 can still provoke anxiety.

Participants in this study indicated they felt distress from audience-related fears such as the mere presence of an audience (#S050, #S053 and #S012), their reactions (#S004), fear of negative evaluation from them (#S050), their notetaking (#S026), and sharing something personal with them (#S012 and #S042). Previous research has also found that time limitations (Aubrey, 2022) and having an audience (Beatty & Behnke, 1991; Tóth, 2019) are sources of anxiety. While Aubrey's (2022) participants grew anxious due to approaching the time limit for completing the task, in this study most of participants pointed out feeling distress because they had to speak for a longer duration, especially when they had run out of ideas (#S005, #S012 and #S032). It was observed that there were very few cases of participants being anxious due to approaching the time limit for completing the task, as reflected by #S053.

As highlighted in previous literature, when learners are given the opportunity to prepare in advance, they are less anxious when performing (Aubrey 2022; Young, 1990; Choi, 2016). Thus, it is very likely that some linguistic and performance-related sources of anxiety would not have been present in this study if participants had had more planning time (e.g., *"the minute and a half to prepare was not enough for me in this topic so I was really, really anxious…"* #S050). Finally, it is interesting that unlike Tóth (2019), in this study the selected participants did not identify physical bodily reactions such as shaking hands and a pounding heart as a source of anxiety. Given that all the stimulated recalls were not analyzed, it is plausible that out of the 86 participants, there is an instance of physical body reactions as a source of anxiety. However, it is also possible that participants do not associate these reactions as a source of anxiety rather than an effect of it.

Further explorations were undertaken in RQ5.1 to compare whether highly and non-anxious learners differed in the sources of L2 speaking anxiety. Two main observations were that firstly, high-anxious learners reported more cognitive-related concerns than low-anxious learners. This finding aligns with observations by MacIntyre (2017) and Kormos (2015), who observed that mental blocks are a common effect of anxiety. Therefore, it is not unexpected that high-anxious participants commonly reported experiencing mental blocks. While the literature suggests that mental blocks are a consequence of anxiety, participants in this study indicated that experiencing mental blocks were one of the sources of their L2 speaking anxiety. This highlights a classic chicken-and-egg scenario, where it is very difficult to determine whether anxiety causes mental blocks or vice versa. Secondly, low-anxious learners expressed greater anxiety over linguistic issues like L1 interference, and pronunciation and vocabulary issues than their anxious counterparts. This finding does not go in line with previous research that has unveiled that high-anxious learners tend to exhibit more concerns over accuracy (Tóth, 2017) and resort to their native language more frequently (Gregersen, 2003) than low-anxious learners.

A possible explanation is that, in this study, given that low-anxious learners did not experience mental blocks frequently, their attention could be directed to linguistic concerns during the presentation. Hence, they were more aware of their linguistic performance, and their anxiety episodes were caused by the linguistic challenges encountered when performing, such as not knowing the word in English or pronunciation issues. On the other hand, as suggested by Kormos (2015), anxiety hinders anxious learners' ability to focus on their own output (e.g., realizing they were making a mistake), so their anxiety episodes are more related to cognitive reasons (e.g., mental block). For high-anxiety learners, anxiety overloading the functioning of cognitive processes hindered their ability to retrieve the necessary words or ideas (Eysenck et al., 2007). In fact, previous research by Gregersen (2003) and Sheen (2008) has also observed that anxious learners recognize fewer errors in their speeches and are less able to self-repair production errors.

Nevertheless, two linguistic-related concerns were more frequent for high-anxious participants: fear of sounding repetitive and pausing excessively. Previous research unveiled that anxious learners tend to use less varied vocabulary (Kormos &

Dörnyei, 2004; MacIntyre & Gardner, 1994b; Steinberg & Horwitz, 1986). Similarly, participants in this study became distressed when they noticed their lack of vocabulary variety during their performances. Regarding pauses, many studies have found a relationship between anxiety and fluency, revealing that anxious learners tend to pause more frequently and for longer (e.g., Aubrey, 2022; Bielak, 2022; Pérez Castillejo, 2019; Simard et al., 2023). Participants in this study indicated that pausing excessively during their speeches triggered their anxiety because it disrupted the flow of their performance and made them feel self-conscious and worried (e.g. "I started to notice that *I was stopping a lot* and saying the word 'basically' too much, so *I just kept thinking 'don't stop' and I got more nervous*" #S032). This self-awareness of pausing may have created a negative effect on speech, where the fear of pausing too much or for too long triggered their anxiety, leading to mental blocks and even more frequent pauses.

The differences between high- and low-anxious participants in performance-related factors and audience-related fears were not as notable as in the linguistic-related and cognitive-related concerns. However, it is interesting to highlight two aspects. Firstly, regarding performance-related factors, highly anxious learners more frequently reported feeling distress due to finding the topic difficult or unfamiliar, facing time-related challenges and having to speak for an extended period. The latter is connected to previous research findings, which have observed that when experiencing higher levels of anxiety, learners are less willing to communicate (MacIntyre, 2017; Steinberg & Horwitz, 1986; Tóth, 2011). Secondly, related to fear of negative evaluation, the mere audience presence and their reactions were mentioned more frequently by highly anxious participants.

Although it was expected that participants would generally experience distress due to fear of negative evaluation and the presence of an audience, it is interesting that low-anxious participants did not report this source of anxiety as often. A possible explanation could be related to personality traits. In other words, high-anxious learners in this study may also be introverts or exhibit tendencies toward perfectionism, which can shape their experiences with anxiety (Dewaele, 2002b; Gregersen & Horwitz, 2002; MacIntyre & Charos, 1996; Tóth 2011). For instance, Tóth (2011) found that anxious learners often exhibited perfectionist traits and reported that one of their main sources of anxiety was the fear of appearing less competent than their peers.

In this study, it seems that low-anxious participants were more confident and less worried about others' judgements. Instead, they identified sharing personal stories as a source of L2 speaking anxiety. This suggests that while low-anxious learners generally felt more confident, they were still susceptible to anxiety, for example, when retelling personal stories, possibly due to the emotional nature of such tasks. All in all, while high-anxious participants may have been more focused on external judgments, low-anxious learners were more sensitive to linguistic aspects of communication in the L2.

CHAPTER 7: Conclusions and Limitations

The final chapter summarizes the main findings of the study, provides pedagogical implications (Section 7.1), acknowledges the study's limitations and paves the way for further research on L2 speaking anxiety (Section 7.2). The chapter ends with some concluding remarks (Section 7.3).

7.1. Conclusions and pedagogical implications

The present study aimed to investigate five key areas related to L2 speaking anxiety: (1) the effects of task type, context and language; (2) the alignment between self-reported anxiety and others' perceptions, and the most prominent anxiety non-verbal cues; (3) the alignment between subjective and physiological measures; (4) anxiety effects on L2 speech; and (5) L2 speaking anxiety sources.

With this objective, 86 EFL learners performed a total of eight oral presentations under two presentation modes (one-to-one vs. group, *N*= 7) and two task types (descriptive vs. argumentative). Four of these presentations were performed in their L2 (English) and the remaining four in their L1. Focusing on L2 speaking anxiety, this dissertation included self-reports, physiological measures and others' perceptions. Moreover, IDs on proficiency, vocabulary size, WM and attention were controlled. The main findings and contributions of this study can be summarized as follows:

1. To the best of the author's knowledge, this is one of the first attempts to analyze together the effects of *task type, context,* and *language* on anxiety levels. Physiological measures were largely unaffected by *Task Type* and *Context*, except for HRV. HRV revealed that participants were more relaxed during descriptive tasks and in one-to-one contexts. *Language* influenced physiological measures suggesting that participants experienced faster beats per minute in the L2 performances. On the other hand, subjective measures showed significant effects of *Task Type, Context*, and *Language*. Subjective reports indicated greater anxiety levels during argumentative tasks, especially in group setting when performed in the L2.

2. This dissertation investigated the *alignment between self-reported anxiety and others' perceptions* as well as *non-verbal anxiety cues*, finding that participants'

self-assessments did not always match external evaluations. The disagreement scores were higher for the descriptive task, indicating that others perceived participants as more anxious than the participants reported. Contrary to "illusion of transparency" (King & Smith, 2017; Savitsky & Gilovich, 2003), this dissertation revealed that it was the audience members, particularly the helper, who perceived participants as more anxious than the participants themselves reported. Regarding non-verbal anxiety cues, the most frequently observed cue was avoiding eye contact. This suggests that judges were particularly sensitive to whether participants looked at them, interpreting the lack of eye contact as a sign of experiencing anxiety. This was followed by participants being "out of words" and participants moving their hand excessively, among many other non-verbal cues.

3. This research explored the *relationship between subjective and physiological measures of anxiety*, revealing weak significant correlations between subjective and physiological measures of L2 anxiety. Although most of previous findings suggest that anxiety experienced by learners does not always manifest in observable physiological changes during L2 speaking tasks (Gallego et al., 2021; Shachter et al., 2022; Schwerdtfeger, 2004; Sosa-López, 2019), this study has observed weak alignment between the measures. This indicates that more intensely anxiety-provoking tasks than standard oral presentations might be necessary to observe a stronger association.

4. This study assessed the *effects of anxiety on L2 speech*, showing that higher self-reported anxiety levels were negatively associated with fluency, comprehensibility, accentedness, accuracy and lexical variety. Physiological measures (i.e., HR and EDA) did not show significant correlations with L2 speech performance. Despite this detrimental effect of anxiety, as found in Bielak (2022) and Mora et al. (2023), proficiency emerged as the strongest predictor of L2 speech performance explaining variance in comprehensibility, accentedness, accuracy, lexical complexity, as well as speed and breakdown fluency. However, as in Mora et al. (2023), subjective measures of anxiety were found to significantly explain an additional 1.6% – 4% of the variance in pause frequency rather than pause duration, indicating that, independent of individual differences in L2

proficiency and cognitive factors, anxiety mainly affects the number of pauses learners make rather than their duration.

5. Regarding *sources of L2 speaking anxiety*, cognitive-related concerns were the most frequently reported sources of anxiety followed by linguistic-related concerns. To illustrate, mental blocks, including difficulties in recalling ideas, words, or instructions, were common experiences that caused distress among participants. In contrast, performance-related concerns and audience-related fears were less frequently reported. Additionally, this dissertation observed that high-anxiety learners were particularly worried by cognitive-related concerns, whereas low-anxiety learners were more focused on linguistic issues, such as pronunciation and vocabulary.

Taken together, the findings of this study suggest several important implications for pedagogical practice involving both learners and teachers. First, this thesis has shown the effect of task type, context, and language on subjective anxiety levels. More specifically the L2 argumentative tasks in the group audience setting can cause a lot of distress. Argumentative tasks can be more complex than other types since they require more linguistic resources. For example, when justifying a point of view, learners must effectively utilize a range of vocabulary and grammar structures to present their arguments clearly and persuasively. This involves not only articulating their own viewpoints but also engaging in critical thinking to address counterarguments.

Furthermore, empirical studies on FLA have shown that experiencing anxiety has a detrimental effect on L2 speech (e.g., increased pause frequency). One way to gradually help students feel less anxious and improve L2 performance is by familiarizing students with oral presentations through repeated practice. As the saying goes, "practice makes perfect"; thus, repeatedly practicing oral presentations (ideally of different task types) can build learners' confidence (Rubio-Alcalá, 2017) by reducing the fear of the unknown. This could potentially allow students to experience anxiety to a lesser degree, even if it is not fully mitigated. Educators should consider starting with low-stakes, informal speaking exercises, such as peer discussions, and gradually increasing the task difficulty to help students feel more at ease over time. Additionally, rather than high-pressure oral presentations, teachers could offer multiple modes of assessment, like recorded presentations, providing anxious students with opportunities to demonstrate their proficiency without excessive pressure.

This study has revealed many sources of L2 speaking anxiety. However, some of them could be prevented or mitigated by considering the following factors: providing enough planning time, allowing the use of outlines (e.g. in bullet points), and encouraging rehearsing (e.g., rehearsing in a smaller, less intimidating setting before presenting to a larger audience). In contrast to L2 classrooms in which teachers typically provide students with resources and preparation time before performing, this study was designed to give participants only one minute and a half to plan their presentations, without prior knowledge of the topic or the use of outlines. As a result, it is understandable that participants expressed distress due to cognitive-related and linguistic concerns which could have been avoided (e.g., not knowing the necessary vocabulary for the presentation or topic unfamiliarity).

On the other hand, anxiety episodes triggered by audience-related fears and cognitive-related issues (e.g., mental blocks and problems expressing ideas) can be more challenging to address. Teachers can support their students in coping with anxiety by introducing relaxation techniques such as deep breathing and positive visualization. These strategies can help learners calm themselves (both before or during speaking tasks), allowing them to approach tasks with greater confidence and resilience. Yet, anxiety may persist and emerge in high-pressure situations. While there is no instant solution to eliminate anxiety "once and for all", we can provide our students with techniques to manage it and take steps to minimize potential stressors through ongoing practice and gradual exposure to more demanding speaking tasks, helping to build confidence over time.

Another key aspect is feedback, which was not present in this study. In this study, the audience and the helper were instructed to behave as naturally as possible, but they were restricted from giving any feedback. However, this is not the case in L2 classrooms: teachers provide feedback either while performing or after the oral. As this study has observed, making mistakes can be a source of anxiety, especially for perfectionist learners. Previous literature considers it essential that teachers correct learners' mistakes while also highlighting their "signature strengths" (Oxford, 2017), as this approach avoids hindering students' enjoyment, motivation and willingness to communicate. By balancing positive reinforcement with effective anxiety management, teachers can create a more supportive learning environment that fosters both skill development and confidence in learners.

A supportive learning environment is necessary to overcome the fear of negative evaluation and mitigate anxiety. It is essential that teachers open communication and promote positive peer interactions. As outlined by Rubio-Alcalá (2017), when learners find themselves in a classroom environment where they do not feel accepted or connected to their peers, they will be prone to developing a lower self-concept and heightened FLA. Indeed, anxious students tend to be self-critical and overthink how others might perceive them, often worrying they will be seen as less competent (Beatty, 1988; Gregersen, 2003; Kitano, 2001; Yan & Horwitz, 2008; Young, 1990). Although this study confirms previous literature that self-assessments do not always align with external evaluations, it is natural for students to fear negative judgement and experience heightened anxiety as a result. Hence, to reduce the pressure of performance and encourage students to take risks without fear of failure, teachers could provide not only additional rehearsal time before presentations, but also implement peer support activities. By establishing small groups where students can practice together, share constructive feedback, and support one another, teachers can create a safe space and a collaborative environment. Moreover, as Savitsky & Gilovich (2003) observed, when the learners are explicitly informed about their negatively biased thoughts, they tend to self-evaluate their speeches more positively and be more composed.

Another important consideration is monitoring learners' anxiety. While this study included sensors of physiological arousal, most schools do not have access to such equipment. Although the use of smartwatches, which could allow students to monitor their HR, is becoming increasingly common, teachers can use instead quick and simple questionnaires or the idiodynamic task after speaking performances. The use of these tools is ideal to identify students' anxiety and develop strategies for improvement. However, the reality in schools is that even the use of the idiodynamic task may be avoided due to logistic constraints, such as a lack of computers or limited time.

Gathering students' insights is highly valuable not only for teachers but also for students themselves. It allows them to reflect on and identify potential anxiety triggers and take an active role in managing their anxiety (e.g., peer rehearsal, preparing outlines and visual aids preparation, breathing exercises). For teachers, this feedback provides valuable data to tailor instruction, offer targeted support, and create a learning environment that addresses both emotional and cognitive needs. Moreover, this approach can be used to foster intrapersonal and interpersonal competencies. In the former, through self-reflections, students work on self-awareness and reflect on their emotional responses. In the latter, by incorporating peer feedback, teachers promote a collaborative learning environment which can help overcome the fear of judgment by building on positive group dynamics, considering others' emotions, and managing feedback in a constructive way.

All in all, with these pedagogical recommendations, educators can better support L2 learners in alleviating anxiety and improving their overall L2 performance. By fostering a supportive learning environment and addressing emotional challenges, learners will be able to perform with less distress and, as a result, more effectively in real-world communication scenarios.

7.2. Limitations and future research

This dissertation is not exempt from limitations. By carefully assessing methodological limitations and external factors affecting data collection, we are hoping to contribute methodological aspects of researching L2 speaking fluency and to pave the way for further research of L2 speaking anxiety.

First, an important limitation lies in the fact that this study did not measure L2 speaking anxiety and L2 speech measures in a classroom. Although the data collection process resembled what L2 learners might go through in their classes (i.e., presenting in front of their peers and teacher in the L2), especially university students, this study did not replicate an authentic classroom dynamic that might intensify L2 speaking anxiety. Still, this study sets a precedent for examining the intricate relationship between anxiety and L2 speech performance. Classroom environments often include additional stressors such as real-time feedback and peer reactions (e.g., frowning, laughter), and the pressure of being graded by the teacher. These dynamics can increase anxiety and affect performance more than in semi-controlled experimental settings. Hence, future research could address this limitation by conducting studies in natural classroom settings to provide a more accurate representation of the challenges learners face in classroom settings with different contextual variables, such as teaching practices and peer interactions.

Another limitation in this study, related to the design and data collection processes, was to record two tasks per day. Over the course of two years, data was collected within a three-to-four-month window to ensure maximum participation.

However, considering the involvement of the helper and the audience members, the design included two tasks for each communicative context. This decision may have impacted participants, as they could have adapted after the first task, resulting in reduced anxiety during the second task. Moreover, given the similarities between each data collection day in terms of the procedures, this could have potentially obscured the effects of anxiety on performance, as familiarity with the setting and format might lead to less anxious responses in subsequent tasks.

With regards to the design, due to convenience, this study uses a cross-sectional design, which provides an insight into participants' anxiety and speech performance at a given moment of the participants' lives. Future research could adopt a longitudinal approach to better understand how affective factors, such as anxiety, evolve over time. By tracking participants across a year, either in a lab setting or real classroom performances, researchers could observe how anxiety, L2 speech, and other individual differences change with practice and exposure. This would allow for testing hypotheses on whether training or increased familiarity with speaking tasks reduces anxiety, improves speech performance, or shifts the relationship between affective factors and L2 outcomes. Furthermore, it would facilitate the development of more effective strategies for both teachers and learners to reduce FLA in the classroom (Oxford, 2017).

Moving on to the instruments, as outlined in *Section 6.3*, it is possible that tasks elicited other emotions, especially descriptive tasks as participants had to explain memorable memories or frustrating challenges. These additional emotional states may have influenced physiological responses in the L2. Although these types of tasks were chosen because they allowed participants to extend their presentations from three to five minutes, future research may consider more neutral topics or picture-based retelling tasks to minimize emotional interference and better isolate anxiety's effects on speech performance.

Following up with emotions, this study focused solely on anxiety experiences and did not control for other emotions, such as enjoyment (Dewaele & MacIntyre, 2014; MacIntyre et al., 2016; Simsek & Dörnyei2017) and willingness to communicate (MacIntyre & Charos, 1996), which could have provided valuable insights. Including these emotional factors could have offered a more comprehensive understanding of the participants' overall emotional experience during the tasks as well as making it possible to explore its possible influence on physiological responses. By controlling a broader range of emotions, future research could examine how negative and positive emotional states shape physiological responses and give a more nuanced understanding of how participants perceive the speaking tasks themselves, such as whether they found them engaging, challenging, or stressful. Future findings will help researchers to understand the extent to which emotional states influence language production.

Furthermore, future research using sensors to capture physiological arousal should consider paying attention to external factors such as participants' movements, room temperature and humidity. This study controlled for room temperature and explicitly informed the participants not to excessively move the hand on which the sensor was placed. As outlined in *Section 1.6.3.2*, EDA is especially sensitive to the location of the electrodes (Boucsein, 1992), as well as room temperature and humidity (Barry & Sokolov, 1993; Empatica support on EDA, n.d.; Roos et al., 2021). Unfortunately, humidity was not controlled in this study and Empatica E4 has its electrodes on the wrist, a place which does not contain as many eccrine sweat glands as in the palms and fingers (Boucsein, 1992; van Dooren & Janssen, 2012). Although Empatica E4 is a professional and reliable device (Borrego et al., 2019; Milstein & Gordon, 2020; Menghini et al., 2019; McCarthy et al., 2016; Schuurmans et al., 2020), it is not comparable to high-end medical equipment (e.g., an ECG), which provides precise measurements down to the millisecond level. Hence, future laboratory-based research may consider having more advanced devices (or different devices) that can record HR and EDA with greater precision and from more optimal body locations (e.g., HR from the chest and EDA from the fingers).

Future studies including physiological sensors should bear in mind a wide range of factors, including positive interest (Roos et al., 2021), food consumption, caffeine intake, lack of sleep and general fatigue (Rivers, 2022). Unfortunately, none of the previously mentioned factors were controlled in this study. Furthermore, physiological measures can be sensitive to a broad range of emotions rather than anxiety in isolation. If participants experienced emotions such as excitement from practicing in their L2 (the language in which they are studying their degree) or frustration when struggling to express their ideas in the L2, these emotions could have overlapped, making it harder to isolate the physiological signals specifically linked to anxiety. It was beyond the scope of this dissertation to focus on other emotions; however, while subjective measures specifically targeted anxiety, physiological measures might have captured a wider spectrum of emotional states over time. Hence, future studies should pay close attention to and control for factors such as participants' emotional states and physical well-being.

Following up with the sensors, provided that this study used means and standard deviations to analyze the data, these measurements do not capture the dynamicity of anxiety well. Thus, with this dissertation's data, a future study could take another approach to analysis focusing specifically on anxiety peaks and L2 speech. For instance, as in Gregersen et al. (2014), the next study could scrutinize in a one-minute window of L2 speech if a long pause or a mistake coincides with an anxiety peak in the idiodynamic task, HR and EDA for high-anxious learners. Indeed, scrutinizing individual cases and shorter time periods in which there is an anxious spike might avoid the emotional dips and spikes cancelling each other out when averaged. Furthermore, future research could include a dynamic assessment of comprehensibility and fluency, following those of Nagle et al. (2019) and Wood (2016). This would allow for a more fine-grained analysis of how these aspects of speech evolve in real time and whether they align with anxiety peaks. By tracking fluctuations in L2 speech (i.e., comprehensibility, accentedness, accuracy and fluency) in relation to anxiety levels, researchers could better understand the extent to which anxiety can disrupt L2 performance.

Continuing with the instruments, due to time constraints stimulated recalls could not be conducted as an extended interview, which could have been useful to further explore learners' experiences (Gass & Mackey, 2017). Future studies which use the idiodynamic task and stimulated recalls could follow MacIntyre (2012) and MacIntyre & Ducker (2022) by printing the resulting graph of the participant's ratings and using it as a reference during the stimulated recall. By integrating this visual output into the recall session, researchers could focus on specific moments while participants reflect on their fluctuating anxiety levels and even other emotions. Moreover, it would be interesting to include a pre-performance questionnaire with Likert-scale questions on affective factors right before performing the task. This would give the researchers valuable insights into participants' emotional states before the task and assess how anxiety, confidence, and other emotions fluctuate right before and after task performance. To conclude on instruments, this study initially included pronunciation anxiety (Baran-Łucarz, 2014) in the LBQ, which later was later removed as it was deemed beyond the focus of this research. However, future studies could further explore this type of anxiety. For instance, using the data from this dissertation, it would be possible to correlate pronunciation anxiety LBQ scores with pronunciation accuracy, offering deeper insights into how anxiety specifically related to pronunciation impacts L2 performance.

Turning to the results, many of the correlations should be interpreted with caution due to their weakness. Nevertheless, the findings in this dissertation highlight promising directions for future exploration, particularly regarding the alignment between 1) self-reports and others' perspectives, 2) subjective and physiological measures of anxiety and 3) L2 speaking anxiety and L2 speech effects. However, further research is always needed to validate these findings, explore these connections more deeply, and refine our understanding of how anxiety manifests and impacts L2 learners in various contexts.

Concerning the alignment between self- and others' assessments of L2 speaking anxiety, the current study obtained the others' assessments data from Likert-scale questions. It was beyond the scope of this study to adopt Gregersen et al.'s (2017) procedure and make either audience members at a later stage or other collaborators use the idiodynamic task to assess participants' L2 speaking anxiety. Hence, in the future, it would be very interesting to use the participants' video recordings with the idiodynamic task to compare whether teachers, non-practitioners and L2 learners peer (e.g., L1 Spanish or other nationalities) ratings of anxiety align with participants' ratings. Such a comparison could provide valuable insights into how different groups perceive anxiety in L2 speakers and further expand research on these possible assessments' alignments.

This project used a convenience sample (i.e., first-year university students), so learners were similar ages. Further research should examine how the results of this study can be extrapolated to other learner populations from different settings (e.g., EFL classes, immersive contexts), different ages (e.g., younger and older students, > 25 years old) and different proficiency levels (e.g., beginners and advanced students). As outlined in *Section 1.2.3*, socio-biographical factors have yielded mixed results, and more research needs to be conducted to be able to generalize these findings. Moreover, most research on anxiety focuses on L2 learners, but the scope could be broadened to language teacher anxiety in L2 classrooms (Gkonou et al., 2017; Goetze, 2013; Horwitz, 1996) since their experience of language anxiety can affect their teaching performance, classroom dynamics, and the overall learning environment.

7.3. Concluding remarks

Affective factors such as anxiety are prevalent in our L2 classrooms and require our attention and intervention. The present investigation has been a first attempt at examining the detrimental effects of anxiety on L2 speech during oral presentations of different types (descriptive vs. argumentative) and under different contexts (one-to-one vs. group audience). Additionally, this study incorporated self-reports of anxiety alongside physiological measures and external evaluations, explored the sources of L2 speaking anxiety and identified the most prominent non-verbal cues observed during these performances. Therefore, it provides a thorough analysis of L2 speaking anxiety, integrating multiple perspectives and measures. On one hand, the current findings add to a growing body of literature on FLA (Horwtiz et al., 1986; MacIntyre, 2017), demonstrating the detrimental effect on L2 speech, particularly in relation to fluency (Bielak, 2022; Mora et al., 2023; Szyszka & Lintunen, 2023), identifying sources (Aubrey, 2022; Tóth, 2009) and non-verbal cues of L2 speaking anxiety (Gregersen, 2005; Gregersen, 2007b; Gregersen et al., 2017). On the other hand, this study contradicts previous findings on the illusion of transparency (King & Smith, 2017; Savitsky & Gilovich, 2003) and the alignment between subjective and objective measures of anxiety (Gallego et al., 2021), and contributed additional evidence on the effects of task type (Sosa-López & Mora, 2022a; Tóth, 2017), context (Beatty & Behnke, 1991) and language on L2 speaking anxiety. However, many questions in FLA research still need to be explored empirically so effective strategies can be developed to mitigate anxiety in language learning contexts.

8. References

Aida, Y. 1994. Examination of Horwitz, Horwitz, and Cope's Construct of Foreign Language Anxiety: The Case of Students of Japanese. *Modern Language Journal*, *78*(2), pp. 155-167. DOI:10.2307/329005

Alpert, R. & R. Haber. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, *61*, pp. 207- 215. https://doi.org/10.1037/h0045464

Arnhold, Anja. (2014). “Anja Arnhold's Praat scripts” https://sites.ualberta.ca/~arnhold/scripts/TextgridsOnly/Print_all_intervals.prat

Arbuthnott, K., & Frank, J. (2000). Trail Making Test, Part B as a measure of executive control: Validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology*, *22*, pp. 518–528. DOI: 10.1076/1380-3395(200008)22:4;1-0;FT518

Arnaiz, P. & Guillén, F. (2012). Foreign language anxiety in a Spanish university setting: Interpersonal differences. *Revista de Psicodidáctica*, *17*(1), pp. 5-26. https://ojs.ehu.eus/index.php/psicodidactica

Arnaiz, P. & Pérez-Luzardo, J. (2014). Anxiety in Spanish EFL university lessons: Causes, responsibility attribution and coping. *Studia Anglica Posnaniensia, 49*(1), pp. 57-76. DOI:10.2478/stap-2014-0003

Aubrey, S. (2022). The relationship between anxiety, enjoyment, and breakdown fluency during second language speaking tasks: An idiodynamic investigation. *Frontiers in Psychology, 13*, pp. 1-17. DOI: 10.3389/fpsyg.2022.968946

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders, 36*(3), pp. 189-208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Bailey, K. M. (1983). Competitiveness and Anxiety in Adult Second Language Learning: Looking at and through the Diary Studies. In H.W. Seliger and M.H. Long (Eds.), *Classroom Oriented Research in Second Language Acquisition*, pp. 67-103. Rowley, Mass.: Newbury House. https://files.eric.ed.gov/fulltext/ED367166.pdf

Bailey, P., Onwuegbuzie, A. J., & Daley, Ch. E. (2000). Correlates of anxiety at three stages of the foreign language learning process. *Journal of Language and Social Psychology, 19*, pp. 474 – 490. https://doi.org/10.1177/0261927X00019004005

Baran-Łucarz, M. (2011). The relationship between language anxiety and the actual and perceived levels of foreign language pronunciation. *Studies in Second Language Learning and Teaching (SSLLT)*, *1*(4), pp. 491-514. DOI: 10.14746/SSLLT.2011.1.4.3

Baran-Łucarz, M. (2013). Phonetics learning anxiety: Results of a preliminary study. *Research in Language*, *11*(1), pp. 57–79. DOI: https://doi.org/10.2478/v10015-012-0005-9

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign- language classroom: the Polish EFL context. *The Canadian Modern Language Review*, *70*(4), pp. 445-473. DOI: https://doi.org/10.3138/cmlr.2666

Baran-Łucarz, M. (2016). Conceptualizing and measuring the construct of pronunciation anxiety: Results of a pilot study. In: Pawlak M. (eds) *Classroom-Oriented Research. Second Language Learning and Teaching* (pp. 39-56). Springer, Cham. DOI:10.1007/978-3-319-30373-4_3

Baran-Łucarz, M. (2017). FL pronunciation anxiety and motivation: Results of a preliminary mixed-method study. In E. Szymańska-Czaplak, M. Szyszka & E. Kuciel-Piechurska (Eds.), *At the Crossroads: Challenges in FL Learning* (pp. 107–133). Heidelberg & New York: Springer.

Baran-Łucarz, M., & Lee, J. H. (2021). Selected Determinants of Pronunciation Anxiety. *International Journal of English Studies*, *21*(1), 93-113. DOI: https://doi.org/10.6018/ijes.426411

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, *10*(1), pp. 20- 46. DOI: 10.1207/s15327957pspr1001_2

Barry, R. J., & Sokolov, E. N. (1993). Habituation of phasic and tonic components of the orienting reflex. *International Journal of Psychophysiology*, *15*(1), pp. 39-42. DOI: 10.1016/0167-8760(93)90093-5

Bassett, R., Behnke, R. R., Carlile, L. W., & Rogers, J. (1973). The effects of positive and negative audience responses on the autonomic arousal of student speakers. *The Southern Speech Communication Journal*, *38*, pp. 255-261. DOI: https://doi.org/10.1080/10417947309372195

Beatty, M. J. (1988). Situational and predispositional correlates of public speaking anxiety. *Communication Education*, *37*(1), pp. 28-39. DOI: ttps://doi.org/10.1080/03634528809378701

Beatty, M. J & Behnke, R. R. (1991). Effects of public speaking trait anxiety and intensity of speaking task on heart rate during performance. *Human Communication Research*, *18*(2), pp. 147-176. DOI: https://doi.org/10.1111/j.1468-2958.1991.tb00542.x

Behnke, R. R. & Beatty, M. J. (1981). A cognitive-physiological model of speech anxiety. *Communication Monographs*, *48*, pp. 158-163. DOI: https://doi.org/10.1080/03637758109376055

Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of accentedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals*, *50*(3), pp. 547-566. DOI: https://doi.org/10.1111/flan.12285

Bielak, J. (2022). To what extent are foreign language anxiety and foreign language enjoyment related to L2 fluency? An investigation of task-specific emotions and breakdown and speed fluency in an oral task. *Language Teaching Research*, pp. 1-31. https://doi.org/10.1177/13621688221079319

Birkett, M. A. (2011). The trier social stress test protocol for inducing psychological stress. *Journal of Visualized Experiments* (56), pp. 32-38. DOI: 10.3791/3238

Blasco, J., & Peris-Lopez, P. (2018). On the feasibility of low-cost wearable sensors for multi-modal biometric verification. *Sensors*, *18*(9), pp. 1- 20. DOI: https://doi.org/10.3390/s18092782

Bloom, B. (1953). Thought processes in lectures and discussions. *Journal of General Education 7*(3), pp. 160-169. https://www.jstor.org/stable/27795429

Bodie, G. D. (2010). A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety. *Communication education, 59*(1), pp. 70-105. https://doi.org/10.1080/03634520903443849

Boersma, P. & Weenink, D. (2020). *Praat*: doing phonetics by computer [Computer program]. Version 6.1.09, retrieved from https://www.fon.hum.uva.nl/praat/

Booth-Butterfield, S. (1987). Action assembly theory and communication apprehension: A psychophysiological study. *Human Communication Research*, *13*, pp. 386-398. DOI: https://doi.org/10.1111/j.1468-2958.1987.tb00111.x

Borrego, A., Latorre, J., Alcañiz, M., & Llorens, R. (2019, July). Reliability of the Empatica E4 wristband to measure electrodermal activity to emotional stimuli. In 2019 *International Conference on Virtual Rehabilitation (ICVR)* (pp. 1-2). IEEE. DOI: 10.1109/ICVR46560.2019.8994546

Botes, E., Dewaele, J. M., & Greiff, S. (2020). The foreign language classroom anxiety scale and academic achievement: an overview of the prevailing literature and a meta-analysis. *Journal for the Psychology of Language Learning*, *2*(1), pp. 26-56. https://jpll.org/index.php/journal/article/view/botesetal

Botes, E., Dewaele, J.-M., Greiff, S., & Goetz, T. (2024). Can personality predict foreign language classroom emotions? The devil's in the detail. *Studies in Second Language Acquisition*, *46*(1), pp. 51–74. DOI:10.1017/S0272263123000153

Botes, E., Van der Westhuizen, L., Dewaele, J. M., MacIntyre, P., & Greiff, S. (2022). Validating the short-form foreign language classroom anxiety scale. *Applied Linguistics*, *43*(5), pp. 1006-1033. https://doi.org/10.1093/applin/amac018

Boucsein, W. (1992). *Electrodermal activity*. Springer Science & Business Media. New York: Springer.

Boucsein, W. (1999). Electrodermal activity as an indicator of emotional processes. *Science of Emotion and Sensibility*, *2*(1), pp. 1-25. https://koreascience.kr/article/JAKO199911919730230.page

Boudreau, C., MacIntyre, P. D., & Dewaele, J.-M. (2018). Enjoyment and anxiety in second language communication: An idiodynamic approach. *Studies in Second Language Learning and Teaching*, *8*, pp. 149–970. DOI: https://doi.org/10.14746/ssllt.2018.8.1.7

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106* (2), pp. 707-729. DOI: 10.1016/j.cognition.2007.04.005

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, *49*(1), pp. 1017-1034. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/psychology/ABriefGuideforAnalysingElectrodermalActivityv4.pdf

Bronfenbrenner, U. (1993). The ecology of cognitive development: Research models and fugitive findings. In R. H. Wozniak, & K. W. Fischer, (Eds.), *Development in context: Acting and thinking in specific environments* (pp. 3–44). Hillsdale: Erlbaum.

Buchanan, T.W., Laures-Goreb, J. S. & Duff, M. C. (2014). Acute stress reduces speech fluency. *Biological Psychology*, *97*, pp. 60- 66. DOI: 10.1016/j.biopsycho.2014.02.005

Çagatay, S. (2015). Examining EFL students' Foreign Language Speaking Anxiety: the case at a Turkish state university. *Procedia - Social and Behavioral Sciences*, *199*(3), pp. 648-656. DOI: https://doi.org/10.1016/j.sbspro.2015.07.594

Carroll, J. B. & Sapon, S. (1959). *Modern languages aptitude*. *Test-Form A*. New York: Psychological Corporation.

