UNIVERSITAT DE BARCELONA

Individual Differences in the Development of the Oral Fluency of American learners of Spanish Studying Abroad

Sapna Sehgal

Individual Differences in the Development of the Oral Fluency of American learners of Spanish Studying Abroad

Tesi presentada per

Sapna Sehgal per obtenir el titol de doctor per la Universitat de Barcelona

Universitat de Barcelona/2021

Directors:

Dr. Joan Carles Mora

Dra. Raquel Serrano

Tutor:

Dr. Joan Carles Mora

Programa de doctorat: Ciència Cognitiva i Llenguatge

Linea de recerca: Lingüística teòrica i aplicada

Facultat de Filologia i Comunicació

Universitat de Barcelona

"The brick walls are there for a reason. The brick walls are not there to keep us out. The brick walls are there to give us a chance to show how badly we want something. Because the brick walls are there to stop the people who don't want it badly enough. They're there to stop the other people."

- Randy Paush, *The Last Lecture*

七転び八起き

"Fall seven times. Rise eight."

- *Ancient Japanese proverb*

DEDICATION

This dissertation is dedicated to all the ambitious, motivated, unwearied first year PhD candidates out there who are about to embark on this wild journey we call a doctoral dissertation. May you have the courage to face the brick walls, the strength to climb over them, the tenacity to rise again when you fall multiple times, and the opportunity to witness your own success.

ACKNOWLEDGEMENTS

It is with humility and gratefulness to those supporting my journey that I have reached the end of this doctoral thesis. To my PhD supervisors, Dr. Joan C. Mora and Dr. Raquel Serrano, thank you for your patience and for pushing me to be and achieve more, especially in the final stages of this endeavour. I am also thankful for APIF, the Generalitat of Catalonia grant that funded three years of this project.

To Prof. Dr. Nivja de Jong, I am grateful for your expert guidance in the field of fluency research and the opportunity to do a research stay at Utrecht University with you; it greatly enhanced my doctoral experience and academic presentation skills (and my knowledge of PRAAT scripting!). To Dr. Burçin Guclu, for your expertise in statistics and all things math, and Dr. Elisa Ruiz Tada for your patience and tips and tricks in my penultimate summer of this project, thank you.

While it would be challenging to mention everyone who has excused my absence as I disappeared periodically from their lives over the last 6+ years to delve deeper into my thesis, there are several individuals who deserve my eternal gratitude:

To Camille Garewal, Sean Malden, Meghan-Cree Smith, Janice Kung, Pamela Osborne, Julie Shepherd and my other Canadian friends always there for me even if I lose touch, and even from 10,000 km away - thank you. To Alexandra Tigan, my close friend who is always there to lend a helping hand – whether in reviewing my thesis, taking wild travel breaks with me, or supporting

me in my other endeavours-- I am grateful that I met you at Oxford, and we that have grown closer since. To Dr. Kimberly Marsh, whose Oxford PhD story and continual kindness reminded me never to give up. To Kristin Greminger (and Sean Malden) for our pre-pandemic annual summer road trips – spending time travelling, pondering life and laughing with both of you rejuvenated me more than you will ever know.

Especially to Kiara, without whose bounding, puppy-like energy and uplifting attitude towards life I would never have made it - all the love in the world is not enough. And to Brielle, who has added even more energy, love and light into my life.

To my fellow doctoral student cohort and my friends in Barcelona - Dr. Marina Ruiz Tada, Dr. Vita Kogan, Dr. Angelica Carlet, Dr. Joanna Corey, Robby de Menezes Montefusco, Rob Muñoz, Meritxell Agyuasanosa, Carly Collins, Valeria Galimberti and Anya Grak - for sharing my successes, failures, pain, tears, struggles, for believing in me and for reminding me that the light at the end of the tunnel is not always an oncoming train. For being my shoulder to cry on, my sounding board, my constant moral support, and truthfully, my family. To my personal trainer Mario Melero and Be Legend gym's professional kickboxing team, for being there when it felt like it was too much to handle - thank you.

To my many, many ESL students who have brightened my days making my side job a fun-filled venture, supporting me and sending me more students, and patiently letting me move their classes when it came to “crunch time” for my PhD – thank you for being a joy in my life and for choosing me to help you reach your goals.

To *The Teaching Cove* blog readers and community- your patience in my sporadic blogging periods over the last 2 years is a blessing and with this project behind me, I am truly excited for the future my blog, online store, and business hold.

To Vita, for always believing in me, from the day our friendship began.

To Robby, especially, for introducing me to the dark, dark world of programming and Python and for, with wizard-like data skills, stopping me (twice) from falling into an emotional breakdown...

...To Dr. Gisela Pi Casaus and especially Helena Garcia Alonso for catching me when I eventually did, and helping me heal. Your patience, humility, dedication, kindness, empathy, and care go beyond your professional duty, and touch my heart. (Again, thanks to my dear friend Marina without whom I would not have found either).

-- To the renowned authors and psychologists whose written words pushed me to survive the final two years of a harrowing journey- Dr. Susan David, Dr. Stan Tankin, Dr. Richard Schwartz, Dr. Lorna Benjamin-Smith, Dr. Jean Shinoda Bolen, Dr. Steven Gilligan, your books have helped me more than you'll ever know.

Thank you. I am eternally grateful.

Now, to a new chapter of my life...

ABSTRACT

This doctoral dissertation examines the effects of a short stay abroad (SA) experience on second language (L2) oral fluency development, taking into consideration individual differences in cognition and the study abroad experience. Relatively little attention has been paid to the cognitive aspects that influence L2 oral fluency gains. Following Segalowitz's (2010) model, which emphasizes the underlying cognitive processes that affect L2 speech development, the present study investigates both utterance and cognitive fluency, specifically looking at the relationship between inhibitory control (an under-researched aspect of L2 cognition in the SA context) on L2 fluency gains. Forty-nine American students studying Spanish abroad in Barcelona were tested. First, this study investigates whether L2 oral fluency improves over a short stay abroad period, using an in-depth battery of L2 oral fluency measures. Second, we examine the relationship between L2 oral fluency gains and inhibitory control in the SA context, as we hypothesized that greater inhibitory control ability was related to L2 oral fluency gains. Third, we assess the extent to which individual differences in SA experience factors (such as the amount of self-reported language use, living situation and hours of language classes taken) affect L2 oral fluency gains over the SA period.

Picture based speech elicitation tasks were used to collect both L1 and L2 oral fluency data, and L2 oral fluency results were adjusted for the L1 where possible. To measure inhibitory control, both non-linguistic and linguistic inhibitory control tasks were used: a Simon task, a Letter (Phoneme) Decision task, and an L1-L2 and L2-L1 language switching task. Participants also completed a post-test questionnaire about their study abroad language use and experience.

Results showed little improvement in L2 speed, breakdown, and repair fluency development after a stay abroad. Participants spoke for a longer duration and had longer fluent runs. Contrary to our predictions, after adjusting for the L1, speed fluency data showed that participants spoke significantly slower at the end of their stay, while breakdown fluency results indicate that participants exhibited more silent and filled pauses at the end of their day, only becoming significantly less disfluent on one breakdown fluency measure (between clause silent pauses). A composite disfluency measure showed participants, on average, were less fluent in their L2 at post-test.

Inhibitory control ability was not found to relate to L2 oral fluency gains for most measures. A decrease in one disfluency measure –filled pauses between speech units (ASUs) – was significantly related to performance on the Letter (Phoneme) Decision test. Interestingly, participants were able to switch significantly faster into Spanish at the end of their stay than the beginning, which may be one indicator of increased inhibitory control.

Most fluency gains were not related to self-reported language use or other experience factors. Only 4 (of 25) fluency measures were significantly related to self-reported language use, one measure was significantly related to classroom instruction hours, and one was significantly related to living situation. The results imply that these specific experiential factors themselves do not have a large impact on participants' L2 oral fluency gains while abroad, although questionnaire data indicates participants perceived these factors to greatly affect their L2 oral fluency development.

Taken together, our findings suggest that a study abroad experience does not necessarily lead to L2 oral fluency gains. Measuring frequency and duration in L2 oral fluency measures and adjusting for the L1 brings insight to the data. Experience factors do not relate to L2 oral fluency gains, contrary to participant expectations and beliefs. Findings from the present study could be applied to pedagogy in the development of future SA programs.

RESUM

Aquesta tesis doctoral investiga els efectes d'una estada a l'estranger en el desenvolupament de la fluïdesa en la llengua castellana d'un grup de 49 americans que van estudiar durant tres mesos a Barcelona. Seguint el model cognitiu de Segalowitz (2010), que es concentra en els processos cognitius darrere de la producció oral en les segones llengües (L2), aquest estudi investiga tan el concepte de fluïdesa en el discurs com la fuïdesa cognitiva que influeix en el discurs. Primerament, aquest estudi avalua si una estada curta a l'estranger resulta en una millora de la fluïdesa oral, utilitzant una bateria de mesures, ajustant-les a la manera individual de parlar l'anglès. En segon lloc, investiguem la relació entre les millores en la fluïdesa i el control de la inhibició, un aspecte de la cognició poc estudiat en la lingüística. Terecrament, avaluem com les diferències individuals en la seva experiència afecten en la millora de l'idioma durant l'estada a l'estranger. Els participants van realitzar proves del discurs en castellà i en anglès, proves d'inhibició lingüística i no lingüística, i un qüestionari sobre l'experiència després de l'estada.

Els resultats mostren molt poca millora en la fluïdesa en la L2 en les tres categories de fluïdesa: la rapidesa, la disrupció per disfluències en el discurs i la repetició, i l'autocorrecció d'errors. Contrari a les expectatives, els participants parlen amb *menys* rapidesa i fluïdesa al final de la seva estada, en la majoria de les mesures, menys una -tenien significativament menys pauses en la meitat de les clàusules en comparació al començament de l'estada. L'inhibició no estava relacionada amb els canvis i millores en la fluïdesa, encara que els participants puguessin canviar de l'anglès al castellà més ràpidament al final de l'estada. Molts d'aquests factors no influeixen en la fluïdesa, encara que el qüestionari mostra que els participants ho creuen. En conclusió, els resultats d'aquest estudi poden ser útils per la pedagogia amb la finalitat de crear nous programes d'estudi en l'estranger.

Table of Contents

DEDICATION i
ACKNOWLEDGEMENTS ii
ABSTRACT v
RESUM viii
LIST OF FIGURES xii
LIST OF TABLES xii
1 Introduction 1
1.1 What is Study Abroad? 2
1.2 Outline of this dissertation 6
2 Literature review 7
2.1 Context of learning in SLA 9
2.1.1 Formal instruction 12
2.1.2 SA as a learning context 17
2.2 L2 performance abroad 19
2.2.1 Cognitive variables 24
2.2.2 Emotional variables 25
2.2.3 Contextual variables in SA 28
2.3 What is oral fluency and why is it important? 33
2.3.1 Speech production models 33
2.3.2 Definitions of fluency 50
2.3.3 How to measure utterance fluency 53
2.4 Oral fluency development in SLA 63
2.4.1 L1 role in L2 fluency 64
2.4.2 Fluency development in L2 learners at home 69
2.4.3 Fluency development in L2 learners abroad 73
2.5 Cognitive individual differences in SLA 81
2.5.1 Attention 87
2.5.2 Working memory 89
2.6 What is inhibitory control and why is it important in L2 speech? 93
2.6.1 Inhibitory control in SLA research 94
2.6.2 Language switching and inhibition 101

2.7 Study abroad experience factors and L2 fluency 104
2.8 Gaps in the literature 107
3 Aims and research questions 111
3.1 Aims of the present study 111
3.2 Research questions 114
4 Methodology 118
4.1 Study design 118
4.2 Participants 121
4.2.1 Excluded participants 126
4.3 Tasks and measures 127
4.3.1 General procedure 127
4.3.2 Speech elicitation tasks (English and Spanish) 128
4.3.3 Inhibitory control task- language switching (English and Spanish) 136
4.3.4 Inhibitory control task (Simon) 144
4.3.5 Inhibitory control task (Letter (Phoneme) Decision) 147
4.3.6 Questionnaire 151
4.4 Data Analysis 156
4.4.1 Pre-processing of speech data 156
4.4.2 Summary of all measures used in analysis 164
4.4.3 Normality of the distribution 166
5 Results 168
5.1 L2 Oral Fluency Behaviour 168
5.1.1 Speed fluency 169
5.1.2 Breakdown fluency 173
5.1.3 Repair fluency 195
5.1.4 Composite disfluency measure 199
5.1.5 Summary of results 201
5.2 Inhibitory Control 201
5.2.1 Non-linguistic inhibitory control 202
5.2.2 Linguistic inhibitory control 203
5.2.3 Language switching as inhibitory control 204
5.2.4 Inhibitory control and fluency gains 210

5.2.5 Summary of results ... 220
5.3 Language use and study abroad experience ... 222
5.3.1 Self-reported language use percentage ... 222
5.3.2 Classroom instruction hours and fluency gains ... 226
5.3.3 Other language use factors and fluency gains ... 228
5.3.4 Other study abroad experience data ... 229
5.3.5 Summary of results ... 238
6 Discussion ... 240
6.1 L2 Fluency ... 241
6.1.1 Speed Fluency ... 242
6.1.2 Breakdown Fluency ... 244
6.1.3 Repair Fluency ... 248
6.2 Inhibitory Control ... 252
6.3 Study Abroad experience factors ... 257
6.4 Limitations of the present study and future research ... 266
7 Conclusion and Pedagogical Implications ... 272
References ... 276
Appendix A - Oral fluency picture prompts ... 287
Appendix B - Language switching task picture prompts and vocabulary lists ... 291
Appendix C - Letter (Phoneme) decision task words ... 294
Appendix D - LEXTALE-ESP vocabulary test ... 295
Appendix E - Questionnaire ... 298
Appendix F - PRAAT ... 303
Appendix G - Segalowitz’s L2 speech model (2010) ... 304
Appendix H – Normality tests ... 305
Appendix I – L1 and L2 Oral fluency correlation tables ... 311
Appendix J – L2 oral fluency correlations tables ... 314
Appendix K –Questionnaire and fluency gains tables (correlations) ... 319

LIST OF FIGURES

Figure 1 - Blueprint of a Speaker (Levelt, 1999, p.87).. 34
Figure 2 - Kormos, 2006, p.168.. 36
Figure 3 – L2 Speech Production Model (Encoding) Segalowitz 2010 p.9 37
Figure 4 – Fluency Relevant Experiences – Segalowitz, 2010, p.131... 40
Figure 5 – Research Design Schema 119
Figure 6 - Speech Elicitation Task 1 Version 1 133
Figure 7 - Language Switching Task Diagram 142
Figure 8 - Simon Task Diagram... 145
Figure 9 - Picture Selection Task Diagram.. 148
Figure 10 - Speed Fluency Relative Gains (n=49) 171
Figure 11 - Relative Gains on Breakdown Fluency – Silent Pauses (n=49).............................. 184
Figure 12 - Breakdown Fluency Relative Gains -Silent Pause Location & Duration (n=49) 186
Figure 13 -Breakdown Fluency Relative Gains -Filled Pause Location & Duration (n=49) 192
Figure 14 - Repair Fluency Mean Relative Gains (n=49) 198
Figure 15 - Pre & Post Test Switch Scores.. 208
Figure 16 - Individual Differences in Switch Gain.. 209
Figure 17- Effect of SA on Spanish level (text-mining analysis).. 235
Figure 18 - Most helpful aspect of SA experience in learning Spanish (self-reported) 236
Figure 19 - Perceived aspects of SA experience that helped L2 Spanish learning.................... 237

LIST OF TABLES

Table 1 - Participants' Study Abroad Demographic and Experience Factor Information (n=52) 125
Table 2 - Participants per task.. 126
Table 3 - Speed Fluency (Raw Counts, Rates & L2 Gains) 172
Table 4 - Breakdown Fluency (Raw Counts, Rates and L2 Gains - Silent Pause Measures) n = 49 175
Table 5- Breakdown Fluency (Raw Counts, Rates and L2 Gains - Filled Pause Measures n=49 187
Table 6 - Repair Fluency Descriptive Statistics (n=49)... 197
Table 7 - Total Disfluency Measures (n=49)... 200
Table 8 - Simon Test Results (n=52) 202
Table 9 - Letter (Phoneme) Decision Test Results (n=52) 204
Table 10 - Descriptive Statistics for Overflow Measures.. 205
Table 11 - Descriptive Statistics & Switch Score & Switch Gain Measures (n=35).................. 207
Table 12 – Pearson Correlations of Inhibitory Control Measures and Speed Fluency Gains (n=49).. 210

Table 13 – Spearman Rank correlations of PICS Inhibitory Control Measure and Speed Fluency Gains (n=49) ... 211
Table 14 - Pearson Correlations of Inhibitory Control Measures and Breakdown Fluency Gains (n=49)... 213
Table 15 - Spearman Rank Correlations of Inhibitory Control Measures and General Breakdown Fluency Gains (n=49) ... 214
Table 16 - Pearson Correlations of Inhibitory Control Measures and Silent Pause Breakdown Fluency Gains (n=49) ... 215
Table 17 - Spearman Rank Correlations of Inhibitory Control Measures and Silent Pause Breakdown Fluency Gains (n=49) ... 216
Table 18 - Pearson Correlations of Inhibitory Control Measures and Filled Pause Breakdown Fluency Gains (n=49) ... 217
Table 19 - Spearman Rank Correlations of Inhibitory Control Measures and Filled Pause Breakdown Fluency Gains (n=49) ... 218
Table 20 - Pearson Correlations of Inhibitory Control Measures and Repair Fluency Gains (n=49)... 219
Table 21 -Spearman rank correlations of PICS Inhibitory Control Measure and Repair Fluency Gains (n=49) ... 220
Table 22 - Self Reported L2 Use ... 222
Table 23 - Spanish weekly classroom contact hours ... 227
Table 24 – Descriptive Statistics for Self-Reported Confidence Level ... 230
Table 25 - Shapiro Wilk test for ICS scores ... 305
Table 26 - Shapiro Wilk test for PICS scores ... 305
Table 27- Kruskal Wallis Test for Significant Differences in Letter (Phoneme) Decision Versions ... 306

1 Introduction

Studying abroad is a popular educational opportunity for students learning a second language, and the number of study abroad (SA) students continues to increase every year. According to the NAFSA Association of International Educators, this number increased 3.8% from 2015 to 2016 for American students alone (Murphy, 2017). The OECD (2012) reports the number of students studying outside their home countries has multiplied five-fold since the 1990s. More recent data from the Institute of International Education points to the importance of SA as a learning context, as it shows the number of American students studying abroad each year is growing by 3% (IIE, 2018), a trend that, up until 2020, was expected to continue. Due to the global pandemic and related travel restrictions, the decreasing number of students studying abroad in the 2020-21 and 2021-22 academic years is expected to temporarily slow this trend. Popular belief holds that an immersion environment, especially abroad, is superior to other contexts for learning a second language, though second language acquisition (SLA) literature only partially supports this idea. As language learners, many students believe they will simply become surrounded by native speakers and "pick up" the language while abroad, yet as applied linguists we know that a plethora of individual differences influence that language learning experience. For students, becoming fluent in the foreign language classroom is often seen as an impossibility (Hessel, 2017), and an SA experience is considered the optimal context.

In fact, my interest in this topic originally arose from the misguided assumption that studying abroad is always the best way to learn a language, regardless of individual or contextual differences. My personal experience as an undergraduate student fully immersed in a French speaking environment with a host family resulted in a marked increase in oral fluency, seemingly confirming my belief; however, several peers in SA environments for the same length of stay (about 4 months) reported little opportunity or motivation for L2 language use. Many returned with second language (L2) French speaking skills similar to those they had had before leaving for France. The stark individual differences in the effectiveness of our short SA experience in terms of oral fluency development, a skill I found to be one least practised in formal instruction (FI) language learning classrooms, fueled my research interest in this research area.

1.1 What is Study Abroad?

Sometimes referred to as “stay abroad”, “language soujourn”, “foreign language exchange” or “residence abroad”, *study abroad* is a general term used for sojourners who spend an amount of time away from their home country to learn a foreign language or experience a cultural exchange living in another country. It is often understood as “a period of time immersed in the target speech community” (LaBrozzi, 2012, p. 229). While these students are generally university-aged students, cultural and language exchange programs also exist for younger and older learners, and range in duration from a few weeks in the summer to an 11 to 13-week semester, to an academic year. Some students complete a semester-long or year-long exchange to satisfy requirements for their degree; this particularly applies to foreign language majors at American and British universities who are often required to spend one to two semesters in a foreign country that speaks the L2 and complete courses in their chosen area of study. Others go on SA experiences to a country where their L2 is spoken as a part of their university experience; most take classes in the L2 and live in university residences in their host country, though some opt to live with a host family. Importantly, some academic institutions count the marks students receive in their SA classes as “Pass/Fail”, removing their impact on the students’ overall Grade

Point Average as they prefer that students focus on their experience and not their academics (e.g.“University of Alberta Exchange Programs,” n.d.). Conversely, other programs include the exchange institutions’ marks on students’ transcripts, which may directly affect their motivation to do well in their SA classes (Baker-Smemoe, Dewey, Bown, & Martinsen, 2014).

Despite the importance and growing frequency of the SA context we do not currently have a clear picture of the impact of the SA experience on second language (L2) speaking fluency development. Some studies in the Second Language Acquisition (SLA) field find the SA experience to positively affect language learning, though they are not consistent in their findings as to *which* learners improve, *what* linguistic gains they acquire abroad, and *how* best to make the most of the SA experience. In fact, many reviews of the literature find inconsistencies and “wild variability” (Tullock & Ortega, 2017, p. 3) in participants’ experiences even while on the same SA program. Furthermore, there are “serious methodological inconsistencies” (Tullock & Ortega, 2017, p. 3) in the studies themselves, such as the use of different fluency measures and a lack of justification for why particular measures are chosen, making it difficult to draw clear comparisons of results across studies.

In a comprehensive meta-analysis of 66 quantitative and 11 qualitative studies, Yang (2016) finds the SA context is generally superior to the at-home (AH) context (a term used in the field to refer to in-class foreign language instruction in learners’ home country) with respect to oral proficiency, grammatical skills and vocabulary, though intensive instruction at home sometimes leads to greater L2 fluency development than longer SA stays. Yang (2016) observes that “longer time spent overseas does not relate to L2 linguistic development” (p.84); this is a finding contrary to popular opinion that spending longer periods of time abroad leads to superior L2 gains.

Some studies focus on SA students' grammatical development and find individual differences in acquisition of grammatical structures (Marqués-Pascual, 2011) and a lack of proceduralized grammar knowledge, leading to difficulty in both accuracy and fluency when speaking (DeKeyser, 2010). Others focus on listening comprehension skills, finding that while there are no significant differences in gains in SA students compared to those at home, SA learners do use different listening strategies (Cubillos, Chieffo, & Fan, 2008). Still others look at native-likeness in oral performance, finding improvements after an SA experience (Mora & Valls-Ferrer, 2012), or even phonology, finding, for example, that phonological awareness can predict learners' vocabulary development and oral fluency (O'Brien, Segalowitz, Freed, & Collentine, 2007). With regards to L2 oral fluency development, we know that speaking fluency generally improves during a SA experience. However, we also know that while speaking fluency studies outside the SA context often conduct an in-depth analysis, speaking fluency research done within the SA context is often limited to a handful of measures, as fluency is often not the main focus of the overall study. This dissertation aims to bring that depth to the study of fluency within the SA context.

While parents, students and educators may assume that being immersed in a second language context fosters L2 oral fluency, in reality, we know the outcome of a SA experience is a function of both individual learner differences (aptitude, motivation, and previous language experience, for example) and contextual differences (such as the amount of language use).

The majority of empirical research on SA effectiveness looks at American students studying foreign languages abroad (Tullock & Ortega, 2017) and most research to date focuses on the acquisition of English or Spanish (Yang, 2016). This limits the majority of our knowledge about SA to a specific social or cultural group (Block, 2003). However, it is also true that European countries receive 48% of students studying abroad worldwide, more than any other region (OECD, 2012). The Institute of International Education reports that 54% of American students studying abroad go to a European destination (IEE, 2018). Given the varied linguistic focus, and the lack of in-depth fluency measurement in SA studies, looking at Americans learners of Spanish abroad in Europe is valuable as it can provide both interesting comparisons to existing research and new additions to current findings. Individual differences among participants and their linguistic performance abroad, too, provide a challenge when assessing the effectiveness of the context.