Chastain, K. (1975). Affective and ability factors in second language acquisition. *Language Learning*, *25*, pp. 153-161. DOI: https://doi.org/10.1111/j.1467-1770.1975.tb00115.x

Cheng, Y. S. (2002). Factors associated with foreign language writing anxiety. *Foreign Language Annuals*, *35*(5), pp. 647-56. DOI: https://doi.org/10.1111/j.1944-9720.2002.tb01903.x

Cheng, Y. S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, *13*, pp. 313-335. https://doi.org/10.1016/j.jslw.2004.07.001

Cheng, Y. S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language learning*, *49*(3), 417-446. https://doi.org/10.1111/0023-8333.00095

Choi, J. (2016). English speaking classroom apprehension: A study of the perceptions held by Hong Kong university students. *The Journal of Teaching English for Specific and Academic Purposes*, *4*(2), pp. 293-308. http://espeap.junis.ni.ac.rs/index.php/espeap/article/view/275

Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology*, *70*(3), pp. 451–460. DOI: 10.1037//0022-3514.70.3.451

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), pp. 73-111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clément, R., Gardner, R. C., & Smythe, P. C. (1977). Motivational variables in second language acquisition: A study of Francophones learning English. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 9*(2), pp. 123–133. https://doi.org/10.1037/h0081614

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, pp. 769-786. https://doi.org/10.3758/BF03196772

Corrigan, J., & Hinkeldey, N. (1987). Relationships between Parts A and B of the Trail Making Test. *Journal of Clinical Psychology, 43*, pp. 402–409. https://doi.org/10.1002/1097-4679(198707)43:4<402::AID JCLP2270430411>3.0.CO;2-E

Coulombe, D. (2000). Anxiety and beliefs of French-as-a-second-language learners at the university level. Unpublished doctoral dissertation, University of Laval, Quebec, Canada. https://corpus.ulaval.ca/jspui/handle/20.500.11794/40683

Croft, R. J., Gonsalvez, C. J., Gander, J., Lechem, L., & Barry, R. J. (2004). Differential relations between heart rate and skin conductance, and public speaking anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *35*(3), pp. 259-271. DOI: 10.1016/j.jbtep.2004.04.012

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *The Modern Language Journal*, *99* (1), pp. 80-95. DOI: https://doi.org/10.1111/modl.12185

Csíkszentmihályi, M. (2008) *Flow: The Psychology of Optimal Experience*. New York: Harper.

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences, 40*, pp. 63-72. https://doi.org/10.1016/j.lindif.2015.04.005

Denley, P., & Bishop, K. (2010). The potential of using stimulated recall approaches to explore teacher thinking. *Using Analytical Frameworks for Classroom Research: Collecting Data and Analysing Narrative*. Abingdon: Routledge, pp. 109-124.

De Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics, 13*, pp. 1–24.

De Jong, N. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching (IRAL), 54*(2), pp. 113–132. DOI: https://doi.org/10.1515/iral-2016-9993

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS)* (pp. 17-20). Retrieved from: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_1900232

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36*(2), pp. 223–243. DOI:10.1017/S0142716413000210

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34*(1), pp. 5-34. DOI: 10.1017/S0272263111000489

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), pp. 893-916. DOI: 10.1017/S0142716412000069

De Jong, N. & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), pp. 385-390. https://doi.org/10.3758/BRM.41.2.385

Derwing, T. M., Rossiter, M. J. Munro, M. J. & Thomson, R. I. (2004). Second Language Fluency: Judgments on different tasks. *Language Learning*, *54*(4), pp. 655–679. DOI: https://doi.org/10.1111/j.1467-9922.2004.00282.x

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language teaching*, *42*(4), pp. 476. DOI:10.1017/S026144480800551X

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*, pp. 533– 557. DOI:10.1017/S0272263109990015

Dewaele, J.-M. (2007). The effect of multilingualism and socio-situational factors on communicative anxiety and foreign language anxiety of mature language learners. *The International Journal of Bilingualism, 11*(4), pp. 391– 410. DOI:10.1177/13670069070110040301

Dewaele, J.-M. (2002a). Individual Differences in L2 Fluency: The Effect of Neurobiological Correlates. In V. Cook (Ed.), *Portraits of the L2 User* (pp. 219-250). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi-org./10.21832/9781853595851-011

Dewaele, J.-M. (2002b). Psychological and sociodemographic correlates of communicative anxiety in L2 and L3 production. International Journal of *Bilingualism, 6*(1), pp. 23-38.https://doi.org/10.1177/13670069020060010201

Dewaele, J.-M. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi-and multilinguals. *The Modern Language Journal, 97*(3), pp. 670-684. https://doi.org/10.1111/j.1540-4781.2013.12036.x

Dewaele, J.-M. (2017). Are perfectionists more anxious foreign language learners and users. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 70-90). Bristol, UK: Multilingual Matters.

Dewaele, J. M., & Al-Saraj, T. (2015). Foreign language classroom anxiety of Arab learners of English: The effect of personality, linguistic and sociobiographical variables. *Studies in Second Language Learning and Teaching, 5*(2), pp. 205-228. Available at: https://www.ceeol.com/search/article-detail?id=297379

Dewaele, J.-M., Botes, E., & Greiff, S. (2023). Sources and effects of foreign language enjoyment, anxiety, and boredom: A structural equation modeling approach. *Studies in Second Language Acquisition, 45*(2), pp. 461–479. DOI:10.1017/S0272263122000328

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in second language learning and teaching, 4*(2), pp. 237-274. https://doi.org/10.14746/ssllt.2014.4.2.5

Dewaele, J.-M., MacIntyre, P. D., Boudreau, C., & Dewaele, L. (2016). Do girls have all the fun? Anxiety and enjoyment in the foreign language classroom. *Theory and practice of second language acquisition, 2*(1), pp. 41-63. Retrieved from https://journals.us.edu.pl/index.php/TAPSLA/article/view/3941

Dewaele J.-M., Petrides, K.V., Furnham, A. (2008). Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning, 58*(4), pp. 911–960 912. DOI:10.1111/j.1467-9922.2008.00482.x

Dewaele, J.-M., & Dewaele, L. (2017). The dynamic interactions in Foreign Language Classroom Anxiety and Foreign Language Enjoyment of pupils Aged 12 to 18. A Pseudo-longitudinal Investigation. *Journal of the European Second Language Association, 1*(1), pp. 12–22. DOI:10.22599/jesla.6

Donovan, L. A., & MacIntyre, P. D. (2005). Age and sex differences in willingness to communicate, communication apprehension and self-perceived competence. *Communication Research Reports, 21*, pp. 420– 427. DOI:10.1080/08824090409360006

Dewaele, J.-M. & Meftah, R. (2023). The Trajectory of English Foreign Learners' Emotions and Motivation from the Start to the End of their Learning Journey: A Pseudo-Longitudinal Investigation. *Journal of the European Second Language Association, 7* (1), pp. 16-30. https://doi.org/10.22599/ jesla.94

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends, 66,* pp. 810–824. https://doi.org/10.1007/s11528-022-00717-w

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. New Jersey: Mahwah: L. Erlbaum.

Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language learning, 59*, pp. 230-248. https://doi.org/10.1111/j.1467-9922.2009.00542.x

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research, 4*(3), pp. 275-300. https://doi-org.sire.ub.edu/10.1177/136216880000400305

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in Second Language research: Construction, Administration, and Processing* (2nd Ed.). Routledge.

Draine, S. (1998). *Inquisit* [Computer software]. Seattle, WA: Millisecond Software.

EDA Explorer. (2015). Algorithm and website created by Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., and Picard, R. Available online at < https://eda-explorer.media.mit.edu/ >

Ellis, R. (2008). *The study of second language acquisition*. 2nd edition. Oxford: Oxford University Press.

Ely, C.M. (1986). An analysis of discomfort, risk-taking, sociability, and motivation in the L2 classroom. *Language Learning*, *36*(1), pp. 1-25. DOI:10.1111/j.1467-1770.1986.tb00366.x

Empatica E4 wristband user's manual (n.d.). Available online at https://empatica.app.box.com/v/E4-User- Manual

Empatica Support (n.d.). "Issues with EDA Values". https://support.empatica.com/hc/en-us/articles/204676479-Issues-with-EDA-values

Empatica Support (n.d.). "E4 data - IBI expected signal". https://support.empatica.com/hc/en-us/articles/360030058011-E4-data-IBI-expected-signal

Empatica Support (n.d.). "E4 data - EDA expected signal". https://support.empatica.com/hc/en-us/articles/360030048131-E4-data-EDA-Expected-signal

Empatica Support (n.d.). "What should I know to use EDA data in my experiment?". https://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment

Electrocardiography (ECG). The Complete Pocket Guide. (2020) iMotions. Available online at < https://imotions.com/guides/>

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, *89*(2), pp. 206-220. DOI: https://doi.org/10.1111/j.1540-4781.2005.00275.x

Ewald, J. D. (2007). Foreign language learning anxiety in upper-level classes: Involving students as researchers. *Foreign Language Annals*, *40*, pp. 122-142. DOI: https://doi.org/10.1111/j.1944-9720.2007.tb02857.x

Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality,* *13*(4), pp. 363-385. DOI: https://doi.org/10.1016/0092-6566(79)90001-1

Eysenck et al., 2007

Felnhofer, A., Kothgassner, O. D., Hetterle, T., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2014). Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task.

Cyberpsychology, Behavior, and Social Networking, 17(5), pp. 310-316. https://doi.org/10.1089/cyber.2013.0472

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America, 76*(3), pp. 692–707. https://doi.org/10.1121/1.391256

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, *21*(3), pp. 354-375. DOI: https://doi.org/10.1093/applin/21.3.354

Frantzen, D., & Magnan, S. S. (2005). Anxiety and the true beginner-false beginner dynamic in beginning French and Spanish classes. *Foreign Language Annals*, *38*, pp. 171–190. DOI: https://doi.org/10.1111/j.1944-9720.2005.tb02483.x

Gallardo del Puerto, F., García Lecumberri, M. L., & Gómez Lacabex, E. (2015). The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics, 25*(2), pp. 202-224. https://doi-org.sire.ub.edu/10.1111/ijal.12063

Gallego, A., McHugh, L., Penttonen, M., & Lappalainen, R. (2021). Measuring Public Speaking Anxiety: Self-report, behavioral, and physiological. *Behavior Modification*, pp. 1 - 17. https://doi.org/10.1177/0145445521994308

Galvanic Skin Response. The Complete Pocket Guide. (2020) iMotions. Available online at < https://imotions.com/guides/>

Garbarino, M., Lai, M., Bender, D., Picard, R. W., & Tognetti, S. (2014). Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)* (pp. 39-42). IEEE. DOI: 10.1109/MOBIHEALTH.2014.7015904

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitude and motivation*. London, United Kingdom: Edward Arnold.

Gardner, R. C., & Smythe, P. C. (1975). Motivation and second-language acquisition. *Canadian Modern Language Review, 31*(3), pp. 218-233. https://doi.org/10.3138/cmlr.31.3.218

Gardner, R. C., Smythe, P. C., Clément, R., & Gliksman, L. (1976). Social and psychological factors in second language acquisition. *Canadian Modern*

Language Review/ La Revue canadienne des langues vivantes, 32(3), pp. 198–213. DOI:10.3138/cmlr.32.3.198

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal, 81*(3), pp. 344–362. https://doi.org/10.1111/j.1540-4781.1997.tb05495.x

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: an overview. *AILA Review, 19*, pp. 3-17. DOI: https://doi.org/10.1075/aila.19.03gas

Gass, S. M., & Mackey, A. (2013). *Stimulated recall methodology in second language research*. Routledge. ISBN: 0805832246

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, *34*(1), pp. 65-87. DOI: https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Gkonou, C. (2013). A diary study on the causes of English language classroom anxiety. *International Journal of English Studies, 13*(1), pp. 51–68. https://doi.org/10.6018/ijes/2013/1/134681

Gkonou, C. (2014). The sociolinguistic parameters of L2 speaking anxiety. In M. Pawlak et al. (Eds.), *Classroom-oriented Research* (pp. 15-32). Switzerland: Second Language Learning and Teaching. Springer International Publishing. DOI: 10.1007/978-3-319-00188-3_2

Gkonou, C. (2017). Towards an ecological understanding of language anxiety. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 135-155). Bristol, UK: Multilingual Matters.

Gkonou, C., Jean-Marc Dewaele, J-M. & Daubney, M. (2017). Conclusion. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 217-223). Bristol, UK: Multilingual Matters.

GSR logger sensor NUL-217 (n.d.). *Neulog.* Retrieved from <https://neulog.com/gsr/#specs>

Goberman, A. M., Hughes, S. & Haydock, T. (2011). Acoustic characteristics of public speaking: Anxiety and practice effects. *Speech Communication*, *53*(6), pp. 867–876. DOI: https://doi.org/10.1016/j.specom.2011.02.005

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), pp. 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Granena, G. (2016). Elicited imitation as a measure of implicit L2 knowledge. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 185-204). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*(3), pp. 311–343. DOI: https://doi.org/10.1177/0267658312461497

Granena, G., Jackson, D.O. & Yilmaz, Y. (2016). Introduction: Cognitive individual differences in second language learning and processing. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 1-14). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Gregersen, T. S. (2003). To err is human: A reminder to teachers of language-anxious students. *Foreign Language Annals, 36* (1), pp. 25-32. DOI: https://doi.org/10.1111/j.1944-9720.2003.tb01929.x

Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals, 38*(3), pp. 388-400. DOI:10.1111/j.1944-9720.2005.tb02225.x

Gregersen, T. S. (2007a). Breaking the code of silence: A study of teachers' nonverbal decoding accuracy of foreign language anxiety. *Language Teaching Research, 11*(2), pp. 209-221. DOI:10.1177/1362168807074607

Gregersen, T. S. (2007b). Language learning beyond words: Incorporating body language into classroom activities. *Reflections on English Language Teaching, 6*(1), pp. 51-64. Retrieved from <http://www.nus.edu.sg/celc/research/books/relt/vol6/no1/51-64gregersen.pdf>

Gregersen, T. S. (2009). Recognising visual and auditory cues in the detection of foreign language anxiety. *TESL Canada Journal, 26*(2), pp. 46-64. DOI:10.18806/tesl.v26i2.414

Gregersen, T. (2020). Dynamic properties of language anxiety. Studies in Second *Language Learning and Teaching, 10*(1), pp. 67-87. http://dx.doi.org/10.14746/ssllt.2020.10.1.4

Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal*, *86*(4), pp. 562-570. DOI: https://doi.org/10.1111/1540-4781.00161

Gregersen, T., MacIntyre, P. D., & Meza, M. D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *The Modern Language Journal*, *98*(2), pp. 574 -588. DOI: https://doi.org/10.1111/modl.12084

Gregersen, T. S., MacIntyre, P. D., & Olson, T. (2017). Do you see what I feel? An idiodynamic assessment of expert and peer's reading of nonverbal language anxiety cues. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 110-134). Bristol, UK: Multilingual Matters.

Harris, S. R., Kemmerling, R. L., & North, M. M. (2002). Brief virtual reality therapy for public speaking anxiety. *CyberPsychology & Behavior*, *5*(6), pp. 543–550. DOI: 10.1089/109493102321018187

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P.G.M., Neerincx, M.A. & Brinkman, W-P. (2014). Controlling social stress in virtual reality environments. *PLoS ONE*, *9*(3), pp. 1-17. DOI: https://doi.org/10.1371/journal.pone.0092804

Heart Rate & Pulse Logger Sensor NUL-208. (n.d.). *Neulog*. Retrieved from NeuLog <https://neulog.com/heart-rate-pulse/>

Held, J., Vîslă, A., Wolfer, C., Messerli-Bürgy, N., & Flückiger, C. (2021). Heart rate variability change during a stressful cognitive task in individuals with anxiety and control participants. *BMC psychology*, *9:44*, pp. 1-8. https://doi.org/10.1186/s40359-021-00551-4

Hewitt, E., & Stephenson, J. (2012). Foreign Language Anxiety and oral exam performance: A replication of Phillips's MLJ study. *The Modern Language Journal*, *96*, pp. 170–189. DOI: https://doi.org/10.1111/j.1540-4781.2011.01174.x

Horwitz, E. K. (1986). Preliminary Evidence for the Reliability and Validity of a Foreign Language Anxiety Scale. *TESOL Quarterly*, *20*(3), pp. 559–562. https://doi.org/10.2307/3586302

Horwitz, E. K. (1996). Even teachers get the blues: Recognizing and alleviating language teachers' feelings of foreign language anxiety. *Foreign Language Annals, 29*(3), pp. 365–372. https://doi.org/10.1111/j.1944-9720.1996.tb01248.x

Horwitz, E. K. (2000). It ain't over 'til it's over: On foreign language anxiety, first language deficits and the confounding of variables. *The Modern Language Journal*, *84,* pp. 256 – 259. DOI: https://doi.org/10.1111/0026-7902.00067

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21,* pp. 112-126. DOI:10.1017/S0267190501000071

Horwitz, E. K. (2010). Research timeline. Foreign and second language anxiety. *Language Teaching, 43* (2), pp. 154–167. DOI: 10.1017/S026144480999036X

Horwitz, E. K. (2017). On the Misreading of Horwitz, Horwitz and Cope (1986) and the Need to Balance Anxiety Research and the Experiences of Anxious Language Learners. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31-47). Bristol, UK: Multilingual Matters

Horwitz, E. K., Horwitz, M. B. & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, *70*(2), pp. 125–132. DOI: https://doi.org/10.2307/327317

Horwitz, E. K., Tallon, M., & Luo, H. (2009). Foreign language anxiety. In Cassady, Jerrell, C. (Ed.), *Anxiety in schools: the causes, consequences, and solutions for academic anxiety* (pp. 97-115). New York: Peter Lang Publishing.

Inquisit Visual Digit Span Test User Manual. (2022). Millisecond Software, LLC. <https://www.millisecond.com/download/library/v6/digitspan/digitspan_visual/digitspanvisual/digitspanvisual.manual>

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), pp. 135-159. DOI:10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2011). Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics, 32*(1), pp. 113-140. DOI:10.1017/S0142716410000317

Jee, M. J. (2014). Affective factors in Korean as a Foreign Language: Anxiety and beliefs. *Language, Culture and Curriculum*, *27*(2), pp. 182-195. DOI: 10.1080/07908318.2014.918626

Jin, Y. X., & Dewaele, J. M. (2018). The effect of positive orientation and perceived social support on foreign language classroom anxiety. *System*, *74*, pp. 149-157. https://doi.org/10.1016/j.system.2018.01.002

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language teaching*, *44*(2), pp. 137-166. DOI: 10.1017/S0261444810000509.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), pp. 809–854. DOI: https://doi.org/10.1111/lang.12084

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), pp. 569-591. DOI:10.1017/S0142716417000534

Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, *41*(2), pp. 457- 480. DOI:10.1017/S0142716420000065

Kaplan-Rakowski, R., Gruber, A. (2023) The impact of high-immersion virtual reality on foreign language anxiety. *Smart Learning Environments, 10* (46), pp. 1-18. https://doi.org/10.1186/s40561-023-00263-9

Kazmi, S. Z. H., Zhang, H., Aziz, W., Monfredi, O., Abbas, S. A., Shah, S. A., ... & Butt, W. H. (2016). Inverse correlation between heart rate variability and heart rate demonstrated by linear and nonlinear analysis. *PloS one*, *11*(6), pp. 1-10. DOI: 10.1371/journal.pone.0157557

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*(3), pp. 459- 489. DOI: https://doi.org/10.3138/cmlr.64.3.459

Kennedy, S. & Trofimovich, P. (2019). Comprehensibility: A useful tool to explore listener understanding. *The Canadian Modern Language* Review / *La revue canadienne des langues vivantes*, *75*(4), pp. 275-284. DOI: 10.3138/cmlr.2019-0280

Kim, Y., Tracy–Ventura, N. & Jung, Y. (2016). A Measure of Proficiency or Short-Term Memory? Validation of an Elicited Imitation Test for SLA Research. *The*

Modern Language Journal, 100, pp. 655-673. https://doi-org.sire.ub.edu/10.1111/modl.12346

King, J., & Smith, L. (2017). Social anxiety and silence in Japan's tertiary foreign language classrooms. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 91-109). Bristol, UK: Multilingual Matters

Kirschbaum C., Pirke KM & Hellhammer DH. (1993). The 'trier social stress test'. A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28* (1-2), pp. 76-81. DOI: 10.1159/000119004

Kitano, K. (2001). Anxiety in the college Japanese language classroom. *Modern Language Journal*, *85*(4), pp. 549-66. DOI: https://doi.org/10.1111/0026-7902.00125

Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning, 27,* pp. 93-17. DOI:10.1111/j.1467-1770.1977.tb00294.x

Koch, A. S., & Terrell, T. D. (1991). Affective reactions of foreign language students to Natural Approach activities and teaching techniques. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 109–126). Englewood Cliffs, NJ: Prentice Hall.

Kocielnik, R., Sidorova, N., Maggi, F. M., Ouwerkerk, M., & Westerink, J. H. (2013). Smart technologies for long-term stress monitoring at work. In *proceedings of the 26th IEEE international symposium on computer-based medical systems*, pp. 53-58). DOI: 10.1109/CBMS.2013.6627764

Kormos, J. (1999). Monitoring and self-repair in L2. *Language learning, 49*(2), pp. 303-342. https://doi.org/10.1111/0023-8333.00090

Kormos, J. (2006). *Speech Production and Second Language Acquisition,* 1st edition. New York: Routledge.

Kormos, J. (2015). Individual differences in second language speech production. In J. W. Schwieter (Ed.) *The Cambridge handbook of bilingual processing,* pp. 369–388. (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107447257.017

Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *Systems, 32*, pp. 145-164. DOI: https://doi.org/10.1016/j.system.2004.01.001

Kormos, J., & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift fur Interkulturellen Fremdsprachenunterricht, 9*(2), pp. 1–19. Available at http://zif.spz.tu-darmstadt.de/jg-09-2/beitrag/kormos2.htm

Kormos, J., & Préfontaine, Y. (2017). Affective factors influencing fluent performance: French learners' appraisals of second language speech tasks. *Language Teaching Research, 21*(6), pp. 699-716. DOI: https://doi.org/10.1177/1362168816683562

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition, 11*(2), pp. 261-271. DOI: 10.1017/S1366728908003416

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In *Second Language Task Complexity* (pp. 267-286). John Benjamins.

Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: A meta-analysis. *Studies in Second Language Acquisition, 44*(3), pp. 886-911. DOI:10.1017/S0272263121000395

Kralova, Z., Tirpakova, A. & Skorvagova, E. (2018). Personality factors and foreign language pronunciation anxiety: The effect of psycho-social training. *European Journal of Contemporary Education*, *7*(4), pp. 728- 740. DOI:10.13187/ejced.2018.4.728

Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Prentice-Hall. ISBN 0137100477.

Kurniawan, H., Maslov, A. K. & Pechenizkiy, M. (2013). Stress detection from speech and galvanic skin response signals. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pp. 209-214. DOI: 10.1109/CBMS.2013.6627790

Kwon, O., Jeong, J., Kim, H. B., Kwon, I. H., Park, S. Y., Kim, J. E., & Choi, Y. (2018). Electrocardiogram sampling frequency range acceptable for heart rate variability analysis. *Healthcare informatics research*, *24* (3), 198- 206. DOI: 10.4258/hir.2018.24.3.198

Larsen-Freeman, D. and Cameron, L. (2008). *Complex Systems and Applied Linguistics*. Oxford, UK: Oxford University Press.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied linguistics, 18*(2), pp. 141-165. https://doi.org/10.1093/applin/18.2.141

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), pp. 35-63. https://doi-org.sire.ub.edu/10.1177/0267658307082981

Lee, H., & Kleinsmith, A. (2019, May). Public speaking anxiety in a real classroom: Towards developing a reflection system. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-6). DOI:10.1145/3290607.3312875

Lee, H., Mandalapu, V., Kleinsmith, A., & Gong, J. (2020a). Distinguishing anxiety subtypes of English language learners towards augmented emotional clarity. In *International Conference on Artificial Intelligence in Education* (pp. 157-161). Springer, Cham. DOI: 10.1007/978-3-030-52240-7_29

Lee, H., Mandalapu, V., Gong, J., Kleinsmith, A., & Kuber, R. (2020b). Using physiological cues to determine levels of anxiety experienced among deaf and hard of hearing English language learners. In *Companion Publication of the 2020 International Conference on Multimodal Interaction* (pp. 72-76) https://doi.org/10.1145/3395035.3425259

Lee, H. (2020, October). Supporting Instructors to Provide Emotional and Instructional Scaffolding for English Language Learners through Biosensor-based Feedback. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 733-737). https://doi.org/10.1145/3382507.3421159

Lee, S., Mendel, L. L., & Bidelman, G. M. (2019). Predicting speech recognition using the speech intelligibility index and other variables for cochlear implant users. *Journal of Speech, Language, and Hearing Research, 62*(5), pp. 1517-1531. https://doi.org/10.1044/2018_JSLHR-H-18-0303

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), pp. 387-417. https://doi.org/10.1111/j.1467–1770.1990.tb00669.x

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), pp. 1-38. DOI:10.1017/S0140525X99001776

Lindberg, R., McDonough, K. & Trofimovich, P. (2021). Investigating verbal and nonverbal indicators of physiological response during second language interaction. *Applied Psycholinguistics*, *42* (6), pp. 1403-1425. DOI: https://doi.org/10.1017/S014271642100028X

Linck, J.A., Osthus, P., Koeth, J.T. & Bunting, M.F. (2014) Working memory and second language comprehension and production A meta-analysis. *Psychonomic bulletin & review*, *21*, pp. 861–883. https://doi-org.sire.ub.edu10.3758s13423-013-0565-2

Lintunen, P., Mutta, M., & Peltonen. P. (2020). Synthesising approaches to second language fluency: Implications and future directions. In P. Lintunen, M. Mutta, & P. Peltonen (Eds.), *Fluency in L2 learning and use* (pp. 219-236). Bristol: Multilingual Matters. https://doi.org/10.21832/9781788926317

Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System, 34*(3), pp. 301-316. https://doi.org/10.1016/j.system.2006.04.004

Liu, M. & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and Foreign Language Anxiety. *The Modern Language Journal*, *92*, pp. 71-86. https://doi.org/10.1111/j.1540-4781.2008.00687.x

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British educational research journal*, *29*(6), pp. 861-878. https://doi-org.sire.ub.edu/10.1080/0141192032000137349

MacIntyre, P.D. (1995a). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal*, *79*, pp. 90 - 99. DOI: https://doi.org/10.2307/329395

MacIntyre, P. D. (1995b). On seeing the forest and the trees: A rejoinder to Sparks and Ganschow. *The Modern Language Journal*, *79*, pp. 245 - 248. https://doi.org/10.2307/329623

MacIntyre, P. D. (1999). Language anxiety: a review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language teaching: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). Boston: McGraw-Hill.

MacIntyre, P. D. (2012). The Idiodynamic method: a closer look at the dynamics of communication traits. *Communication Research Reports*, 29(4), pp. 361–367. https://doi.org/10.1080/08824096.2012.723274

MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development in: Gkonou, C., Daubney, M. & Dewaele, J.-M. eds. *New Insights into Language Anxiety: Theory, Research and educational implications*. Bristol: Multilingual Matters, pp. 11–30. DOI:10.21832/9781783097722-003

MacIntyre, P. D. 2002. Motivation, anxiety and emotion in second language acquisition. In *Individual Differences in Second Language Acquisition*, P. Robinson (ed.), pp. 45–68. Amsterdam: John Benjamins

MacIntyre, P.D. (2020). *Anion Variable Tester* [Computer program]. Version 2 https://petermacintyre.weebly.com/idiodynamic-software.html

MacIntyre, P. D., Baker, S. C., Clément, R., y Donovan, L. A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. *Language Learning, 52*(3), pp. 537-564. DOI: https://doi.org/10.1111/1467-9922.00226

MacIntyre, P. D., & Charos, C. (1996). Personality, Attitudes, and Affect as Predictors of Second Language Communication. *Journal of Language and Social Psychology, 15*(1), pp. 3-26. https://doi.org/10.1177/0261927X960151001

MacIntyre, P., & Ducker, N. (2022). The idiodynamic method: A practical guide for researchers. *Research Methods in Applied Linguistics, 1*(2), pp. 1-10. https://doi.org/10.1016/j.rmal.2022.100007

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning*, *39*(2), pp. 251-275. DOI: https://doi.org/10.1111/j.1467-1770.1989.tb00423.x

MacIntyre, P. D. & Gardner, R. (1991a). Methods and results in the study of anxiety and language learning: A review of literature. *Language Learning*, *41*, pp. 85-117. DOI: https://doi.org/10.1111/j.1467-1770.1991.tb00677.x

MacIntyre, P. D. & Gardner, R. (1991b). Language anxiety: Its relation to other anxieties and to processing in native and second languages. *Language Learning, 41*, pp. 513–34. DOI:10.1111/J.1467-1770.1991.TB00691.X

MacIntyre, P. D. & Gardner, R. (1991c). Investigating language class anxiety using the focused essay technique. *The Modern Language Journal, 75*, pp. 296-304. DOI: https://doi.org/10.2307/328723

MacIntyre, P., & Gardner, R. (1994a). The Effects of Induced Anxiety on Three Stages of Cognitive Processing in Computerized Vocabulary Learning. *Studies in Second Language Acquisition, 16*(1), 1-17. DOI: 10.1017/S0272263100012560

Maclntyre, P. D., & Gardner, R. C. (1994b). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44,* pp. 283-305. DOI: https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Maclntyre, P. D., & Gardner, R. C. (1988). The measurement of anxiety and applications to second language learning: An annotated bibliography. *Research Bulletin, 672*, pp. 1-39. https://eric.ed.gov/?id=ED301040

MacIntyre, P. D., & Gregersen, T. (2022). The idiodynamic method: Willingness to communicate and anxiety processes interacting in real time. *International Review of Applied Linguistics in Language Teaching, 60*(1), pp. 67-84. https://doi.org/10.1515/iral-2021-0024

MacIntyre, P. D., & Legatto, J. J. (2011). A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics, 32*(2), pp. 149-171. https://doi.org/10.1093/applin/amq037

MacIntyre, V. A., MacIntyre, P. D. & Carre, G. (2010). Heart rate variability as a predictor of speaking anxiety. *Communication Research Reports*, *27*(4), pp. 286-297. DOI: https://doi.org/10.1080/08824096.2010.496323

MacIntyre, P. D., & Thivierge, K. A. (1995). The effects of audience pleasantness, audience familiarity, and speaking contexts on public speaking anxiety and willingness to speak. *Communication Quarterly, 43*(4), pp. 456–466. https://doi-org.sire.ub.edu/10.1080/01463379509369992

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs* (3rd ed.). Lawrence Erlbaum Associates Publishers.

Mak, B. 2011. An exploration of speaking-in-class anxiety with Chinese ESL learners. *System 39*, pp. 202–214. DOI: https://doi.org/10.1016/j.system.2011.04.002

Marcos-Llinás, M. & Garau, M. J. (2009). Effects of language anxiety on three proficiency-level courses of Spanish as a foreign language. *Foreign Language Annals, 42*(1), pp. 94- 111. DOI: https://doi.org/10.1111/j.1944-9720.2009.01010.x

Mauss, I., Wilhelm, F., & Gross, J. (2004). Is there less to social anxiety than meets the eye? Emotion experience, expression, and bodily responding. *Cognition and emotion, 18*(5), pp. 631-642. https://doi-org. /10.1080/02699930341000112

Masrai, A. (2022). The development and validation of a lemma-based yes/no vocabulary size test. *SAGE Open, 12*(1), pp.1-17. https://doi.org/10.1177/21582440221074355

Matsuda, S. & P. Gobel (2004). Anxiety and predictors of performance in the foreign language classroom. *System* 32 (1), pp. 21–36. https://doi.org/10.1016/j.system.2003.08.002

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016, May). Validation of the Empatica E4 wristband. In 2016 *IEEE EMBS international student conference (isc)* (pp. 1-4). IEEE. DOI: 10.1109/EMBSISC.2016.7508621

McCroskey, J. C. (1970). Measures of communication-bound anxiety. *Speech Monograph*, *37*(4), pp. 269-277. DOI: https://doi.org/10.1080/03637757009375677

McCroskey, J. C. (1978). Validity of the PRCA as an index of oral communication apprehension. *Communication Monographs*, *45,* pp .192-203. DOI: https://doi.org/10.1080/03637757809375965

McCroskey, J. C., Fayer, J. M., & Richmond, V. P. (1985). Don't speak to me in English: Communication apprehension in Puerto Rico. *Communication Quarterly*, *33*, pp. 185-192. DOI: https://doi.org/10.1080/01463378509369597

McManus, K., & Liu, Y. (2022). Using elicited imitation to measure global oral proficiency in SLA research: A close replication study. *Language Teaching, 55*(1), pp. 116-135. DOI:10.1017/S026144482000021X

Meara, P. M., & Miralpeix, I. (2015a). *V_YesNo*. Retrieved from _lognostics: https://www.lognostics.co.uk/

Meara, P. M., & Miralpeix, I. (2015b). *V_YesNo v1.0 manual*. Retrieved from _lognostics: https://www.lognostics.co.uk/tools/Manuals/V_YesNo_Manual.pdf

Meara, P., & Miralpeix, I. (2017). *Tools for researching vocabulary*. Bristol, Blue Ridge Summit: Multilingual Matters.

Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M., & Sarlo, M. (2019). Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*, *56*(11), pp. 1-15. DOI:10.1111/psyp.13441

Melchor-Couto, S. (2016). Foreign language anxiety levels in second life oral interaction. *ReCALL: Journal of Eurocall*, *29* (1), pp. 99–119. DOI: https://doi.org/10.1017/S0958344016000185

Milstein, N., & Gordon, I. (2020). Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience, 14* (Article 148), pp. 1- 13. DOI: https://doi.org/10.3389/fnbeh.2020.00148

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

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science, 21*(1), pp.8-14. https://doi-org.sire.ub.edu/10.1177/0963721411429458

Mejias, H., R. L. Applbaum, S. J. Applbaum and R. T. Trotter. (1991). Oral communication apprehension and Hispanics: An exploration of oral communication apprehension among Mexican American students in Texas. In *Language anxiety: From theory and research to classroom implications*, eds. E. K. Horwitz and D. J. Young, 87–97. Upper Saddle River, NJ: Prentice Hall. https://eric.ed.gov/?id=ED274010

Mora, J. C. (2006). Age Effects on Oral Fluency Development. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 65-88). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781853598937-005

Mora, J. C. (2022). Aptitude and individual differences. In Tracey M. Derwing, Murray J. Munro & Ron I. Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 68-82). New York, NY, USA: Routledge.

Mora, J. C., & Darcy, I. (2023). Individual differences in attention control and the processing of phonological contrasts in a second language. *Phonetica, 80*(3-4), pp. 153-184. https://doi-org.sire.ub.edu/10.1515/phon-2022-0020

Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2023). Speaking anxiety and task complexity effects on second language speech. International Journal of Applied Linguistics, 1-24. https://doi.org/10.1111/ijal.12494

Montero-Pérez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition, 42*(4), pp. 749 – 773. DOI:10.1017/S0272263119000706

Moyer, A. (2014a). Exceptional Outcomes in L2 Phonology: The Critical Factors of Learner Engagement and Self-Regulation. *Applied Linguistics, 35*(4), pp. 418–440. https://doi-org.sire.ub.edu/10.1093/applin/amu012

Moyer, A. (2014b). The Social Nature of L2 Pronunciation. In J. Levis & A. Moyer (Ed.), *Social Dynamics in Second Language Accent* (pp. 11-30). Berlin, Boston: De Gruyter Mouton. https://doi-org.sire.ub.edu/10.1515/9781614511762.11

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), pp. 139-154. DOI:10.1017/S0272263198002022

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning, 45*(1), pp. 73-97. https://doi-org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech*, *38*(3), pp.289-306. DOI:10.1177/002383099503800305

Muñoz, C. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. *IRAL-International Review of Applied Linguistics in Language Teaching*, *46*(3), 197-220. DOI: https://doi.org/10.1515/IRAL.2008.009

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, *44*(1), 1-35. DOI: 10.1017/S0261444810000327

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. Studies in Second Language Acquisition, 41(4), pp. 647-672. https://doi.org/10.1017/S0272263119000044

Nagle, C. L., Trofimovich, P., O'Brien, M. G., & Kennedy, S. (2022). Beyond linguistic features: Exploring behavioral and affective correlates of comprehensible second language speech. *Studies in Second Language Acquisition, 44(1),* pp. 1-16. DOI: 10.1017/S0272263121000073

Nagle, C., Trofimovich, P., Tekin, O., & McDonough, K. (2023). Framing second language comprehensibility: Do interlocutors' ratings predict their perceived communicative experience? *Applied Psycholinguistics, 44(1)*, pp. 131-156. https://doi.org/10.1017/S0142716423000073

Norris, R., & French, A. (2014). *Ready for advanced Student's Book* (3rd ed.). Macmillan Education. ISBN 9781786327574.

Nourbakhsh, N., Chen, F., Wang, Y. & Calvo, R. (2017). Detecting users' cognitive load by galvanic skin response with affective interference. *ACM Transactions on Interactive Intelligent Systems*, *7*(3), pp. 1-20. DOI:10.1145/2960413

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*(4), 557–581. DOI:10.1017/S027226310707043X

Onwuegbuzie, A., Bailey, P., & Daley, C. E. (1999). Factors associated with foreign language anxiety. *Applied Psycholinguistics*, *20*(2), pp. 217–239. DOI: 10.1017/S0142716499002039

Ortega, L., Iwashita, N., Rabie, S. & Norris, J. M. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. *Conference Handout from Paper Presented at the Meeting of the Second Language Research Forum*, Toronto, Canada. DOI:10.21832/9781783092291-011

Oxford, R. L. (2017). Anxious language learners can change their minds: Ideas and strategies from traditional psychology and positive psychology. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 179-199). Bristol, UK: Multilingual Matters.

Owren, M. (2008). GSU Praat Tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, *40*(3), pp. 822-829. DOI: 10.3758/BRM.40.3.822

Pace-Schott, E. F., Amole, M. C., Aue, T., Balconi, M., Bylsma, L. M., Critchley, H., et al. (2019). Physiological Feelings. *Neuroscience Biobehavioral Reviews 103*, pp.267–304. https://doi.org/10.1016/j.neubiorev.2019.05.002

Parmaxi, A. (2023). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*(1), pp. 172–184. https://doi.org/10.1080/10494820.2020.1765392

Pérez-Castillejo, S. (2019). The role of foreign language anxiety on L2 utterance fluency during a final exam. *Language Testing, 36*(3), pp. 327-345. DOI: https://doi.org/10.1177/0265532218777783

Pérez Castillejo, S. (2023). Prior processing, foreign language classroom anxiety, and L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *61*(2), 519-544. https://doi-org.sire.ub.edu/10.1515/iral-2021-0091

Phillips, E. M. (1992). The effects of language anxiety on student oral test performance and attitudes. *The Modern Language Journal*, *76*, pp. 14-26. DOI: https://doi.org/10.2307/329894

Piechurska-Kuciel, E. (2008). *Language anxiety in secondary grammar school students*. Opole: Wydawnictwo Uniwersytetu Opolskiego

Pinget, A. F., Bosker, H. R., Quené, H., & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), pp. 349 - 365. DOI: https://doi.org/10.1177/0265532214526177

Préfontaine, Y. (2013). Perceptions of French fluency in second language speech production. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, *69*(3), pp. 324–348. DOI: https://doi.org/10.3138/cmlr.1748

Préfontaine, Y., & Kormos, J. (2015). The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal*, *99*(1), pp. 96-112. DOI: https://doi.org/10.1111/modl.12186

Papi, M. (2010). The L2 motivational self-system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, *38*, pp. 467–479. https://doi.org/10.1016/j.system.2010.06.011

Park, G. P. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports 115*, pp. 261–75. https://doi.org/10.2466/28.11.PR0.115c10z2

Park, G. P., & French, B. F. (2013). Gender differences in the foreign language classroom anxiety scale. *System*, *41*(2), pp. 462-471. https://doi.org/10.1016/j.system.2013.04.001

Pérez-Paredes, P. F., & Martínez-Sánchez, F. (2000). A Spanish version of the foreign language classroom anxiety scale: Revisiting Aida's factor analysis. *Revista española de lingüística aplicada (RESLA)*, *14*, pp. 337-352. Available at https://dialnet.unirioja.es/servlet/articulo?codigo=871315

Price, L. M. (1991). The subjective experience of foreign language anxiety: Interviews with highly anxious students. In *Language anxiety: From theory and research to classroom implications*, eds. E. K. Horwitz and D. J. Young, 101–108. Upper Saddle River, NJ: Prentice Hall.

Qiu, X. B., Shan, C., Yao, J., & Fu, Q. K. (2024). The effects of virtual reality on EFL learning: A meta-analysis. *Education and Information Technologies*, *29*, pp. 1379-1405. https://doi.org/10.1007/s10639-023-11738-0

Rivers, D. J. (2022). The Situation Specific Arousal Analyzer: Innovation in the Physiological Assessment of Foreign Language Education Anxiety. *Frontiers in education*, 7, pp. 1-8. https://doi.org/10.3389/feduc.2022.802639

Reitan, R. (1958). Validity of the Trail Making Test as an indication of organic brain damage. *Perceptual and Motor Skills, 8*, pp. 271–276. https://doi.org/10.2466/pms.1958.8.3.271

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*(1), pp. 162–181. https://doi.org/10.1111/j.1540-4781.2011.01241.x

Révész, A., & Gurzynski-Weiss, L. (2016). Teachers' Perspectives on Second Language Task Difficulty: Insights from Think-Alouds and Eye Tracking. *Annual Review of Applied Linguistics, 36*, pp. 182-204. DOI: 10.1017/S0267190515000124

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring Cognitive Task Demands Using Dual-Task Methodology, Subjective Self-Ratings, and Expert Judgments: A Validation Study. *Studies in Second Language Acquisition, 38*(4), pp. 703-737. DOI:10.1017/S0272263115000339

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second language research*, *17*(4), pp. 368-392. DOI:10.1191/026765801681495877

Robinson, P. (2003). Attention and Memory during SLA. In C.J. Doughty and M.H. Long (Eds.) *The Handbook of Second Language Acquisition* (pp. 631-678). Newark: Wiley- Blackwell. https://doi-org.sire.ub.edu/10.1002/9780470756492.ch19

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, 45*(3), pp. 193-213. https://doi.org/10.1515/iral.2007.009

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review, 65*, pp. 395–412. DOI:10.3138/cmlr.65.3.395

Roos, L. G., & Slavich, G. M. (2023). Wearable technologies for health research: Opportunities, limitations, and practical and conceptual considerations. *Brain, Behavior, and Immunity, 113*, pp. 444-452. https://doi.org/10.1016/j.bbi.2023.08.008

Roos, A. L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2021). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review, 33*, pp. 579-618. https://doi.org/10.1007/s10648-020-09543-z

Rood, M. M. S. J., & de Jong, N. H. (2023). A mixed-methods study on the effects of manipulated speaking anxiety on L2 utterance and L2 cognitive fluency. *Journal of the European Second Language Association, 7*(1), pp. 60–74. https://doi.org/10.22599/jesla.96

Rubio-Alcalá, F. D. (2017). The Links Between Self-Esteem and Language Anxiety and Implications for the Classroom. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 198-216). Bristol, UK: Multilingual Matters.

Shachter, J., Kangas, M., Sweller, N., & Stewart, J. (2022). Affective States and Heart Rate Response: Measuring Foreign Language Speaking Performance Reactions in a Japanese University Classroom. *Journal for the Psychology of Language Learning, 4*(2), pp. 1-23. Retrieved from https://jpll.org/index.php/journal/article/view/84

Saito, K. (2015). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition, 37*(4), pp. 713–743. doi:10.1017/S0272263115000248

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-asecond-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly, 55*, pp. 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., Dewaele, J. M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning, 68* (3), pp. 709-743. DOI:10.1111/lang.12297

Saito, K., Dewaele, J. M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech, 60*, pp. 614–632. https://doi.org/10.1177/0023830916687793

Saito, Y., Garza, T. J. & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal*, *83*(2), pp. 202-218. DOI: https://doi.org/10.1111/0026-7902.00016

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning, 69*(3), pp.652–708. https://doi.org/10.1111/lang.12345

Saito, Y. & Samimy, K. (1996). Foreign language anxiety and language performance: A study of learner anxiety in beginning, intermediate, and advanced-level college students of Japanese. *Foreign Language Annals*, *29*(2), pp. 239-249. DOI: https://doi.org/10.1111/j.1944-9720.1996.tb02330.x

Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do L2 listeners perceive the comprehensibility of foreign-accented speech? Roles of L1 profiles, L2 proficiency, age, experience, familiarity and metacognition. *Studies in Second Language Acquisition, 41*(5), pp. 1133-1149. DOI:10.1017/S0272263119000226

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*(2), pp. 217-240. DOI:10.1017/S0142716414000502

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*(4), pp. 439-462. https://doi.org/10.1093/applin/amv047

Saito, K., Trofimovich, P., Abe, M., & In'nami, Y. (2020). Dunning-Kruger effect in second language speech learning: How does self-perception align with other perception over time? *Learning and Individual Differences, 79*, pp. 1-10. https://doi.org/10.1016/j.lindif.2020.101849

Sanaei, O., Zafarghandi, A. M., & Sabet, M. K. (2015). The effect of classroom anxiety on EFL learners' oral narratives fluency: The case of intermediate level students. *Theory and Practice in Language Studies*, *5*(7), pp. 1390–1400. DOI:10.17507/tpls.0507.11

Sarason I.G., Sarason B.R. (1990). Test Anxiety. In: Leitenberg H. (eds) *Handbook of Social and Evaluation Anxiety*. Springer, Boston, MA.

Savitsky, K., & Gilovich, T. (2003). The illusion of transparency and the alleviation of speech anxiety. *Journal of experimental social psychology, 39*(6), pp. 618-625. https://doi.org/10.1016/S0022-1031(03)00056-8

Schuurmans, A. A., de Looff, P., Nijhof, K. S., Rosada, C., Scholte, R. H., Popma, A., & Otten, R. (2020). Validity of the Empatica E4 wristband to measure Heart Rate Variability (HRV) Parameters: A Comparison to electrocardiography (ECG). *Journal of medical systems*, *44*(11), pp. 1-11. DOI: 10.1007/s10916-020-01648-w

Schwerdtfeger, A. (2004). Predicting autonomic reactivity to public speaking: don't get fixed on self-report data! *International Journal of Psychophysiology*, *52* (3), pp. 217-224. DOI: 10.1016/j.ijpsycho.2003.10.008

Schlenker, B. R. and M. R. Leary. 1982. Social anxiety and self-presentation: A conceptualization and model. *Psychological Bulletin*, *92*, pp. 641–669. DOI: 10.1037/0033-2909.92.3.641

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, *28*, pp. 129-142. DOI:10.1111/j.1467-1770.1978.tb00309.x

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. Routledge: New York.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching, 54*(2), pp. 79-95. https://doi-org.sire.ub.edu/10.1515/iral-2016-9991

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & cognition, 33*(4), pp. 644-653. https://doi.org/10.3758/BF03195331

Sevinç, Y. (2018). Language anxiety in the immigrant context: Sweaty palms? *International Journal of Bilingualism, 22*(6), pp. 717-739. https://doi-org.sire.ub.edu/10.1177/1367006917690914

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health, 258,* pp. 1-17. DOI: 10.3389/fpubh.2017.00258

Sharma, M., Kacker, S. & Sharma, M. (2016). A brief introduction and review on galvanic skin response. *International Journal of Medical Research Professionals, 2* (6), pp. 13-17. DOI:10.21276/ijmrp.2016.2.6.003

Shi, R., Sharpe, L., & Abbott, M. (2019). A meta-analysis of the relationship between anxiety and attentional control. *Clinical psychology review, 72*, pp. 1-24. https://doi.org/10.1016/j.cpr.2019.101754.

Shintani, N, Saito, K. & Koizumi, R. (2019). The relationship between multilingual raters' language background and their perceptions of accentedness and comprehensibility of second language speech. *International Journal of Bilingual Education and Bilingualism, 22*(7), pp. 849-869. DOI: https://doi.org/10.1080/13670050.2017.1320967

Sheen, Y. (2008). Recasts, FLA, modified output, and L2 learning. *Language Learning, 58*(4), pp. 835-874. https://doi.org/10.1111/j.1467-9922.2008.00480.x

Simard, D. (2022). Psycholinguistic processes in L2 oral production. In Tracey M. Derwing, Murray J. Munro & Ron I. Thomson (Eds.), *The Routledge Handbook of Second Language Acquisition and Speaking* (pp. 24-38). New York, NY, USA: Routledge.

Simard, D., Zuniga, M., & Hameau, F. (2023). How do attention-shift and foreign language anxiety interact with objective and subjective measures of fluency? *Language Teaching Research*, pp.1-21. https://doi.org/10.1177/13621688221146379

Simsek, E., & Dörnyei, Z. (2017). Anxiety and L2 self-images: The "anxious self.". In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 51-69). Bristol, UK: Multilingual Matters

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36*(1), pp. 1–14. DOI:10.1017/S026144480200188X

Soehn, A. (2017, March 21). Measuring the Heart – How do ECG and PPG Work? iMotions. https://imotions.com/blog/learning/measuring-the-heart-how-does-ecg-and-ppg-work/

Sosa-López, G. (2019). The role of anxiety in L2 speech production. Unpublished MA Thesis. Universitat de Barcelona.

Sosa-López, G., & Mora, J.C. (2022a). Different tasks, different levels of anxiety? A pilot study. *Speak out!* by PronSIG - Pronunciation Special Interest Group at IATEFL, issue 66, pp. 37-49.

Sosa-López, G., & Mora, J. (2022b). The role of speaking anxiety on L2 English speaking fluency, accuracy and complexity. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13362

Spada, N., Shiu, J. L. & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*(3), pp. 723–751. DOI: https://doi.org/10.1111/lang.12129

Sparks, R. L. & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *The Modern Language Journal*, *75* (1), pp. 3-16. DOI: https://doi.org/10.1111/j.1540-4781.1991.tb01076.x

Sparks, R. L. & Ganschow, L. (1996). Anxiety about foreign language learning among high school women. *The Modern Language Journal*, *80*, pp.199 - 212. DOI: https://doi.org/10.2307/328636

Sparks, R. L., Ganschow, L. & Javorsky, J. (2000). Déjà vu all over again. A response to Saito, Horwitz, and Garza. *The Modern Language Journal*, *84*, pp. 251 - 255. DOI: https://doi.org/10.1111/0026-7902.00066

Spielberger, C. D. (1966). Theory and research on anxiety. In C.D. Spielberger (Ed.), *Anxiety and Behavior*. New York: Academic Press.

Spielberger, C. D. (1983) *Manual for the State-Trait Anxiety Inventory (STAI)*. Palo Alto, CA: Consulting Psychologists Press.

Steinberg, F.S. & Horwitz, E.K. (1986). The Effect of Induced Anxiety on the Denotative and Interpretive Content of Second Language Speech. *TESOL Quarterly*, *20*(1), pp. 131-136. DOI: https://doi.org/10.2307/3586395

Strachan, L., Kennedy, S., & Trofimovich, P. (2019). Second language speakers' awareness of their own comprehensibility: Examining task repetition and self-assessment. *Journal of Second Language Pronunciation*, *5* (3), pp. 347-373. DOI: https://doi.org/10.1075/jslp.18008.str

Sudina, E. (2023). A primer on measurement invariance in L2 anxiety research. *Annual Review of Applied Linguistics, 43*, pp. 140–146. DOI:10.1017/S0267190523000089

SUDS image (n.d.) in Concurrent treatment of PTSD and substance use disorders using prolonged exposure (COPE): Patient Workbook by Back, S. E., Foa, E. B., Killeen T. K., Mills K. L., Teesson M., Cotton B. D., Carroll K. M., & Brady K. T. Oxford University Press, 2015 <http://www.oxfordclinicalpsych.com/view/10.1093/med:psych/9780199334513.001.0001/ med-9780199334513-appendix-13>

Sutarsyah, C. (2017). An analysis of student's speaking anxiety and its effect on speaking performance. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics), 1*(2), pp. 143- 152. DOI: http://dx.doi.org/10.21093/ijeltal.v1i2.14

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), pp. 143-167. DOI:10.1017/S0272263119000421

Suzukida, Y. (2021). Roles of Cognitive and Sociopsychological Individual Differences in Second Language Pronunciation Development in Classroom Settings: A Dynamic Systems Theory Approach. [Unpublished Doctoral dissertation, UCL (University College London]. Retrieved from https://discovery.ucl.ac.uk/id/eprint/10122329/

Suzukida, Y., & Saito, K. (2023). Detangling experiential, cognitive, and sociopsychological individual differences in second language speech learning: Cross-sectional and longitudinal investigations. *Bilingualism: Language and cognition, 26*(4), pp. 762-775. DOI:10.1017/S1366728922000700

Swain, M. (2005). The output hypothesis: Theory and research. In Eli Hinkel (Ed.) *Handbook of research in second language teaching and learning* (pp. 471-483). Routledge

Szyszka, M. (2011). Foreign language anxiety and self-perceived English pronunciation competence. *Studies in Second Language Learning and Teaching (SSLLT), 1* (2), pp. 283-300. DOI: https://doi.org/10.14746/ssllt.2011.1.2.7

Szyszka, M. (2017). *Pronunciation learning strategies and language anxiety: In search of an interplay*. Springer. https://doi.org/10.1007/978-3-319-50642-5

Szyszka, M., & Lintunen, P. (2023). Zooming into the L2 Speech Fluency Markers of Anxious and Non-Anxious Advanced L2 Learners–an Extreme Case Sampling Report. *Research in Language, 21*(4), pp. 357-376. Available at https://www.ceeol.com/search/article-detail?id=1209436

Tanner, B. A. (2012). Validity of global physical and emotional SUDS. *Applied Psychophysiology and Biofeedback*, *37*(1), 31-34. DOI: 10.1007/s10484-011-9174-x

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239-273). Amsterdam: John Benjamins.

Tavakoli, P., & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge University Press. DOI: https://doi.org/10.1017/9781108589109

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015, August). Automatic identification of artifacts in electrodermal activity data. In 2015 *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1934-1937. DOI: 10.1109/EMBC.2015.7318762

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), pp. 363-387. DOI:10.1017/S0272263118000311

Tian, C. (2019). Anxiety in classroom English presentations: A case study in Korean tertiary educational context. *Higher Education Studies*, *9* (1), pp. 132- 143. DOI:10.5539/hes.v9n1p132

Tobias, S. (1979). Anxiety research in educational psychology. *Journal of Educational Psychology, 71* (5), pp. 573-582. DOI: https://doi.org/10.1037/0022-0663.71.5.573

Tobias, S. (1986). Anxiety and cognitive processing of instruction. In Schwarzer, R. (Ed.) *Self-related cognitions in anxiety and motivation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Trofimovich, P. & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15* (4), pp. 905-916. DOI:10.1017/S1366728912000168

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self-and other-perception of second language speech.

Bilingualism: Language and Cognition, 19(1), pp. 122-140. https://doi.org/10.1017/S1366728914000832

Trotman, G. P., Veldhuijzen van Zanten, J. J., Davies, J., Möller, C., Ginty, A. T., & Williams, S. E. (2019). Associations between heart rate, perceived heart rate, and anxiety during acute psychological stress. *Anxiety, Stress, & Coping, 32*(6), pp. 711-727. https://doi.org/10.1080/10615806.2019.1648794

Tóth, Z. (2007). Predictors of foreign-language anxiety: Examining the relationship between anxiety and other individual learner variables. In József Horváth and Marianne Nikolov (Eds.), *UPRT*: *Empirical Studies in English Applied Linguistics* (pp. 123-148). Pécs: Lingua Franca Csoport. Available at https://mek.oszk.hu/07500/07557/07557.pdf

Tóth, Z. (2008). A foreign language anxiety scale for Hungarian learners of English. *WoPaLP, 2*(1), pp. 55-77. Available at https://langped.elte.hu/WoPaLParticles/W2TothZs.pdf

Tóth, Z. (2009). Foreign language anxiety: For beginners only? In R. Lugossy, J. Horváth, & M. Nikolov (Eds.), *UPRT 2008: Empirical studies in English applied linguistics* (pp. 225-246). Pécs: Lingua Franca Csoport. Available at http://mek.niif.hu/07500/07558/07558.pdf

Tóth, Z. (2010). *Foreign language anxiety and the advanced language learner: A study of Hungarian students of English as a foreign language*. Newcastle upon Tyne (UK): Cambridge Scholars Publishing.

Tóth, Z. (2012). Foreign language anxiety and oral performance: Differences between high-vs. low-anxious EFL students. *US-China Foreign Language, 10*(5), pp. 1166-1178. ISSN 1539-8080.

Tóth, Z. (2011) Foreign language anxiety and advanced EFL learners: An interview study. *Working Papers in Language Pedagogy, 5*, pp. 39–57. Available at https://langped.elte.hu/WoPaLParticles/W5Toth.pdf

Tóth, Z. (2017). Exploring the Relationship between Anxiety and Advanced Hungarian EFL Learners' Communication Experiences in the Target Language: A Study of High-vs Low-Anxious Learners. In C. Gkonou, M. Daubney, & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 156-174). Bristol, UK: Multilingual Matters.

Tóth, Z. (2019). Under the magnifying glass': students' perspectives on oral presentations and anxiety in the EFL classroom. *European Journal of Foreign*

Language Teaching 4 (2), pp. 126-146. DOI: https://doi.org/10.5281/zenodo.3370709

Trail Making Test available at CENTER-TBI https://www.center-tbi.eu/files/approved-translations/English/ENGLISH_TMT.pdf

Tucker, G. R., Hamayan, E., & Genesee, F. H. (1976). Affective, cognitive and social factors in second-language acquisition. *Canadian modern language review*, *32*(3), pp. 214-226. https://doi.org/10.3138/cmlr.32.3.214

Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language teaching research, 24*(4), 540-556. https://doi.org/10.1177/1362168818799371

Val-Calvo, M., Álvarez-Sánchez, J. R., Ferrández-Vicente, J. M., & Fernández, E. (2020). Affective Robot Story-Telling Human-Robot Interaction: Exploratory Real-Time Emotion Estimation Analysis Using Facial Expressions and Physiological Signals. *IEEE Access*, *8*, pp. 134051-134066. DOI: 10.1109/ACCESS.2020.3007109.

Van Dooren, M., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & behavior*, *106*(2), pp. 298-304. https://doi.org/10.1016/j.physbeh.2012.01.020

Watson, D., & Friend, R. (1969). Measurement of social-evaluative anxiety. *Journal of Consulting and Clinical Psychology*, *33* (4), pp. 448-457. DOI: https://doi.org/10.1037/h0027806

Westenberg, P.M., Bokhorst, C.L., Miers, A.C., Sumter, S.R., Kallen, V.L., Van Pelt, Johannes, Blö, A.W. (2009). A prepared speech in front of a pre-recorded audience: Subjective, physiological, and neuroendocrine responses to the Leiden Public Speaking Task. *Biological Psychology*, *82*, pp. 116–124. DOI: 10.1016/j.biopsycho.2009.06.005

Wood, D. (2016). Willingness to communicate and second language speech fluency: An idiodynamic investigation. *System 60*, pp. 11–28. https://doi.org/10.1016/j.system.2016.05.003

Woodrow, L. (2001). Towards a Model of Adaptive Language Learning: A Pilot Study. *ERIC Documents*: ED456645, pp. 1-25. Available at https://eric.ed.gov/?id=ED456645

Woodrow, L. (2003). A Model of Adaptive Learning of International English for Academic Purposes (EAP) at Australian Universities' (Unpublished doctoral thesis, University of Sydney). Available at https://ses.library.usyd.edu.au/handle/2123/27874

Woodrow, L. (2006). Anxiety and speaking English as a second Language. *RELC Journal*, *37* (3), pp. 308-328. https://doi.org/10.1177/0033688206071315

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R.F. & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. Journal of clinical and experimental neuropsychology, 33(1), 101-111. DOI: 10.1080/13803395.2010.493149

Wu, S. L., Tio, Y. P., & Ortega, L. (2021). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, *44*(1) pp. 271-300. DOI: 10.1017/S0272263121000103

Yan, J. X. & Horwitz, E. K. (2008). Learners' perceptions of how anxiety interacts with personal and instructional factors to influence their achievement in English: A qualitative analysis of EFL learners in China. *Language Learning, 58 (1)*, pp. 151-183. https://doi-org. /10.1111/j.1467-9922.2007.00437.x

Yan, X., Maeda, Y., Lv, J. & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*(4), pp. 497- 528. DOI: https://doi.org/10.1177/0265532215594643

Young, D. J. (1986). The relationship between anxiety and foreign language oral proficiency ratings. *Foreign Language Annals, 19*, pp. 439–445. https://doi-org. /10.1111/j.1944-9720.1986.tb01032.x

Young, D. J. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals, 23*, pp. 539–553. DOI: https://doi.org/10.1111/j.1944-9720.1990.tb00424.x

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *Modern Language Journal, 75*, pp. 426–439. https://doi.org/10.2307/329492

Young, D. J. (1992). Language anxiety from the foreign language specialist's perspective: Interviews with Krashen, Omaggio Hadley, Terrell, and Rardin. *Foreign Language Annals, 25*, pp. 157-172. DOI:10.1111/j.1944-9720.1992.tb00524.x

Zhou, Y. (2024). Aptitude, Anxiety, and Success in L2 Speech Development: A Longitudinal Study of Chinese EFL College-Level Learners. *Language and Speech*, pp. 1-23. https://doi.org/10.1177/00238309241281741

Zuniga, M., & Simard, D. (2019). Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguistic Research*, *48*, pp. 43–59. https://doi-org./10.1007/s10936-018-9587-2

Zuniga, M., & Simard, D. (2022). Exploring the intricate relationship between foreign language anxiety, attention and self-repairs during L2 speech production. *System*, *105*, pp. 1-10. https://doi.org/10.1016/j.system.2022.102732

9. Appendices

Appendix A1: Set up / Layout

Subject Group 1, Type A:												
Time Slots	Time		DAY 1	Time		DAY 2	Time		DAY 3	Time		DAY 4
	15'		LBQ (HOME)	2'		Welcome	2'		Welcome	2'		Welcome
	5'		Welcome & Sign Informed Consent	1.5'		Preparation Time	1.5'		Preparation Time	1.5'		Preparation Time
	1.5'		Preparation Time	5'		L2 Group Or.Presentation_D2	5'		L1 One-To-One_D3	5'		L1 Group Oral Presentation_D4
	5'		L2 One-To-One_D1	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire
	5'		Post-Task Questionnaire	5'		Idiodynamic Task	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		Idiodynamic Task				1.5'		Preparation Time	1.5'		Preparation Time
8:00	10'	S1	Stimulated Recall	10'	S2	Stimulated Recall	5'	S3	L1 One-To-One_A3	5'	S4	L1 Group Oral Presentation_A4
				1.5'		Preparation Time	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire
	1.5'		Preparation Time	5'		L2 Group Or.Presentation_A2	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		L2 One-To-One_A1	5'		Post-Task Questionnaire	10'		Elicited Imitation Task (Eit)	10’		Yes/No Vocabulary Size
	5'		Post-Task Questionnaire	5'		Idiodynamic Task				10’		Woking Memory Task (Digit)
	5'		Idiodynamic Task	10'		Stimulated Recall	10'		Attention Switch-Trail			
	10'		Stimulated Recall									

Subject Group 1, type B:												
Time slots	Time		DAY 1	Time		DAY 2	Time		DAY 3	Time		DAY 4
8:00	15'	S1	LBQ (HOME)	2'	S2	Welcome	2'	S3	Welcome	2'	S4	Welcome
	5'		Welcome & Sign Informed Consent	1.5'		Preparation time	1.5'		Preparation time	1.5'		Preparation time
	1.5'		Preparation time	5'		L2 Group or.presentation_A2	5'		L1 one-to-one_A3	5'		L1 Group oral presentation_A4
	5'		L2 One-To-One_A1	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire
	5'		Post-Task Questionnaire	5'		Idiodynamic Task	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		Idiodynamic Task	10'		Stimulated Recall	1.5'		Preparation Time	1.5'		Preparation Time
	10'		Stimulated Recall	1.5'		Preparation Time	5'		L1 One-To-One_D3	5'		L1 Group Oral Presentation_D4
	1.5'		Preparation Time	5'		L2 Group Or.Presentation_D 2	5'		Post-Task QUESTIONNAIRE	5'		Post-Task QUESTIONNAIRE
	5'		L2 One-To-One_D1	5'		Post-Task Questionnaire	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		Post-Task Questionnaire	5'		Idiodynamic Task	10'		Elicited Imitation	10’		Yes/No Vocabulary Size
	5'		Idiodynamic Task	10'		Stimulated Recall	10'		Attention Switch-Trail	10’		Woking Memory Task (Digit)
	10'		Stimulated Recall									

Subject Group 2, type A:												
Time slots	Time		DAY 1	Time		DAY 2	Time		DAY 3	Time		DAY 4
8:00	15'	S1	LBQ (HOME)	2'	S2	Welcome	2'	S3	Welcome	2'	S4	Welcome
	5'		Welcome & Sign Informed Consent	5'		Preparation Time	1.5'		Preparation Time	1.5'		Preparation Time
	1.5'		Preparation Time	1.5'		L2 One-To-One_D1	5'		L1 Group Or.Presentation_D4	5'		L1 One-To-One_D3
	5'		L2 Group Or.Presentation_D2	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire
	5'		Post-Task Questionnaire	5'		Idiodynamic Task	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		Idiodynamic Task	5'		Stimulated Recall	1.5'		Preparation Time	1.5'		Preparation Time
	10'		Stimulated Recall	10'		Preparation Time	5'		L1 Group Or.Presentation_A4	5'		L1 One-To-One_A3
	1.5'		Preparation Time	1.5'		L2 One-To-One_A1	5'		Post-Task QUESTIONNAIRE	5'		Post-Task QUESTIONNAIRE
	5'		L2 Group Or.Presentation_A2	5'		Post-Task Questionnaire	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		Post-Task Questionnaire	5'		Idiodynamic Task	10'		Elicited Imitation Task	10'		Yes/No Vocabulary Size
	5'		Idiodynamic Task	5'		Stimulated Recall	10'		Attention Switch-Trail	10'		Woking Memory Task (Digit)
	10'		Stimulated Recall	10'								

Subject Group 2, type B:

Time slots	Time		DAY 1	Time		DAY 2	Time		DAY 3	Time		DAY 4
	15'		LBQ	2'		Welcome	2'		Welcome	2'		Welcome
	5'		Welcome & Sign Informed Consent	5'		Preparation Time	1.5'		Preparation Time	1.5'		Preparation Time
	1.5'		Preparation Time	1.5'		L2 One-To-One_A1	5'		L1 Group Or.Presentation_A4	5'		L1 One-To-One_A3
	5'		L2 Group Or.Presentation_A2	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire	5'		Post-Task Questionnaire
	5'		Post-Task Questionnaire	5'		Idiodynamic Task	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		Idiodynamic Task	5'		Stimulated Recall	1.5'		Preparation Time	1.5'		Preparation Time
							5'		L1 Group Or.Presentation_D4	5'		L1 One-To-One_D3
8:00	10'	S1	Stimulated Recall	10'	S 2	Preparation Time	5'	S3	Post-Task QUESTIONNAIRE	5'	S4	Post-Task QUESTIONNAIRE
	1.5'		Preparation Time	1.5'		L2 One-To-One_D1	5'		Idiodynamic Task	5'		Idiodynamic Task
	5'		L2 Group Or.Presentation_D2	5'		Post-Task Questionnaire	10'		Elicited Imitation Task	10'		Yes/No Vocabulary Size
	5'		Post-Task Questionnaire	5'		Idiodynamic Task				10'		Woking Memory Task (Digit)
	5'		Idiodynamic Task	5'		Stimulated Recall	10'		Attention Switch-Trail			
	10'		Stimulated Recall	10'								

Appendix A2: Bioethics Committee of the Universitat de Barcelona Resolution

UNIVERSITAT DE BARCELONA

Oficina de Gestió de la Recerca
Pavelló Rosa (recinte Maternitat) primer pis
Travessera de les Corts, 131-159 93-4035398
08028 Barcelona

Bioethics Committee of the Universitat de Barcelona

Gemma Marfany Nadal, as Academic Secretary of the Committee of Bioethics of Universitat de Barcelona (CBUB).

CERTIFIES:

that the application for a project Thesis presented by Ms. **Gisela Sosa López**, PhD student of the Department of Modern Languages and Literatures and English Studies, at Universitat de Barcelona, entitled **"The Role of L2 Speaking Anxiety on L2 Speaking Fluency, Pronunciation and Comprehensibility in Instructed English"** and supervised by Dr. **Joan Carles Mora Bonilla**, has been evaluated by the CBUB. This Committee agreed on February 15, 2021 that this PhD project conforms to the bioethics directives of Universitat de Barcelona, and therefore, the CBUB considers favourably the Thesis project.

To this end, the Secretary of the CBUB signs this document, with the acknowledgement of the President of this Committee, in Barcelona, March 04, 2021.

Firmado por MARFANY NADAL, GEMMA (FIRMA) el día 04/03/2021 con un certificado emitido por AC DNIE 005

Gemma Marfany

Acknowledged by
President of the Bioethics Committee of the Universitat de Barcelona

35030054L
JORDI GARCIA
(R: Q0818001J)

Firmado digitalmente por 35030054L JORDI GARCIA (R: Q0818001J)
Fecha: 2021.03.05 11:08:25 +01'00'

Jordi Garcia-Fernàndez

Institutional Review Board (IRB00003099)

Appendix B: Summary of Participants' demographics

Table B.1: Participants' background information. *M= Mean, SD= Standard Deviation, AOL= Age of Learning, Cat= Catalan, Sp= Spanish, FCE= First Certificate, CAE= Certificate in Advanced English, CPE= Certificate of Proficiency in English, RP= Received Pronunciation, GA= General American*

***N* = 86**			
Sex (N)	Male = 12 (13.95%)	Female = 74 (86.04%)	
Pathologies	No = 84 (97%)	Yes = 2 (3%)	
Pathology type	Hearing = 1	Speech = 1	
Clinical anxiety	No= 86 (100%)	Yes= 0 (0%)	
Mother Tongue (N)	Cat/Sp = 75 (87%)	Other L1s= 11 (13%)	
L1 Daily Use (%)	Cat = 70.49%	Sp= 29.50%	Cat. Range = 20-100 Sp. Range = 10-95
L2 (N)	English = 86 (100%)		
L2 Academy exposure (h/week)	*M* = 2.69	*SD* = 0.96	
L2 certificates (N)	No = 27 (31.39%)	Yes = 73 (84.88%)	FCE → *N* = 29, 33.72% FCE and CAE → *N* = 11, 12.79% CAE (only) → *N* = 8, 9.30% TOEFL → *N* = 2, 2.32% IELTS → *N* = 3, 3.48% CPE → *N* = 1, 1.16% B1 → *N* = 5, 5.81%
British exposure	*M* = 41.68	*SD* = 19.93	
American exposure	*M* = 56.04	*SD* = 20.50	
Accent production	RP = 24 (27.90%)	GA = 56 (65.11%)	Don't know = 5 (5.81%)
Stays Abroad (+3 weeks)	No = 54 (62.79%)	Yes = 14(16.27%)	
Stay Abroad duration	*M* = 19.40 weeks	*SD* = 26.11	Range = 3-96 weeks

Table B.2: Participants' proficiency, WM and attention information.