Research in SA is a "relatively young field of inquiry" (Yang, 2016, p. 73) that needs further investigation. Several researchers caution against attributing speaking fluency success or failure to language learning contexts (such as SA), without taking cognitive individual differences into account (Freed, Segalowitz, & Dewey, 2004) Delving into this scarcely explored area, this dissertation will examine the effectiveness of the SA context in the development of *oral fluency*, with a focus on individual differences primarily cognitive differences, with some focus on differences that pertain to related to the SA experience, such as motivational and contextual factors. Where are we in our quest to understand the impact of studying abroad on L2 oral fluency? What individual and contextual differences in studying abroad affect language performance, specifically speaking fluency? If, as research suggests, studying abroad does not lead to equal L2 oral fluency gains for all students, what cognitive and contextual individual differences affect participants' oral fluency progress?

1.2 Outline of this dissertation

This dissertation seeks to contribute to closing the gap in our knowledge on the impact of individual differences on L2 oral fluency progress when studying abroad. Our study examines individual differences in L2 oral fluency development over the course of a SA period, focusing on individuals' inhibitory control ability, as well as individual differences in the SA experience, such as language input and language use. Chapter 2 reviews the literature on learning contexts, and findings in SA literature, establishing a basis for studying *oral fluency* more in depth in this context, and providing reasoning for investigating *inhibitory control* – a cognitive skill that has been overlooked in SA research – as an individual difference that could impact L2 fluency. Chapter 3 outlines the aims and research questions and Chapter 4 details the methodology of our study. Chapter 5 presents the results of the study and Chapter 6 discusses these results in the context of relevant current literature. Finally, Chapter 7 provides conclusions, ideas for further research and pedagogical implications of the present study's findings.

2 Literature review

This chapter is an explanation of the state of the art – a review of the literature in the SA context. First, in Section 2.1, we will review what we know about context of learning in SLA and establish the importance of understanding the SA context. Section 2.2 will look at linguistic performance – other than L2 *oral fluency* - in the SA context, to present a picture of what we know with respect to L2 development in different linguistic areas, with respect to contextual and emotional variables that affect performance. Studies examined look at L2 performance on various linguistic measures, students' attitudes, motivation and anxiety levels, and their experiences learning an L2 within the SA context. In 2.3, we will define oral fluency and explain L2 speech processing models. We will also explore what is known in the field of L2 oral fluency research *not* conducted in the SA context, to understand findings in L2 oral fluency development in general. In section 2.4, we will also review findings of L2 oral fluency studies in the same context (SA) as the present study.

In 2.5, we will establish the importance of individual differences, specifically the role of cognitive individual differences in SLA. Key findings in cognitive research lines will be reviewed to present a holistic review of what we know about cognitive skills in L2 acquisition, especially in the SA context, and especially in relation to attention and working memory, two cognitive differences often studied in the SA context. Then, in 2.6, we will define and discuss inhibitory control and its impact on L2 ability in other language learning contexts, as very few studies have addressed this cognitive skill in the SA context. Findings in studies on *language switching*, (part of the inhibitory control function, defined as the ability to smoothly switch from L1 to L2 and vice versa) will be discussed, both in general and in the SA context. Section 2.7

will discuss what we know about study abroad experience factors and their effect on L2 oral fluency. Finally, in 2.8 gaps in the literature will be identified, leading to research questions presented in Chapter 3.

Individual differences often play a significant role in one's L2 learning development abroad. As individuals' lexical knowledge, perception and self-identity in the L2 can differ vastly (Hessel, 2017), so, too can individual speaking performance even within the same L2 learning context. For example, Dewaele and Furnham (2000) investigate the effect of personality on L2 speech in the SA context, and find that "interpersonal stress" (p.362) in formal speaking situations has a strong effect on introverts, and leads to their decreased L2 fluency in these situations.

What is interesting about what we know about research on individual differences and language contact is that while personality and motivation may be relatively static traits, seemingly small changes in program design (e.g. cultural sensitivity training and setting expectations prior to departure) have the potential to greatly affect L2 gains while abroad. Gaining a deeper understanding of the complexities in individual differences in cognitive ability and experience factors that influence L2 oral fluency development can help linguists and educators design SA programs which provide students with the optimal context in which to improve their L2 oral fluency and linguistic skills.

2.1 Context of learning in SLA

Approaches to language learning and teaching have evolved greatly over the last few decades, as researchers in applied linguistics have strived to identify both *how* languages are learned, and the optimal conditions in which to develop one's various linguistic skills. Theories on how an L2 is learned have arisen since the 1960s, when SLA emerged as a young sub-field of applied linguistics. Optimal context of learning in SLA has been a continual point of debate, leading to investigation into context effects. SLA researchers often explore changes in the classroom and teacher practice that would lead to optimal language learning conditions, including the effectiveness of *implicit* vs. *explicit* approaches to SLA, a focus on form vs. meaning, and the prevalence of grammar-based or authentic conversation and listening activities in the classroom (Larsen-Freeman, 2000).

Other theories focus on factors external to the classroom. Krashen's Input Hypothesis states that *comprehensible input*, and more specifically *i+1*, input slightly beyond the students' comprehension, is crucial to language acquisition (Krashen and Gregg, 2012; Krashen, 1998). Learners must be exposed to relevant, challenging input to be able to acquire the language, for both receptive and productive knowledge, rather than be made to study grammar explicitly. However, studies have shown that *comprehensible input* itself is not sufficient for L2 learners to acquire a language (e.g. Lightbown & Spada, 2006), and advocate for a combination of explicit, direct elements (*focus on form)* with a communicative approach. Swain's (1993) *output hypothesis* posited that input is not enough; learners must also have comprehensible output. They must notice a gap in their speech, pushing them to modify their output so it becomes comprehensible. The context in which students are learning, then, must provide this opportunity for noticing.

Following Swain's hypothesis, Long (1996) introduced the *interaction hypothesis*, a foundational SLA theory building upon the idea of comprehensible input stating face-to-face communication and interaction lead to language acquisition. The L2 classroom context often focuses on linguistic, grammar-based learning, rather than real-life contextual learning, even today, years after communicative learning and the *interaction hypothesis* have been widely accepted by SLA researchers. In practice, classroom learners rarely engage in meaningful conversation in their L2 while in class. As Swain (1996) notes in her observations of the L2 classroom, even in immersion classrooms teachers focus "more on manipulating and categorizing language forms than on relating forms to their meaningful use in communicative contexts" (p.531), leading to the question as to whether or not the formal instruction context allows for optimal L2 fluency development.

Opp-Beckman and Klinghammer (2006) advocate a "whole, integrated communication system" (p.9) for L2 instruction, which uses contextual situations to educate students, noting that body language, intonation, facial expressions, pronunciation and even the cultural interpretation of a concept such as "customer service" affect the meaning of the message. A learner who understands only the textbook definition of a clerk's speech may be ill-equipped to understand how customer service works in a particular cultural context. Dornyei (2013) sums it up nicely as he suggests that for effective L2 learning we need "creative integration of meaningful communication with relevant declarative input and the automatization of both linguistic rules and lexical items" (p. 42).

Given these theories, we may expect the immersion or SA context to be superior to foreign language classroom learning for language gains. Not only does SA present the potential for *comprehensible input* (though perhaps not for all learners), *comprehensible output, interaction,* and both explicit and implicit learning, but it may create a stimulating environment where learners are exposed to cultural norms and better equipped to practise the language they learn in the classroom in a communicative context. Yet, this is not always the case. Behavioural expectations based on cultural norms, age or gender roles also change the intended meaning of the message. The SA context in its nature is ideal for learning these expectations. A focus on both linguistic features and the context in which the language is spoken is necessary to become an efficient and effective learner (Opp-Beckman & Klinghammer, 2006). Liddicoat (2008) also supports the idea that students must "engage with culture" (p.278) when speaking a language, regardless of the lesson format. Fluency in a foreign language requires communicative approaches "just as much as grammar or vocabulary" (Liddicoat, 2008, p. 279), and it is crucial that the context in which one learns an L2 fosters the development of these communicative approaches.

Pérez-Vidal (2014) observes that different contexts of learning accumulate within a person's learning experience; the contexts are not simply separate experiences but work in tandem with previous learning experiences. Language learners tend to experience multiple contexts of learning throughout their quest to become fluent in an L2, whether or not they choose to immerse themselves in a situation where their L1 is not spoken. For example, many learners learn in a classroom setting or take language courses online. Some complete classroom instruction before embarking upon a SA experience, while others choose to increase their exposure to the L2 through domestic immersion or foreign language housing. Thus, it is not necessarily crucial to investigate which language learning context is *best*, but it is important to understand the value and benefits of each type of context and how individual differences in learners' attributes affect L2 linguistic development.

For example, in the SALA project, a longitudinal SA study of Spanish learners of English, the researchers note that their participants' notable gain in oral fluency and their knowledge transfer from one context to another may be due to the fact that they were "well-prepared to benefit" (Pérez-Vidal, 2014, p. 466) from their SA experience. Not only had they had classroom training on explicit metalinguistic knowledge, but they had done so in a formal-instruction context, a "safe environment" (Pérez-Vidal, 2014, p. 466) where they felt less pressure to communicate and could focus on phonological development. The idea that a *threshold of language learning* is necessary before a stay abroad can be beneficial has also been introduced in the literature (Collentine, 2009; DeKeyser, 2007). The subsections below details what can be theoretically expected from different contexts.

2.1.1 *Formal instruction*

Linguistic development in a formal instruction context (classroom learning) may vary with the method of language instruction. The extent to which teachers use an approach to focus on form (FoF) rather than a focus on meaning (FoM) have long been of debate in SLA literature (Ellis, Basturkmen, & Loewen, 2001; Swain & Lapkin, 2001) with some research suggesting that teachers should focus on form in the classroom to enhance *noticing*, but without interrupting language use (Long, 1991). An entire body of literature has looked at the effects of task-based classroom learning in SLA, which suggests that foreign language classroom learning is a diverse context and the L2 learner's experience is not necessarily consistent among different FI settings.

Long (1983) suggested that SLA studies be classified as those dealing with instruction, exposure or instruction + exposure, and claimed that "instruction is good for you" (p.112), regardless of learning context, whether it be solely foreign language instruction (FI), or FI in combination with another context. Studies on different aspects of language instruction have found varying results on this claim. For example, some do find that FI provides significant linguistic gains such as those in phonetic discrimination (Mora, 2014) or listening ability (Beattie, Valls-Ferrer, & Pérez-Vidal, 2014). Others find it to be a less effective context for speaking fluency, though helpful "as a solid lexical and grammatical foundation" (Juan-Garau, 2014, p.106) that later allows participants to progress in oral fluency in the SA context. Lord (2010) too, finds a significant difference in pronunciation ability in learners of Spanish with FI and without FI before they embark on a SA experience. While both groups improved after the SA, the FI group retained higher scores than their peers, supporting the case for FI.

2.1.1.1 *Immersion Context*

The original idea behind the immersion context – a context in which students in primary, secondary and tertiary education are educated in a second language – was to provide meaningful exposure to the L2 through the environmental context (and not language) focused studies, following a communicative approach (Genesee, 1987; Swain, 2000). While foreign language classroom-based instruction is the norm in many school systems, immersion contexts are available in some schools. In immersion, students are exposed almost exclusively to the target language, taking most school subjects in that language. One example of these types of schools are French immersion programs at schools in Canada, where the immersion curriculum parallels the L1 curriculum, the culture is that of the L1 local community, students enter with limited L2 proficiency and all teachers are bilingual (Swain, 2000, p. 200).

While foreign language instruction may provide the "safe", unpressured environment for speaking that Perez-Vidal (2015) refers to, it is also limited in its ability to expose learners to *comprehensible input*, colloquial language often not found in textbooks, native speaker contact, or cultural context; the lack of these components in the regular foreign language instruction classroom may advocate for a more immersive context. In an immersion context, students are immersed in their L2 within the domestic classroom context and learn subject matter such as social science, geography, mathematics, etc. in their L2. One would expect the continuous, daily L2 input in a variety of classroom contexts to lead to superior L2 linguistic development than the traditional FI setting. SLA literature shows that immersion programs exhibit several advantages compared the FI classroom (Swain, 2000). Most obviously, students have more L2 input and more opportunities to use their L2 in a natural way. Content they learn both in the target language class (such as French class in a French immersion program) and content classes (such as mathematics) mimics the way native speaker students learn the same content. The Centre for Advanced Research on Language Acquisition at the University of Minnesota, which publishes guidelines on immersion programs, cautions these programs should not only have comprehensive language arts programs, but also encourage extra-curricular activities with other immersion schools to encourage increased "social language use" (Klee, Lynch, & Tarone, 1998, p. 55).