Variables	*M*	*SD*	95% CI	
Age	20.16	4.09	19.28	21.04
AOL English (years)	6.57	3.24	5.88	7.26
INPUT				
Spoken L2 Input*[35]	4.33	0.73	4.17	4.48
Listening to songs in English*	4.88	0.47	4.78	4.98
Watching English language televisión*	3.64	1.42	3.33	3.94
Watching films or videos in English*	4.60	0.60	4.48	4.73
Written L2 Input*	3.24	1.06	3.01	3.46
Read newspapers or magazines in English*	3.08	1.28	2.81	3.36
Read books in English*	3.45	1.30	3.17	3.73
OUTPUT				
Spoken L2 Output*	3.37	0.95	3.17	3.57
Speaking English with natives or fluent speakers*	3.12	1.24	2.85	3.38
Speaking English with non-natives speakers*	3.69	1.15	3.44	3.93
Written L2 Output*	3.18	1.02	2.96	3.40
Writing emails or letters in English*	3.21	1.00	2.99	3.43
Self-estimated Overall Proficiency**[36]	6.86	1.19	6.61	7.12
Command of Reading**	7.5	1.22	7.24	7.76
Command of Listening**	7.16	1.26	6.89	7.43
Command of Speaking**	6.78	1.23	6.52	7.04
Command of Writing**	6.72	1.36	6.43	7.01
Command of Pronunciation**	6.51	1.41	6.21	6.81
Elicited Imitation (0-120)	97.42	13.21	96.43	98.41

[35] * 1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily
[36] **(Reading, Listening, Speaking, Writing): Likert scale 1=VERY POOR and 9=NATIVE-LIKE

Vocabulary size (0-10.000 words)	6546.92	1211.43	6287.19	6806.65
Forward digit-span (TE_ML)	5.86	1.24	5.59	6.12
Forward digit-span (mean span)	6.13	1.10	5.90	6.37
Backward digit-span (TE_ML)	5.53	1.56	5.20	5.86
Backward digit-span (mean span)	5.63	1.42	5.32	5.93
WM (fWM + bWM)	11.40	2.39	11.22	11.57
Attention part A (seconds)	32.84	12.8	31.88	33.80
Attention part B (seconds)	68.08	25.18	66.19	69.96
Time difference	35.23	21.33	33.63	36.83
TMT ratio 1	3.58	12.93	2.61	4.54
TMT ratio 2	2.58	12.93	1.61	3.54

Appendix C: Audience Background Questionnaire

Statement of confidentiality: Your name and other information gathered in this study will be codified and won't be disclosed to any person other than the investigator. It will only be used for statistical purposes without reference to individual participants' personal information.

- Please, write down your Name Surname *
- Age *
- Gender *

Mark only one oval.

Male ⬭
Female ⬭
Prefer not to say ⬭

- Studies *

English Learning Experience

1. Age at which you started learning English: *

2. Have you taken English courses/private classes outside your school/university? *

Mark only one oval.

Yes ⬭ *Skip to question 2.1*
No ⬭ *Skip to question 3*

2.1. Please, specify WHEN –whether it was in Primary/Secondary/High school/University– WHERE (e.g., British Council) and HOW LONG you took those classes (e.g., 2 years, 4 hours a week).

3. Do you have any Certificate(s) of English level? *

Tick all that apply.

No, I don't. ⬭
First Certificate in English (FCE) ⬭
Certificate in Advanced English (CAE) ⬭
Certificate of Proficiency in English (CPE) ⬭

Proficiency in English (IELTS)	⬭
TOEFL	⬭
Other	⬭ ____________________

4. Rate your command of English on the scale: **1=VERY POOR and 9=NATIVE-LIKE** *

	1	2	3	4	5	6	7	8	9
Reading	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Listening	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Speaking	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Writing	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Pronunciation	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

5. For each of the items below, choose the response that corresponds to the amount of time you estimate you spend on average doing each activity in English. *
Please, use this scale **1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily**

Mark only one oval per row.

	1	2	3	4	5
Watching English language television	⬭	⬭	⬭	⬭	⬭
Watching movies or videos in English	⬭	⬭	⬭	⬭	⬭
Reading newspapers/magazines in English	⬭	⬭	⬭	⬭	⬭
Reading books in English	⬭	⬭	⬭	⬭	⬭
Listening to songs in English	⬭	⬭	⬭	⬭	⬭
Speaking English with native or fluent speakers	⬭	⬭	⬭	⬭	⬭
Speaking English with nonnative speakers	⬭	⬭	⬭	⬭	⬭
Writing e-mails/letters in English	⬭	⬭	⬭	⬭	⬭

7. Have you ever lived in an English-speaking country more than 2 weeks? *

Mark only one oval.

Yes ⬭ *Skip to question 7.1*

No ⬭ *Skip to question 8*

7.1. Please, specify HOW MANY TIMES you have lived in an English-speaking country. For each English-speaking country, please specify (1) LOCATION, (2) DATES, (3) DURATION and (4) PURPOSE of your stay(s):

8. Do you have any speech or hearing problems/pathology? *

Mark only one oval.

Yes ⬭ *Skip to question 8.1*

No ⬭

8.1. Please, indicate which kind of speech or problem you have:

To finish, please click on "**SUBMIT**". THANKS!

Appendix D1: Summary of the audience members' demographics in 2021-2022

Table D1.1: Audience members in 2021-2022 linguistic information. *M= Mean, SD= Standard Deviation, Cat= Catalan, Sp= Spanish, FCE= First Certificate, CAE= Certificate in Advanced English.*

***N*=7**			
Sex (N)	Male = 2 (28.57%)	Female = 5 (71.42%)	
Pathologies	No = 7 (100%)	Yes = 0 (0%)	
Mother Tongue (N)	Cat/Sp = 5 (71.42%)	Sp-English = 1 (14.28%)	Hindi-English = 1 (14.28%)
L2 (N)	English = 5 (100%)		
L2 Academy exposure (h/week)	*N* = 4, *M* = 2.87	*SD* = 0.89	Range = 1.5- 4 h/week
L2 certificates (N)	No = 2 (28.57%)	Yes = 5 (71.42%)	FCE → *N* = 4, 80% CAE→ *N* = 1, 20%
Stays Abroad (+3 weeks)	No = 4 (57.14%)	Yes = 3 (42.85%)	
Stay Abroad duration	*M* = 10.66 weeks	*SD* = 11.54	Range = 4-24 weeks

Table D1.2: Helper's linguistic information. *M= Mean, SD= Standard Deviation, Cat= Catalan, Sp= Spanish.*

***N*=1**		
Sex (N)	Female	
Pathologies	No	
Mother Tongue (N)	Marathi, Hindi and English	
Mother's L1	Marathi	
Father's L1	English	
Can read Spanish*[37]	5	
Can understand Spanish*	4	
Can speak Spanish*	4	
SCORE Spanish	*M* = 4.33	*SD* = 0.57
Can read Catalan*	4	
Can understand Catalan*	3	
Can speak Catalan*	2	
SCORE Catalan	*M* = 3.00	*SD* = 1.00
Years learning Spanish	8 months	
Years learning Catalan	0	
L2 Academy exposure (h/week)	5 h/week	
L2 certificates (N)	Yes, C1 DELE	
Stays Abroad (+3 weeks)	No	
Years living in Spain	9	
Familiarity with Spanish-accented English**[38]	9	
Familarity with American English**	9	
Familarity with British English**	9	
Rate knowledge of Spanish***[39]	8	

[37] * 1= HARDLY, 5= VERY WELL
[38] ** 1= not at all familiar, 9= very familiar
[39] *** 1=I don't speak Spanish, 9= I speak Spanish almost like a native

Rate everyday use of Spanish****[40]	7
Rate everyday exposure to Spanish-Accented English*****[41]	8
Hearing impairment	No

Table D1.3: Audience members' proficiency in 2021-2022

Variables	*M*	*SD*	95% CI	
Age	26.71	4.78	22.29	31.14
AOL English (years)	5.86	4.48	1.71	10.01
INPUT				
Spoken L2 Input*[42]	4.19	0.50	2.77	4.78
Listening to songs in English*	4.57	0.78	3.84	5.30
Watching English language televisión*	3.43	1.13	2.38	4.48
Watching films or videos in English*	4.57	0.53	4.08	5.07
Written L2 Input*	3.71	1.29	2.38	4.50
Read newspapers or magazines in English*	3.71	1.25	2.55	4.87
Read books in English*	3.71	1.49	2.33	5.10
OUTPUT				
Spoken L2 Output*	2.57	0.93	1.64	3.13
Speaking English with natives or fluent speakers*	2.71	1.25	1.55	3.87
Speaking English with non-natives speakers*	2.43	0.97	1.53	3.33
Written L2 Output*	3.00	1.53	1.75	3.92
Writing emails or letters in English*	3.00	1.52	1.59	4.41
Self-estimated Overall Proficiency**[43]	7.74	1.26	5.16	8.88
Command of Reading**	8.29	0.75	7.59	8.98

[40] **** 1= I never speak Spanish, 9= I speak Spanish all the time
[41] ***** 1= never, 9= all the time
[42] * 1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily
[43] **(Reading, Listening, Speaking, Writing): Likert scale 1=VERY POOR and 9=NATIVE-LIKE

Command of Listening**	8.14	0.90	7.31	8.97
Command of Speaking**	7.29	2.13	5.31	9.26
Command of Writing**	7.86	0.90	7.03	8.69
Command of Pronunciation**	7.14	2.11	5.19	9.10

Appendix D2: Summary of the audience members' demographics in 2022-2023

Table D2.1: Audience members in 2022-2023 linguistic information. *M= Mean, SD= Standard Deviation, Cat= Catalan, Sp= Spanish, FCE= First Certificate.*

***N*=7**			
Sex (N)	Male = 3 (42.85%)	Female = 4 (57.14%)	
Pathologies	No = 7 (100%)	Yes = 0 (0%)	
Mother Tongue (N)	Cat/Sp = 3 (42.85%)	English = 2 (28.57%)	Polish = 1 (14.28%) Hindi-English = 1 (14.28%)
L2 (N)	English = 4	Spanish = 2	
L2 Academy exposure (h/week)	*N* = 4, *M* = 2.5	*SD* = 1.29	Range = 1-4h/week
L2 certificates (N)	No = 4 (57.14%)	Yes = 3 (42.85%)	FCE → *N* = 2, 66.66% IELTS → *N* = 1, 33.33%
Stays Abroad (+3 weeks)	No = 4 (57.14%)	Yes = 3 (42.85%)	
Stay Abroad duration	*M* = 14.66 weeks	*SD* = 18.47	Range = 4-36 weeks

Table D2.2: Native speakers of English in the Audience 2022-2023 linguistic information. *M= Mean, SD= Standard Deviation, Cat= Catalan, Sp= Spanish.*

***N*=2**		
Sex (N)	Male = 1	Female = 1
Pathologies	No	
Mother Tongue (N)	English	
Mother's L1	English	
Father's L1	English	
Can read Spanish*[44]	*M*= 5	*SD*= 0
Can understand Spanish*	*M*= 5	*SD*= 0
Can speak Spanish*	*M*= 4	*SD*= 0
SCORE Spanish	*M*= 4.67	*SD*= 0
Can read Catalan*	*M*= 3	*SD*= 1.41
Can understand Catalan*	*M*= 3	*SD*= 0
Can speak Catalan*	*M*= 2	*SD*= 1.41
SCORE Catalan	*M*= 2.67	*SD*= 0.94
Years learning Spanish	*M*= 16	*SD*= 15.56
Years learning Catalan	*M*= 3	*SD*= 0
Years living in Spain	*M*= 6.50	*SD*= 2.12
Familiarity with Spanish-accented English**[45]	*M*= 9	*SD*= 0
Familarity with American English**	*M*= 9	*SD*= 0
Familarity with British English**	*M*= 9	*SD*= 0
Rate knowledge of Spanish***[46]	*M*= 7.50	*SD*= 0.71

[44] * 1= HARDLY, 5= VERY WELL
[45] ** 1= not at all familiar, 9= very familiar
[46] *** 1=I don't speak Spanish, 9= I speak Spanish almost like a native

Rate everyday use of Spanish****[47]	*M*= 4	*SD*= 2.83
Rate everyday exposure to Spanish-Accented English*****[48]	*M*= 5	*SD*= 2.83

Table D1.3: Audience members' proficiency in 2022-2023

Variables	*M*	*SD*	*95% CI*	
Age	33.71	9.93	24.53	42.90
AOL English (years)	3.57	3.69	0.16	6.98
INPUT				
Spoken L2 Input*[49]	4.67	0.47	4.23	5.10
Listening to songs in English*	4.86	0.37	4.51	5.21
Watching English language televisión*	4.43	1.13	3.38	5.48
Watching films or videos in English*	4.71	0.48	4.26	5.17
Written L2 Input*	4.29	0.70	3.64	4.93
Read newspapers or magazines in English*	4.71	0.48	4.26	5.17
Read books in English*	3.86	1.21	2.73	4.98
OUTPUT				
Spoken L2 Output*	4.21	0.95	3.33	5.09
Speaking English with natives or fluent speakers*	4.12	1.21	3.02	5.27
Speaking English with non-natives speakers*	4.29	1.11	3.26	5.31
Written L2 Output*	4.43	1.13	3.38	5.48
Writing emails or letters in English*	4.43	1.13	3.38	5.48
Self-estimated Overall Proficiency**[50]	8.26	0.94	7.38	9.13
Command of Reading**	8.71	0.75	8.02	9.41
Command of Listening**	8.29	0.95	7.41	9.17

[47] **** 1= I never speak Spanish, 9= I speak Spanish all the time
[48] ***** 1= never, 9= all the time
[49] * 1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily
[50] **(Reading, Listening, Speaking, Writing): Likert scale 1=VERY POOR and 9=NATIVE-LIKE

Command of Speaking**	8.14	1.21	7.02	9.27
Command of Writing**	8.14	1.21	7.02	9.27
Command of Pronunciation**	8.00	1.00	7.08	8.92

Appendix E: Native Raters' background questionnaire

Questionnaire

Statement of confidentiality:
Your name and other information gathered in this study will not be disclosed to any person other than the investigator and his collaborators and will only be used for statistical purposes without reference to individual participants' personal information.

Participant Name/Code: Date and time:

Personal Information

- Age: Occupation:
- Place of birth:
- Mother's place of birth:
- Father's place of birth:
- Your current place of residence:
- Previous place(s) of residence (where you have lived for at least a few months; indicate when and for how long):

Language information

- Native language (NL):
- Mother's NL: Father's NL:
- Do you speak any other language fluently?
- Knowledge of Spanish and Catalan (use the 1-5 scale to indicate your self-estimated level, where 1=hardly, 5=very well)

	I can read it	I can understand it	I can speak it
Spanish	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Catalan	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

I have been learning Spanish for

I have been learning Catalan for

- Language use (weekly average). Indicate appropriate % for each language you speak.

English:	0-25%	25-49%	50-74%	75-100%	100%
Spanish:	0-25%	25-49%	50-74%	75-100%	100%
Catalan:	0-25%	25-49%	50-74%	75-100%	100%
_______	0-25%	25-49%	50-74%	75-100%	100%

-Age: -Place of Birth: -English accent:

-Profession: -Years living in Spain (or Spanish-speaking country):

-Non-native Languages you speak (specify level 1-10):

-Please describe your accent of English:

-Please answer the following questions:

1a. Rate your familiarity with Spanish-accented English (tick the appropriate box)

1	2	3	4	5	6	7	8	9

not at all familiar **very familiar**

1b. Rate your familiarity with American English (tick the appropriate box)

1	2	3	4	5	6	7	8	9

not at all familiar **very familiar**

1c. Rate your familiarity with British English (tick the appropriate box)

1	2	3	4	5	6	7	8	9

not at all familiar **very familiar**

2. Rate your knowledge of Spanish (tick the appropriate box)

1	2	3	4	5	6	7	8	9

I don't speak Spanish **I speak Spanish almost like a native**

3. Rate your everyday use of Spanish (tick the appropriate box)

1	2	3	4	5	6	7	8	9

I never speak Spanish **I speak Spanish all the time**

4. Rate your everyday exposure to Spanish-accented English (tick the appropriate box)

1	2	3	4	5	6	7	8	9

never **all the time**

I would be willing to participate in this research project rating non-native speech samples

YES	NO

I would be willing to participate in this research project recording speech samples in English

YES	NO

Do you have any hearing impairments?

YES	NO

If yes, please specify:

Appendix F: Summary of Native Raters' demographics

Table F1: Native raters demographics.

***N* = 7**			
Sex (N)	Male = 2 (28.57%)	Female = 5 (71.42%)	
Pathologies	No = 7 (100%)	Yes = 0 (0%)	
Mother Tongue (N)	English = 7 (100%)		
Mother's L1	English = 6 (85.71%)	Portuguese = 1 (14.28%)	
Father's L1	English = 7 (100%)		
Accent	RP = 3 (42.85%)	GA = 4 (57.14%)	
Years living in Spain	*M* = 5.43	*SD* = 2.29	Range = 2-8 years

Table F2: Native raters' Spanish and Catalan proficiency information. *M= Mean, SD= Standard Deviation,*

Variables	*M*	*SD*	*95% CI*	
Age	31.86	3.80	28.34	35.38
Can read Spanish*[51]	4.14	0.90	3.31	4.97
Can understand Spanish*	3.71	1.38	2.44	4.99
Can speak Spanish*	3.43	1.27	2.25	4.61
Score Spanish	3.76	1.16	2.68	4.84
Can read Catalan*	2.57	1.13	1.52	3.62
Can understand Catalan*	2.29	1.11	1.26	3.31
Can speak Catalan*	1.71	0.95	0.83	2.59
Score Catalan	2.19	0.99	1.26	3.11
Years learning Spanish	9.29	10.51	-.44	19.01
Years learning Catalan	1.64	1.59	0.16	3.12
Years living in Spain	5.43	2.29	3.30	7.55
Familiarity with Spanish-accented English**[52]	8.29	1.11	7.26	9.31
Familarity with American English**	8.86	0.37	8.51	9.21

[51] * 1= HARDLY, 5= VERY WELL
[52] ** 1= not at all familiar, 9= very familiar

Familarity with British English**	7.29	1.97	5.46	9.11
Rate knowledge of Spanish***[53]	5.71	2.28	3.60	7.83
Rate everyday use of Spanish****[54]	3.57	2.37	1.38	5.76
Rate everyday exposure to Spanish-Accented English*****[55]	6.43	2.37	4.24	8.62

Note: The score of Spanish and Catalan knowledge was computed from adding the scores and dividing

[53] *** 1=I don't speak Spanish, 9= I speak Spanish almost like a native
[54] **** 1= I never speak Spanish, 9= I speak Spanish all the time
[55] ***** 1= never, 9= all the time

Appendix G: Speaking tasks

L2 tasks:

1. Describe the **hardest and toughest challenge** in your life

   Make sure you speak for a few minutes (5 min.). You should say:

   - When and where?
   - How old were you?
   - Why did you encounter this challenge?
   - Why was it so challenging?
   - Did anybody (e.g., friends, family, etc.) help you?
   - What did you learn from this experience?
   - Would you like to go through the same experience again?

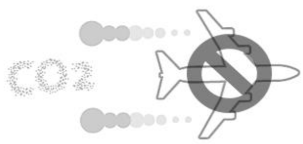
2. **Transport**: Discuss (5 min.) which way of transportation is better. Bear in mind:
   - Accessibility within **<u>and</u>** out of the city
   - Environment and pollution (e.g., of the air)
   - Mobility
   - Other aspects you may consider important.

3. Describe a **memorable (childhood) experience**
   - Make sure you speak for a few minutes (5 approx.). You should say:
     - What happened?
     - When and where?
     - How old were you?
     - Why is it memorable?
     - What did you learn from this experience?
     - Would you like to go through the same experience again?

4. **Global warming**: Discuss (5 min.) which option is better to stop global warming.

   Bear in mind:

   - Cooperation by the citizens.
   - The impact on the environment
   - Whether it is realistic and doable.
   - The economic impact
   - Other aspects you may consider important

		
1	2	3

L1 tasks:

1. Describe un momento en el que hayas pasado mucha **frustración**.

   Tendrás que hablar unos 5 minutos. Deberías explicar:

   – ¿Cuándo y dónde?
   – ¿Cuántos años tenías?
   – ¿Porque te sentiste así?
   – ¿Alguien te ayudó (por ejemplo, amigos, familiares, etc.)?
   – ¿Qué aprendiste de esa experiencia?
   – ¿Te gustaría volver a vivir la misma experiencia?

2. **Learning**: Argumenta (en 5 minutos) cuales de estos 3 tipos de aprendizaje es mejor.

   Ten en cuenta:

   – El rol del alumno
   – Flexibilidad horaria y costes (€)
   – Popularidad
   – Otros aspectos que consideres importantes

		
1	2	3
		(learn by yourself)

3. Describe el **mejor viaje** que hayas hecho

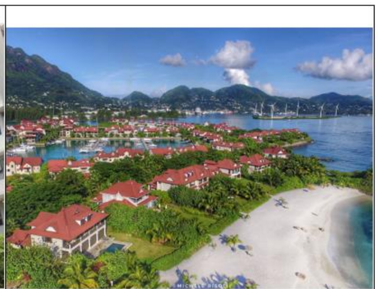
   Tendrás que hablar unos 5 minutos. Deberías explicar:

   – ¿Cuándo y dónde?
   – ¿Qué hiciste? ¿Con quién?
   – ¿Cuántos años tenías?
   – ¿Por qué es memorable?
   – ¿Qué aprendiste de esa experiencia?
   – ¿Te gustaría volver a vivir la misma experiencia?

4. **Opciones de vivienda**:

   Argumenta (en 5 minutos) en cual de estos 3 sitios es mejor vivir. Ten en cuenta:

   – Movilidad
   – Opciones de trabajo/estudios y ocio
   – Calidad de vida
   – Otros aspectos que consideres importantes

		
1	2	3

Appendix H: LBQ

This questionnaire consists of different sections including English learning experience, Speaking anxiety and Language background.

It will take you 15 minutes to complete.

All the data collected in this study will be anonymized and secured in a safe place. Your identity will be held confidentially in reports in which the study may be published, and databases in which results may be stored. Your data will be de-identified and only the researchers of this study will have access to it.

Thank you for your participation!

* Indicates required question

- Please, write down your Name, Surname and NiUB *

e.g., Natalia Mora - 1788969

- Please, write down the group you belong to (for the credit) *

Lengua Anglesa II: G2

- Gender *

Mark only one oval.

Male ◯

Female ◯

Prefer not to say ◯

- Age *

English Learning Experience

1. Age at which you started learning English: *

2. Have you taken English courses/private classes outside your school/university? *

Mark only one oval.

Yes ◯ *Skip to question 2.1*

No Skip to question 3

English courses outside school

2.1. Please, specify WHEN –whether it was in Primary/Secondary/High school/University– WHERE (e.g., British Council) and HOW LONG you took those classes (e.g., 2 years, 4 hours a week). *

3. Do you have any Certificate(s) of English level? *

Tick all that apply.

No, I don't.
First Certificate in English (FCE)
Certificate in Advanced English (CAE)
Certificate of Proficiency in English (CPE)
Proficiency in English (IELTS)
TOEFL
Other ____________________

4. Estimate the % of exposure to British and American English (or other varieties): *
For example: British 50% - American 50% OR Irish 20% - British 30% - American 50%

5. Do you think your pronunciation is more British- or American-like? *

Mark only one oval.

British
American
Other ____________________

6. For each of the items below, choose the response that corresponds to the amount of time you estimate you spend on average doing each activity in English. *
Please, use this scale **1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily**

Mark only one oval per row.

	1	2	3	4	5

Watching English language television					
Watching movies or videos in English					
Reading newspapers/magazines in English					
Reading books in English					
Listening to songs in English					
Speaking English with native or fluent speakers					
Speaking English with nonnative speakers					
Writing e-mails/letters in English					

7. Rate your command of English on the scale: **1=VERY POOR and 9=NATIVE-LIKE** *

	1	2	3	4	5	6	7	8	9
Reading									
Listening									
Speaking									
Writing									
Pronunciation									

8. Have you ever lived in an English-speaking country more than 2 weeks? *

Mark only one oval.

Yes *Skip to question 8.1*

No *Skip to question 9*

Stays in English-speaking countries

8.1. Please, specify HOW MANY TIMES you have lived in an English-speaking country. For each English-speaking country, please specify (1) LOCATION, (2) DATES, (3) DURATION and (4) PURPOSE of your stay(s):

9. Have you previously received any type of Phonetic Training/ Pronunciation instruction in English? *

Mark only one oval.

Yes ⬭ *Skip to question 9.1*

No ⬭ *Skip to question 10*

Phonetic training/Pronunciation instruction in English

9.1. Please, specify WHERE, WHEN and for HOW LONG you received phonetic training/pronunciation instruction in English:

10. Do you suffer from clinical anxiety? *
Excessive feelings of worry or persistent, even intrusive, thoughts about certain fears or constant fear in general.

Mark only one oval.

Yes ⬭

No ⬭

11. Do you have any speech or hearing problems/pathology? *

Mark only one oval.

Yes ⬭ *Skip to question 11.1*

No ⬭ *Skip to question 12*

11.1. Please, indicate which kind of speech or problem you have:

Speaking Anxiety

Now you will be asked about your perceptions and thoughts regarding speaking anxiety (feeling nervous) in relation to English (L2), your mother tongue (L1, for example, Catalan)

Nervousness when speaking your L2 (English)

12. Rate your anxiety levels when speaking ENGLISH in the following contexts. *
1 = Extremely RELAXED (e.g., as if you were talking with friends comfortably), **9= Extremely ANXIOUS** (e.g., having palpitations or feeling your heartbeat accelerated or sweating a lot):

	1	2	3	4	5	6	7	8	9
Talking English to your teachers	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

Talking in English to non-native speakers of English									
Talking in English to native speakers of English									
Doing an oral presentation in English									
Doing a task with a peer (e.g., a dialogue) in English									
Taking part in a group discussion in class in English									

13. Rate the extent to which you agree/disagree with the following statements when speaking ENGLISH. *
1= **STRONGLY DISAGREE** (definitely not true about you), 9= **STRONGLY AGREE** (definitely true about you)

Mark only one oval per row

	1	2	3	4	5	6	7	8	9
I fear others might find my pronunciation of English foreign-accented.									
I fear others might find difficult to understand me when I speak English.									
I get nervous when I can't find the word to say.									
I feel stressed knowing someone is listening to me.									
English pronunciation is difficult for me.									
My pronunciation is at a lower level than that of people around me.									
I'd rather people did not hear me making									

a pronunciation mistake									
I feel stressed knowing that others are listening to me.									
I (would) feel more embarrassed making a pronunciation mistake than any other type of mistake when I speak in English.									
I feel sure of myself when speaking in English									
I am calm when I have to speak in English without preparation time									
I enjoy expressing my ideas in English									
I remember the pronunciation of new words easily									
I am satisfied with my current level of English speaking									
Whenever I think of my future career, I imagine myself using English with native-like pronunciation									
I like speaking to myself in English									
I think I have a talent to acquire English pronunciation									
Being understandable depends on my pronunciation									

14. Rate the extent to which you agree/disagree with the following statements when speaking ENGLISH. *
1= **STRONGLY DISAGREE** (definitely not true about you), 9= **STRONGLY AGREE** (definitely true about you)

Mark only one oval per row.

	1	2	3	4	5	6	7	8	9
I never feel quite sure of myself when I am speaking in English in class									
I don't worry about making mistakes in English									
I tremble when I know that I'm going to be called on in class									
It frightens me when I don't understand what the teacher is saying in the foreign language									
During class, I find myself thinking about things that have nothing to do with the course									
I keep thinking that the other students are better at languages than I am									
I am usually at ease during tests									
I start to panic when I have to speak without preparation in class									
It embarrasses me to volunteer answers in class									
I worry about the consequences of failing my foreign language classes									
I would not be nervous speaking English with native speakers									
I get upset when I don't understand									

what the teacher is correcting									
Even if I am well prepared for class, I feel anxious about it									
I can feel my heart pounding when I'm going to be called on in class									
I would probably feel comfortable around native speakers of English									

Nervousness when speaking your mother tongue or L1 (e.g. Spanish)

15. Rate your anxiety levels when speaking your L1 (e.g. Spanish) in the following contexts. *
1 = **Extremely RELAXED** (e.g. as if you were talking with friends comfortably), 9= **Extremely ANXIOUS** (e.g. having palpitations or feeling your heartbeat accelerated or sweating a lot):

	1	2	3	4	5	6	7	8	9
Talking to your teachers									
Doing an oral presentation									
Doing a task with a peer (e.g., a dialogue)									
Taking part in a group discussion in class									
Talking to your classmates									

16. Rate the extent to which you agree/disagree with the following statements when speaking your L1 (e.g., SPANISH). *
1= **STRONGLY DISAGREE** (definitely not true about you), 9= **STRONGLY AGREE** (definitely true about you)

Mark only one oval per row.

	1	2	3	4	5	6	7	8	9
While participating in a conversation with a new acquaintance I feel nervous									

I have no fear of facing an audience	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I usually talk less because I'm shy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I look forward to expressing my opinions in class	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I am afraid to express myself in a group	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I look forward to an opportunity to speak in public	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I find the prospect of speaking mildly pleasant	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
When communicating, my posture feels strained and unnatural	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I am tense when participating in group discussions	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Although I talk fluently with friends, I am at loss for words when doing a presentation	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I have no fear of expressing myself in a group	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
My hands tremble when I try to handle objects in a presentation	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I always avoid speaking in public	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I feel that I am more fluent when talking to people than most other people	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I'm tense while I'm speaking beore a group of people	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
My thoughts become confused when I speak before an audience	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I like to get involved in group discussions	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

Although I am nervous before a presentation, I soon forget my fears and I enjoy the experience	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Conversing with people who hold positions of authority cause me tension	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I feel relaxed and comfortable while speaking in any context	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I feel nervous when I am called upon to answer a question or give an opinion in class	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I face the prospect of making a speech with complete confidence	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I'm afraid to speak up in a conversation									
I would enjoy presenting a speech on a local television show	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

Language Background

17. Indicate your mother tongue (first language or L1):
Language(s) you learnt to speak from birth. *
Tick all that apply:

Catalan ⬭
Spanish ⬭
English ⬭
Other ⬭ ____________________

18. Indicate which language(s) you normally speak on a daily basis: *
You can tick more than one box.

Catalan ⬭
Spanish ⬭
English ⬭
Other ⬭ ____________________

18.1. Estimate the % of DAILY USE of the languages you speak: *
Type in the number from 0 to 100. Please, make sure that the percentages add up to 100.
For example: Catalan 70, Spanish 20 and English 10 = 100

To finish, please click on "**SUBMIT**". THANKS!

Appendix I: Audience's Language Background Questionnaire

This questionnaire consists of two sections including your Language Background and English learning experience. It will take you 10 minutes to complete.

All the data collected in this study will be anonymized and secured in a safe place. Your identity will be held confidently in reports in which the study may be published, and databases in which results may be stored. Your data will be de-identified and only the researchers of this study will have access to it.

Thank you for your participation!

* Indicates required question

- Please, write down your Name, Surname*

e.g., Natalia Mora

- Gender *

Mark only one oval.

Male ⬭

Female ⬭

Prefer not to say ⬭

- Age *

- Studies *

English Learning Experience

1. Age at which you started learning English: *

2. Have you taken English courses/private classes outside your school/university? *

Mark only one oval.

Yes ⬭ *Skip to question 2.1*

No ⬭ *Skip to question 3*

English courses outside school

2.1. Please, specify WHEN –whether it was in Primary/Secondary/High school/University– WHERE (e.g., British Council) and HOW LONG you took those classes (e.g., 2 years, 4 hours a week). *

3. Do you have any Certificate(s) of English level? *

Tick all that apply.

- No, I don't.
- First Certificate in English (FCE)
- Certificate in Advanced English (CAE)
- Certificate of Proficiency in English (CPE)
- Proficiency in English (IELTS)
- TOEFL
- Other ____________________

4. For each of the items below, choose the response that corresponds to the amount of time you estimate you spend on average doing each activity in English. *
Please, use this scale **1=never; 2=a few times a year; 3=monthly; 4=weekly; 5=daily**

Mark only one oval per row.

	1	2	3	4	5
Watching English language television					
Watching movies or videos in English					
Reading newspapers/magazines in English					
Reading books in English					
Listening to songs in English					
Speaking English with native or fluent speakers					
Speaking English with nonnative speakers					
Writing e-mails/letters in English					

5. Rate your command of English on the scale: **1=VERY POOR and 9=NATIVE-LIKE** *

	1	2	3	4	5	6	7	8	9
Reading									
Listening									
Speaking									
Writing									
Pronunciation									

68. Have you ever lived in an English-speaking country more than 2 weeks? *

Mark only one oval.

Yes *Skip to question 6.1*

No *Skip to question 7*

Stays in English-speaking countries

6.1. Please, specify HOW MANY TIMES you have lived in an English-speaking country. For each English-speaking country, please specify (1) LOCATION, (2) DATES, (3) DURATION and (4) PURPOSE of your stay(s):

To finish please click on "Submit". Thanks!

Appendix J: Post-task questionnaire

Now you will answer some questions about your task performance.

* Indicates required question

Participant ID code:

1. How WELL do you think you have performed the task?

	1	2	3	4	5	6	7	8	9	
Very poorly	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very well

2. How DIFFICULT did you find the task?

	1	2	3	4	5	6	7	8	9	
Not difficult at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

3. Did you feel ANXIOUS when performing the task?

	1	2	3	4	5	6	7	8	9	
I did not feel anxious at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	I felt very anxious

4. From 1 to 9, how did these factors made you feel when performing the tasks?
9 = the **highest anxiety** you have ever felt (e.g. you had palpitations or felt your heartbeat very accelerated, you sweat a lot, you felt sick) 1 = **totally relaxed** (e.g. as if you were talking with friends comfortably)

	1	2	3	4	5	6	7	8	9
The reactions of the people I was speaking to	◯	◯	◯	◯	◯	◯	◯	◯	◯
Speaking in English (regardless to whom)	◯	◯	◯	◯	◯	◯	◯	◯	◯
Speaking English in front of someone/people	◯	◯	◯	◯	◯	◯	◯	◯	◯
Being audio recorded	◯	◯	◯	◯	◯	◯	◯	◯	◯
Being video recorded	◯	◯	◯	◯	◯	◯	◯	◯	◯

Being monitored with the anxiety sensor (the device on your wrist)	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Knowing these tasks are worth a percentage of your grade	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Forgetting what I wanted to say	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

5. Did any other factor that does not appear on the previous list make you anxious? If so, please specify it.

6. How much MENTAL EFFORT did you have to put to perform this task?

	1	2	3	4	5	6	7	8	9	
No mental effort at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extreme mental effort

7. Rate the extent to which you agree/disagree with the following statements when performing the task.

	1	2	3	4	5	6	7	8	9
I found it easy to plan what I wanted to say in this task	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I think it was easy to understand my speech	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I easily recalled all the words I needed	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I felt I hesitated a lot while speaking	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭
I couldn't remember the pronunciation of some words	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

To finish, please click on "**SUBMIT**". THANKS!

Appendix K: Audience's Questionnaire

Participants code: Date:

Select the task:

ENGLISH	L1 (Spanish, Catalan…)
Descriptive – challenge ☐ Argumentative – transport ☐ Descriptive – memorable ☐ Argumentative – global warming ☐	Descriptive – frustration ☐ Argumentative – learning ☐ Descriptive – travel ☐ Argumentative – living ☐

1. How anxious did you see the participant?

	1	2	3	4	5	6	7	8	9	
Extremely relaxed	○	○	○	○	○	○	○	○	○	Extremely anxious

2. What made you think this person was nervous? E.g., constant moving, stuttering...

3. Did you see the participant nervous…

	1	2	3	4	5	
Never	○	○	○	○	○	All the time

4. How easy was it for you to understand the participant?

	1	2	3	4	5	6	7	8	9	
Extremely easy	○	○	○	○	○	○	○	○	○	Extremely difficult

5. How difficult was it for you to rate the student's anxiety?

	1	2	3	4	5	6	7	8	9	
very difficult to understand	○	○	○	○	○	○	○	○	○	very easy to understad

6. How confident are you with your rating of anxiety?

	1	2	3	4	5	6	7	8	9	
Not confident at all	○	○	○	○	○	○	○	○	○	Extremely confident

Appendix L1: Sentences from the Elicited Imitation Task (Ortega et al. 2002; Wu et al., 2021)

1. I have to get a haircut (7)
2. The red book is on the table (8)
3. The streets in this city are wide (8)
4. He takes a shower every morning (9)
5. What did you say you were doing today? (10)
6. I doubt that he knows how to drive that well (10)
7. After dinner I had a long, peaceful nap (11)
8. It is possible that it will rain tomorrow (12)
9. I enjoy movies which have a happy ending (12)
10. The houses are very nice but too expensive (12)
11. The little boy whose kitten died yesterday is sad (13)
12. That restaurant is supposed to have very good food (13)
13. I want a nice, big house in which my animals can live (14)
14. You really enjoy listening to country music, don't you (14)
15. She just finished painting the inside of her apartment (14)
16. Cross the street at the light and then just continue straight ahead (15)
17. The person I'm dating has a wonderful sense of humor (15)
18. She only orders meat dishes and never eats vegetables (15/16)
19. I wish the price of town houses would become affordable (15)
20. I hope it will get warmer sooner this year than it did last year (16)
21. A good friend of mine always takes care of my neighbor’s three children (16)
22. The black cat that you fed yesterday was the one chased by the dog(16)
23. Before he can go outside, he has to finish cleaning his room (16)

24. The most fun I've ever had was when we went to the opera (16)

25. The terrible thief whom the police caught was very tall and thin (17)

26. Would you be so kind as to hand me the book which is on the table? (17)

27. The number of people who smoke cigars is increasing every year (17/18)

28. I don't know if the 11:30 train has left the station yet (18)

29. The exam wasn't nearly as difficult as you told me it would be (18)

30. There are a lot of people who don’t eat anything at all in the morning (19)

Appendix L2: EIT scoring Rubric

Table L2.1. EIT scoring rubric (based on Ortega et al., 2002)

Item Score	Description	Illustration of Item 11 on EIT A & B
4	Perfect repetition	- The little boy whose kitten died yesterday is sad. - The young man whose horse ran away yesterday is mad.
3	Accurate content repetition with some (ungrammatical or grammatical) changes of form	-The little boy whose kitten died yesterday is very sad. -The young man whose horse run away yesterday is mad.
2	Changes in content or in form that affect meaning	- The little boy whose cat died was sad. - The young man whose horse…was mad
1	Repetition of half or less of the stimulus leading to substantial loss of meaning	- The boy is sad. - The young man whose horse
0	Silence, only one word repeated, or unintelligible repetition	-No response -The man

SCORE 0

Criteria	Examples
Nothing (Silence)	
Garbled (unintelligible, usually transcribed as XXX)	
Minimal repetition, then item abandoned: - Only 1 word repeated - Only 1 content word plus function word(s) - Only 1 content word plus function word(s) plus extraneous words that weren't in the original stimulus - Only function word(s) repeated **NOTE: with** only, just, yet (meaningful adverbs), score 1	- The- the street in... in... street... hmm (16/#2) - I wish... comfta-portable (19/#1) - I watch a movie (9/#22) - You don't... don't you? (14/#1) - He just finished (15/#23) (Closed word + Adv + lexical word) (score 1)

SCORE 1

Criteria	Examples
When only about half of idea units are represented in the string but a lot of important information in the original stimulus is left out	- Cross the cross--cross the street ahead and. (16/#4) - I don't have nap (7/#1) - I ...the last year (20/#4) - I have to hair-haircu (1/#24)

When barely half of lexical words get repeated and meaningful content results that is unrelated (or opposed) to stimulus, frequently with hesitation markers	- Would you... the book on the table (26/#7) - I wonder... why he... drive... well (6/#9) - He just finished painting... inside the park (15/#11)
Or when string doesn't in itself constitute a self-standing sentence with some (targetlike or nontargetlike) meaning (This may happen more often with shorter items, where if only 2 of 3 content words are repeated and no grammatical relation between them is attempted, then score 1) Also when half of a long stimulus is left out, and the sentence produced is incomplete	- I enjoy movie what shew have a... have a (9/#3) - She only eats vegetables and have xx- never eat vegetables (18/#4) - I want to big nice house.(13/#25) - A good frien of my take a good my chilren (21/#25) - I wannata .... animalslive (13/#26) - Zu book .... table (2/#26) - I doubt he how to drive (6/#25) -The little boy the kitten... no.. is sad... I can't remember (11/#8) - Before... before he can go outside for (23/#11)

SCORE 2

Criteria	**Examples**
When content of string preserves at least more than half of the idea units in the original stimulus; string in meaningful, and the meaning is close or related to original, but it departs from it in some slight changes in content, which makes content inexact, incomplete, or ambiguous	- The gooda friend take care o- chi- children (left out that it was the neighbor's children, and that they were three) (21/#1) - After dinner I have a long piece [peace?] of a nap (<a long, peaceful nap) (7/#4) - She just finished painting the seaside her apartment (<inside of) (15/#4) - The restaurant was supposed to have ve- good food (<is supposed; meaning changed to past) (12/#4) - I want to big house which... in which... animal can live (left out 'nice' 'my' and made animal into singular) (13/#4) - Would you hand me... the books which are on the table (<book; meaning changed to plural) (26/#4) - It is possible to day tomorrow (from pronunciation problem, it is ambiguous whether 'rain' has been understood, but it is possible) (8/#1)

SCORE 3

Criteria	**Examples**
Original, complete meaning is preserved as in the stimulus. Strings which are quite	- It is possible... the rain tomorrow (8/#11)

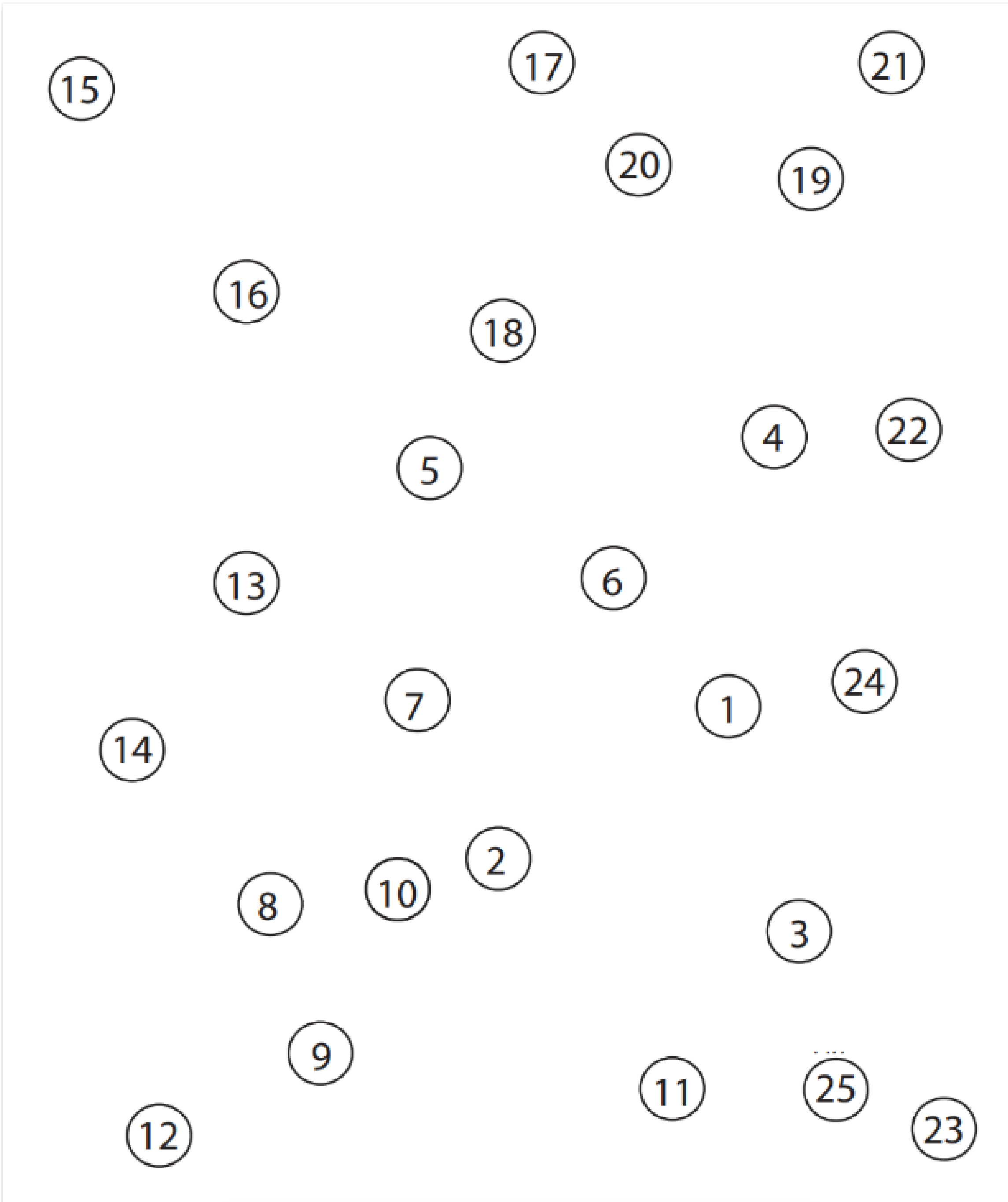
ungrammatical can get a 3 score, as long as exact meaning is preserved. Some synonymous substitutions are acceptable.	- That restaurant ah.... supposed to... ah... very good food (12/#14)
Examples of acceptable substitutions (SCORE 3): hand/give/pass are acceptable synonyms for item 26. Substitutions of and & but are acceptable. A lot of = many, etc. Anything with or without ‘very’ can be considered synonymous.	- Would you pass me the book on the table (26/#21)(Score 3) - Would you be so kind...to bring... the book...on the table (26/#13)(Score 3) - The rest-restaurant is supposed to have **good food** (12/#11)(Score 3)
Examples of unacceptable substitutions or omissions (SCORE 2): - cigar smoking> smoking - apartment >house/room - he<>she - sense of humor> humor - finished cleaning>cleaned - order> eat - nice,big > big - AUX cannot be omitted (can go> go) - a lot of Noun> 0 Noun -too Adj > 0 Adj	- The number of people who **smoke** ...um is increasing every year (27/#10)(Score 2) - **He** just finished painting... inside of a **his** house (15/#5)(Score 2) - She finished **a** painting... inside her apartment (15/#7)(Score 2) - The person I'm dading is ...**wonderful... humour** (17/#11)(Score 2) - Before he get outside...he must **clean** his room (23/#9)(Score 2) - She always **eat**...meat...nev-never eat vegetable (18/#5)(Score 2)
Changes in grammar that don’t affect meaning should be scored as 3. For instance, failure to supply past tense (had>have) and missing articles should be considered grammar change only (score 3).	- After dinner I **have** a long peaceful nap. (7/#17)(Score 3)
By contrast, cases of extra marking or more marked morphology should be considered as meaning change. For example, a present tense repeated as past or as future should be scored as meaning change (score 2).	- The restaurant **was** supposed to have ve- good food.(12/#24)(Score 2) - After the dinner I **will have** a long... sp-peaceful nap. (7/#8)(Score 2)
Similarly, singular/plural differences between stimulus and repeated string change the meaning, not only the grammar (score 2).	- The **street** in the city is wide (3/#8)(Score 2)
Changes of person (he for she or she for he) change the meaning; but problems of agreement (she...her versus she...his) should be considered grammatical change, not meaning change.	- **She** just finished painting ...**his** room inside (15/#14) (Score 2)(apartment is missing)
Ambiguous changes in grammar that COULD be interpreted as meaning changes from a NS perspective should be scored as 2. That is, as a general principle in case of doubt about whether meaning has changed or not, score 2.	- The street**s** on the city **is** wide (3/#23)(Score 2) (We can’t know whether the number agreement is just a grammar problem or an interpretation problem, but string is ambiguous in meaning: (a) a generic plural statement or (b) a statement about one street (score 2).

SCORE 4

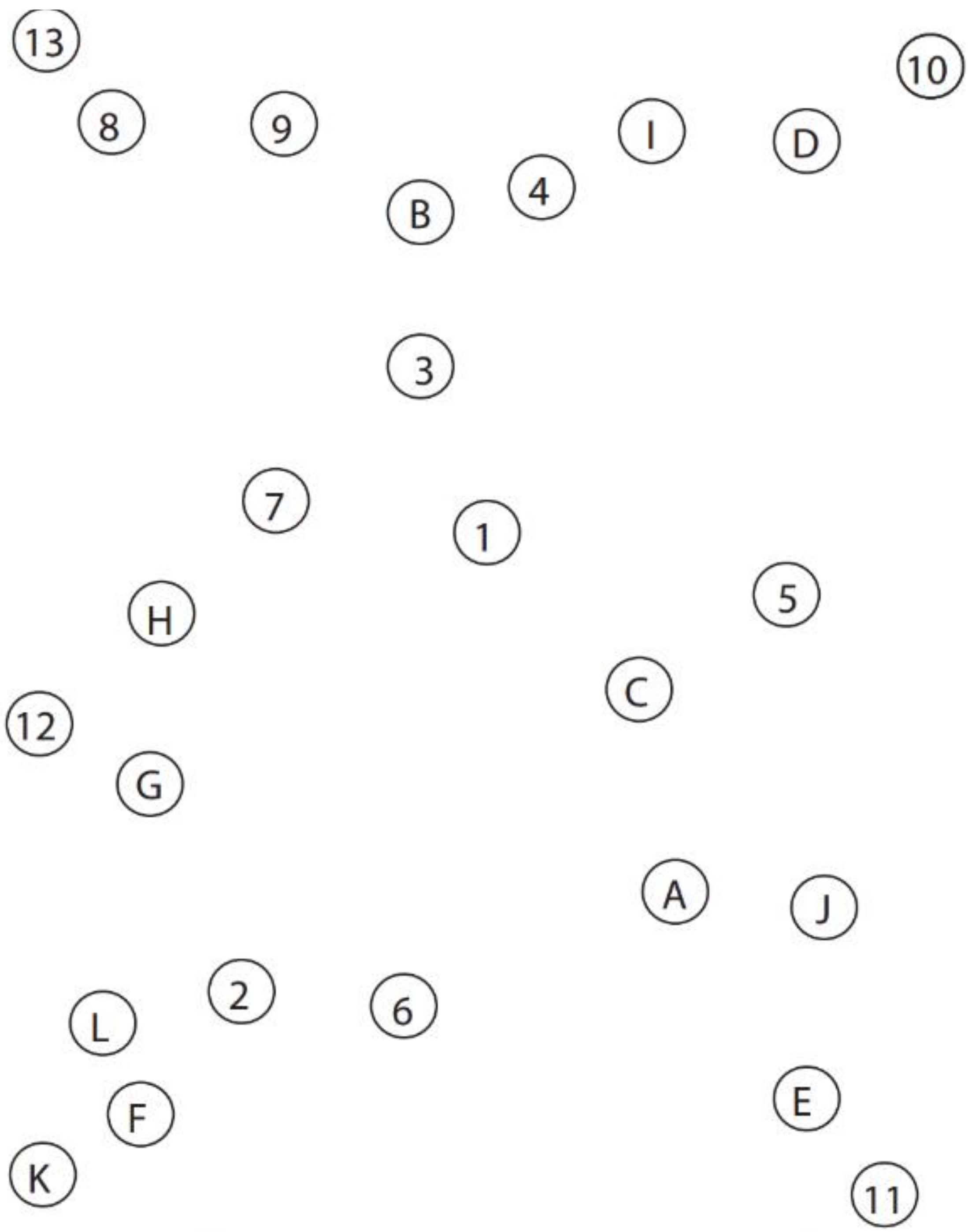
Criteria	**Examples**
Exact repetition: String matches stimulus exactly. Both form and meaning are correct without exception or doubt.	

Appendix M: TMT

Part A



Part B



Appendix N: Consent Form

UNIVERSITAT DE BARCELONA

STUDY INFORMATION SHEET and INFORMED CONSENT for
L2 oral production and psycholinguistic factors

You are invited to participate in a research study on **how psycholinguistic factors affect your speaking skills in English**. Your participation will help the researchers better understand the psycholinguistic processes that underlie second language acquisition. You were selected as a potential participant in this study because you are an **advanced-level learner of English.** We ask that you read this form and ask any questions you may have before agreeing to be in the study.

The study is being conducted by PhD candidate Gisela Sosa and supervised by Dr. Joan Carles Mora.

STUDY PURPOSE:

In this study we aim at better understanding the psycholinguistic processes that underlie second language acquisition by investigating how they may affect your oral production skills in English.

PROCEDURES FOR THE STUDY*:

You will do several tasks in the GRAL LAB and Sala de Professors during the pre-assigned time slots. The remaining tasks will be done at home. The total amount of time you will devote to performing these tasks including the lab sessions and the tasks done at home will be approximately 4h.

If you agree to be in the study*, you will perform the following tasks:

1) A language background questionnaire (~15 mins) [home].
2) A vocabulary size test (~10 mins) [lab].
3) An elicited imitation task (~8 mins) [lab].
4) Two narrative tasks in English (~5 mins each) [lab and Sala de Professors].
5) Two narrative tasks in your L1 (~5 mins each) [lab and Sala de Professors].
6) A short questionnaire about the speaking tasks you have performed (~5 mins) [lab].
7) Two reflections on the speaking tasks you have performed (~5 mins each) [lab].
8) A working memory task (~10 mins) [lab].
9) A global attention task (~7 mins) [lab].

* In order to be able to participate in the study you need to have completed the *Language Background Questionnaire*. If you have not completed it you will be asked to fill it out after doing the tasks described above. Please tick the appropriate box below:

Language Background Questionnaire

YES, I have completed it **NO, I will fill it out after the tasks**

CONFIDENTIALITY

Efforts will be made to keep your personal information confidential. All the data collected in this study will be anonymized and secured in a safe place and your identity will be held in confidence in reports in which the study may be published, and databases in which results may be stored. Only the investigators of this study will have access to your audio recordings, and your recordings and data will be de-identified so that your identity will not be associated with the recordings or your test scores.

COMPENSATION for your participation

The participation in this study is worth an additional 5% of your final *Llengua Anglesa II* course grade.

CONTACTS FOR QUESTIONS OR PROBLEMS

For questions about the study, contact giselasosa@ub.edu

VOLUNTARY NATURE OF STUDY

Taking part in this study is voluntary. You may choose not to take part or may leave the study at any time. Your decision whether or not to participate in this study will not affect your current or future relations with the researchers or the Universitat de Barcelona.

Name and Surnames:

YES, I agree to participate in the study

Signature

Date: ____ / ____ / 2022
day month

Appendix O1: Consent Form for Native Raters

UNIVERSITAT DE BARCELONA

STUDY INFORMATION SHEET and INFORMED CONSENT

You are invited to participate in a research study on **the effects of psycholinguistic factors on L2 English speaking performance.** Your participation will help the researchers better understand the effects of experiencing L2 speaking anxiety on L2 speech. You were selected as a possible participant in this study because **you are a Native Speaker of English.** We ask that you read this form and ask any questions you may have before agreeing to be in the study.

The study is being conducted by a PhD candidate under the supervision of Dr. Joan C. Mora (University of Barcelona).

PROCEDURES FOR THE STUDY: Your participation involves listening to speech samples produced by non-native speakers through headphones, and you will rate them for Comprehensibility and Accentedness. We would also like to have a *short meeting* with you to go over the procedures and answer any questions you may have. The meeting will be held on Zoom. You will be able to complete the rating tasks at home. In total, the meeting and tasks, would take around 5 hours for which you will be compensated *€15 per hour (€75 total).* We are hoping to have the tasks completed in the *next one to two weeks*.

If you are interested in participating, please let us know so that we can schedule a meeting!

Joan C. Mora mora@ub.edu Gisela Sosa López giselasosa@ub.edu

CONFIDENTIALITY

Your personal information will be kept confidential. All the data collected in this study will be anonymized and secured in a safe place and your identity will be held in confidence in reports in which the study may be published, and databases in which results may be stored. Only the investigators of this study will have access to your audio recordings, and your recordings and data will be de-identified so that your identity will not be associated with the recordings or your test scores.

COMPENSATION for your participation

You will receive €15 per hour to compensate for your time.

CONTACTS FOR QUESTIONS OR PROBLEMS

For questions about the study, your rights as a research participant or to discuss problems, complaints or concerns about this research study, or to obtain information, or offer input, contact Joan C. Mora at mora@ub.edu or Gisela Sosa López giselasosa@ub.edu

VOLUNTARY NATURE OF STUDY

Taking part in this study is voluntary. You may choose not to take part or may leave the study at any time. Leaving the study will not result in any penalty or loss of benefits to which you are entitled. Your decision whether or not to participate in this study will not affect your current or future relations with the researchers or the University of Barcelona.

YES, I agree to participate in the study.

Name and Surname: ____________________________________

Signature Date: ___ / ___ / 2023

Appendix O2: Rating Instructions for Native Raters

INSTRUCTIONS for RATERS

COMPREHENSIBILITY and ACCENTEDNESS rating tasks

In order to do this task, you need to have Praat installed on your computer. You can download it from this link: http://www.fon.hum.uva.nl/praat/

In this task we are asking you to rate some speech samples for comprehensibility and accentedness. There are 2 speech samples for practice and 320 speech samples of approximately 45 seconds each that you need to rate twice, once for comprehensibility and once for accentedness. We estimate rating all the speech samples will take you about 4 hours of your time, so we have distributed the speech samples into 4 rating tasks (COMP&ACC_1, COMP&ACC_2, COMP&ACC_3, COMP&ACC_4). Each talk should take you approximately 1 hour. We have also included breaks, each one after 26 speech samples.

• In COMP&ACC_1, students discuss which option is better to stop global warming bearing in mind the cooperation by the citizens, the impact on the environment and the possible economic impact.

• In COMP&ACC_2, students describe the hardest and toughest challenge they have experienced in their life.

• In COMP&ACC_3, students discuss which way of transportation is better bearing in mind its accessibility within and out of the city, environment and pollution, and mobility.

• In COMP&ACC_4, students describe their most memorable (childhood) experience.

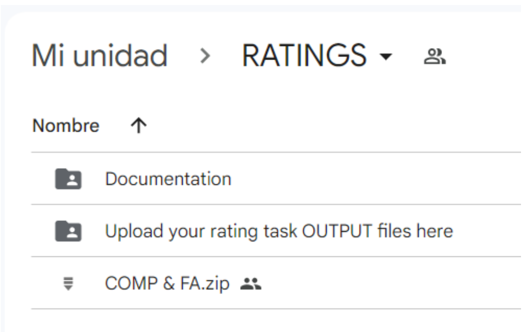
What do you need to pay attention to when rating the speech samples for comprehensibility and accentedness?

- **Comprehensibility** is how much effort it takes to understand what someone is saying. When you listen to someone's speech and you can understand her/him with ease, that is, without putting any effort, then a speaker is highly comprehensible. However, if you struggle with it and must listen very carefully, putting a great amount of effort in listening, or in fact you cannot understand what is being said at all, then a speaker has low comprehensibility.
- **Accentedness** is how much of a non-native accent (for example a Spanish accent) you can perceive in someone's speech. For example, when a Spanish speaker speaks English you may notice that he/she does not sound like a native speaker of English because of the Spanish accent in his/her speech. We would like to know how much of a non-native English accent you perceive in the speech samples you are asked to rate.

Each rating task (COMP&FA_1, COMP&FA_2, COMP&FA_3, COMP&FA_4) will take you 1 hour approximately. After doing the task we ask you to upload your ratings in the link provided.

How do I access the rating tasks and how do I upload my results?

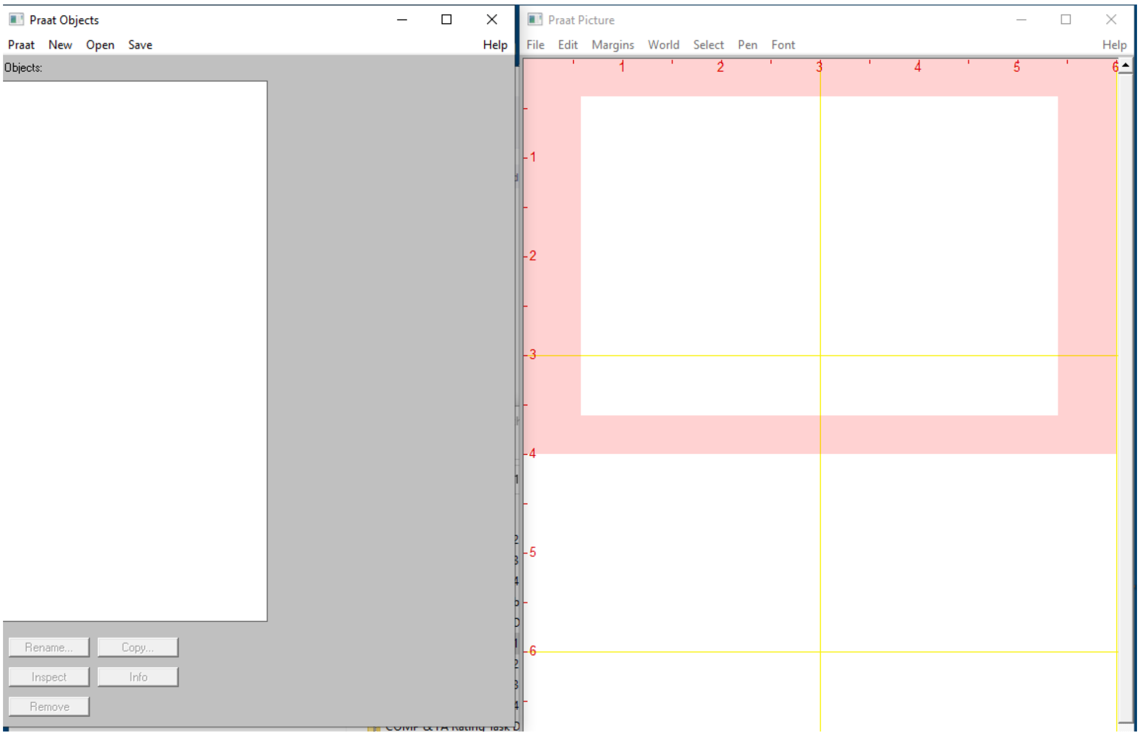
You can download the rating tasks from drive through the following link: https://drive.google.com/drive/folders/1InPAR_4NUSICg0G8m6vPsUIs1YCvV898?usp=sharing



The ratings tasks are in a zipped folder (COMP & FA.zip). Please download this file onto your computer and then uncompress it. Praat can only run the rating tasks if they are uncompressed and saved locally in your computer. If you do not have Praat installed on your computer please do the following:

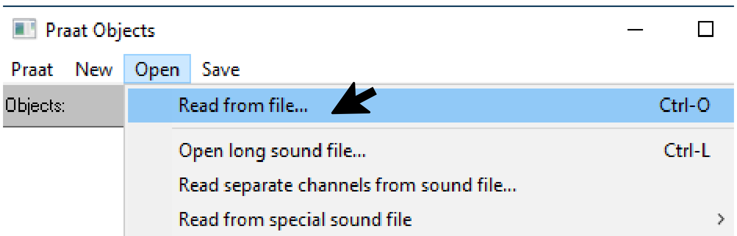
Go to https://www.fon.hum.uva.nl/praat/ and download and install Praat on your computer. This is the programme that we are using to run the Rating Tasks

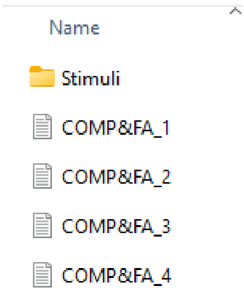
In order to start rating double-click on the Praat icon. You will only need the Praat Objects window, you can close the Praat Picture Window.



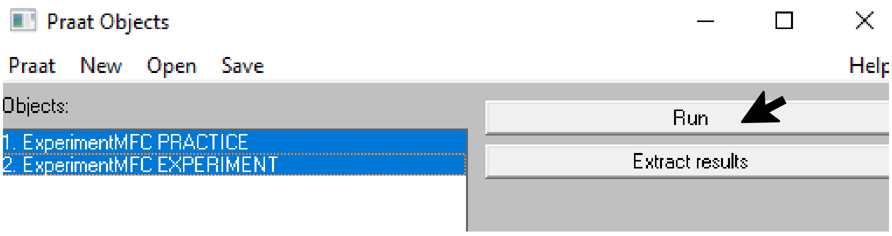
In order to rate the speech samples, please follow the instructions below carefully:

0. In Praat go to Open > Read from file to the COMP & FA folder you have downloaded from drive and open the first rating task “COMP&FA_1”.

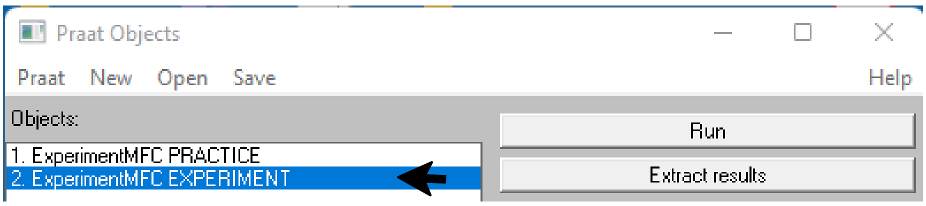




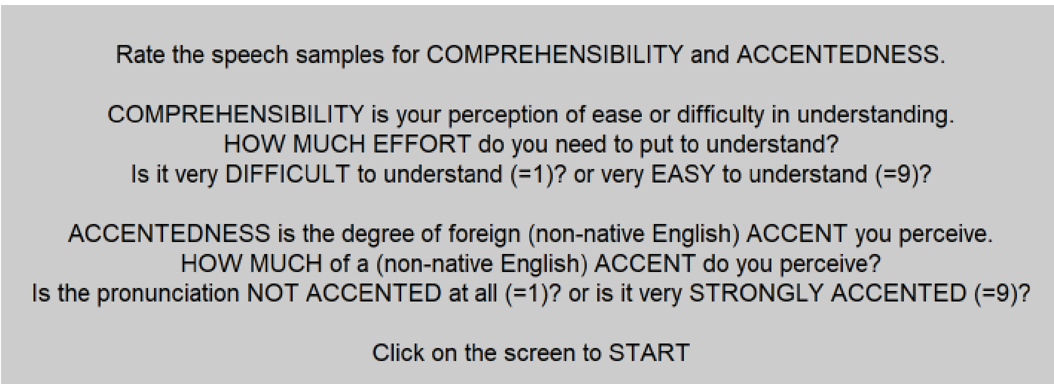
This will put the task in the Praat objects window. For the COMP&FA_1 rating task make sure both the PRACTICE and the EXPERIMENT are highlighted in the Praat objects window. To start with the rating task, click on Run.



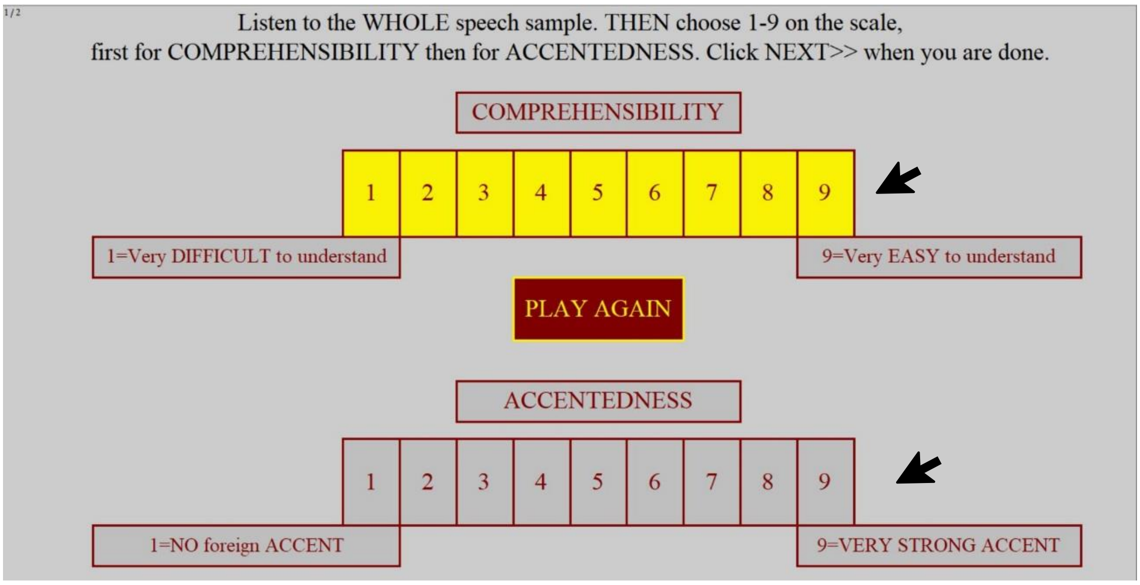
For the other three rating tasks COMP&FA_2, COMP&FA_3, COMP&FA_4, you may want to skip the practice trials. You can skip them by only selecting the EXPERIMENT file in the Praat objects window:



Read the instructions on the first page carefully,



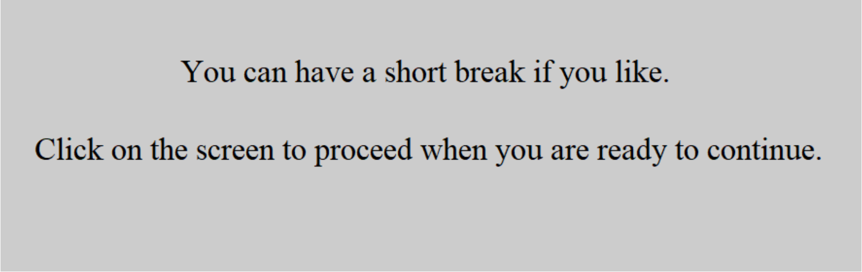
then click on the screen to start the task. There will be three practice trials before the actual rating task begins.