Immersion has also been shown to force students to think of metalinguistic information and to recognize gaps in their language knowledge (between writing and speaking, for example)

(Netelenbous & Swain, 2013), increase the native-like-ness of their VOT in French pronunciation (Netelenbos, Li, & Rosen, 2016), and increases proficiency in the L2. However, individual differences exist even in school-aged children's fluency, language use and desire to invest in following the "no-L1" rules established in the programs (Ballinger, 2017). These individual differences in motivation, attitude, and confidence clearly mean that even within an "optimal" context, participants have different learning experiences, input, and opportunities for output.

Intensive summer language programs are another option for L2 learners. Some American immersion programs are often held at domestic summer institutes (in the participants' home country); they have rigorous entrance requirements and students sign a pledge promising to converse only in the L2 for the duration of the program (Du, 2013; Martinsen, Baker, Bown, & Johnson, 2011). Other programs, such as intensive overseas immersion through summer camps, provide increased input but do not guarantee full immersion. One example is the large number of students attending summer programs in place for EFL learners in Korea; 40,000 students attended in the summer of 2005, yet programs like these remain largely understudied (Collentine, 2009). At present we know little about the effect of immersion programs on the L2 oral fluency of large, diverse groups of learners.

2.1.1.2 CLIL Programs

In foreign language contexts in Europe for learners of English, Content Language Integrated Learning (CLIL) programs have been introduced. These programs adapt the immersion context widely used in Canada on a smaller scale; they all contain immersion-like sessions to increase L2 input, though the number of sessions and level of input and hour of exposure varies from program to program. Research in SLA in the CLIL setting has discovered somewhat varied results in L2 linguistic development, though studies on CLIL generally show positive development, so much so that (Dalton-Puffer, 2008) advocates for the "EFL and the language dimension of CLIL to be integrated into one foreign language curriculum" (p.15). For example, a large study of 360 Swedish learners of English found CLIL students to have superior vocabulary acquisition over their EFL counterparts (Sylvén, 2004), and CLIL contexts have also been found to foster positive attitudes towards the L2 (Lasagabaster, 2011; Lasagabaster & Sierra, 2009). In fact, a special issue on CLIL cites its superiority in terms of native-like listening and reading abilities, cognitive abilities and even brain plasticity, with no evidence supporting any negative effect of CLIL on the L1 (Van de Craen, Mondt, Allain, & Gao, 2007). Interestingly, however, the issue does note that there are "erratic results" (p. 72) in terms of L2 speaking skills in primary school-aged children. In one study on Catalan-Spanish adolescent learners of English showed participants to perform better overall in foreign instruction (FI) + CLIL context than a solely FI context, when tested on written comprehension, oral comprehension and lexico-grammatical ability in both receptive and productive skills (Roquet, Llopis, & Pérez-Vidal, 2016). However, in terms of grammatical ability, reading, and listening skills, there was no significant difference between the CLIL+FI or the FI context. Others find no significant difference in EFL learners' performance on language tests (Pladevall-Ballester & Vallbona, 2016) compared to their non-CLIL peers and suggest CLIL may be more effective "in the long-run with more intensive exposure" (Pladevall-Ballester & Vallbona, 2016, p. 37).

Some research has looked into the domestic immersion context for EFL learners in Spain, finding that university students experience significant gains in oral fluency in two months (Serrano, Llanes, & Tragant, 2011). For younger learners, studies find that 11-13 year olds significantly improved on L2 gains on both oral and written tasks after a short, 4 week stay abroad (Tragant, Serrano, & Llanes, 2017). Both summer camp and intensive instruction groups had significant fluency gains. However, it should be noted that the focus on these studies was much broader than fluency development and only a few measures were used. For example, syllables per minute was taken as a measure of oral fluency and the other measure of oral performance was Guiraud's index, which measures lexical diversity. Syllables per minutes is one measure of fluency, but does not take into consideration many other aspects of the concept of fluency, such as pause frequencies and durations. While the field of L2 oral fluency research in other contexts assess fluency using a variety measures (see 2.3.2), measures used in research on SA and immersion programs is limited. Thus, we do not have a clear picture of learners' L2 fluency progress in these contexts.

2.1.2 *SA as a learning context*

So far, we have been discussing contexts other than the one of interest in the present study– study abroad. In popular opinion, and even with SA participants themselves, SA is viewed as the "optimal" learning context, though research clearly shows it is not the "magical formula" (Kinginger, Wu, Lee, & Tan, 2016, p. 58) learners often expect. It satisfies theoretical conditions for Krashen's *i+1* by exposing learners to new and challenging context, though it must be noted that the immersion-like context of a study abroad experiences may often go much beyond *i+1*; input may be much beyond their comprehension level, which could negatively affect their experience. SA also meets the criteria for Long's Interaction Hypothesis, as students should find abundant opportunities for L2 language use, engaging with both each other and native speakers in their L2. SA research began with early research on undergraduate language majors in which those who studied abroad showed a marked difference in oral communication ability (Carroll, 1967), noting that "even a brief stay abroad had a potential effect" (p.1) on American students' L2 skills.

Participants in the *SALA project*, a Spanish university project dedicated to investigating effects of different learning contexts on learners' linguistic progress, for example, consistently cite that participants believe "you can't learn a language in your country" and that to "speak properly" you must learn a language abroad (Lara, 2014). However, results of SA studies show great variance in individual learner progress. As Dekeyser (2007) notes, the transfer of knowledge from one context to another is not a simple concept, and SA is "not as obviously ideal [a context] as assumed" (p.209). Thus, it is important that we take learners' specific experiences into consideration when evaluating the ability of a context to foster L2 acquisition.

As we have seen, different learning contexts might affect L2 development differently. The SA, CLIL, immersion (domestically and abroad) and foreign language classroom environments provide a possible set of contexts to foster linguistic development. Given the research conducted thus far, we would expect the SA context to be superior to other contexts for L2 oral fluency gains, even on short stays abroad. A combination of contexts could be expected to further increase fluency gains, but this has not been studied extensively. Improving our understanding of how the L2 learner can excel in each context, and what individual differences are at play, will allow both students to better take advantage of their existing L2 learning opportunities and educators to better understand how to create the optimal learning environment for their students.

2.2 L2 performance abroad

This section of Chapter 2 reviews main findings in linguistic performance abroad other than L2 oral fluency, grouped by linguistic area. While L2 oral fluency is the focus of this dissertation, it is important to understand what we know so far about *other* linguistic performance in SA literature to position L2 oral fluency findings in the same context. The particular focus of linguistic development, methodologies used, and results obtained vary greatly from study to study; there is no comprehensive set of measures studies used to measure linguistic development abroad.

In terms of grammatical development, the SA context has generally not been shown to be a superior or particularly effective context (see Collentine 2004; Dekeyser 1990; Gunterman; 1995; Martinsen et al., 2011) with some studies showing no gains in grammatical ability (Segalowitz & Hulstijn, 2005). A few have found improvement on specific features and not on others – such as improvement on past marking and gender agreement (e.g. Howard, 2005) or in Spanish verb tenses for *ser and estar* in students who spend one year, rather than a semester, abroad (Gunterman, 1995; Ryan & Lafford, 1992), or grammatical correctness of the use of subjunctive in speech (Freed, So & Lazar, 2003; Isabelli 2004; Isabelli & Nishida, 2005) and subject-verb agreement (Marques-Pascual, 2011). Martinsen, Baker, Dewey, Bown and Johnson, (2010) find that grammatical development as judged by native speakers in terms of accuracy and complexity to be not significantly different to AH learners. In an interesting case, Grey (2015) administered a grammatical-judgment and lexical-retrieval task to 26 US advanced learners of Spanish studying in Barcelona who had all previously studied the L2 in a classroom context for 6 years and controlled the SA context to provide maximum access to Spanish and minimum use of English in residence halls, dining halls, classroom and extra-curricular activities, finding an improvement in their grammatical skills. LaBrozzi (2012) uses an eye-tracking task on American English learners of Spanish in the SA and AH context and investigates the context effects on learners' ability to process morphological cues in a more native-like manner. The SA participants were found to act in a more native-like manner (relying more on morphological clues rather than lexical ones as Spanish monolinguals would) than their AH counterparts.

One study clearly shows that if SA has a grammatical benefit, it is not within the usual 4 to 12-month period students spend abroad. In a study of 18 Irish learners of French abroad, Howard (2015) found that 3rd person plural was not fully acquired even after a year abroad; past tense acquisition, however, was acquired in stays of up to 3 years abroad, and a longitudinal study confirmed participants retained this information over time. Interestingly, Edmonds and Gudmedstad (2021) find that, at least for the linguistic feature of gender marking in written form, classroom input was sufficient for British learners of Spanish to demonstrate target-like gender accuracy.

Unlike grammar, vocabulary acquisition in SA studies generally shows positive development in the SA context. Early SLA studies in the SA context showed that self-assessments and vocabulary development improve (Meara 1994). However, vocabulary acquisition measures are not consistent across studies – some measure acquisition of certain words using a pre-post test design, some using standardized vocabulary tests, and others looking at lexical diversity in speech or written samples.

For example, much of the literature in the field has used Guiraud's Index (a measure of lexical diversity) as a vocabulary measure, finding gains in the SA context. Serrano, Tragant and Llanes (2012), for example and found that one semester was enough time for 14 Spanish learners of English to significantly improve their oral lexical richness, though lexical richness in written production was slower to develop. Foster (2009) found learners of English studying abroad in London produce significantly more diverse vocabulary than those at home in Tehran. Dewey (2008), too, in a study of Japanese learners of English found that SA learners gained in passive vocabulary, but not significantly more than intensive domestic immersion students.

Fitzpatrick (2012) found SA learners to improve on collocations and native-like word associations, but not other measures of vocabulary knowledge. Laufer and Paribakht (1998) looked at L1-Israeli learners of English in both Israel and Canada and found that learners abroad improved on both active and receptive vocabulary tests. Some research, though, has found no significant differences in SA learners on lexical growth (e.g. Dewey & Foster, 2008; Foster, 2009). Collentine (2004), for example, finds no significant difference in the frequency of vocabulary items in oral production for participants after a SA experience. Briggs (2015), interestingly, found a complex relationship between length of stay and vocabulary acquisition measured through Nation's (1990) receptive vocabulary test for 241 adults mixed-L1 learners of English at private language institutions in Oxford and London. There is often large variability in individual learners' rates of vocabulary acquisition abroad, and this study showed that those who stayed abroad longer gained more passive vocabulary.

With respect to the development of pragmatics in the SA context, we know that participants are generally successful in pragmatic development abroad, such as the acquisition of colloquial language or sociolinguistic competence (Taguchi, 2011; Taguchi 2015; Llanes, 2011) However, their knowledge about routines, such as how to leave a conversation or social setting "remains distinctly non-native-like" (Hassall, 2018).

In sum, what we know linguistic skills development other than oral fluency in the SA context is that while SA may not be the optimal context to acquire grammar, some grammar forms improve while others do not. Vocabulary and lexical richness development are generally positive.

Serrano, Tragant and Llanes (2011), too, found that in a short stay abroad of 15 days the SA context was “more advantageous for the development of written and oral production in terms of fluency and lexical complexity” (p. 140) than the semi-intensive AH context. However, in comparison to AH intensive learners, SA participants did not perform significantly differently on these measures. Others have also found improvements in oral proficiency for American learners of Spanish abroad when measuring phonological gains in reading aloud tasks (Dias-Campos, 2006; Freed & Segalowitz, 2004) and pronunciation (Dias-Campos, 2006; Freed, So & Lazar, 2003).

Pronunciation research has looked at native speaker perception of learner development, which tends to improve even for short stays. For example, Llanes (2016) found that native speakers’ perceptions of eleven-year-old Catalan/Spanish learners of English accent significantly improved after a short, 8-week stay abroad. Juan-Garau (2014), in his study of L2 English learners concludes that a SA “fosters a more native-like performance” (p. 103) than FI when oral accuracy is concerned. Pinget, Bosker and De Jong (2014) find that native speakers can judge fluency as a separate construct to foreign accent.

Surprisingly few studies have dealt with the development of listening skills abroad, though those that do have not found significant differences in listening gains between SA and AH groups. Some researchers did find, however, that SA participants approach L2 listening significantly differently, using top-down and social listening strategies as opposed to bottom-up strategies preferred by their AH counterparts (Cubillos et al., 2008). Listening skills have also been shown to improve significantly more in an SA context than FI, being particularly beneficial for listeners who begin at a lower level (Beattie et al., 2014).