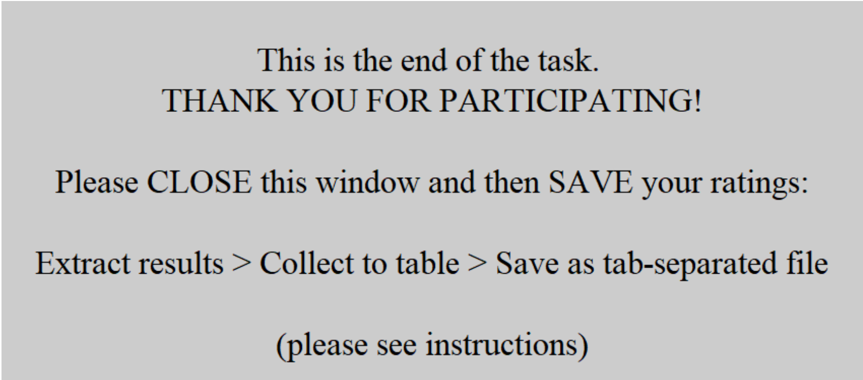
1.FIRST pay attention to HOW EASY THE SPEECH SAMPLE IS TO UNDERSTAND. Listen to the full speech sample carefully and rate it for COMPREHENSIBILITY by clicking on a number 1-9 on the scale.

2.Click on PLAY AGAIN to listen to the speech SAMPLE again. NOW pay attention to HOW STRONGLY ACCENTED THE SPEECH SAMPLE IS. Listen to the speech sample carefully and rate it for ACCENTEDNESS by clicking on a number 1-9 on the scale. You may not want to listen to the full speech sample now. Once you have rated the speech sample for ACCENTEDNESS you can click on "Next" to rate the following speech sample.

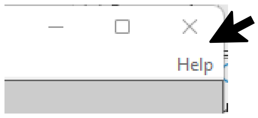
3.CLICK on NEXT to listen to the next speech sample, and rate it in the same way. Every 15 speech samples you will be prompted to take a little break.



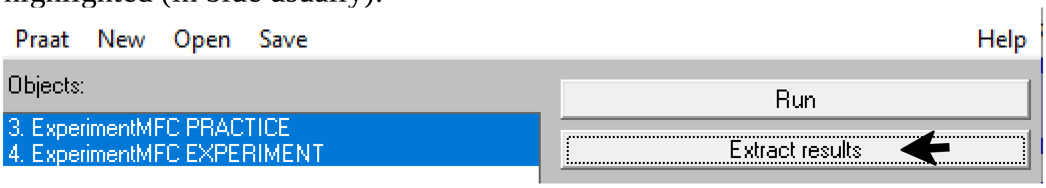
4.When you get to the end of the task you will see the following screen:



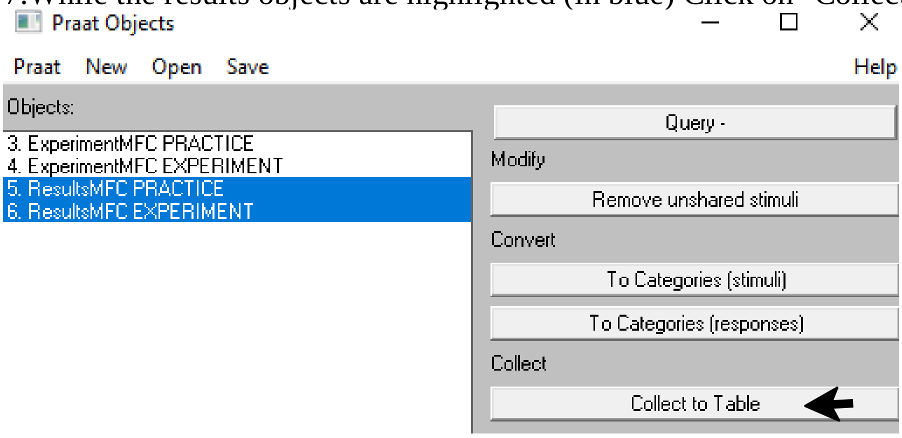
5.Close this screen



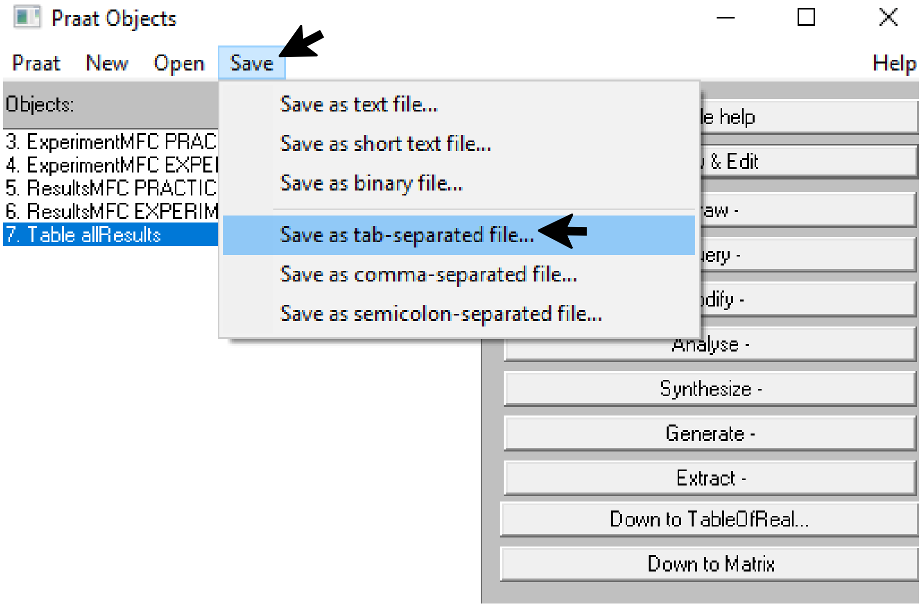
6.Click on Extract results (note that the items in the Praat objects window should be highlighted (in blue usually).



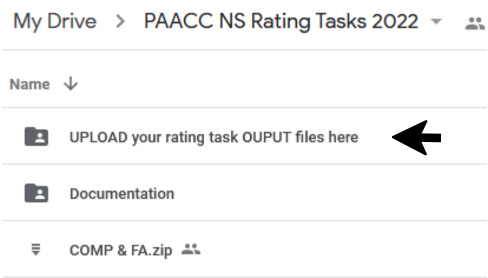
7.While the results objects are highlighted (in blue) Click on "Collect to table"



8. This generates a "Table allResults" object. This file contains your responses. Please save it in your computer as a tab-separated file with the name of the task and your name as follows: "COMP&FA_1_YourName", for example "COMP&FA_1_JohnSmith".



9. Then upload this output file in the folder " UPLOAD your rating task OUPUT files here" in Drive through the link you used to download the rating tasks. You may want to do all rating tasks first and upload your output files when you are done with all rating tasks.



Appendix P: LBQ Section on L1 and L2 Anxiety Descriptives

Table P1. Nervousness when speaking your L2 (English)

	M	*SD*	*95% CI*	*Range (Min-Max)*
1. Talking in English to your teachers	4.42	2.59	3.96 – 5.07	1-9
2. Talking in English to non-native speakers of English	3.08	2.10	2.63 – 3.53	1-9
3. Talking in English to native speakers of English	4.63	2.70	4.04 – 5.20	1-9
4. Doing an oral presentation in English	5.85	2.67	5.27 – 6.42	1-9
5. Doing a task with a peer (e.g., a dialogue) in English	3.58	2.38	3.07 – 4.09	1-9
6. Taking part in a group discussion in class in English	4.31	2.75	3.72 – 4.90	1-9

Table P2. Nervousness when speaking your L2 (English)

	M	*SD*	*95% CI*	*Range Min-Max*
1. I fear others might find my pronunciation of English foreign-accented.	5.02	2.69	4.44 – 5.59	1-9
2. I fear others might find difficult to understand me when I speak English.	4.07	2.59	3.51 – 4.62	1-9
3. I get nervous when I can’t find the word to say.	6.19	2.61	5.62 – 6.74	1-9
4. English pronunciation is difficult for me.	3.76	2.38	3.24 – 4.26	1-9
5. My pronunciation is at a lower level than	3.58	2.23	3.10 – 4.05	1-9

that of people around me.				
6. I'd rather people did not hear me making a pronunciation mistake	6.27	2.57	5.71 – 6.81	1-9
7. I feel stressed knowing that others are listening to me.	5.26	2.71	4.67 – 5.83	1-9
8. I (would) feel more embarrassed making a pronunciation mistake than any other type of mistake when I speak in English.	4.27	2.49	3.73 – 4.80	1-9
9. I feel sure of myself when speaking in English*[56]	5.94	2.31	5.44 – 6.43	1-9
10. I am calm when I have to speak in English without preparation time*	5.37	2.67	4.79 – 5.94	1-9
11. I enjoy expressing my ideas in English*	7.21	1.97	6.78 – 5.94	1-9
12. I remember the pronunciation of new words easily*	6.74	1.82	6.78 – 7.63	1-9
13. I am satisfied with my current level of English speaking*	5.63	2.49	6.35 – 7.13	1-9
14. Whenever I think of my future career, I imagine myself using English with native-like pronunciation*	7.62	1.85	5.09 – 6.16	1-9
15. I like speaking to myself in English*	7.56	2.11	7.22 – 8.01	1-9
16. I think I have a talent to acquire English pronunciation*	5.55	2.26	5.06 – 6.03	1-9
17. Being understandable depends on my pronunciation	5.24	2.47	4.71 – 5.77	1-9

[56] * = The questionnaire items that had to be reversed and recoded for analysis

18. I never feel quite sure of myself when I am speaking in English in class	4.67	2.80	4.07-5.27	1-9
19. I don't worry about making mistakes in English*	4.38	2.29	3.89 – 4.87	1-9
20. I tremble when I know that I'm going to be called on in class	5.02	2.97	4.38 – 5.66	1-9
21. It frightens me when I don't understand what the teacher is saying in the foreign language	4.38	2.65	3.81 – 4.95	1-9
22. During class, I find myself thinking about things that have nothing to do with the course	5.27	2.40	4.75 – 5.78	1-9
23. I keep thinking that the other students are better at languages than I am	5.19	2.62	4.62 – 5. 74	1-9
24. I am usually at ease during tests*	5.26	2.36	4.74 – 5.76	1-9
25. I start to panic when I have to speak without preparation in class	4.80	2.84	4.19 – 5.41	1-9
26. It embarrasses me to volunteer answers in class	5.26	2.82	4.65 – 5.86	1-9
27. I worry about the consequences of failing my foreign language classes	5.33	2.89	4.70 – 5.94	1-9
28. I would not be nervous speaking English with native speakers*	4.74	2.80	4.14 – 5.34	1-9
29. I get upset when I don't understand what the teacher is correcting	4.97	2.55	4.41 – 5.51	1-9

30. Even if I am well prepared for class, I feel anxious about it	5.00	2.74	4.41 – 5.58	1-9
31. I can feel my heart pounding when I'm going to be called on in class	5.69	2.92	5.05 – 6.31	1-9
32. I would probably feel comfortable around native speakers of English*	5.93	2.42	5.41 – 6.44	1-9

Table P3. Nervousness when speaking your L1

	M	*SD*	*95% CI*	*Range (Min-Max)*
1. Talking to your teachers	3.28	2.61	2.71 – 3.83	1-9
2. Doing an oral presentation	4.91	2.90	4.28 – 5.52	1-9
3. Doing a task with a peer (e.g., a dialogue)	2.67	2.40	2.15 – 3.18	1-9
4. Taking part in a group discussion in class	3.74	2.82	3.13 – 4.34	1-9
5. Talking to your classmates	2.41	2.24	1.92 – 2.88	1-9

Table P4. Nervousness when speaking your L1

	M	*SD*	*95% CI*	*Range Min-Max*
1. While participating in a conversation with a new acquaintance I feel nervous	4.02	2.53	3.48 – 4.56	1-9
2. I have no fear of facing an audience*[57]	4.00	2.67	3.42 – 4.57	1-9
3. I usually talk less because I'm shy	5.02	2.79	4.42 – 5.62	1-9
4. I look forward to expressing my opinions in class*	4.86	2.59	4.03 – 5.41	1-9
5. I am afraid to express myself in a group	4.17	2.82	3.57 – 4.77	1-9
6. I look forward to an opportunity to speak in public*	3.45	2.48	2.92 – 3.98	1-9
7. I find the prospect of speaking mildly pleasant*	4.91	2.31	4.41 – 5.40	1-9
8. When communicating, my posture feels	3.79	2.41	3.27 – 4.30	1-9

[57] * = The questionnaire items that had to be reversed and recoded for analysis

strained and unnatural				
9. I am tense when participating in group discussions	4.36	2.59	3.80 – 4.91	1-9
10. Although I talk fluently with friends, I am at loss for words when doing a presentation	4.47	2.72	3.88 – 5.04	1-9
11. I have no fear of expressing myself in a group*	5.22	2.85	4.60 – 5.83	1-9
12. My hands tremble when I try to handle objects in a presentation	4.62	2.82	4.01 – 5.22	1-9
13. I always avoid speaking in public	4.86	2.38	4.25 – 5.46	1-9
14. I feel that I am more fluent when talking to people than most other people	4.70	2.38	4.18 – 5.20	1-9
15. I'm tense while I'm speaking before a group of people	4.99	2.53	4.44 – 5.53	1-9
16. My thoughts become confused when I speak before an audience	4.84	2.57	4.28 – 5.38	1-9
17. I like to get involved in group discussions*	5.10	2.61	4.54 – 5.66	1-9
18. Although I am nervous before a presentation, I soon forget my fears and I enjoy the experience*	5.06	2.70	4.47 – 5.63	1-9
19. Conversing with people who hold positions of authority cause me tension	5.24	2.71	4.66 – 5.82	1-9
20. I feel relaxed and comfortable while speaking in any context*	4.98	5.47	4.45 – 5.49	1-9

21. I feel nervous when I am called upon to answer a question or give an opinion in class	5.47	2.76	4.87 – 6.05	1-9
22. I face the prospect of making a speech with complete confidence*	4.59	2.50	4.05 – 5.12	1-9
23. I'm afraid to speak up in a conversation	3.59	2.38	3.08 – 4.10	1-9
24. I would enjoy presenting a speech on a local television show*	3.62	2.69	3.03 – 4.19	1-9

Appendix Q: Correlations mental effort, task difficulty and anxiety measures

Table Q1: Two-tailed Pearson correlation between mental effort, task difficulty and physiological and subjective anxiety measures

*Note: *p* < .05 (2-tailed); *p* < .001** (2-tailed)

	Mental effort	Task Difficulty
HR	.099*	.137**
HRV avnn	.003	.016
EDA	.125**	.178**
Idiodynamic	.462**	.812**
Peaks per minute	.362**	.254**
Self-reports	.654**	.596**

Appendix R: Parameter estimates RQ1

*Note: Reference categories indicated in brackets

Table R1: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/ L1) on HRm.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.93	.006	305.39	.000	1.92	1.94
Task type	-.006	.008	-.748	.455	-.023	.010
Audience	.004	.008	.431	.667	-.013	.020
Language	.008	.008	.989	.323	-.008	.025
Task Type x Context	.002	.011	.203	.839	-.021	.026
Task Type x Language	.017	.011	1.45	.146	-.006	.041
Context x Language	.005	.011	.418	.676	-.018	.028
Task Type x Context x Language	-.016	.016	-.983	.326	-.049	.016

Table R2: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/ L1) on HRsd.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	.940	.023	39.33	.000	.893	.986
Task type	-.006	.033	-.167	.867	-.071	.060
Audience	.011	.033	.340	.734	-.054	.076
Language	.012	.033	.350	.726	-.054	.077
Task Type x Context	-.006	.046	-.129	.898	-.098	.086
Task Type x Language	.051	.046	1.09	.275	-.041	.144
Context x Language	.060	.046	1.27	.203	-.032	.152
Task Type x Context x Language	-.067	.066	-1.00	.314	-.197	.063

Table R3: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/L1) on HRV avnn.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	.572	.014	38.51	.000	.543	.601
Task type	-.029	.017	-1.64	.100	-.063	.006
Audience	.005	.017	.278	.781	-.029	.039
Language	-.006	.017	-.338	.735	-.040	.028
Task Type x Context	.027	.024	1.09	.274	-.022	.076
Task Type x Language	.003	.024	.129	.897	-.045	.052
Context x Language	.020	.024	.803	.422	-.029	.068
Task Type x Context x Language	-.016	.035	-.448	.654	-.084	.053

Table R4: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/L1) on EDAm.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	-.067	.0615	-1.097	.273	-.188	.053
Task type	-.053	.0708	-.751	.453	-.192	.086
Audience	.009	.0708	.126	.900	-.130	.148
Language	.016	.0708	.220	.826	-.123	.154
Task Type x Context	-.003	.1001	-.032	.974	-.200	.193
Task Type x Language	.096	.1001	.961	.337	-.100	.293
Context x Language	.009	.1001	.089	.929	-.188	.205
Task Type x Context x Language	.001	.1415	.010	.992	-.276	.279

Table R5: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/L1) on EDAsd.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	-.519	.073	-7.05	.000	-.663	-.374
Task type	-.085	.087	-.978	.329	-.256	.086
Audience	-.009	.087	-.108	.914	-.180	.162
Language	.030	.087	.350	.726	-.141	.201
Task Type x Context	.055	.123	.443	.658	-.187	.296
Task Type x Language	.091	.123	.742	.458	-.150	.333
Context x Language	.118	.123	.954	.340	-.124	.359
Task Type x Context x Language	-.105	.174	-.603	.547	-.447	.237

Table R6: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/L1) on Idiodynamic.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	-.899	.083	-10.70	.000	-1.06	-.734
Task type	-.169	.066	-2.56	.010	-.299	-.040
Audience	-.222	.065	-3.40	.001	-.350	-.094
Language	.412	.063	6.47	.000	.287	.537
Task Type x Context	.259	.093	2.77	.006	.075	.442
Task Type x Language	-.030	.089	-.331	.741	-.206	.146
Context x Language	.036	.088	.407	.684	-.138	.211
Task Type x Context x Language	-.195	.126	-1.54	.124	-.443	.053

Table R7: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/L1) on Peaks per minute (Idiodynamic).

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	.394	.045	8.68	.000	.305	.483
Task type	-.079	.043	-1.82	.069	-.164	.006
Audience	-.078	.042	-1.83	.067	-.163	.006
Language	.146	.042	3.48	.001	.064	.229
Task Type x Context	.124	.061	2.01	.044	.003	.245
Task Type x Language	.056	.059	.940	.348	-.061	.173
Context x Language	.048	.059	.813	.417	-.068	.164
Task Type x Context x Language	-.171	.084	-2.03	.043	-.336	-.006

Table R8: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/L1) on Self-reports.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	4.55	.242	18.79	.000	4.08	5.034
Task type	-.860	.257	-3.34	.001	-1.36	-.355
Audience	-.977	.257	-3.79	.000	-1.48	-.471
Language	1.558	.257	6.04	.000	1.05	2.06
Task Type x Context	.826	.364	2.26	.024	.110	1.54
Task Type x Language	.128	.364	.351	.726	-.588	.843
Context x Language	.047	.364	.128	.898	-.669	.762
Task Type x Context x Language	-.198	.515	-.384	.701	-1.20	.814

Appendix S1: Post-task Questionnaire results

Table S1.1: Post-task questionnaire results for L2 Descriptive task in the one-to-one context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	6.24	1.62	5.89 – 6.59	1-9
Task difficulty	3.81	2.01	3.38 – 4.24	1-9
Mental effort	4.65	1.99	4.22 – 5.07	1-9
Level of anxiety	5.08	2.32	4.58 – 5.58	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	3.22	2.44	2.69 – 3.74	1-9
Speaking in English (regardless to whom)	3.72	2.26	3.23 – 4.20	1-9
Speaking English in front of someone/people	4.05	2.45	3.53 – 4.58	1-9
Being audio recorded	4.32	2.67	3.75 – 4.89	1-9
Being video recorded	4.54	2.76	3.95 – 5.13	1-9
Being monitored with the anxiety sensor (the device on your wrist	3.52	2.53	2.97 – 4.06	1-9
Knowing these tasks are worth a percentage of your grade	2.87	2.44	2.34 – 3.39	1-9
Forgetting what I wanted to say	5.03	2.63	4.46 – 5.60	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	5.27	2.55	4.73 – 5.82	1-9
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[58]	6.15	2.05	5.71 – 6.59	1-9
I think it was easy to understand my speech*	6.89	1.37	6.60 - 7	1-9
I easily recalled all the words I needed*	5.65	2.07	5.20 – 6.09	1-9

[58] * = items that were reversed and recoded in the analysis

I felt I hesitated a lot while speaking	5.40	2.13	4.94 – 5.86	1-9
I couldn't remember the pronunciation of some words	3.55	2.25	3.07 – 4.04	1-9

Table S1.2: Post-task questionnaire results for L2 Descriptive task in the group audience context

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	6.05	1.75	5.68 – 6.43	1-9
Task difficulty	3.86	2.12	3.40 – 4.31	1-9
Mental effort	4.90	1.98	4.48 – 5.33	1-9
Level of anxiety	5.38	2.33	4.88 – 5.88	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	3.43	2.25	2.94 – 3.91	1-9
Speaking in English (regardless to whom)	4.04	2.49	3.51 – 4.58	1-9
Speaking English in front of someone/people	4.82	2.67	4.25 – 5.39	1-9
Being audio recorded	3.98	2.85	3.37 – 4.60	1-9
Being video recorded	4.34	2.95	3.71 – 4.98	1-9
Being monitored with the anxiety sensor (the device on your wrist	3.23	2.46	2.70 – 3.76	1-9
Knowing these tasks are worth a percentage of your grade	2.93	2.53	2.38 – 3.47	1-9
Forgetting what I wanted to say	5.27	2.69	4.70 – 5.85	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	5.26	2.64	4.70 – 5.83	1-9
RATINGS: To what extent you agree with the statements				

I found it easy to plan what I wanted to say in this task*[59]	5.76	2.39	5.25 – 6.28	1-9
I think it was easy to understand my speech*	6.90	1.50	6.58 – 7.23	1-9
I easily recalled all the words I needed*	5.27	2.08	4.83 – 5.72	1-9
I felt I hesitated a lot while speaking	5.60	2.21	5.12 – 6.08	1-9
I couldn't remember the pronunciation of some words	3.82	2.35	3.32 – 4.32	1-9

Table S1.3: Post-task questionnaire results for L2 Argumentative task in the one-to-one context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	5.72	1.79	5.33 – 6.10	1-9
Task difficulty	3.75	1.87	3.35 – 4.15	1-9
Mental effort	4.84	1.91	4.43 – 5.25	1-9
Level of anxiety	5.19	2.22	4.71 – 5.66	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	2.96	2.25	2.48 – 3.44	1-9
Speaking in English (regardless to whom)	3.80	2.21	3.32 – 4.27	1-9
Speaking English in front of someone/people	4.02	2.32	3.52 – 4.52	1-9
Being audio recorded	4.22	2.77	3.62 – 4.81	1-9
Being video recorded	4.58	2.84	3.97 – 5.19	1-9
Being monitored with the anxiety sensor (the device on your wrist	3.55	2.61	2.99 – 4.11	1-9
Knowing these tasks are worth a percentage of your grade	3.01	2.51	2.47 – 3.55	1-9
Forgetting what I wanted to say	5.39	2.64	4.82 – 5.96	1-9
Realizing that I made mistakes	5.54	2.48	5.01 – 6.07	1-9

[59] * = items that were reversed and recoded in the analysis

(mispronouncing a word, wrong structure...)				
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[60]	5.75	2.24	5.27 – 6.23	1-9
I think it was easy to understand my speech*	6.33	1.87	5.93 – 6.73	1-9
I easily recalled all the words I needed*	4.86	1.99	4.43 – 5.28	1-9
I felt I hesitated a lot while speaking	5.79	2.14	5.33 – 6.25	1-9
I couldn't remember the pronunciation of some words	4.26	2.34	3.76 – 4.76	1-9

Table S1.4: Post-task questionnaire results for L2 Argumentative task in the group audience context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	5.05	2.00	4.62 – 5.48	1-9
Task difficulty	5.37	2.12	4.91 – 5.82	1-9
Mental effort	6.02	1.89	5.61 – 6.43	1-9
Level of anxiety	6.12	2.40	5.60 – 6.63	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	3.75	2.24	3.27 – 4.23	1-9
Speaking in English (regardless to whom)	4.18	2.47	3.65 – 4.71	1-9
Speaking English in front of someone/people	4.89	2.79	4.29 – 5-49	1-9
Being audio recorded	4.15	2.69	3.57 – 4.72	1-9
Being video recorded	4.47	2.88	3.85 – 5.09	1-9
Being monitored with the anxiety sensor (the device on your wrist	3.39	2.63	2.83 – 3.96	1-9

[60] * = items that were reversed and recoded in the analysis

Knowing these tasks are worth a percentage of your grade	2.98	2.66	2.41 – 3.56	1-9
Forgetting what I wanted to say	6.06	2.49	5.53 – 6.60	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	5.81	2.52	5.27 – 6.35	1-9
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[61]	4.43	2.42	3.91 – 4.94	1-9
I think it was easy to understand my speech*	5.96	1.96	5.54 – 6.38	1-9
I easily recalled all the words I needed*	4.39	2.19	3.92 – 4.86	1-9
I felt I hesitated a lot while speaking	6.13	2.32	5.64 – 6.63	1-9
I couldn't remember the pronunciation of some words	4.22	2.30	3.72 – 4.71	1-9

Table S1.5: Post-task questionnaire results for L1 Descriptive task in the one-to-one context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	7.32	1.37	7.03 – 7.62	1-9
Task difficulty	3.04	1.75	2.67 – 3.42	1-9
Mental effort	3.56	1.91	3.15 – 3.98	1-9
Level of anxiety	3.55	2.13	3.09 – 4.00	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	2.51	1.94	2.09 – 2.92	1-9
Speaking in English (regardless to whom)	2.01	2.02	1.57 – 2.44	1-9
Speaking English in front of someone/people	2.27	2.00	1.84 – 2.70	1-9

[61] * = items that were reversed and recoded in the analysis

Being audio recorded	2.80	2.44	2.27 – 3.32	1-9
Being video recorded	3.18	2.53	2.64 – 3.72	1-9
Being monitored with the anxiety sensor (the device on your wrist	2.41	2.07	1.97 – 2.86	1-9
Knowing these tasks are worth a percentage of your grade	2.44	2.33	1.94 – 2.94	1-9
Forgetting what I wanted to say	3.53	2.51	2.99 – 4.07	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	2.61	2.09	2.16 – 3.06	1-9
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[62]	6.77	2.04	6.34 – 7.21	1-9
I think it was easy to understand my speech*	7.48	1.53	7.15 – 7.81	1-9
I easily recalled all the words I needed*	6.96	1.89	6.55 – 7.37	1-9
I felt I hesitated a lot while speaking	3.80	2.36	3.29 – 4.30	1-9
I couldn't remember the pronunciation of some words	2.03	1.79	1.65 – 2.41	1-9

Table S1.6: Post-task questionnaire results for L1 Descriptive task in the group audience context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	7.09	1.49	6.77 – 7.41	1-9
Task difficulty	2.80	1.79	2.41 – 3.18	1-9
Mental effort	3.32	1.85	2.92 – 3.72	1-9
Level of anxiety	3.70	2.15	3.23 – 4.16	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	2.97	2.33	2.47 – 3.47	1-9

[62] * = items that were reversed and recoded in the analysis

Speaking in English (regardless to whom)	2.06	1.99	1.64 – 2.49	1-9
Speaking English in front of someone/people	3.05	2.51	2.51 – 3.59	1-9
Being audio recorded	3.05	2.55	2.51 – 3.60	1-9
Being video recorded	3.41	2.70	2.83 – 3.99	1-9
Being monitored with the anxiety sensor (the device on your wrist	2.47	2.22	2.00 – 2.95	1-9
Knowing these tasks are worth a percentage of your grade	2.44	2.33	1.94 – 2.94	1-9
Forgetting what I wanted to say	4.17	2.50	3.63 – 4.71	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	3.05	2.38	2.54 – 3.56	1-9
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[63]	7.00	1.84	6.60 – 7.39	1-9
I think it was easy to understand my speech*	7.63	1.43	7.33 – 7.94	1-9
I easily recalled all the words I needed*	7.10	1.84	6.70 – 7.50	1-9
I felt I hesitated a lot while speaking	4.03	2.30	3.54 – 4.52	1-9
I couldn't remember the pronunciation of some words	2.00	1.76	1.62 – 2.37	1-9

Table S1.7: Post-task questionnaire results for L1 Argumentative task in the one-to-one context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	7.03	1.35	6.74 – 7.32	1-9
Task difficulty	3.04	1.82	2.65 – 3.43	1-9
Mental effort	3.61	1.82	3.22 – 4.00	1-9
Level of anxiety	3.58	1.99	3.15 – 4.00	1-9

[63] * = items that were reversed and recoded in the analysis

RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	2.29	1.83	1.89 – 2.68	1-9
Speaking in English (regardless to whom)	2.08	1.94	1.66 – 2.49	1-9
Speaking English in front of someone/people	2.67	2.21	2.19 – 3.14	1-9
Being audio recorded	2.91	2.38	2.40 – 3.42	1-9
Being video recorded	3.23	2.48	2.70 – 3.76	1-9
Being monitored with the anxiety sensor (the device on your wrist	2.46	2.13	2.00 – 2.92	1-9
Knowing these tasks are worth a percentage of your grade	2.37	2.28	1.88 – 2.86	1-9
Forgetting what I wanted to say	3.98	2.37	3.47 – 4.49	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	2.83	2.24	2.35 – 3.31	1-9
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[64]	6.97	1.68	6.61 – 7.33	1-9
I think it was easy to understand my speech*	7.51	1.42	7.20 – 7.81	1-9
I easily recalled all the words I needed*	6.50	1.89	6.09 – 6.90	1-9
I felt I hesitated a lot while speaking	4.11	2.38	3.60 – 4.62	1-9
I couldn't remember the pronunciation of some words	2.04	1.76	1.66 – 2.42	1-9

[64] * = items that were reversed and recoded in the analysis

Table S1.8: Post-task questionnaire results for L1 Argumentative task in the group audience context.

	M	*SD*	*95% CI*	*Range (Min-Max)*
How well	6.60	1.52	6.27 – 6.93	1-9
Task difficulty	3.82	2.08	3.37 – 4.27	1-9
Mental effort	4.43	2.03	3.99 – 4.86	1-9
Level of anxiety	4.56	2.39	4.04 – 5.07	1-9
RATINGS: How these factors made you feel when performing the task				
The reactions of the people I was speaking to	3.24	2.30	2.74 – 3.73	1-9
Speaking in English (regardless to whom)	2.16	2.06	1.71 – 2.60	1-9
Speaking English in front of someone/people	3.17	2.61	2.61 – 3.73	1-9
Being audio recorded	3.15	2.57	2.59 – 3.70	1-9
Being video recorded	3.47	2.71	2.89 – 4.05	1-9
Being monitored with the anxiety sensor (the device on your wrist	2.50	2.16	2.03 – 2.96	1-9
Knowing these tasks are worth a percentage of your grade	2.39	2.22	1.91 – 2.87	1-9
Forgetting what I wanted to say	4.94	2.64	4.37 – 5.50	1-9
Realizing that I made mistakes (mispronouncing a word, wrong structure...)	3.37	2.45	2.84 – 3.89	1-9
RATINGS: To what extent you agree with the statements				
I found it easy to plan what I wanted to say in this task*[65]	5.93	2.14	5.47 – 6.39	1-9
I think it was easy to understand my speech*	7.23	1.71	6.86 – 7.59	1-9
I easily recalled all the words I needed*	6.41	2.08	5.97 – 6.86	1-9
I felt I hesitated a lot while speaking	4.68	2.52	4.14 – 5.22	1-9

[65] * = items that were reversed and recoded in the analysis

I couldn't remember the pronunciation of some words	1.93	1.83	1.53 – 2.32	1-9

Appendix S2: Audience Questionnaire descriptives

Table S2.1: Ratings for Descriptive task in the one-to-one context (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	6.45	0.94	6.25 – 6.65	1-9
Did you see the participant nervous (never – all the time)	3.38	0.84	3.20 – 3.56	1-5
How easy was it for you to understand the participant?	8.16	1.17	7.91 – 8.41	1-9
How difficult it was for you to rate the student's anxiety?	5.57	1.07	5.34 – 5.80	1-9
How confident are you with your rating of anxiety?	6.10	1.19	5.84 – 6.35	1-9

Table S2.2: Ratings for Descriptive task in the group context (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	4.81	1.29	4.53 – 5.08	1-9
Did you see the participant nervous (never – all the time)	2.88	0.74	2.72 – 3.04	1-5
How easy was it for you to understand the participant?	8.00	0.71	7.84 – 8.15	1-9
How difficult it was for you to rate the student's anxiety?	3.97	0.86	3.78 – 4.16	1-9
How confident are you with your rating of anxiety?	6.70	0.64	6.56 – 6.83	1-9

Table S2.3: Ratings for Descriptive task in the one-to-one context (L1)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	5.45	1.46	5.12 – 5.78	1-9
Did you see the participant nervous (never – all the time)	3.21	0.83	3.03 – 3.40	1-5
How easy was it for you to understand the participant?	8.24	1.85	7.82 – 8.65	1-9
How difficult it was for you to rate the student's anxiety?	5.56	1.41	5.24 – 5.87	1-9

How confident are you with your rating of anxiety?	5.79	1.29	5.49 – 6.08	1-9

Table S2.4: Ratings for Descriptive task in the group context task (L1)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	3.99	1.19	3.74 – 4.25	1-9
Did you see the participant nervous (never – all the time)	2.62	0.72	2.46 – 2.77	1-5
How easy was it for you to understand the participant?	7.79	2.35	7.29 – 8.30	1-9
How difficult it was for you to rate the student's anxiety?	3.69	0.93	3.49 – 3.89	1-9
How confident are you with your rating of anxiety?	6.75	0.81	6.58 – 6.93	1-9

Table S2.5: Ratings for Argumentative task in the one-to-one context (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	6.29	1.05	6.07 – 6.52	1-9
Did you see the participant nervous (never – all the time)	3.28	0.84	3.10 – 3.46	1-5
How easy was it for you to understand the participant?	8.17	1.14	7.93 – 8.42	1-9
How difficult it was for you to rate the student's anxiety?	5.55	1.04	5.33 – 5.78	1-9
How confident are you with your rating of anxiety?	5.98	1.20	5.72 – 6.24	1-9

Table S2.6: Ratings for Argumentative task in the group audience context (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	5.39	1.43	5.09 – 5.70	1-9
Did you see the participant nervous (never – all the time)	3.20	0.79	3.03 – 3.37	1-5
How easy was it for you to understand the participant?	7.70	0.80	7.53 – 7.88	1-9

How difficult it was for you to rate the student's anxiety?	3.82	0.91	3.63 – 4.02	1-9
How confident are you with your rating of anxiety?	6.86	0.53	6.74 – 6.97	1-9

Table S2.7: Ratings for Argumentative task in the one-to-one context (L1)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	5.45	1.20	5.19 – 5.72	1-9
Did you see the participant nervous (never – all the time)	3.20	0.78	3.03 – 3.38	1-5
How easy was it for you to understand the participant?	8.20	1.90	7.78 – 8.63	1-9
How difficult it was for you to rate the student's anxiety?	5.69	1.22	5.42 – 5.96	1-9
How confident are you with your rating of anxiety?	5.66	1.23	5.39 – 5.93	1-9

Table S2.8: Ratings for Argumentative task in the group audience context (L1)

	M	*SD*	*95% CI*	*Range (Min-Max)*
How anxious did you see the participant?	4.25	1.35	3.96 – 4.54	1-9
Did you see the participant nervous (never – all the time)	2.71	0.71	2.55 – 2.86	1-5
How easy was it for you to understand the participant?	7.72	2.34	7.21 – 8.22	1-9
How difficult it was for you to rate the student's anxiety?	3.79	0.87	3.60 – 3.98	1-9
How confident are you with your rating of anxiety?	6.66	0.69	6.51 – 6.81	1-9

Appendix S3: Parameter estimates RQ2

**Note*: Reference categories indicated in brackets

Table S1: Parameter estimates of the GLMM on the effects of Task Type (descriptive/argumentative), Context (one-to-one/group audience) and Language (L2/ L1) and their interaction on the Disagreement score.

Predictor	*β*	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	.303	.250	1.20	.227	-.189	.794
Task type	-.603	.283	-2.12	.034	-.1.15	-.047
Audience	-2.15	.283	-7.59	< .001	2.70	-.159
Language	.416	.283	1.46	.142	-.140	.972
Task Type x Context	501	.400	1.25	.211	-.285	1.28
Task Type x Language	.457	.400	1.14	.254	-.329	1.24
Context x Language	.32	.400	.799	.424	-.466	1.10
Task Type x Context x Language	-.615	.566	-1.08	.278	-1.72	.497

Note: SE = standard error

Appendix T: Correlations HRV measures

Table T1: Two-tailed Pearson correlation between HRV measures and other anxiety measures

Note: N = 688; *p < .05 (2-tailed); **p < .001 (2-tailed)

	EDA	IDm	Peaks per minute	Self-reports
HRV avnn	-.035	.051	.044	.028
HRV avrr	-.045	.049	.041	.022
HRV sdnn	-.081	.025	.027	-.001
HRV sdrr	-.036	.032	-.080	-.045
RMSSD	-.092	-.027	-.051	.000

Appendix U1: L2 speech Descriptives

Table U1.1: Descriptives in the L2 Descriptive task One-to-one

		M	*SD*	*95% CI*
	SR	136.46	20.39	131.86 – 141.06
	ASD	.30	.053	.292 - .316
	AR	3.40	.769	3.23 – 3.58
	MLoR	.075	.023	.070 - .080
	PhonTimeR	68.42	10.85	65.97 - 70.87
	Durpse	.742	.174	.702 - .781
	Durpsi	.617	.101	.594 - .640
	Durpfe	.575	.144	.542 - .608
	Durpfi	.528	.088	.508 - .548
	PDur	.664	.127	.635 - .693
Fluency	pTimeR	31.57	10.85	29.12 - 34.02
	pFreq	28.81	8.65	26.85 - 30.76
	pseFreq	8.65	1.88	8.22 - 9.07
	psiFreq	14.07	3.66	13.24 - 14.89
	pfeFreq	1.67	1.21	1.40 - 1.95
	pfiFreq	3.26	1.83	2.85 - 3.68
	psfeFreq	10.11	2.49	9.58 - 10.65
	psfiFreq	17.20	4.62	16.21 - 18.19
	ExtpTimeR	12.39	3.94	11.50 - 13.28
	IntpTimeR	17.44	5.36	16.23 - 18.65
	P100Syll	21.87	8.57	19.94 - 23.81

	Ep100Syll	6.50	1.78	6.10 - 6.90
	Ip100Syll	10.63	3.61	9.82 - 11.45
	Fpe100Syll	1.27	.956	1.05 - 1.48
	Fpi100Syll	2.47	1.47	2.14 - 2.80
	DysR	2.69	1.60	2.33 - 3.06
	ErrASU	10.19	9.73	8.00 - 12.39
	LexErr	2.63	3.86	1.76 - 3.51
	GramErr	5.72	6.94	4.15 - 7.28
	PronErr	1.83	2.87	1.18 - 2.48
Accuracy and Complexity	ErrASUsyll	1.11	1.07	.868 - 1.35
	LexErrSyll	.284	.422	.189 - .379
	GramErrSyll	.626	.796	.446 - .805
	PronErrSyll	.199	.305	.130 - .268
	TTR	.400	.048	.389 - .411
	Guiraud	7.58	.704	7.42 - 7.74
	WDxASU	8.35	1.12	8.10 - 8.61

Table U.2: Descriptives in the L2 Descriptive task Group

		M	*SD*	*95% CI*
Fluency	SR	141.98	21.61	137.11 - 146.85
	ASD	.287	.049	.276 - .299
	AR	3.58	.658	3.43 - 3.73
	MLoR	.076	.022	.071 - .081
	PhonTimeR	66.82	7.81	65.06 - 68.58
	Durpse	.679	.137	.648 - .710

	Durpsi	.601	.098	.578 - .623
	Durpfe	.564	.141	.532 - .596
	Durpfi	.555	.134	.524 - .585
	PDur	.654	.101	.631 - .677
	pTimeR	33.17	7.81	31.41 - 34.93
	pFreq	30.65	6.82	29.11 - 32.19
	pseFreq	9.19	1.85	8.77 - 9.61
	psiFreq	13.12	3.75	12.27 - 13.96
	pfeFreq	1.75	1.16	1.49 - 2.01
	pfiFreq	3.26	2.07	2.80 - 3.73
	psfeFreq	10.70	2.60	10.14 - 11.26
	psfiFreq	16.29	4.48	15.33 - 17.25
	ExtpTimeR	12.08	3.50	11.29 - 12.87
	IntpTimeR	16.19	4.81	15.11 - 17.28
	P100Syll	22.02	5.572	20.76 - 23.27
	Ep100Syll	6.59	1.55	6.24 - 6.94
	Ip100Syll	9.53	3.38	8.77 - 10.29
	Fpe100Syll	1.29	.982	1.07 - 1.51
	Fpi100Syll	2.40	1.62	2.03 - 2.76
	DysR	3.36	1.89	2.93 - 3.79
	ErrASU	.577	9.74	8.37 - 12.77
Accuracy and Complexity	LexErr	1.58	2.41	1.04 - 2.13
	GramErr	7.38	7.32	5.73 - 9.03
	PronErr	1.60	3.47	.821 - 2.38
	ErrASUsyll	1.22	.996	1.00 - 1.44

	LexErrSyll	.182	.276	.120 - .245
	GramErrSyll	.865	.786	.688 - 1.04
	PronErrSyll	1.04	.896	.846 - 1.25
	Guiraud	7.44	.740	7.27 - 7.61
	TTR	.389	.042	.379 - .398
	WDxASU	7.97	.941	7.76 - 8.18

Table U.3: Descriptives in the L2 Argumentative task One-to-one

		M	*SD*	*95% CI*
Fluency	SR	136.08	19.71	131.60 - 140.55
	ASD	.308	.052	.296 - .320
	AR	3.36	.766	3.18 - 3.53
	MLoR	.075	.020	.071 - .075
	PhonTimeR	68.95	10.08	66.66 - 71.24
	Durpse	.765	.234	.711 - .818
	Durpsi	.643	.125	.614 - .671
	Durpfe	.581	.151	.546 - .615
	Durpfi	.541	.101	.518 - .564
	PDur	.645	.091	.625 - .666
	pTimeR	31.04	10.08	28.75 - 33.33
	pFreq	29.01	8.92	26.98 - 31.04
	pseFreq	8.46	1.74	8.06 - 8.85
	psiFreq	14.37	3.75	13.52 - 15.22
	pfeFreq	1.85	1.20	1.57 - 2.12

	pfiFreq	4.17	2.21	3.66 - 4.67
	ExtpTimeR	12.61	4.38	11.61 - 13.61
	IntpTimeR	19.18	5.88	17.85 - 20.52
	P100Syll	21.96	8.14	20.11 - 23.81
	Ep100Syll	6.33	1.49	5.99 - 6.67
	Ip100Syll	10.95	3.88	10.07 - 11.83
	Fpe100Syll	1.38	.912	1.18 - 1.59
	Fpi100Syll	3.17	1.80	2.76 - 3.58
	DysR	2.55	1.45	2.22 - 2.88
	ErrASU	14.90	11.26	12.34 - 17.45
	LexErr	3.99	4.23	3.02 - 4.95
	GramErr	8.98	7.76	7.22 - 10.74
	PronErr	1.92	2.76	1.29 - 2.55
Accuracy and Complexity	ErrASUsyll	1.38	.977	1.16 - 1.61
	LexErrSyll	.375	.386	.287 - .463
	GramErrSyll	.834	.680	.680 - .989
	PronErrSyll	.179	.246	.123 - .235
	Guiraud	7.28	.819	7.10 - 7.47
	TTR	.400	.064	.386 - .415
	WDxASU	8.92	1.51	8.57 - 9.26

Table U.4: Descriptives in the L2 Argumentative task Group

		M	*SD*	*95% CI*
Fluency	SR	134.71	23.00	129.46 - 139.97

	ASD	.329	.074	.312 - .346
	AR	3.16	.616	3.02 - 3.30
	MLoR	.068	.026	.062 - .073
	PhonTimeR	71.96	9.70	69.74 - 74.17
	Durpse	.722	.164	.685 - .760
	Durpsi	.660	.130	.630 - .690
	Durpfe	.589	.149	.555 - .624
	Durpfi	.567	.111	.541 - .592
	PDur	.617	.092	.596 - .638
	pTimeR	28.03	9.70	25.82 - 30.25
	pFreq	27.14	8.18	25.27 - 29.01
	pseFreq	7.40	2.03	6.94 - 7.87
	psiFreq	14.93	3.39	14.16 - 15.71
	pfeFreq	1.90	1.11	1.64 - 2.15
	pfiFreq	4.95	2.53	4.37 - 5.53
	ExtpTimeR	10.63	3.29	9.88 - 11.38
	IntpTimeR	21.17	5.83	19.83 - 22.50
	P100Syll	21.25	8.72	19.16 - 23.14
	Ep100Syll	5.64	1.84	5.22 - 6.06
	Ip100Syll	11.59	4.08	10.66 - 12.52
	Fpe100Syll	1.48	.931	1.26 - 1.69
	Fpi100Syll	3.84	2.10	3.36 - 4.32
	DysR	3.35	1.83	2.93 - 3.77
Accuracy and Complexity	ErrASU	19.63	15.51	16.09 - 23.18
	LexErr	4.63	5.22	3.44 - 5.82

GramErr	11.72	9.69	9.50 - 13.93
PronErr	3.27	4.67	2.21 - 4.34
ErrASUsyll	1.80	1.30	1.50 - 2.10
LexErrSyll	.432	.464	.326 - .538
GramErrSyll	1.06	.804	.882 - 1.25
PronErrSyll	.302	.422	.206 - .399
Guiraud	7.59	.859	7.40 - 7.79
TTR	.420	.059	.407 - .434
WDxASU	9.08	1.13	8.82 - 9.34

Appendix U2: L1 speech descriptives

Table U2.1: L1 Speech data for each task and context.

			Descriptive one-to-one	Descriptive group	Argumentative one-to-one	Argumentative group
PTR	*M*		73.11	73.16	76.18	76.30
	95% CI	*Lower Bound*	71.25	71.28	74.58	74.51
		Upper Bound	74.97	75.05	77.78	78.09
	SD		7.69	8.25	6.85	7.77
SR	*M*		167.37	167.53	169.78	167.89
	95% CI	*Lower Bound*	162.34	162.26	164.47	162.39
		Upper Bound	172.39	172.81	175.10	173.39
	SD		2.75	23.09	22.77	23.91
AR	*M*		3.83	3.84	3.72	3.67
	95% CI	*Lower Bound*	3.73	3.72	3.62	3.57
		Upper Bound	3.92	3.97	3.83	3.78
	SD		.40	.54	.44	.44
ASD	*M*		.26	.26	.27	.28
	95% CI	*Lower Bound*	.26	.26	.26	.27
		Upper Bound	.27	.27	.28	.28
	SD		.03	.03	.04	.03
MLoR	*M*		.059	.063	.057	.057
	95% CI	*Lower Bound*	.055	.059	.053	.052
		Upper Bound	.063	.068	.061	.062
	SD		.020	.021	.018	.023
durpse	*M*		.70	.70	.71	.70

	95% CI	*Lower Bound*	.67	.67	.67	.66
		Upper Bound	.74	.74	.75	.74
	SD		.14	.17	.17	.18
durpsi	*M*		.61	.60	.61	.61
	95% CI	*Lower Bound*	.59	.56	.58	.58
		Upper Bound	.64	.63	.64	.64
	SD		.10	.13	.13	.13
durfpe	*M*		.58	.64	.58	.61
	95% CI	*Lower Bound*	.54	.60	.55	.57
		Upper Bound	.62	.67	.61	.64
	SD		.17	.17	.14	.16
durfpi	*M*		.55	.57	.57	.57
	95% CI	*Lower Bound*	.52	.54	.54	.54
		Upper Bound	.58	.59	.59	.60
	SD		.12	.12	.12	.13
PDur	*M*		.63	.64	.66	.64
	95% CI	*Lower Bound*	.61	.61	.63	.61
		Upper Bound	.65	.67	.69	.66
	SD		.10	.12	.13	.12
PTimeR	*M*		26.89	26.84	23.82	23.70
	95% CI	*Lower Bound*	25.03	24.95	22.22	21.91
		Upper Bound	28.75	28.72	25.42	25.49
	SD		7.69	8.25	6.85	7.77
pFreq	*M*		25.68	25.19	21.85	22.32
	95% CI	*Lower Bound*	24.17	23.72	2.58	21.04
		Upper Bound	27.18	26.67	23.13	23.60

	SD		6.22	6.45	5.47	5.57
pseFreq	*M*		8.58	8.95	8.30	7.78
	95% CI	*Lower Bound*	7.96	8.33	7.83	7.29
		Upper Bound	9.21	9.58	8.78	8.26
	SD		2.58	2.74	2.04	2.12
psiFreq	*M*		1.55	9.97	11.27	11.63
	95% CI	*Lower Bound*	9.70	9.20	1.53	1.91
		Upper Bound	11.41	1.74	12.00	12.34
	SD		3.55	3.37	3.16	3.11
pfpeFreq	*M*		1.51	1.60	1.62	1.75
	95% CI	*Lower Bound*	1.29	1.36	1.39	1.45
		Upper Bound	1.73	1.84	1.85	2.05
	SD		.91	1.04	.98	1.32
pfpiFreq	*M*		2.31	2.14	2.84	2.73
	95% CI	*Lower Bound*	1.93	1.73	2.39	2.31
		Upper Bound	2.69	2.54	3.29	3.15
	SD		1.58	1.78	1.93	1.82
DysR	*M*		2.52	2.44	2.37	2.31
	95% CI	*Lower Bound*	2.20	2.15	2.10	1.98
		Upper Bound	2.84	2.74	2.64	2.63
	SD		1.34	1.30	1.16	1.42
Px100Syll	*M*		15.76	15.46	13.25	13.79
	95% CI	*Lower Bound*	14.52	14.31	12.25	12.61
		Upper Bound	16.99	16.61	14.25	14.97
	SD		5.09	5.05	4.29	5.14
EPx100Syll	*M*		5.20	5.42	4.97	4.76

	95% CI	*Lower Bound*	4.78	5.01	4.65	4.39
		Upper Bound	5.62	5.84	5.30	5.12
	SD		1.74	1.80	1.38	1.59
IPx100Syll	*M*		6.52	6.16	6.85	7.21
	95% CI	*Lower Bound*	5.87	5.57	6.29	6.57
		Upper Bound	7.17	6.74	7.42	7.85
	SD		2.69	2.55	2.42	2.77
FPEx100Syll	*M*		.93	.99	.98	1.10
	95% CI	*Lower Bound*	.78	.83	.83	.88
		Upper Bound	1.08	1.14	1.12	1.32
	SD		.60	.68	.63	.95
FPIx100Syll	*M*		1.44	1.33	1.74	1.71
	95% CI	*Lower Bound*	1.18	1.07	1.46	1.40
		Upper Bound	1.69	1.59	2.03	2.02
	SD		1.06	1.14	1.24	1.35
TTR	*M*		.46	.46	.45	.42
	95% CI	*Lower Bound*	.44	.45	.44	.41
		Upper Bound	.47	.48	.46	.44
	SD		.06	.06	.05	.05
Guiraud	*M*		9.21	9.33	9.02	8.66
	95% CI	*Lower Bound*	8.99	9.12	8.80	8.46
		Upper Bound	9.44	9.54	9.24	8.86
	SD		.92	.92	.94	.87
ErrxASU	*M*		2.58	7.71	7.49	1.23
	95% CI	*Lower Bound*	-8.68	4.45	4.58	6.27
		Upper Bound	49.84	1.97	1.39	14.19

	SD		12.88	14.26	12.46	17.22
LexErrR	*M*		1.36	3.97	3.29	4.81
	95% CI	*Lower Bound*	-4.33	2.31	1.16	2.87
		Upper Bound	25.06	5.63	5.41	6.74
	SD		6.70	7.28	9.10	8.43
GrErrR	*M*		1.19	3.59	4.06	5.05
	95% CI	*Lower Bound*	-4.44	1.56	2.73	2.75
		Upper Bound	24.82	5.63	5.40	7.34
	SD		6.43	8.90	5.73	9.96
PrErrR	*M*		.02	.15	.14	.38
	95% CI	*Lower Bound*	-.02	-.06	.00	.04
		Upper Bound	.07	.35	.27	.72
	SD		.20	.88	.58	1.47
ErrxASU_100syll	*M*		.55	.57	.58	.74
	95% CI	*Lower Bound*	.38	.41	.42	.55
		Upper Bound	.72	.73	.74	.92
	SD		.71	.69	.68	.82
LexErrR_100syll	*M*		.26	.30	.23	.35
	95% CI	*Lower Bound*	.14	.20	.12	.22
		Upper Bound	.38	.41	.33	.47
	SD		.49	.45	.46	.55
GrErrR_100syll	*M*		.28	.26	.34	.37
	95% CI	*Lower Bound*	.19	.16	.24	.25
		Upper Bound	.37	.36	.44	.48
	SD		.38	.42	.41	.49
PrErrR_100syll	*M*		.00	.01	.01	.02

	95% CI	*Lower Bound*	.00	.00	.00	.01
		Upper Bound	.01	.02	.02	.04
	SD		.02	.04	.05	.07
WdxASU	*M*		14.24	1.90	1.88	12.10
	95% CI	*Lower Bound*	6.45	9.68	1.07	11.11
		Upper Bound	22.03	12.11	11.70	13.09
	SD		32.19	5.31	3.47	4.29

Appendix U3: Native Speakers' ratings

Table U3.1: Ratings by Native Speakers of English in the Descriptive task, one-to-one context (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
Comprehensibility	6.96	0.86	6.77 – 7.14	1-9
Accentedness	4.85	1.26	4.58 – 5.12	1-9

Table U3.2: Ratings by Native Speakers of English in the Descriptive task, group audience (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
Comprehensibility	6.88	0.94	6.68 – 7.08	1-9
Accentedness	4.86	1.32	4.58 – 5.15	1-9

Table U3.3: Ratings by Native Speakers of English in the Argumentative task, one-to-one (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
Comprehensibility	6.62	1.01	6.40 – 6.84	1-9
Accentedness	5.05	1.29	4.77 – 5.33	1-9

Table U3.4: Ratings by Native Speakers of English in the Argumentative task, group audience (L2)

	M	*SD*	*95% CI*	*Range (Min-Max)*
Comprehensibility	6.86	1.02	6.64 – 7.08	1-9
Accentedness	4.91	1.28	4.64 – 5.19	1-9

Appendix U4: Correlations on L2 speech, proficiency, cognitive factors and anxiety

Table U4.1: Two-tailed Pearson correlation between EIT, WM, attention, L2 speaking anxiety and L2 speech measures

Note: *N* = 344; **p* < .05 (2-tailed); ***p* < .001 (2-tailed).
Comp. = Comprehensibility; Acc. = Accentedness

		EIT	WM	Att.	HRm	EDAm	IDm	Self-reports
	EIT	-	.284**	.053	-.041	-.047	-.165**	-.144*
	WM	.284**	-	.127*	-.038	-.034	-.113*	-.070*
	Attention	.053	.127*	-	-.014	-.031	-.085*	-.115**
	Comp.	.303*	.022	.031	.020	-.040	-.052	-.275**
	Acc.	-.227*	-.051	-.050	.012	.013	.027	.234**
Speed and breakdown fluency	SR	.334	.135*	.142**	-.015	-.040	-.271*	-.231**
	ASD	.142*	.089	.034	.006	.056	.184**	.092
	AR	-.142**	-.089	-.034	-.006	-.056	-.184**	-.092
	MLoR	.128*	.063	.250*	-.055	-.125*	-.099	-.171**
	PhonTimeR	.164*	-.047	.112	.040	.102	-.091	-.175*
	Durpse	-.070	-.054	-.102	.010	.072	.062	.021
	Durpsi	-.052	-.100	-.078	.058	.013	.192**	.116*
	Durpfe	-.089	-.058	-.146*	.060	.001	.231**	.134*
	Durpfi	-.127*	-.049	-.114*	.020	.001	.145**	.091
	pDur	-.094	-.050	-.122*	.002	.025	.062	.017
	pTimeR	-.147**	-.081	-.124*	.043	.102	.095	.170**
	ExtpTimeR	-.022	-.034	-.243**	.028	.068	-.005	.029
	IntpTimeR	-.204	-.106*	-.183*	.089	.103	.264**	.238*
	pFreq	-.112*	-.042	-.137*	.049	.100	.072	.176**
	pseFreq	-.005	-.127*	-.186*	.032	.123*	.140*	.036
	psiFreq	-.209	-.039	-.123*	.078	.148*	.137*	.266*
	pfeFreq	-.101	-.091	-.211*	.066	.167*	.122*	.123*
	pfiFreq	-.067	-.111*	-.135*	.060	.003	.186*	.111*
	P100Syll	-.250**	-.107*	-.123*	.054	.095	.192*	.219**
	Ep100Syll	-.203**	-.065	-.230**	.053	.124**	.045	.123*
	Ip100Syll	-.305**	-.107*	-.155*	.075	.150**	.230*	.268**
	Fpe100Syll	-.022	-.056	-.226*	.073	.160*	.178**	.167**
	Fpi100Syll	-.201**	-.126	-.124*	.003	.042	.214**	.149**
	DysR	-.098	-.193*	-.110*	.010	.045	.031	.061

Accuracy and Complexity	ErrASU	-.495**	-.023	-.033	.007	.024	.141*	.173*
	LexErr	-.391**	-.037	-.049	.021	.015	.074	.012
	GramErr	-.453**	-.064	-.047	.082	.068	.175*	.171**
	PronErr	-.347**	-.010	-.066	.064	.123	.144	.118
	ErrASUsyll	-.502**	-.016	-.021	.010	.010	.132*	.170**
	LexErrSyll	-.359**	-.024	-.031	.043	.003	.039	.022
	GramErrSyll	-.450**	-.034	-.036	.081	.058	.162*	.110
	PronErrSyll	-.338**	-.048	-.047	.067	.166*	.108	.065
	TTR	.080	.122*	.047	-.103	-.085	-.168**	-.082
	Guiraud	.385	.012	.159*	-.065	-.015	-.260**	-.354**
	WDxASU	.032	.129*	.070	-.059	-.023	-.008	-.123*

Appendix V1: Regressions with physiological measures

Table V1.1: Hierarchical multiple regression models on fluency, accuracy and complexity measures with L2 proficiency, cognitive factors, and physiological measures of anxiety as predictor variables.

		EIT		WM and attention		HRm and EDAm	
		ΔR^2	*p*	ΔR^2	*p*	ΔR^2	*p*
Speed fluency	Comp.	0.92	< .001	.008	.220	.000	.993
	Acc.	.052	< .001	.010	.153	.003	.556
	SR	.112	< .001	.056	< .001	.002	.645
	ASD	.020	.009	.014	.081	.003	.580
	MLoR	.016	.018	.060	< .001	.017	.041
Breakdown fluency	PhonTimeR	.027	.002	.017	.051	.011	.134
	Durpse	.005	.193	.014	.085	.005	.387
	Durpsi	.003	.332	.018	.044	.004	.496
	Durpfe	.008	.117	.026	.017	.003	.603
	Durpfi	.016	.019	.017	.056	.000	.918
	pDur	.009	.083	.024	.015	.001	.857
	pTimeR	.022	.006	.020	.029	.011	.130
	ExtpTimeR	.000	.682	.059	< .001	.005	.399
	IntpTimeR	.042	< .001	.021	.022	.018	.055
	pFreq	.012	.039	.005	.406	.011	.147
	pseFreq	.000	.927	.044	< .001	.015	.073
	psiFreq	.044	< .001	.007	.287	.024	.014
	pfeFreq	.010	.076	.049	< .001	.032	.005
	pfiFreq	.005	.212	.020	.030	.004	.486
	P100Syll	.062	< .001	.027	.008	.011	.122
	Ep100Syll	.041	.061	.050	< .001	.016	.045
	Ip100Syll	.093	< .001	.017	.043	.025	.008
	Fpe100Syll	.000	.699	.051	< .001	.031	.006
	Fpi100Syll	.040	< .001	.045	< .001	.002	.705
Accuracy and Complexity	DysR	.010	.073	.057	< .001	.001	.804
	ErrASU	.245	< .001	.017	.027	.001	.855
	LexErr	.153	< .001	.019	.113	.003	.682
	GramErr	.205	< .001	.017	.043	.012	.108
	PronErr	.120	< .001	.023	.134	.032	.057
	ErrASUsyll	.252	< .001	.010	.109	.001	.831
	LexErrSyll	.129	< .001	.009	.383	.006	.495
	GramErrSyll	.203	< .001	.009	.176	.010	.146
	PronErrSyll	.114	< .001	.013	.321	.051	.011

TTR	.006	.139	.006	.346	.016	.061
Guiraud	.148	< .001	.020	.018	.003	.588
WDxASU	.001	.555	.016	.067	.004	.473

Table V1.2: Hierarchical multiple regression models on fluency, accuracy and complexity measures with L2 proficiency, cognitive factors, and physiological measures of anxiety as predictor variables.

Note: Cells highlighted in grey indicate significant beta coefficients (β), representing the strongest unique contributions of the predictor variables to the model. Other significant values are reported but not highlighted.

	R^2 (%)	*R^2 Change (%)*	*F Change*	*Sig. F Change*	*B*	*95% CI*		*t*	*p*
Comprehensibility									
Proficiency	.092 (9.2%)	.092 (9.2%)	*F* (1, 342) = 33.67	< .001	.320	.016	.031	6.07	.000
WM –	.100 (10%)	.008 (0.8%)	*F* (2, 340) = .1.52	.220	-.081	-.075	.009	-1.54	.124
Attention					-.041	-.006	.003	-.795	.427
HR –	.100 (10%)	.000 (0%)	*F* (2, 338) = .007	.993	-.006	-1.683	1.49	-.118	.906
EDA					.003	-.174	.183	.052	.959
Accentedness									
Proficiency	.052 (5.2%)	.052 (5.2%)	*F* (1, 342) = 18.61	< .001	-.246	-.034	-.014	-4.59	.000
WM –	.062 (6.2%)	.010 (1%)	*F* (2, 340) = 1.89	.153	.096	-.005	.108	1.78	.075
Attention					-.037	-.008	.004	-.699	.485
HR –	.065 (6.5%)	.003 (0.3%)	*F* (2, 338) = .589	.556	-.056	-3.28	1.01	-1.03	.300
EDA					.026	-.181	.302	.493	.623
PTR									
Proficiency	.027 (2.7%)	.027 (2.7%)	*F* (1, 342) = 9.49	.002	.172	.049	.209	3.17	.002
WM –	.044 (4.4%)	.017 (1.7%)	*F* (2, 340) = 2.99	.051	-.087	-.802	.081	-1.60	.109
Attention					.105	.000	.098	1.98	.048
HR –	.055 (5.5%)	.011 (1.1%)	*F* (2, 338) = 2.01	.134	.030	-11.94	21.48	.561	.575
EDA					.097	-.158	3.59	1.80	.072
SR									

Proficiency	.112 (11.2%)	.112 (11.2%)	F (1, 342) = 43.07	< .001	.365	.443	.774	7.22	.000
WM –	.167 (16.7%)	.056 (5.6%)	F (2, 340) = 11.35	< .001	-.206	-2.81	-.987	-4.08	.000
Attention					.125	.028	.230	2.52	.012
HR –	.170 (17%)	.002 (0.2%)	F (2, 338) = .438	.645	.024	-26.16	43.14	.482	.630
EDA					.036	-2.49	5.29	.706	.480
ASD									
Proficiency	.020 (2%)	.020 (2%)	F (1, 342) = 6.99	.009	-.165	-.002	.000	-3.02	.003
WM –	.034 (3.4%)	.014 (1.4%)	F (2, 340) = 2.53	.081	.116	.000	.008	2.12	.034
Attention					-.025	-.001	.000	-.471	.638
HR –	.038 (3.8%)	.003 (0.3%)	F (2, 338) = .546	.580	-.011	-.155	.125	-.208	.835
EDA					.057	-.007	.024	1.04	.297
MLoR									
Proficiency	.016 (1.6%)	.016 (1.6%)	F (1, 342) = 5.69	.018	-.118	.000	.000	-2.23	.026
WM –	.077 (7.7%)	.060 (6%)	F (2, 340) = 11.08	< .001	.038	-.001	.001	.722	.471
Attention					-.246	.000	.000	-4.74	.000
HR –	.094 (9.4%)	.017 (1.7%)	F (2, 338) = 3.22	.041	-.037	-.052	.025	-.698	.485
EDA					-.120	-.009	-.001	-2.28	.023
Durpse									
Proficiency	.005 (0.5%)	.005 (0.5%)	F (1, 342) = 1.70	.193	-.082	-.001	.000	-1.492	.137
WM –	.019 (1.9%)	.014 (1.4%)	F (2, 340) = 2.47	.085	.065	-.002	.007	1.187	.236
Attention					-.097	-.001	.000	-1.796	.073
HR –	.025 (2.5%)	.005 (0.5%)	F (2, 338) = .953	.387	-.026	-.209	.128	-.475	.635
EDA					.074	-.006	.032	1.359	.175
Durpsi									
Proficiency	.003 (0.3%)	.003 (0.3%)	F (1, 342) = .943	.332	-.073	-.001	.000	-1.32	.185
WM –	.021 (2.1%)	.018 (1.8%)	F (2, 340) = 3.15	.044	.113	.000	.007	2.06	.039
Attention					-.075	-.001	.000	-1.39	.165

HR –	.025 (2.5%)	.004 (0.4%)	*F* (2, 338) = .703	.496	-.064	-.207	.052	-1.17	.239
EDA					.018	-.012	.017	.336	.737
Durpfe									
Proficiency	.008 (0.8%)	.008 (0.8%)	*F* (1, 342) = 2.47	.117	-.094	-.002	.000	-1.65	.099
WM –	.034 (3.4%)	.026 (2.6%)	*F* (2, 340) = 4.11	.017	.074	-.002	.012	1.29	.197
Attention					-.143	-.002	.000	-2.54	.011
HR –	.037 (3.7%)	.003 (0.3%)	*F* (2, 338) = .506	.603	.057	-.126	.385	.995	.320
EDA					-.018	-.034	.024	-.318	.751
Durpfi									
Proficiency	.016 (1.6%)	.016 (1.6%)	*F* (1, 336) = 5.53	.019	-.135	-.002	.000	-2.46	.014
WM –	.033 (3.3%)	.017 (1.7%)	*F* (1, 334) = 2.91	.056	.074	-.002	.008	1.35	.177
Attention					-.107	-.001	.000	-1.98	.048
HR –	.034 (3.4%)	.000 (0%)	*F* (1, 332) = .086	.918	-.022	-.231	.151	-.411	.681
EDA					.006	-.020	.022	.115	.908
pDur									
Proficiency	.009 (0.9%)	.009 (0.9%)	*F* (1, 342) = 3.02	.083	-.105	-.001	.000	-1.92	.055
WM –	.033 (3.3%)	.024 (2.4%)	*F* (2, 340) = 4.24	.015	.106	.000	.006	1.93	.054
Attention					-.118	-.001	.000	-2.20	.028
HR –	.034	.001 (0.1%)	*F* (2, 338) = .155	.857	.004	-.114	.122	.065	.948
EDA					-.030	-.017	.010	-.556	.579
pTimeR									
Proficiency	.022 (2.2%)	.022 (2.2%)	*F* (1, 342) = 7.51	.006	-.153	-.003	.000	-2.84	.005
WM –	.042 (4.2%)	.020 (2%)	*F* (2, 340) = 3.57	.029	.089	-.001	.011	1.64	.101
Attention					-.119	-.001	.000	-2.23	.026
HR –	.053 (5.3%)	.011 (1.1%)	*F* (2, 338) = 2.05	.130	-.032	-.299	.162	-.587	.557
EDA					-.097	-.050	.002	-1.80	.072

pFreq									
Proficiency	.012 (1.2%)	.012 (1.2%)	F (1, 342) = 4.31	.039	-.113	-.002	.000	-2.05	.040
WM –	.018 (1.8%)	.005 (0.5%)	F (2, 340) = .903	.406	.040	-.003	.008	.723	.470
Attention					-.066	-.001	.000	-1.22	.220
HR –	.029 (2.9%)	.011 (1.1%)	F (2, 338) = 1.92	.147	-.038	-.280	.135	-.689	.491
EDA					-.092	-.043	.003	-1.69	.092
pseFreq									
Proficiency	.000 (0%)	.000 (0%)	F (1, 342) = .008	.927	.027	-.012	.020	.495	.621
WM –	.044 (4.4%)	.044 (4.4%)	F (2, 340) = 7.74	< .001	-.089	-.164	.014	-1.65	.098
Attention					-.188	-.028	-.008	-3.56	.000
HR –	.058 (5.8%)	.015 (1.5%)	F (2, 338) = 2.64	.073	-.007	-3.60	3.14	-.134	.894
EDA					-.120	-.809	-.052	-2.23	.026
psiFreq									
Proficiency	.044 (4.4%)	.044 (4.4%)	F (1, 342) = 15.62	< .001	-.219	-.090	-.032	-4.09	.000
WM –	.051 (5.2%)	.007 (0.7%)	F (2, 340) = 1.25	.287	.050	-.083	.238	.946	.345
Attention					.071	-.005	.030	1.36	.174
HR –	.074 (7.4%)	.024 (2.4%)	F (2, 338) = 4.30	.014	-.066	-9.91	2.25	-1.23	.217
EDA					-.128	-1.52	-.152	-2.40	.017
pfeFreq									
Proficiency	.010 (1%)	.010 (1%)	F (1, 310) =	.076	.119	.000	.005	2.14	.033
WM –	.059 (5.9%)	.049 (4.9%)	F (1, 308) =	< .001	-.041	-.021	.009	-.745	.457
Attention					-.220	-.005	-.002	-4.04	.000
HR –	.091 (9.1%)	.032 (3.2%)	F (1, 306) =	.005	-.034	-.726	.380	-.616	.539
EDA					-.170	-.161	-.035	-3.06	.002
pfiFreq									
Proficiency	.005 (0.5%)	.005 (0.5%)	F (1, 342) = 1.56	.212	-.091	-.010	.001	-1.65	.098
WM –	.025 (2.5%)	.020 (2%)	F (2, 340) = 3.52	.030	.129	.006	.064	2.36	.019