Some research has looked at complexity, accuracy and fluency (CAF) in oral performance, suggesting an inverse relationship between fluency and complexity and accuracy (Housen 2012). Some research has found that while the SA context helps learners move towards more target-like lexical diversity in their speech, it does little to improve their oral fluency (Lara, 2014). Serrano, Tragant, and Llanes, (2012), in a longitudinal study, found that participants benefitted first from oral fluency earlier in their stay, and later had accuracy gains.

Taken together, we can see that linguistic gains in the SA context are anything but consistent; we see neither consistent improvement, nor consistent failures. In fact, it is clear that more research into the effect of learning context and development of specific skills within the context is needed. Language, as Piller and Pavlenko (2009) note, is "not contagious"; the environment itself will not necessarily facilitate acquisition. As Tullock and Ortega (2017) observe in their scoping review, results from SA studies over the years with respect to linguistic gains are "disappointingly mixed" (p.2), leaving a wide gap for further study necessary to fully understand the impact the SA context has on learners' linguistic ability.

In sum, this section has shown that both linguistic skills measures and improvements in linguistic skills gained abroad vary greatly in the literature; we still do not have a clear picture of how participants will improve when they study abroad. Linguistic gains abroad differ from study to study, and among participants. Participants are likely to improve on vocabulary development, though perhaps less likely to improve their grammar. Individual differences and specific contextual or experiential differences then may be key in determining how exactly SA participants will progress during their stay. The three subsections that follow summarize what we know about different variables and their impact on linguistic development abroad; Section 2.2.1 discuses the impact of cognitive individual differences on SA, while section 2.2.1.2 focuses on our knowledge of the impact of emotional variables, and 2.2.1.2. discusses differences in contextual differences, such as language contact specifically.

2.2.1 *Cognitive variables*

Individual differences both specific to the individual as well as differences in the SA context itself with respect to the type and amount of language contact have been claimed to influence linguistic outcomes in the SA experience; cognitive variables are one such type of individual difference. Collentine (2009) shows concern that SA studies do not take into consideration cognitive individual differences or L2 experience factors in evaluating L2 learner programs. In a 2004 special issue of Studies in SSLA dedicated to the SA context, researchers also caution that a complex relationship between "oral, cognitive, and contextual variables…[that] explain individual variation" exist (Freed, Dewey, Segalowitz, & Halter, 2004, p. 174); the L2 SA experience is neither uniform between programmes, nor the same for each individual learner who undergoes the experience.

What we know about the impact of cognitive variables in SA is limited, as this is an area of individual differences in study abroad that has been sparsely explored. Segalowitz and Freed (2004) did measure the cognitive fluency variable of attention in the SA context, finding a relationship between cognitive fluency and L2 oral fluency gains. They found efficiency and speed of lexical access to be related to one measure of speed fluency (mean length of runs). Attention control, the cognitive variable tested in the study was found to be negatively related to gains in speech rate; the more participants paid attention to their speech, the lower their speech fluency gains were. Section 2.5 elaborates on cognitive individual differences usually explored in SLA, namely attention and working memory, though most research in this area has been dedicated to contexts other than SA, revealing a large gap in the literature.

2.2.2 *Emotional variables*

While affective factors such as motivation, anxiety and personality are not the focus of the present study, we cannot ignore the fact that these variables may influence the linguistic development of our participants. An individual's personal motivation for language use, for example, could also greatly impact the effectiveness of a SA experience.

Dornyei (2013) speaks of the *L2 Self*, a psychological concept that refers to the perceived differences in oneself when conversing in the L2 - the "individual's self-knowledge related to how they see themselves at present" (Zolta, 2006, p. 11). Several studies report a profound impact of perceived L2 self on language development. Wood (2007) found that even in the FI context with L2 native speakers running a workshop, L2 interaction has a different impact on different individuals. As an example, one participant gained confidence in writing abilities while another previously confident student displayed more *emotional distance* at the end of the course. (Chirkov, Safdar, De Guzman, & Playford, 2008) find that autonomous motivation predicts adjustment outcomes (how well students feel they adjust) in international students at Canadian universities; L2 learners have individual differences in their personality and confidence that cause them to adjust to new situations at different rates. Although not investigated in many SA studies, it is possible that L2 development, like any learning, would be more easily facilitated in an environment where students feel well-adjusted. Cigliana and Serrano (2016) is one such study within the SA context in which the researchers find that American students studying abroad are highly motivated in both instrumental and integrative motivation, and also have perceptions of high achievement and improvement in their linguistic skills (self-reported). Students with high integrative motivation also tend to interact with the culture more (rather than remaining isolated with their L1 peers), and have more proficiency gains (see Hernandez, 2010).

Foreign language anxiety is often associated with the development of the L2 self (Dewaele & MacIntyre, 2016; Hessel, 2017) and language contact while abroad (Trentman, 2017). Hessel (2017) suggests that these psychological, background and cultural differences need to serve as a "vantage point for study abroad research on differential linguistic attainment" (p.50). In a study of 59 Belgian and Swedish SA participants Arvidsson, Eyckmans, Rosiers, and Lundell (2018) found a moderate correlation between both self-reported language use and perceived gain in speaking ability, with changes in cultural openness and empathy by the end of their stay, suggesting the importance of the cultural link. Individual differences in anxiety have also been found to negatively affect language learning; in fact there is an inverse relationship between anxiety levels and language attainment (Gardner & MacIntyre, 1993) and even L2 pronunciation (Szyszka, 2017). Tracy-Ventura, Dewaele, Köylü, and McManus (2016) reported that 77% of participants in a study of 58 British learners of French and Spanish studying abroad reported increased confidence and lower anxiety in speaking their L2 after a year abroad. Many participants also scored higher on cultural openness and emotional stability measures of a personality test. Some studies also document learner attitudes and self-reported willingness to communicate, noting that most believe acquiring L2 English skills "happen simply by being present in an English speaking country" (Jackson, 2016, p.159), even though participants had been formally exposed to a range of language-learning strategies.

Another line of research in SA looks at *self-efficacy* (how well students believe they do), beliefs, and expectations. Generally, we know that there are mixed results as to students' individual experiences by the end of their stay, importantly **"sometimes leading to decreased confidence"** (Surtees, 2016, p.5), especially in speaking proficiency (Amuzie & Winke, 2009) and a realization of their low proficiency and unrealistic expectations (Allen, 2010; Allen, Dristas & Mills, 2006; Mendelson, 2002; Wilkinson, 1998 and others).

Given the stressful environment that being away from one's home country and culture can create, it is no wonder that participants in SA programs seem to have their own *unique* experiences, not necessarily shared by others in the group. Additionally, it is possible that some *cognitive* individual differences (some, if not all of which are beyond the individual's conscious control) exist between those with greater and less linguistic gains abroad (see Section 2.4.1.3 for more information on fluency of cognitive processes). Thus, when considering gains (or the lack thereof) in linguistic attainment in a SA setting we must be conscious of the psychological, cultural and cognitive individual forces that are at play and likely affect each individual participants' results. As Juan-Garau (2015) reminds us, it is entirely possible to "unfold the potential [in SA]…if well supported by all stakeholders, including learners themselves" (p. 55).

While some students may expect vocabulary, grammar or writing skills to improve during their stay abroad, the majority arrive expecting to *speak* more fluently after their stay abroad. Furthermore, the development of speaking skills is often seen as more difficult than reading or writing as speaking requires online, almost instant output. As Juan-Garau (2015) notes, learners who experience "uneasiness, bordering on anxiety…with respect to relationships" (p.105) have a less successful oral performance abroad and sustain less conversation than their less-anxious peers. What we know about motivation and anxiety in SA is that higher levels of anxiety and lower levels of motivation can negatively affect linguistic outcomes while abroad.

2.2.3 *Contextual variables in SA*

Individual differences such as those in cognitive skills, motivation, or experience factors while abroad can affect L2 oral fluency gains, but contextual variables also play an important role in providing opportunities for authentic input and output. While most SA studies focus on quantitative results, some focus on the SA experience as a whole, from the perspective of the language learner. The majority of data of this nature is self-reported through questionnaires and interviews (e.g. Grey, Cox, Serafini, & Sanz, 2015), and self-reported data is not objective. However, it does provide some insight on the perceived value of the SA experience from the perspective of an L2 learner.

Interestingly, while the SA context may be assumed to be linguistically rich, several researchers suggest that SA students often have little control over the amount L2 exposure they experience during their stay (Morita, 2004; Ranta & Meckelborg, 2013). Conditions of the stay may vary considerably from one program to the next, thus varying the learning conditions for one learner to the next. For example, features such as length of stay, living conditions, rules regarding language use, classroom contact hours change the L2 learning context, amount of input, willingness (and necessity) to communicate in the L2, and may therefore substantially affect L2 development.

In Nation (1990), however, out-of-class language contact (self-reported through a questionnaire) was found to have no relationship to vocabulary gain. This prompted the researcher suggested that a wider variety of language contact types may have a different relationship to vocabulary gain and advocates for more guidance for SA students as to how to maximize their language contact while abroad. Although learners may arrive abroad expecting to have many opportunities to converse with native speakers, this is not always the case. For example, students may arrive and live with a group of their L1 speaking peers, or feel excluded from L2 social groups (Grey et al., 2015). Moreover, learners do not always acquire more fluent speaking skills while abroad simply because comprehensible input or face-to-face interaction opportunities exist in these environments. In reality, the acquisition of linguistic skills and specifically L2 oral fluency while abroad is a complex, often oversimplified endeavour.

Language contact can be defined as the use of more than one language at a time, not requiring fluent bilingualism or multilingualism, but requiring some communication between speakers of different languages (Thomason, 2001). Within the SA context, for the purposes of this dissertation, we define language contact as the contact with the target language, both in terms of exposure (input) and production (output). Conditions of the SA stay may vary considerably from one program to the next, thus varying the learning conditions for one learner to the next. For example, features such as length of stay, living conditions, rules regarding language use, classroom contact hours change the L2 learning context, amount of input, willingness (and necessity) to communicate in the L2, and may therefore substantially affect L2 development.

Research on living situations and other environmental factors which could potentially affect an individual's progress has been done in the SA context, finding that the right environment has the potential to improve linguistic gains. Martinsen et al., (2011), for example, focus on the effect of the foreign language housing (FLH) environment on learners' oral proficiency. Although these students are on their at-home American university campus, the environments studied consisted of apartments with language learners and one native speaker in which participants are encouraged to use the L2 whenever possible. Regardless of the L2 they studied (French, German, Japanese or Russian), the researchers found significant improvement in oral proficiency gains, and significantly higher overall language use, and variety of tasks in which the L2 was used for FLH participants compared to classroom students. Though both were in a domestic context taking formal instruction, informal language contact was higher for the group participating in foreign language housing.

In terms of stays abroad with host families, what we know about their impact on linguistic gains is that this type of accommodation does not necessarily lead to linguistic gains. Taguchi (2008) found that participants in a one-month homestay abroad gained in comprehension speed, but not accuracy, while Hernandez (2016) asserts that much of the language contact gained abroad is "superficial cultural contact" (p.42) and that language stay and accommodation with a host family had little effect on proficiency gains (p.52). In fact, Hernandez (2016) found that the majority of encounters in which L2 learners spoke the L2 were service encounters, as the majority of their time abroad was spent with other native L1 speakers.

Briggs (2015) found no relationship between informal language contact and vocabulary gain in learners of English. In a later mixed-methods study on informal language contact of 241 learners of English in Oxford and London, she found that when placed in a variety of simulated L2 contexts, the choice of individuals' communicative strategy when using new vocabulary words depends on the learners' perception of the target community (L2) culture. For example, if a learner believes the target community is likely to also speak their L1, it may affect the effort and strategy use the learner employs to communicate their message. The extent to which language contact has an effect on vocabulary gain and may be influenced by a variety of factors.

There is another component that might be responsible for language contact and individual differences in SA experience – program design. Evaluating the specifics of SA programs for American students, Kinginger (2011) advocates for carefully designed SA programs that take into consideration pre-departure level, allowing participants the opportunity for level-appropriate exposure to the L2 in authentic environments. Less accurate students at pre-departure sometimes gain the most from the SA experience – see Juan-Garau (2014) – though this claim does not always hold true (Kinginger et al., (2016), for example). A carefully designed program would take into account students' pre-departure levels, motivation for the stay and personal goals, accommodation types and provide a variety of language contact opportunities. Kinginger (2011) also suggests programs require language-related projects that require communication with host families and other native speakers (Knight and Schmidt-Rinehart (2010) also found this task-based approach beneficial for language use and cultural learning), and provide participants with opportunities to engage in volunteer or work activities in their host communities.

Accommodation types present another difference in the SA experience that can impact linguistic development. Though grammar generally does not improve abroad, Dewey (2018) did find learners *living in domestic foreign language housing for a year* showed significantly more gains in past tense grammar ability than their AH counterparts. Garcia-Amaya (2012) explored two living situations – an SA context, and a domestic immersion context, both which required a language pledge in which participants promised not to speak their native language. Despite only a 6 week stay in the program, both participants abroad and participants in a domestic immersion context experienced significant oral fluency gains. As Garcia-Amaya (2012) rightly points out, what remains to be seen in future research is whether it the language pledge, the SA (or immersion) experience overall, or other parts of language contact that truly make a difference in language gains abroad.

With respect to pre-departure level, Kinginger et al. (2016) provide evidence in case studies of American learners of Mandarin on an 8-week homestay program that a relatively advanced level is necessary for proficiency development, though even an initial level can result in cultural understanding and warm relationships. Researchers do caution, however, that initial proficiency level can greatly impact individual progress (Marqués-Pascual, 2011), and the nature of the SA environment may not be questioning the extent to which the SA context facilitates language acquisition for all learners. Pre-departure cultural sensitivity levels have also been of interest in the literature and these levels have often been found to predict gains in L2 proficiency while abroad (e.g. Baker-Smemoe et al, 2014), with many advocating for pre-departure cultural sensitivity training on SA programs (Goldoni, 2013; Jackson, 2013; Baker-Smemoe et. al, 2014; Martinsen, 2010; Jackson, 2006). Given language contact research in the SA contact thus far, then, we can see there are various factors that can drastically change the individuals' experience.

One of these factors is the amount of authentic input students gain from their time abroad. The amount of time students spend abroad varies greatly; most studies look at semester abroad stays of approximately 11 to 13 weeks (Yang, 2016). Furthermore, the amount of exposure and L2 use can be vastly different from program to program, as most programs rely almost exclusively on student motivation alone to control language use. Even in short stay abroad programs, such as intensive 20-hours per week for 3 weeks courses, programs requiring language pledges and limiting L1 exposure both within and outside the L2 classroom are rare.

Furthermore, L2 language use in the SA context is often self-reported and therefore not accurately tracked, casting doubt on the true “immersion” like nature of the learning context.

As Amuzie and Winke (2009) note, some studies find that upon returning from their stay abroad, learners better understand that learner autonomy and studying outside of class lead to more linguistic gains; studies continue to find that regardless of learning context, it is “increasingly clear… [that] those who are proactive in their pursuit of language learning appear to learn best” (Macaro, 2001, p. 294). While this finding may seem obvious, it is directly contrary to popular belief that the SA environment is automatically conducive to language acquisition and in itself allows students to “pick-up” the language. Participants in SA programs must make a concerted effort to make the most of their individual language learning experience.

2.3 What is oral fluency and why is it important?

The following section first presents definitions of oral fluency and then discusses its importance through the explanation of speech production and speech processing models. A discussion of measures of fluency then follows.

2.3.1 *Speech production models*

To understand and operationalize the concept of *oral fluency,* and subsequently measure fluency progress, we must first look at speech production models to understand the components that comprise fluent speech.

According to the widely accepted speech production model, *the blueprint of a speaker*, (Levelt, 1989; 1999) (Figure 1), both the semantic-syntactic and the phonological-phonetic systems are involved in articulation. The speaker conceptually plans a *pre-verbal* message at both a *macro* (what to say) and *micro* (how to say it) level, in a piece of the model called the *conceptualizer*; the conceptualizer is the component of the model, where *planning* occurs. The speaker then encodes the pre-verbal message with grammatical knowledge, morpho-phonological knowledge, and phonetic encoding in the *formulator* before producing overt speech. Parsed speech is processed through the speech comprehension system and also becomes a part of the *monitoring* process that occurs in the conceptualizer; the speaker hears and processes speech in their environment.

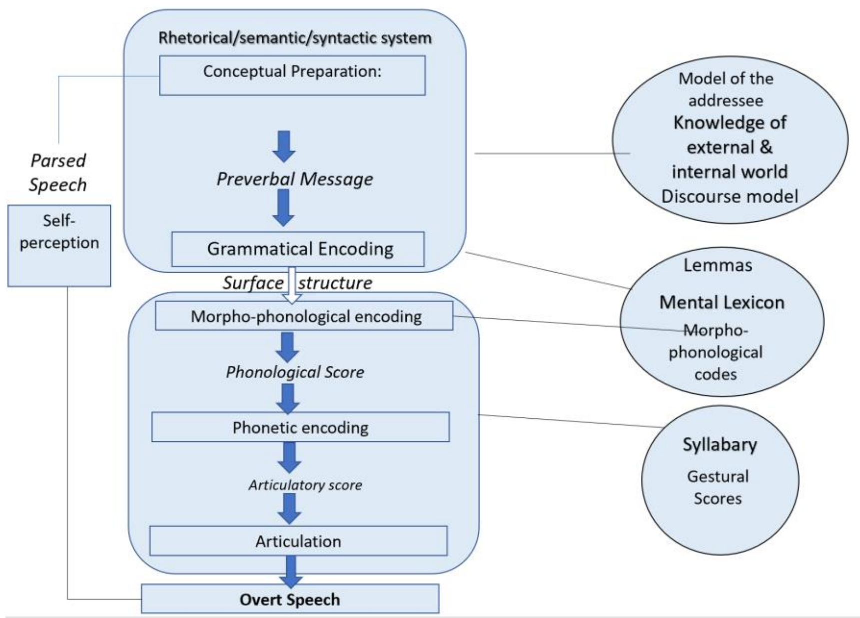


Figure 1 - Blueprint of a Speaker (Levelt, 1999, p.87)

Levelt's (1999) model was later adapted to L2 speech. De Bot (1992) first adapted Levelt's (1989) model to bilingual speech production, as he asserted that the model must accommodate for separate, language specific formulators and a conceptualizer that takes into account the speaker's multiple languages; macroplanning is not language specific, whereas microplanning is. The formulators then "submit their speech plan" (de Bot, 1992, p. 21) to an articulator that is not specific to a language, and produces the overt message. Before the message is produced, it must pass through morpho-phonological and phonetic encoding, where a phonetic plan (internal speech) is developed in the phonetic system. Importantly, *self-perception, the mental lexicon, knowledge of the internal and external world and even gestures* are components that influence the environment in which speech production occurs.

Kormos (2006) proposes an integrated approach to understanding L2 speech fluency, adding to Levelt's conceptualizer, formulator, and articulator model. In the conceptualizer, Kormos (2006) emphasizes the importance of monitoring and its direct relationship to message generation; L2 speakers monitor their own speech with the help of the speech comprehension system. The lexicon, syllabary and L1 & L2 episodic memory fall within a category called long-term memory, as Kormos (2006) proposes one memory store, rather than separate storage of syllabary, internal and external knowledge and the mental lexicon. The speech comprehension system is also emphasized in this version of the model; a speaker's ability to parse speech directly affects their spoken output next processed in the conceptualizer.

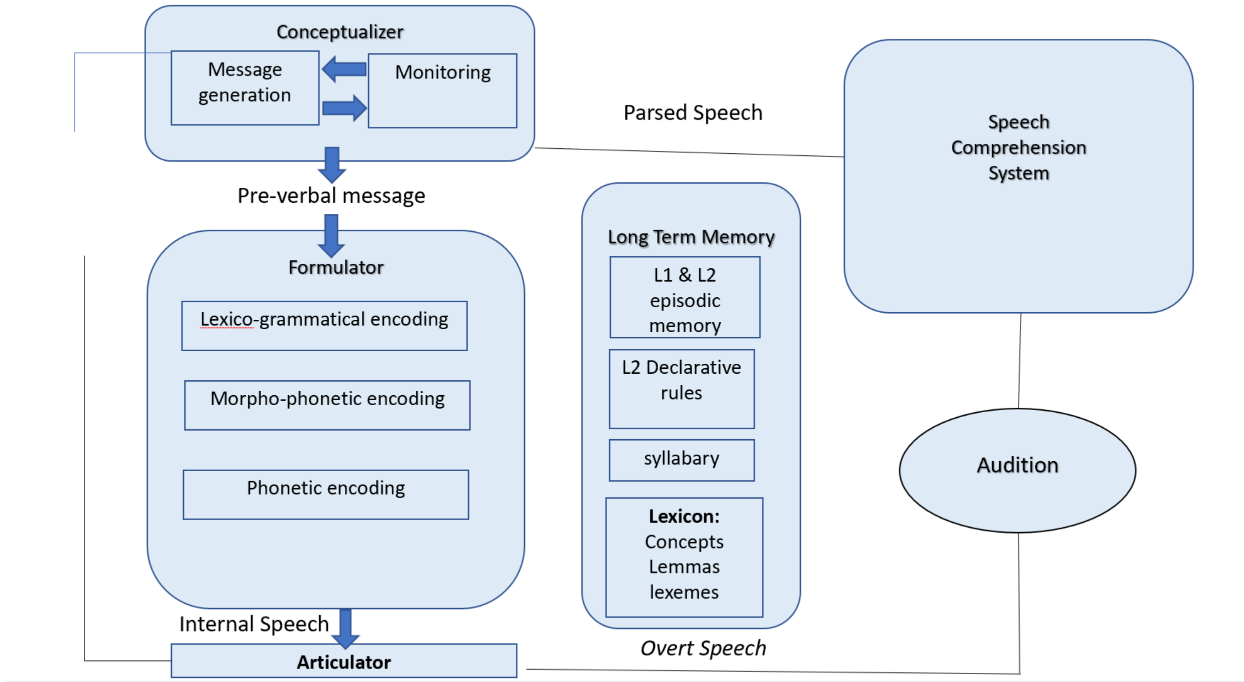


Figure 2 - Kormos, 2006, p.168

The present study uses this widely accepted *blueprint of the L2 speaker*, and more specifically, Segalowitz (2010)'s additions to this model, as illustrated in Figure 2 below. It also investigates a more complete construct of *fluency* discussed below, as the latter provides the most comprehensive, complete model of L2 speech production to date.