Attention					-.065	-.005	.001	-1.21	.224
HR –	.029 (2.9%)	.004 (0.4%)	F (2, 338) = .722	.486	-.065	-1.77	.433	-1.19	.233
EDA					.005	-.118	.130	.093	.926
EpTimeR									
Proficiency	.000 (0%)	.000 (0%)	F (1, 342) = .168	.682	-.002	-.005	.004	-.036	.971
WM –	.060 (6%)	.059 (5.9%)	F (2, 340) = 10.75	< .001	-.026	-.031	.019	-.481	.631
Attention					-.244	-.009	-.004	-4.62	.000
HR –	.065 (6.5%)	.005 (0.5%)	F (2, 338) = .923	.399	-.013	-1.07	.834	-.249	.804
EDA					-.068	-.177	.038	-1.27	.204
IpTimeR									
Proficiency	.042 (4.2%)	.042 (4.2%)	F (1, 342) = 14.84	< .001	-.231	-.004	-.001	-4.34	.000
WM –	.063 (6.3%)	.021 (2.1%)	F (2, 340) = 3.85	.022	.155	.003	.015	2.91	.004
Attention					-.009	-.001	.001	-.164	.869
HR –	.080 (8%)	.018 (1.8%)	F (2, 338) = 3.22	.055	-.084	-.426	.045	-1.59	.112
EDA					-.089	-.049	.004	-1.67	.096
DysR									
Proficiency	.010 (1%)	.010 (1%)	F (1, 336) = 3.22	.073	-.134	-.005	-.001	-2.46	.014
WM –	.067 (6.7%)	.057 (5.7%)	F (1, 334) = 10.23	< .001	.217	.012	.035	4.01	.000
Attention					-.104	-.003	.000	-1.96	.050
HR –	.068 (6.8%)	.001 (0.1%)	F (1, 332) = .218	.804	-.002	-.449	.432	-.038	.970
EDA					.035	-.033	.066	.656	.512
P100Syll									
Proficiency	.062 (6.2%)	.062 (6.2%)	F (1, 342) = 22.77	< .001	-.266	-.004	-.002	-5.04	.000
WM –	.089 (8.9%)	.027 (2.7%)	F (2, 340) = 4.95	.008	.129	.002	.015	2.44	.015
Attention					-.111	-.002	.000	-2.14	.032
HR –	.100 (10%)	.011 (1.1%)	F (2, 338) = 2.11	.122	-.050	-.371	.128	-.961	.337
EDA					-.085	-.051	.005	-1.62	.106

Ep100Syll									
Proficiency	.041 (4.1%)	.041 (4.1%)	*F* (1, 342) = 14.75	.061	-.198	-.039	-.012	-3.767	.000
WM – Attention	.092 (9.2%)	.050 (5%)	*F* (2, 340) = 9.38	< .001	.052	-.037	.112	.996	.320
					-.222	-.026	-.010	-4.31	.000
HR – EDA	.108 (10.8%)	.016 (1.6%)	*F* (2, 338) = 3.12	.045	-.039	-3.89	1.73	-.754	.451
					-.116	-.672	-.040	-2.21	.027
Ip100Syll									
Proficiency	.093 (9.3%)	.093 (9.3%)	*F* (1, 342) = 35.20	< .001	-.328	-.005	-.003	-6.36	.000
WM – Attention	.110 (11%)	.017 (1.7%)	*F* (2, 340) = 3.16	.043	.140	.002	.016	2.71	.007
					.011	-.001	.001	.217	.828
HR – EDA	.135 (13.5%)	.025 (2.5%)	*F* (2, 338) = 4.88	.008	-.067	-.414	.084	-1.30	.193
					-.132	-.065	-.009	-2.57	.011
Fpe100Syll									
Proficiency	.000 (0%)	.000 (0%)	*F* (1, 310) = .149	.699	.032	-.002	.004	.572	.568
WM – Attention	.052 (5.2%)	.051 (5.1%)	*F* (2, 308) = 8.36	< .001	.006	-.015	.017	.103	.918
					-.233	-.006	-.002	-4.25	.000
HR – EDA	.083 (8.3%)	.031 (3.1%)	*F* (2, 306) = 5.13	.006	-.046	-.835	.338	-.834	.405
					-.162	-.165	-.032	-2.90	.004
Fpi100Syll									
Proficiency	.040 (4%)	.040 (4%)	*F* (1, 336) = 14.17	< .001	-.230	-.009	-.003	-4.30	.000
WM – Attention	.085 (8.5%)	.045 (4.5%)	*F* (2, 334) = 8.18	< .001	.180	.011	.042	3.36	.001
					-.112	-.004	.000	-2.13	.033
HR – EDA	.087 (8.7%)	.002 (0.2%)	*F* (2, 332) = .349	.705	-.013	-.660	.515	-.242	.809
					.044	-.038	.093	.829	.408
TTR									
Proficiency	.006 (0.6%)	.006 (0.6%)	*F* (1, 342) = 2.20	.139	-.097	-.001	.000	-1.77	.077

WM – Attention	.013 (1.3%)	.006 (0.6%)	*F* (2, 340) = 1.06	.346	.083 .021	-.001 .000	.005 .000	1.51 .392	.130 .695
HR – EDA	.029 (2.9%)	.016 (1.6%)	*F* (2, 338) = 2.82	.061	-.096 -.070	-.189 -.019	.011 .004	-1.75 -1.27	.081 .204
Guiraud									
Proficiency	.148 (14.8%)	.148 (14.8%)	*F* (1, 342) = 59.40	< .001	.395	.017	.029	7.81	.000
WM – Attention	.168 (16.8%)	.020 (2%)	*F* (2, 340) = 4.09	.018	-.102 .101	-.065 .000	-.001 .007	-2.01 2.04	.045 .042
HR – EDA	.171 (17.1%)	.003 (0.3%)	*F* (2, 338) = .532	.588	-.052 .015	-1.85 -.116	.583 .157	-1.02 .293	.306 .770
ErrxASU									
Proficiency	.245 (24.5%)	.245 (24.5%)	*F* (1, 326) = 105.60	< .001	-.520	-.017	-.012	-1.65	.000
WM – Attention	.261 (26.1%)	.017 (1.7%)	*F* (2, 324) = 3.66	.027	.118 .058	.003 -.001	.032 .003	2.42 1.20	.016 .228
HR – EDA	.262 (26.2%)	.001 (0.1%)	*F* (2, 322) = .157	.855	-.024 -.009	-.681 -.067	.408 .056	-.493 -.176	.622 .860
LexErr									
Proficiency	.153 (15.3%)	.153 (15.3%)	*F* (1, 193) = 34.84	< .001	-.425	-.011	-.006	-6.25	.000
WM – Attention	.172 (17.2%)	.019 (1.9%)	*F* (2, 191) = 2.20	.113	.120 .081	-.002 -.001	.029 .003	1.77 1.22	.077 .223
HR – EDA	.175 (17.5%)	.003 (0.3%)	*F* (2, 189) = .384	.682	-.057 .026	-.853 -.056	.343 .082	-.842 .379	.401 .705
GrErr									
Proficiency	.205 (20.5%)	.205 (20.5%)	*F* (1, 298) = 76.82	< .001	-.478	-.015	-.010	-9.19	.000
WM – Attention	.222 (22.2%)	.017 (1.7%)	*F* (2, 296) = 3.17	.043	.117 .066	.002 -.001	.033 .003	2.24 1.28	.025 .200
HR –	.233 (23.3%)	.012 (1.2%)	*F* (2, 294) = 2.23	.108	-.086	-1.02	.092	-1.64	.101

EDA					-.052	-.094	.031	-1.00	.317
PrErr									
Proficiency	.120 (12%)	.120 (12%)	*F* (1, 155) = 21.23	< .001	-.394	-.011	-.005	-5.14	.000
WM –	.143 (14.3%)	.023 (2.3%)	*F* (2, 153) = 2.04	.134	.119	-.004	.032	1.54	.125
Attention					-.036	-.003	.002	-.485	.628
HR –	.175 (17.5%)	.032 (3.2%)	*F* (2, 151) = 2.91	.057	-.094	-1.075	.244	-1.24	.215
EDA					.175	.012	.173	2.26	.025
ErrSyll									
Proficiency	.252 (25.2%)	.252 (25.2%)	*F* (1, 326) = 110.11	< .001	-.523	-.016	-.011	-1.72	.000
WM –	.263 (26.3%)	.010 (1%)	*F* (2, 324) = 2.23	.109	.091	-.001	.026	1.86	.062
Attention					.046	-.001	.002	.970	.333
HR –	.264 (26.4%)	.001 (0.1%)	*F* (2, 322) = .185	.831	-.029	-.663	.351	-.606	.545
EDA					.008	-.053	.062	.159	.873
LexErrSyll									
Proficiency	.129 (12.9%)	.129 (12.9%)	*F* (1, 193) = 28.61	< .001	-.386	-.010	-.005	-5.58	.000
WM –	.138 (13.8%)	.009 (0.9%)	*F* (2, 191) = .963	.383	.080	-.006	.023	1.16	.247
Attention					.059	-.001	.002	.870	.386
HR –	.144 (14.4%)	.006 (0.6%)	*F* (2, 189) = .707	.495	-.077	-.891	.243	-1.12	.261
EDA					.039	-.047	.084	.562	.575
GrErrSyll									
Proficiency	.203 (20.3%)	.203 (20.3%)	*F* (1, 298) = 75.72	< .001	-.469	-.014	-.009	-8.96	.000
WM –	.212 (21.2%)	.009 (0.9%)	*F* (2, 296) = 1.74	.176	.083	-.003	.026	1.59	.113
Attention					.056	-.001	.002	1.07	.282
HR –	.222 (22.2%)	.010 (1%)	*F* (2, 294) = 1.93	.146	-.086	-.960	.088	-1.63	.103
EDA					-.040	-.081	.036	-.769	.442
PrErrSyll									

Proficiency	.114 (11.4%)	.114 (11.4%)	*F* (1, 155) = 19.99	< .001	-.383	-.010	-.004	-5.02	.000
WM – Attention	.127 (12.7%)	.013 (1.3%)	*F* (2, 153) = 1.14	.321	.074	-.008	.024	.970	.334
					-.023	-.002	.002	-.311	.756
HR – EDA	.178 (17.8%)	.051 (5.1%)	*F* (2, 151) = 4.67	.011	-.109	-1.02	.159	-1.44	.150
					.225	.034	.179	2.91	.004
WDxASU									
Proficiency	.001 (0.1%)	.001 (0.1%)	*F* (1, 342) = .349	.555	-.051	-.001	.000	-.927	.355
WM – Attention	.017 (1.7%)	.016 (1.6%)	*F* (2, 340) = 2.72	.067	.106	.000	.005	1.93	.054
					.070	.000	.001	1.29	.196
HR – EDA	.021 (2.1%)	.004 (0.4%)	*F* (2, 338) = .751	.473	.061	-.045	.164	1.12	.262
					-.037	-.016	.008	-.680	.497

Appendix V2: Regressions with subjective measures

Table V2.1: Hierarchical multiple regression models on fluency, accuracy and complexity measures with L2 proficiency, working memory, and subjective measures of anxiety as predictor variables.

		EIT		WM and attention		IDm and self-reports	
		ΔR^2	*p*	ΔR^2	*p*	ΔR^2	*p*
Speed fluency	Comp.	.091	< .001	.006	.339	.041	< .001
	Acc.	.057	< .001	.009	.219	.034	.003
	SR	.120	< .001	.051	< .001	.036	< .001
	ASD	.015	.027	.013	.108	.022	.023
	MLoR	.017	.019	.064	< .001	.013	.108
	PhonTimeR	.040	< .001	.016	.063	.005	.407
Breakdown fluency	Durpse	.001	.524	.016	.073	.002	.768
	Durpsi	.002	.424	.016	.037	.028	.059
	Durpfe	.009	.094	.027	.018	.039	.052
	Durpfi	.021	.008	.016	.071	.013	.118
	pDur	.006	.150	.025	.016	.001	.793
	pTimeR	.032	.001	.020	.034	.003	.555
	ExtpTimeR	.000	.884	.067	< .001	.001	.772
	IntpTimeR	.049	< .001	.016	.060	.050	.051
	pFreq	.025	.004	.005	.419	.018	.049
	pseFreq	.000	.977	.045	< .001	.024	.016
	psiFreq	.055	< .001	.006	.342	.036	.002
	pfeFreq	.015	.057	.054	< .001	.020	.038
	pfiFreq	.006	.172	.021	.029	.024	.019
	P100Syll	.086	< .001	.025	.011	.021	.020
	Ep100Syll	.043	.071	.053	< .001	.008	.256
	Ip100Syll	.112	< .001	.010	.157	.040	< .001
	Fpe100Syll	.001	.513	.057	< .001	.032	.006
	Fpi100Syll	.045	< .001	.045	< .001	.022	.022
	DysR	.012	.053	.050	< .001	.001	.889
Accuracy and Complexity	ErrASU	.263	< .001	.012	.071	.003	.519
	LexErr	.142	< .001	.012	.114	.002	.840
	GramErr	.224	< .001	.014	.078	010	.147
	PronErr	.120	< .001	.022	.156	.008	.507
	ErrASUsyll	.273	< .001	.007	.238	.003	.565
	LexErrSyll	.124	< .001	.009	.382	.002	.827

GramErrSyll	.222	< .001	.007	.261	.009	.192
PronErrSyll	.109	< .001	.014	.319	.002	.860
TTR	.006	.172	.007	.337	.022	.025
Guiraud	.149	< .001	.022	.014	.033	.001
WDxASU	.001	.641	.016	.074	.021	.030

Note: Cells in grey indicate significance; Comp. = Comprehensibility, Acc. = Accentedness

Table V2.2: Hierarchical multiple regression models on fluency, accuracy and complexity measures with L2 proficiency, working memory, and subjective measures of anxiety as predictor variables.

Note: Cells highlighted in grey indicate significant beta coefficients (β), representing the strongest unique contributions of the predictor variables to the model. Other significant values are reported but not highlighted.

	R^2 (%)	R^2 Change (%)	*F Change*	*Sig. F Change*	*B*	*95% CI*		*t*	*p*
Comprehensibility									
Proficiency	.091 (9.1%)	.091 (9.1%)	*F* (1, 328) = 32.77	< .001	.263	.012	.028	4.77	.000
WM – Attention	.097 (9.7%)	.006 (0.6 %)	*F* (2, 326) = 1.08	.339	-.047	-.062	.023	-.891	.373
					-.047	-.007	.002	-.901	.368
ID – Self-reports	.138 (13.8%)	.041 (4.1%)	*F* (2, 324) = 7.77	< .001	.085	-.035	.262	1.50	.134
					-.229	-.010	-.003	-3.94	.000
Accentedness									
Proficiency	.057 (5.7%)	.057 (5.7%)	*F* (1, 328) = 19.72	< .001	-.209	-.032	-.010	-3.717	.000
WM – Attention	.065 (6.5%)	.009 (0.9%)	*F* (2, 326) = 1.52	.219	.073	-.018	.097	1.351	.178
					-.036	-.008	.004	-.677	.499
ID – Self-reports	.099 (9.9%)	.034 (3.4%)	*F* (2, 324) = 6.03	.003	-.101	-.378	.023	-1.746	.082
					.204	.003	.012	3.429	.001
PTR									
Proficiency	.040 (4%)	.040 (4%)	*F* (1, 328) = 13.71	< .001	.198	.068	.236	3.55	.000
WM – Attention	.056 (5.6%)	.016 (1.6%)	*F* (2, 326) = 2.78	.063	-.070	-.738	.158	-1.27	.203
					.096	-.005	.092	1.76	.078
ID – Self-reports	.061 (6.1%)	.005 (0.5%)	*F* (2, 324) = .901	.407	-.008	-1.777	1.555	-.131	.896
					-.069	-.831	.230	-1.11	.266

SR									
Proficiency	.120 (12%)	.120 (12%)	*F* (1, 328) = 44.54	< .001	.333	.398	.742	6.51	.000
WM – Attention	.171 (17.1%)	.051 (5.1%)	*F* (2, 326) = 10.10	< .001	-.161	-2.391	-.560	-3.17	.002
					.112	.014	.213	2.25	.025
ID – Self-reports	.207 (20.7%)	.036 (3.6%)	*F* (2, 324) = 7.44	< .001	-.154	-8.029	-1.217	-2.67	.008
					-.068	-1.738	.431	-1.18	.237
ASD									
Proficiency	.015 (1.5%)	.015 (1.5%)	*F* (1, 328) = 4.90	.027	-.110	-.001	.000	-1.96	.051
WM – Attention	.028 (2.8%)	.013 (1.3%)	*F* (2, 326) = 2.24	.108	.088	-.001	.007	1.58	.113
					-.023	.000	.000	-.420	.675
ID – Self-reports	.050 (5%)	.022 (2.2%)	*F* (2, 324) = 3.79	.023	.158	.004	.031	2.48	.013
					-.009	-.005	.004	-.139	.890
MLoR									
Proficiency	.017 (1.7%)	.017 (1.7%)	*F* (1, 328) = 5.55	.019	-.104	.000	.000	-1.90	.058
WM – Attention	.081 (8.1%)	.064 (6.4%)	*F* (2, 326) = 11.38	< .001	.012	-.001	.001	.229	.819
					-.242	.000	.000	-4.53	.000
ID – Self-reports	.093 (9.3%)	.013 (1.3%)	*F* (2, 324) = 2.23	.108	.008	-.004	.004	.128	.898
					.109	.000	.002	1.79	.073
Durpse									
Proficiency	.001 (0.1%)	.001 (0.1%)	*F* (1, 328) = .406	.524	-.034	-.001	.001	-.596	.552
WM – Attention	.017 (1.7%)	.016 (1.6%)	*F* (2, 326) = 2.64	.073	.067	-.002	.007	1.18	.236
					-.100	-.001	.000	-1.81	.071
ID – Self-reports	.019 (1.9%)	.002 (0.2%)	*F* (2, 324) = .265	.768	.034	-.012	.021	.535	.593
					.011	-.005	.006	.181	.857
Durpsi									
Proficiency	.002 (0.2%)	.002 (0.2%)	*F* (1, 328) = .640	.424	-.026	-.001	.000	-.470	.639
WM –	.022 (2.2%)	.020 (2%)	*F* (2, 326) = 3.33	.037	.090	-.001	.006	1.62	.105

Attention					-.069	-.001	.000	-1.26	.206
ID – Self-reports	.050 (5%)	.028 (2.8%)	*F* (2, 324) = 4.79	.059	.159	.004	.029	2.51	.012
					.025	-.003	.005	.396	.693
Durpfe									
Proficiency	.009 (0.9%)	.009 (0.9%)	*F* (1, 298) = 2.82	.094	-.059	-.002	.001	-1.01	.312
WM – Attention	.036 (3.6%)	.027 (2.7%)	*F* (2, 296) = 4.07	.018	.040	-.004	.009	.699	.485
					-.132	-.002	.000	-2.33	.020
ID – Self-reports	.075 (7.5%)	.039 (3.9%)	*F* (2, 294) = 6.21	.052	.202	.014	.065	3.09	.002
					.004	-.008	.008	.068	.946
Durpfi									
Proficiency	.021 (2.1%)	.021 (2.1%)	*F* (1, 322) = 7.01	.008	-.128	-.002	.000	-2.26	.024
WM – Attention	.037 (3.7%)	.016 (1.6%)	*F* (2, 320) = 2.63	.073	.044	-.003	.007	.780	.436
					-.107	-.001	.000	-1.95	.052
ID – Self-reports	.050 (5%)	.013 (1.3%)	*F* (2, 318) = 2.15	.118	.132	.001	.039	2.05	.041
					-.047	-.008	.004	-.740	.460
pDur									
Proficiency	.006 (0.6%)	.006 (0.6%)	*F* (1, 328) = 2.07	.150	-.087	-.001	.000	-1.54	.123
WM – Attention	.031 (3.1%)	.025 (2.5%)	*F* (2, 326) = 4.16	.016	.104	.000	.006	1.85	.064
					-.116	-.001	.000	-2.11	.035
ID – Self-reports	.032 (3.2%)	.001 (0.1%)	*F* (2, 324) = .232	.793	.041	-.008	.016	.640	.523
					-.032	-.005	.003	-.510	.611
pTimeR									
Proficiency	.032 (3.2%)	.032 (3.2%)	*F* (1, 328) = 10.89	.001	-.175	-.003	-.001	-3.13	.002
WM – Attention	.052 (5.2%)	.020 (2%)	*F* (2, 326) = 3.41	.034	.072	-.002	.010	1.29	.195
					-.112	-.001	.000	-2.06	.040
ID – Self-reports	.055 (5.5.%)	.003 (0.3%)	*F* (2, 324) = .589	.555	.028	-.018	.028	.438	.662
					.041	-.005	.010	.667	.506

pFreq									
Proficiency	.025 (2.5%)	.025 (2.5%)	*F* (1, 328) = 8.25	.004	-.117	-.002	.000	-2.01	.044
WM – Attention	.030 (3%)	.005 (0.5%)	*F* (2, 326) = .873	.419	.008	-.005	.006	.143	.886
					-.061	-.001	.000	-1.11	.264
ID – Self-reports	.048 (4.8%)	.018 (1.8%)	*F* (2, 324) = 3.05	.049	-.011	-.021	.018	-.190	.850
					.145	.000	.001	2.36	.019
pseFreq									
Proficiency	.000 (0%)	.000 (0%)	*F* (1, 328) = .001	.977	-.002	-.017	.017	-.033	.974
WM – Attention	.045 (4.5%)	.045 (4.5%)	*F* (2, 326) = 7.68	< .001	-.067	-.146	.034	-1.21	.225
					-.198	-.028	-.008	-3.67	.000
ID – Self-reports	.069 (6.9%)	.024 (2.4%)	*F* (2, 324) = 4.16	.016	-.180	-.825	-.153	-2.86	.004
					.065	-.050	.164	1.05	.294
psiFreq									
Proficiency	.055 (5.5%)	.055 (5.5%)	*F* (1, 328) = 18.96	< .001	-.179	-.081	-.019	-3.172	.002
WM – Attention	.061 (6.1%)	.006 (0.6%)	*F* (2, 326) = 1.07	.342	-.019	-.190	.133	-.347	.729
					.090	-.002	.032	1.693	.091
ID – Self-reports	.097 (9.7%)	.036 (3.6%)	*F* (2, 324) = 6.51	.002	.045	-.343	.787	.773	.440
					.180	.007	.031	3.016	.003
pfeFreq									
Proficiency	.015 (1.5%)	.015 (1.5%)	*F* (1, 298) = 4.38	.057	.169	.001	.007	2.92	.004
WM – Attention	.069 (6.9%)	.054 (5.4%)	*F* (2, 296) = 8.58	< .001	-.086	-.027	.003	-1.52	.129
					-.210	-.005	-.002	-3.74	.000
ID – Self-reports	.089 (8.9%)	.020 (2%)	*F* (2, 294) = 3.30	.038	.134	.003	.115	2.07	.039
					.025	-.014	.021	.391	.696
pfiFreq									
Proficiency	.006 (0.6%)	.006 (0.6%)	*F* (1, 328) = 1.87	.172	-.064	-.009	.002	-1.13	.256

WM – Attention	.027 (2.7%)	.021 (2.1%)	*F* (2, 326) = 3.56	.029	.106	-.001	.059	1.91	.057
					-.057	-.005	.002	-1.04	.298
ID – Self-reports	.051 (5.1%)	.024 (2.4%)	*F* (2, 324) = 4.02	.019	.154	.026	.249	2.42	.016
					.009	-.033	.038	.145	.885
EpTimeR									
Proficiency	.000 (0%)	.000 (0%)	*F* (1, 328) = .021	.884	.025	-.004	.006	.450	.653
WM – Attention	.067 (6.7%)	.067 (6.7%)	*F* (2, 326) = 11.68	< .001	-.027	-.031	.019	-.499	.618
					-.256	-.009	-.004	-4.73	.000
ID – Self-reports	.068 (6.8%)	.001 (0.1%)	*F* (2, 324) = .259	.772	-.037	-.122	.065	-.596	.552
					.039	-.020	.039	.639	.524
IpTimeR									
Proficiency	.049 (4.9%)	.049 (4.9%)	*F* (1, 328) = 16.75	< .001	-.198	-.003	-.001	-3.66	.000
WM – Attention	.065 (6.5%)	.016 (1.6%)	*F* (2, 326) = 2.84	.060	.096	-.001	.012	1.79	.074
					.011	-.001	.001	.210	.834
ID – Self-reports	.115 (11.5%)	.050 (5%)	*F* (2, 324) = 9.21	.051	.181	.012	.057	2.96	.003
					.080	-.002	.012	1.32	.186
DysR									
Proficiency	.012 (1.2%)	.012 (1.2%)	*F* (1, 323) = 3.77	.053	-.140	-.005	-.001	-2.49	.013
WM – Attention	.061 (6.1%)	.050 (5%)	*F* (2, 321) = 8.51	< .001	.203	.010	.033	3.64	.000
					-.110	-.003	.000	-2.00	.046
ID – Self-reports	.062 (6.2%)	.001 (0.1%)	*F* (2, 319) = .118	.889	-.022	-.051	.036	-.344	.731
					-.009	-.015	.013	-.136	.892
P100Syll									
Proficiency	.086 (8.6%)	.086 (8.6%)	*F* (1, 328) = 30.74	< .001	-.279	-.005	-.002	-5.21	.000
WM – Attention	.111 (11.1%)	.025 (2.5%)	*F* (2, 326) = 4.60	.011	.093	-.001	.012	1.757	.080
					-.100	-.001	.000	-1.91	.056
ID –	.132 (13.2%)	.021 (2.1%)	*F* (2, 324) = 3.95	.020	.082	-.008	.041	1.35	.176

Self-reports					.091	-.002	.014	1.52	.128
Ep100Syll									
Proficiency	.043 (4.3%)	.043 (4.3%)	*F* (1, 328) = 14.75	.071	-.201	-.041	-.013	-3.69	.000
WM – Attention	.096 (9.6%)	.053 (5.3%)	*F* (2, 326) = 9.47	< .001	.043	-.046	.108	.799	.425
					-.221	-.026	-.009	-4.17	.000
ID – Self-reports	.103 (10.3%)	.008 (0.8%)	*F* (2, 324) = 1.36	.256	-.060	-.426	.145	-.970	.333
					.099	-.015	.166	1.63	.102
Ip100Syll									
Proficiency	.112 (11.2%)	.112 (11.2%)	*F* (1, 328) = 41.32	< .001	-.316	-.005	-.002	-6.01	.000
WM – Attention	.122 (12.2%)	.010 (1%)	*F* (2, 326) = 1.85	.157	.076	-.002	.011	1.45	.148
					.036	.000	.001	.695	.487
ID – Self-reports	.162 (16.2%)	.040 (4%)	*F* (2, 324) = 7.68	< .001	.109	-.002	.046	1.82	.068
					.128	.001	.016	2.18	.030
Fpe100Syll									
Proficiency	.001 (0.1%)	.001 (0.1%)	*F* (1, 298) = .430	.513	.088	-.001	.005	1.53	.127
WM – Attention	.058 (5.8%)	.057 (5.7%)	*F* (2, 296) = 8.93	< .001	-.046	-.023	.009	-.811	.418
					-.221	-.005	-.002	-3.94	.000
ID – Self-reports	.090 (9%)	.032 (3.2%)	*F* (2, 294) = 5.16	.006	.163	.017	.135	2.52	.012
					.038	-.013	.025	.603	.547
Fpi100Syll									
Proficiency	.045 (4.5%)	.045 (4.5%)	*F* (1, 322) = 15.30	< .001	-.206	-.009	-.003	-3.75	.000
WM – Attention	.090 (9%)	.045 (4.5%)	*F* (2, 320) = 7.93	< .001	.158	.008	.039	2.91	.004
					-.104	-.003	.000	-1.96	.051
ID – Self-reports	.112 (11.2%)	.022 (2.2%)	*F* (2, 318) = 3.86	.022	.137	.007	.124	2.20	.028
					.026	-.015	.022	.418	.676
TTR									

Proficiency	.006 (0.6%)	.006 (0.6%)	*F* (1, 328) = 1.87	.172	-.062	-.001	.000	-1.08	.277
WM – Attention	.012 (1.2%)	.007 (0.7%)	*F* (2, 326) = 1.09	.337	.056	-.001	.004	1.00	.315
					.027	.000	.000	.486	.627
ID – Self-reports	.034 (3.4%)	.022 (2.2%)	*F* (2, 324) = 3.71	.025	.159	.003	.023	2.48	.013
					-.013	-.003	.003	-.202	.840
Guiraud									
Proficiency	.149 (14.9%)	.149 (14.9%)	*F* (1, 328) = 57.57	< .001	.363	.016	.028	7.07	.000
WM – Attention	.171 (17.1%)	.022 (2.2%)	*F* (2, 326) = 4.29	.014	-.077	-.058	.007	-1.51	.130
					.090	.000	.007	1.80	.072
ID – Self-reports	.204 (20.4%)	.033 (3.3%)	*F* (2, 324) = 6.76	.001	-.162	-.294	-.051	-2.79	.006
					-.045	-.054	.023	-.784	.433
ErrxASU									
Proficiency	.263 (26.3%)	.263 (26.3%)	*F* (1, 312) = 111.54	< .001	-.522	-.017	-.012	-1.45	.000
WM – Attention	.276 (27.6%)	.012 (1.2%)	*F* (2, 310) = 2.66	.071	.084	-.002	.027	1.68	.092
					.067	.000	.003	1.37	.171
ID – Self-reports	.279 (27.9%)	.003 (0.3%)	*F* (2, 308) = .657	.519	.054	-.029	.081	.937	.350
					.006	-.017	.019	.114	.909
LexErr									
Proficiency	.142 (14.2%)	.142 (14.2%)	*F* (1, 190) =	< .001	-.400	-.011	-.005	-5.81	.000
WM – Attention	.162 (16.2%)	.020 (2%)	*F* (2, 188) =	.114	.115	-.002	.028	1.67	.095
					.087	-.001	.003	1.27	.205
ID – Self-reports	.163 (16.3%)	.002 (0.2%)	*F* (2, 186) =	.840	.047	-.041	.075	.590	.556
					-.025	-.022	.016	-.323	.747
GrErr									
Proficiency	.224 (22.4%)	.224 (22.4%)	*F* (1, 285) =	< .001	-.474	-.016	-.010	-8.90	.000
WM – Attention	.238 (23.8%)	.014 (1.4%)	*F* (2, 283) =	.078	.074	-.005	.027	1.39	.165
					.082	.000	.003	1.57	.118

ID –	.248 (24.8%)	.010 (1%)	*F* (2, 281) =	.147	.102	-.009	.104	1.66	.097
Self-reports					.003	-.018	.019	.041	.967
PrErr									
Proficiency	.120 (12%)	.120 (12%)	*F* (1, 148) = 20.18	< .001	-.361	-.011	-.004	-4.49	.000
WM –	.142 (14.2%)	.022 (2.2%)	*F* (2, 146) = 1.88	.156	.140	-.002	.035	1.76	.079
Attention					-.024	-.003	.002	-.309	.758
ID –	.150 (15%)	.008 (0.8%)	*F* (2, 144) = .683	.507	.031	-.056	.080	.340	.735
Self-reports					.071	-.013	.031	.790	.431
ErrSyll									
Proficiency	.273 (27.3%)	.273 (27.3%)	*F* (1, 312) = 116.93	< .001	-.528	-.016	-.011	-1.59	.000
WM –	.279 (27.9%)	.007 (0.7%)	*F* (2, 310) = 1.44	.238	.058	-.006	.022	1.16	.245
Attention					.051	-.001	.002	1.05	.293
ID –	.282 (28.2%)	.003 (0.3%)	*F* (2, 308) = .572	.565	.061	-.024	.079	1.06	.288
Self-reports					-.024	-.020	.013	-.422	.673
LexErrSyll									
Proficiency	.124 (12.4%)	.124 (12.4%)	*F* (1, 190) = 26.82	< .001	-.370	-.010	-.004	-5.29	.000
WM –	.133 (13.3%)	.009 (0.9%)	*F* (2, 188) = .967	.382	.078	-.006	.023	1.12	.264
Attention					.058	-.001	.002	.837	.404
ID –	.134 (13.4%)	.002 (0.2%)	*F* (2, 186) = .190	.827	.027	-.046	.064	.337	.736
Self-reports					-.049	-.023	.012	-.616	.539
GrErrSyll									
Proficiency	.222 (22.2%)	.222 (22.2%)	*F* (1, 285) = 81.16	< .001	-.469	-.014	-.009	-8.74	.000
WM –	.229 (22.9%)	.007 (0.7%)	*F* (2, 283) = 1.35	.261	.041	-.009	.021	.774	.439
Attention					.068	-.001	.002	1.30	.193
ID –	.238 (23.8%)	.009 (0.9%)	*F* (2, 281) = 1.65	.192	.106	-.007	.099	1.70	.089
Self-reports					-.021	-.020	.014	-.334	.739

PrErrSyll									
Proficiency	.109 (10.9%)	.109 (10.9%)	*F* (1, 148) = 18.06	< .001	-.344	-.010	-.003	-4.22	.000
WM –	.123 (12.3%)	.014 (1.4%)	*F* (2, 146) = 1.15	.319	.113	-.005	.029	1.41	.161
Attention					-.014	-.002	.002	-.176	.861
ID –	.124 (12.4%)	.002 (0.2%)	*F* (2, 144) = .151	.860	.025	-.054	.070	.264	.792
Self-reports					.026	-.017	.023	.281	.779
WDxASU									
Proficiency	.001 (0.1%)	.001 (0.1%)	*F* (1, 328) = .217	.641	-.043	-.001	.000	-.765	.445
WM –	.017 (1.7%)	.016 (1.6%)	*F* (2, 326) = 2.63	.074	.104	.000	.005	1.85	.064
Attention					.088	.000	.001	1.61	.108
ID –	.038 (3.8%)	.021 (2.1%)	*F* (2, 324) = 3.56	.030	-.087	-.017	.003	-1.36	.174
Self-reports					.167	.001	.008	2.66	.008

Appendix W1: Table with codes and frequencies (%) between highly, mid and low anxious learners

Factors	Sources	*TOTAL (%)*	Total instances (%) High anxious	Mid	Low anxious
Performance	Being recorded	4 (0.55%)	1 (0.33%)	1 (0.52%)	2 (0.90%)
	Topic difficulty/ unfamiliarity	19 (2.63%)	10 (3.26%)	3 (1.55%)	6 (2.71%)
	Time constraints	39 (5.40%)	20 (6.51%)	9 (4.64%)	10 (4.52%)
	Using English	2 (0.28%)	1 (0.33%)	0 (0%)	1 (0.45%)
Linguistic	Grammar	12 (1.66%)	4 (1.30%)	4 (2.06%)	4 (1.81%)
	Pronunciation	25 (3.46%)	5 (1.63%)	12 (6.19%)	8 (3.62%)
	Not knowing the word in English	29 (4.02%)	8 (2.61%)	10 (5.15%)	11 (4.98%)
	L1 interference	13 (1.80%)	2 (0.65%)	3 (1.55%)	8 (3.62%)
	Sounding repetitive	25 (3.46%)	13 (4.23%)	6 (3.09%)	6 (2.71%)
	Realizing making a mistake	55 (7.62%)	14 (4.56%)	20 (10.31%)	21 (9.50%)
	Realizing pausing excessively	18 (2.49%)	12 (3.91%)	3 (1.55%)	3 (1.36%)
Cognitive	Mental blocks	53 (7.34%)	26 (8.47%)	18 (9.28%)	9 (4.07%)
	Lack of ideas	97 (13.44%)	44 (14.33%)	30 (15.46%)	23 (10.41%)
	Problems to recall ideas	30 (4.15%)	10 (3.26%)	12 (6.19%)	8 (3.62%)
	Problems to recall task instructions	11 (1.52%)	6 (1.95%)	0 (0%)	5 (2.26%)
	Problems to recall words	120 (16.62%)	60 (19.54%)	22 (11.34%)	38 (17.19%)
	Problems to express themselves	100 (13.85%)	38 (12.38%)	30 (15.46%)	32 (14.48%)
Audience	Fear of negative evaluation	36 (4.99%)	17 (5.54%)	9 (4.64%)	10 (4.52%)
	Notetaking	2 (0.28%)	1 (0.33%)	0 (0%)	1 (0.45%)
	Having an audience	8 (1.11%)	7 (2.28%)	1 (0.52%)	0 (0%)
	Sharing something personal	22 (3.05%)	6 (1.95%)	1 (0.52%)	15 (6.79%)
	Audience reactions	2 (0.28%)	2 (0.65%)	0 (0%)	0 (0%)
	TOTAL	*722 (100%)*	*307 (100%)*	*194 (100%)*	*221 (100%)*

Appendix W2: Graph with codes and frequencies (%) for highly and low anxious group