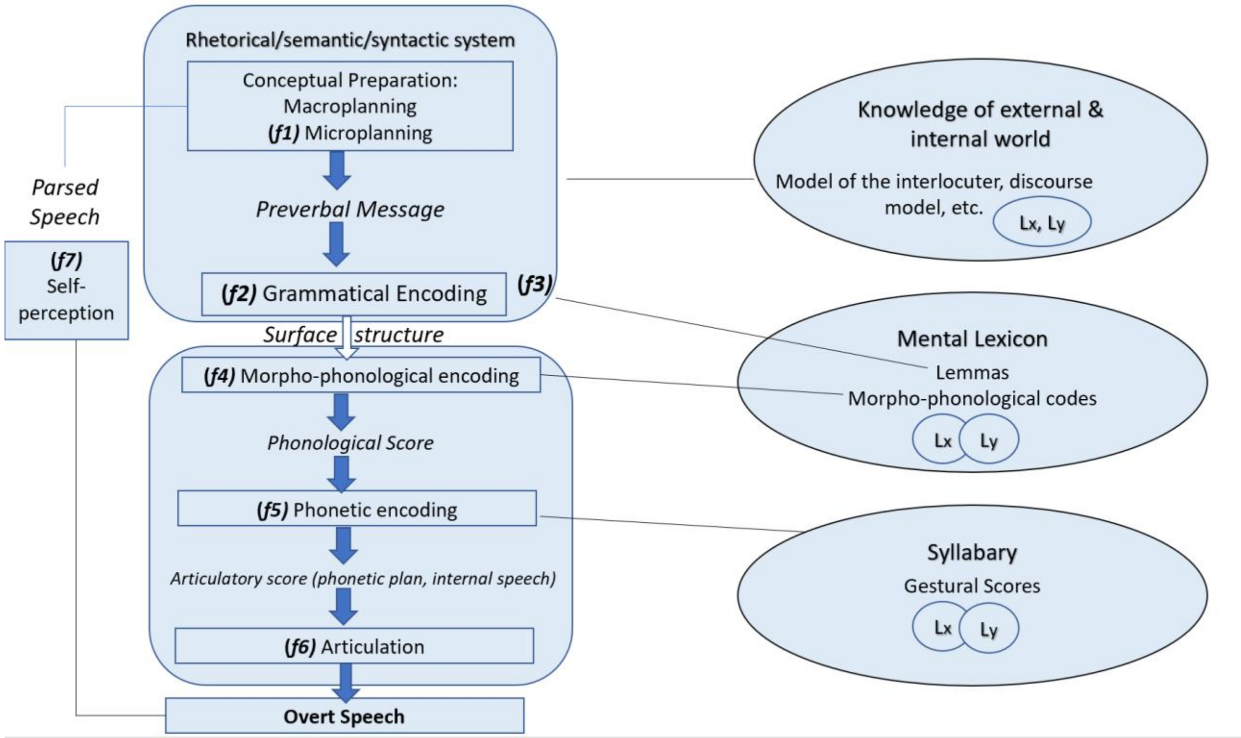


Figure 3 – L2 Speech Production Model (Encoding) Segalowitz 2010 p.9

Given the complexity of the speech production models, we can see that L2 speech production and L2 fluency involve multiple processes and sub-processes, some of which are conscious and others of which are not. As Chambers (1997) states, oral fluency is "a multi-faceted construct" not limited to oral production. The *macroplanning* and *microplanning* stages may be conscious, but the stages that follow from encoding until articulation are not noticeable to the speaker themselves. Rather, they rely on automated processes, Segalowitz's (2010) discussion of fluency delves further into automaticity to focus on the *cognitive bases* that underlie linguistic developments, and this in itself is key to a more comprehensive understanding of fluency. In the *conceptual planning* stage of forming the pre-verbal message, systems other than the linguistic system are at play; this extends the field of research from simply linguistics to psychology and psycholinguistics to truly understand forces at play in speech production. Specifically, Segalowitz (2010) adds *fs* and *Ls* to Levelt's (1999) model, labelled in Figure 2. *Fs* points are “fluency vulnerability points” that refer to points in the system where underlying processing difficulties could result in speech disfluencies. F_1, microplanning, is the first vulnerability point, where individuals may differ with respect to the way they process speech. An individual’s knowledge of the internal and external world affects the planning process. The person with whom one is communicating, for example, may change the message. Once a preverbal message is formed, individuals can differ on several fluency vulnerability points. The mental lexicon (F_3) affects the way a message progresses through grammatical (F_2) encoding before a basic surface structure of the message is made. Morpho-phonological (F_4) and phonetic (F_5) encoding depends on information that filters in from one’s mental lexicon and knowledge of syllabary. Thus, the way an individual articulates a message after encoding is another point of vulnerability (F_6); a point where speech can become less fluent. Importantly, self-perception of speech that has just been produced (F_7) also affects microplanning for the next utterance. L_x and L_y refer to information outside the target language, or target register that also have an effect on L2 speech production. Clearly, there are many points in the L2 speech process that are vulnerable to disfluencies and allow for vast individual differences in speech production.

In L1 speech, conceptualizing a pre-verbal message requires the speaker's attention, but once speakers have an idea of what they want to say, formulation and articulation are automatic. In the L1, many parts of the speech process are virtually effortless. Speakers can easily access and retrieve lexical items, perform grammatical, morphological, phonological, and phonetic encoding, articulate and gesture, and monitor their own speech due to their automatized lexical, morphosyntactic, and phonological knowledge (Levelt, 1989; 1992; 1999). In the L2, however, each of these fluency “vulnerability points” represent parts of the speech process that can break down, resulting in slower, more inaccurate, and lexically less diverse speech. L2 learners are then obligated to use the linguistic knowledge they have – which may still be declarative knowledge, or in the process of becoming proceduralized and not yet automatic (Segalowitz & Hulstijn, 2005). This often affects the comprehensibility of one's L2 speech, making it more difficult to understand (Isaacs & Trofimovich, 2012; Saito, Trofimovich, & Isaacs, 2015) and therefore harder to process for both native (Munro & Derwing, 1995) and non-native (Ludwig & Mora, 2017) listeners.

Segalowitz (2010) advocates a "cognitive science perspective" (p.7) on oral fluency since it is a multi-dimensional construct that still lacks a systematic framework. In his comprehensive book on *cognitive bases* of L2 fluency, Segalowitz (2010) makes two significant theoretical contributions to the study of cognitive aspects of L2 oral fluency in addition to his adaptation of the *blueprint* model. He provides a broader L2 fluency framework in which the *blueprint of the speaker* fits, and he identifies and discusses three categories of fluency that can be operationalized in further studies, in terms of an L2 speaker's speech process.

First, he introduces what he calls as *provisional framework* for discussing relationships between sources that influence L2 fluency, including the *blueprint*. He refers to the latter as the *cognitive-perceptual processing systems* (see Figure 3), which are "characterized by cognitive fluency, or processing efficiency" (Segalowitz 2010, p.131). The *L2 Speech Production* system is central to this framework and characterized by utterance fluency (temporal measures, such as speed, pausing, etc. discussed below). However, it is important to note that while the *blueprint* and cognitive factors such as processing speed, lexical access and flexibility in speech planning feed into speech production, they are far from being only factor affecting it. Motivation to Communicate (such as beliefs about communication, language identity and the L2 self), the

Interactive Communicative Context (social context and cognitive task demands), and relevant perceptual and cognitive experiences (opportunities for repetition practice, L2 exposure, etc.) affect one another and contribute to speech production. Thus, Segalowitz (2010) presents us with a broader context for L2 fluency development and reminds us that L2 fluency development cannot be assessed in isolation from the learner's context.

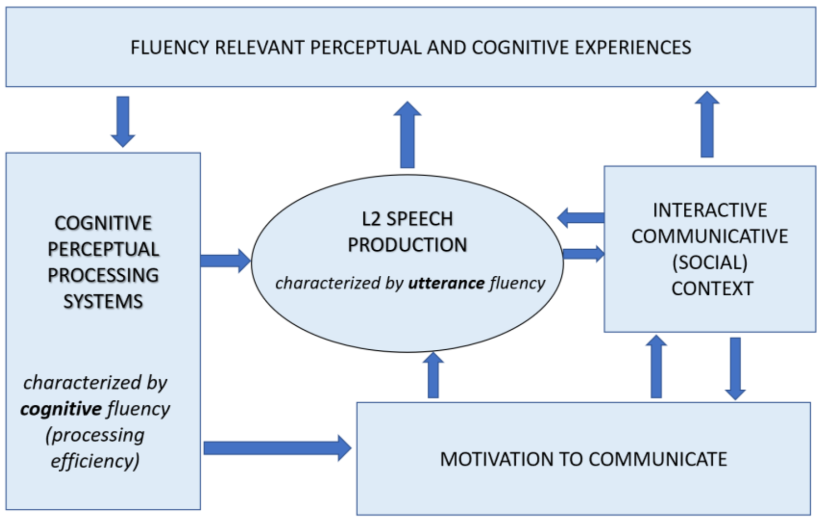


Figure 4 – Fluency Relevant Experiences – Segalowitz, 2010, p.131

Second, Segalowitz (2010) defines and discusses three types of fluency L2 speakers exhibit - *cognitive fluency*, *perceived fluency*, and *utterance fluency*, all of which contribute to his more complete definition of L2 oral fluency. As Segalowitz notes, stating that an L2 speaker is *fluent* can be interpreted in three "generally mutually supportive" ways: first, the ability to mobilize cognitive systems efficiently to speak, similar to the way a native speaker would; second, the ability to exhibit characteristics of speech that render speech fluid; and third, to be interpreted as having efficient cognitive systems and fluid speech by native speaker listeners. Specifically, Segalowitz (2010) defines *utterance fluency* as a set of objective measures, *cognitive fluency* as the efficiency of the underlying speech production system and *perceived fluency* as the listener's interpretation of connection between the first two senses.

The following three sub-sections define and detail these constructs of fluency and their importance to the development of oral fluency for L2 learners, and application to the data collected in the present study.

2.3.2.1 Utterance Fluency

Utterance fluency is the most commonly explored of the three types of fluency, as it is operationalized through a set of quantitative measures that can be compared across studies. Segalowitz (2010) defines the term as the characteristics of a speech sample that is produced, such as speed, pausing, repetition or repair characteristics of the sample. De Jong (2013) points out that utterance fluency is the "objective acoustic measure of an utterance" (p.3).

Utterance fluency is the type of fluency most learners are likely to imagine when they think of fluent L2 speech. It most clearly aligns with definitions of fluency as *ease* of speech, or a flowing, fluid movement. The *utterance* is the output of speech production; it includes words spoken, the speed at which they are spoken, hesitations and attempts the learner makes to correct speech they realize is incorrect after its production. While Segalowitz (2010) states that there are theoretically a large number of "possible-fluency-relevant" (p. 48) characteristics, SLA research is beginning to narrow down critical items in understanding fluency.

To operationalize the concept of *utterance fluency*, it is necessary to assess changes in temporal measures in learners' speech data. Tavakoli and Skehan (2005) developed a now widely accepted grouping of utterance fluency into three categories – *breakdown, speed and repair fluency*, which has been the centre point of the majority of utterance fluency studies over the last decade. These measures are objective, calculable, and comparable over time, and to other studies in the field. A discussion of various measurements of utterance fluency research can be found in Section 2.3.2 below.

The present study focuses on utterance fluency and changes in L2 learners' speech over the SA period. Temporal measures have been commonly used in the literature, allowing for comparisons between the present study and others in SLA literature. For example, changes in speech rate, pausing and disfluency behaviour (repetitions, repairs and elongations) are of interest. Taking into consideration Segalowitz's (2010) model of fluency, this study investigates both *cognitive* and *utterance* aspects of L2 learner fluency development while abroad.

2.3.2.2 Cognitive Fluency

Cognitive fluency, the first sense of fluency discussed in Segalowitz's model, is "the automatization and coordination of 'integrating cognitive processes'" (p.48) behind speech production, which affects speaking fluency. The underlying processes are unconscious processes that occur within the cognitive systems before an utterance is produced. They include lexical search and access, the formulation of the message, and articulation; cognitive fluency is the efficient use of all of these processes resulting in a speech act.

Segalowitz (2007) coined the term *access fluidity*, referring to the "smoothness of flow of mental traffic" (p.183), as learners access the necessary vocabulary to produce their message. *Attention control* is another important aspect in producing a fluent stream of speech, as speakers must continually and efficiently focus on the relationships between words and their meanings. However, lexical links are only one of two concepts that are automatized in speech production; the second is attention-based mechanisms used in language processing (Segalowitz, 2007). For example, continuous shifting of attention from one aspect of language (e.g. content words) to another (e.g. linking words) may hinder fluency if this shifting is not yet automatic for the L2 learner. It is this concept of *automaticity* that is crucial to a comprehensive definition of fluency.

Automaticity is perhaps the most relevant cognitive construct in producing fluent speech, as researchers agree that high levels of fluency require automatization of cognitive processes (Segalowitz & Hulstijn, 2005). Segalowitz (2010) also explains that processing happens regardless of the "load" of information given and the extent to which processing is *effortless* (no more additional resources are used). Furthermore, Segalowitz refers to the U-shaped behaviour curve, a psychological concept that may be relevant to automaticity, as the "student reverts to using incorrect forms of language that had recently been under control" (p.397) because back-up, automatic processes are in motion when controlled processes are not operating (e.g. because of fatigue). Delays in cognitive fluency may affect accuracy in L2 speech.

Importantly, Segalowitz (1993) distinguishes between speeding up of performance and automaticity; the former is performance the learner initiates while the latter is an automatic cognitive process. New knowledge must be proceduralized before becoming automatized and therefore requiring less attention. Towell, Hawkins and Bazergui (1996) argue that "increases in fluency are attributable mainly to increases in the degree of proceduralization of knowledge" (p.84), which leads to automaticity. Thus, if a second language learner has only declarative knowledge of a word or phrase and does not yet have procedural knowledge – whether it be grammatical, lexical, or phonetic – they are unlikely to have fluid, fluent speech. This *automaticity* of cognitive processes is key for fluent speech to occur. Operationalizing cognitive fluency presents a challenge, as the construct itself is comprised of many processes. Few studies have investigated L2 cognitive fluency, and those that do focus on different aspects of the construct. Kahng (2020) is one example which shows the direction in which L2 oral fluency development research is heading, incorporating both cognitive individual difference factors and L1 measures into the analysis of L2 fluency. In a study of 44 Chinese learners of English, for example, L1 utterance fluency measures (mid-clause pauses) explain L2 oral fluency development. Interestingly, the duration of silent pauses, and number of filled pauses only explained L1, not L2, fluency. While studies of this type in L2 oral fluency research have been limited, they have been even more limited within the SA context. Within the SA context, studies have focused on only a few measures of breakdown or repair fluency, limiting our understanding of the impact of an SA experience on L2 learners' oral fluency development, and their progress.

It is important to differentiate between the process of *cognitive fluency*, the underlying speech processing mechanism that is present in all speech, and *individual differences in cognition* that have an impact on the speech production process. When we investigate the way in which cognitive fluency works, we are researching the automatic process underlying speech mechanisms and what underlies speaking fluency at a cognitive level. Segalowitz (2010) explains the concept of automaticity that lies behind speech processing, and studies have measured parts of this process, such as the ability to quickly activate and retrieve L2 lexical items from memory (Sunderman et al, 2009; Costa & Santeseban, 2004).

Investigating cognitive fluency, however, is not the same as investigating cognitive individual differences, which are factors specific to individuals that may affect the way in which processing happens. Cognitive fluency vs. individual differences in cognition are important constructs to disentangle in SLA literature, especially in research that looks at L2 oral fluency development. For example, ability levels in attention, inhibition, and working memory are cognitive factors which may affect one's utterance fluency – speech rate, pausing patterns and disfluencies, to name a few. Furthermore, the methodology and procedures for measures one of these constructs instead of the other may be different. Individual differences may need to be measured in the L1. Cognitive processing advantages, however, would need to be measured in the L2, or as an L2-L1 specific measure (to account for individual differences in L2 speech that are derived from differences in one's native language), following Segalowitz (2010; 2016). In this dissertation, we try to avoid the amalgamation of these two, separate constructs, and focus on the cognitive individual difference of inhibitory control, as discussed below.

The present study investigates one type of *cognitive individual difference* within its exploration of utterance fluency, the main focus of this dissertation: the role of inhibitory control in L2 speaking performance and its relationship to L2 utterance fluency in the SA context. *Inhibitory control* is a cognitive individual difference the researcher choses to focus on in this study because it refers to a psychological control mechanism that allows one to suppress competing information. As learners seek to gain proficiency and speak more fluently in their L2, they struggle with competing information from their L1 – lexical, phonetic, or grammatical, for example. When examining cognitive fluency and the efficiency at which automatized processes occur, understanding one's ability to suppress competing information is therefore of interest to the researcher. If one can efficiently and easily suppress competing information from the L1, L2 speed production may occur more fluently. (See Section 2.6 for a discussion on inhibitory control).

2.3.2.3 Perceived Fluency

Perceived fluency is the second "sense" of fluency that Segalowitz (2010) discusses. It refers to the inferences that listeners make regarding L2 speech; that is, how *fluent* native speaker raters judge the learner to be. It is a subjective measure dependent upon the listener's impressions of how L2 speech sounds.

Many second language learners strive to speak more like native speakers (and associate fluency with the perception of speaking this way). Chambers (1997) notes that "fluency…often consists of…an overall aim of native-like speech" (p.543). Syder and Pawley (2014) identify two components of fluency - the ability of a learner to produce "fluent stretches of highly connected discourse" and to "convey an expression [that is] … not only grammatical, but native-like" (p.191). There is disagreement in the literature as to whether native-like proficiency is a fair and achievable ultimate goal. Fluency as perceived by one or two native speakers may not be the best measure of gains in fluency for L2 learners. Nevertheless, it may be a realistic measure by which individuals are assessed in authentic L2 situations, such as social settings in the SA context. Some studies investigate perceived fluency alongside temporal or utterance measures of fluency to gain a broader picture of second language proficiency, and most find a strong correlation between the two.

The construct of *perceived fluency* has been operationalized in a variety of ways - some with explicit instruction on how to judge native speech, and some without. For example, Yager (1998) leaves ratings to the judges' discretion, except to remind them that "nativeness does not mean academic correctness" (p.903). Different categories of raters are used to evaluate speech – while most studies use native speakers in general Freed (1995), some use native speaker teachers, phoneticians or even speech therapists to increase the quality of their expert ratings (Cucchiarini, Strik, & Boves, 2000). Usually, native speaker raters give a general rating of 1 to 10, or "not at all fluent" to "extremely fluent" as to their opinion of the quality of speech (Llanes, 2011). Research thus far has investigated how speech rate and pause frequency relate to listeners' perceived fluency (Bosker et al., 2013; Kahng, 2018; Suzuki & Kormos, 2019), finding strong relationships between these measures and perceived fluency. There is also some evidence of a relationship between speech rate and speakers' cognitive fluency (Kahng, 2020).

Interestingly, while Pinget et al., (2014) also find acoustic (temporal) measures of fluency are good predictors of perceived fluency, they find that judgments of fluency are only weakly related to judgements of accent. Perceived fluency, then, while related to measures such as speed, pausing and hesitations, does not necessarily mean the extent to which speech sounds like a native speaker.

Perceived fluency is often examined with respect to its relationship to temporal (utterance) fluency measures, and the research often finds an improvement over time, and sometimes finding a relationship to some temporal measures such as speech rate. Early research showed SA participants had positive gains on speaking fluency in terms of judgment by native speakers (Freed, 1995). Kormos and Dénes (2004) studied Hungarian learners of English who completed structured narrative tasks based on actions occurring in a given cartoon strip. MLR and phonation time ratio were the best predictors of both native and non-native speaker teachers' ratings of fluency. While both native-speaker and non-native-speaker teachers take speed and breakdown fluency measures (such as pausing behaviour) into consideration, raters differed in how much importance they gave to lexical diversity, accuracy and pause length when assessing fluency; ratings, however, did not correlate with quantitative measures of speaking fluency. Diaz-Campos (2004) finds a high correlation between temporal and perceived fluency measures for intermediate L2 learners of Spanish; others also find speech rate, phonation rate and MLR correlate with NS ratings (Kormos & Dénes, 2004).

De Jong et al. (2012) conclude that raters' fluency may be related to more than fluency measures; raters only partially do as they are told when given in instructions are "probably not the best starting point" (p.36) for fluency assessments. Eisenstein (2008) finds inconclusive results in her study of native reactions to non-native speech; NS are not necessarily more accurate judges of fluency than proficient NNS. In a unique study comparing *perceived fluency* by NS judges to participants' L1 and L2 speech samples, Derwing, Munro, Thomson, and Rossiter (2009a) found speech rate, pausing rates and syllables per second to be related to perceived fluency in Mandarin and Slavic learners of English. However, there was no statistically significant evidence that L1 baseline fluency measures predicted L2 fluency improvement over time. Comparing *perceived fluency* and *utterance fluency* in Canadian English learners of French, Préfontaine and Kormos (2016) find that in fact, false starts, repetitions and repairs do not hinder perception of L2 fluency, but rather make it appear more native-like. Raters' comments showed that L2 speakers need to find a "fine balance between speech rate, rhythm, pausing and lexical retrieval" (Préfontaine & Kormos, 2016, p. 12) to appear fluent.

De Jong et al. (2012) emphasize the importance of following Segalowitz's idea of investigating oral fluency with respect to both the speaker's speech (*utterance fluency*) and the listener's perception (*perceived fluency)*. In their study, they assess participants' performances on the Pearson Test of English for learners of English; they rated 397 speaking performances on a fluency scale. For learners of Dutch, they rated 354 performances on the Common European Reference (CEFR) framework's spoken fluency scale after they completed an image description task. They found 30% of the variance in fluency ratings could be explained by temporal fluency measures of MLR, silent pauses, filled pauses and speech rate. Regression analysis to compare low proficiency and high proficiency learners found temporal measures explained 30% of the variance in low proficiency learners, but only 6% in high proficiency learners, a surprisingly low percentage of variance explained, suggesting that temporal measures more effectively measure fluency in lower proficiency learners. Other factors (for example, accent or intonation) could have an impact on perceived fluency, more for lower proficiency than high proficiency learners. Préfontaine, Kormos, and Johnson (2016), too, find MLR and articulation rate to be the best temporal measures for predicting how native speaker raters assess "fluent" speech. Pausing frequency was found to be a much weaker predictor of how "fluent" native speakers thought participants' speech was.

The present study does not investigate perceived fluency, as first, it has continually been shown to correlate positively with utterance fluency and second, only moderate changes in fluency were expected over the course of a three-month stay, that perhaps would not have been detectable by the native speaker listener. While there are a vast number of studies on utterance fluency, few examine the relationship between *utterance fluency* to *cognitive fluency*. Kahng (2020) is one recent example that does. Following Levelt's (1999) *blueprint of the speaker*, and Segalowitz's (2010) adaptation of the model, it is clear that *cognitive fluency* is crucial to understanding speech production, as are the cognitive individual differences that may affect utterance fluency.

Rather than perceived fluency, the present study focuses on utterance fluency and cognitive fluency, examining one individual difference in cognition, namely inhibitory control. Several studies have shown perceived fluency to be closely related to utterance fluency, finding that perceived fluency highly correlates with temporal measures of speed fluency, pausing behaviour and self-corrections (Préfontaine & Kormos, 2016; Valls Ferrer, 2011). Furthermore, in her assessment of fluency testing across disciplines De Jong (2018) argues that perceived fluency ratings themselves show what is salient to raters, and previous research shows that even when not instructed to focus on fluency, measures of speed fluency highly correlate with perceived fluency ratings. Furthermore, due to the present study's objective to examine fluency changes over a relatively short length of the stay, measuring fluency instrumentally offers a strong advantage –we are able to detect small changes in specific components of fluency such as pausing or speed, and are thus able to discuss which aspects of fluency have changed, rather than focusing on a holistic measure of perceived fluency which would merge speed, breakdown and repair fluency into one measure. In addition, moderate, rather than drastic changes in utterance fluency were expected in L1 American English learners of Spanish over the 13-week SA program. Changes in fluency may therefore not have been easily detectible to native speaker listeners, so the researcher decided to focus measuring utterance fluency to clearly identify changes. Thus, the present study focuses instead on Segalowitz's (2010) two other senses of fluency (utterance and cognitive fluency).

2.3.2 *Definitions of fluency*

Lennon (1990) was one of the first to explore L2 fluency development in SLA, and differentiated between fluency in the broad sense, as a "cover term" (p. 390) for oral proficiency and as used in general parlance, and a narrow sense of fluidity, consisting of an "isolatable component…of proficiency" (p. 391). He was also the first to use both temporal measures and native speaker raters to assess fluency, thus exploring both quantitative measures of fluency, and perceived fluency, two important components in L2 fluency measurement (Segalowitz, 2010) The study looked at speech rate, filled pauses per T-unit (a measure of meaning used in writing or transcribed speech research), the percentage of T-units followed by a pause, and native speaker perceptions of speech recordings. Native speaker ratings of their perceptions of the fluidity of EFL learners' speech did increase after a stay abroad in Britain, but there were individual differences in temporal measures of fluency performance, even in his small-scale study of 4 German learners of English majoring in English and studying abroad at a British university.

Sometimes defined as "the ability to speak at length without pausing" (Fillmore, 1991), speed and effortlessness of speech (Brumfit, 1984) or "automaticity" (Segalowitz, 2010), oral fluency is a "confused, multi-layered construct"(Chambers, 1997) *Fluency* comes from the Middle French term *fluidité,* which originally meant "abundance" ("The chambers dictionary," 1998) and in the mid-15th century, came to mean "a smooth and easy flow" (p.394). Riggenbach (1991) suggests that fluency is the concept of "easiness, ease, … [and even] willingness" (p.435) to communicate. In terms of L2 fluency, some researchers examine the extent to which speech is perceived to be fluent or native-like to the listener (Guillot, 1999), the ability to express an idea in both one's L1 and L2, a lack of accent (Pinget, Bosker, Quene, & De Jong, 2014), the amount of perceived L2 rhythm (Valls Ferrer, 2011), or the ability to speak appropriately in a variety of social contexts and use language to express ideas in creative or novel ways(Fillmore, 1979). Colloquially, when learners strive to be "fluent" in a language, they often refer to oral proficiency skills – correct grammar and sentence structure, the ability to have their message understood (communicative competence, rather than the flow of language) or, as one Japanese learner of English put it, prior to departing on a SA experience, to "have normal conversation with native speakers" (Jackson, 2016).

As an online and productive rather than receptive skill oral fluency requires immediate action; learners do not have the same time to think and process as they may have in a written activity in the L2. Neither do they have the luxury of time or avoidance as present with a receptive skill; one can easily ignore part of a message when listening or reading. Arguably, then, learners may find that speaking fluently and smoothly in an L2 is one of the most difficult skills to master, and one the most challenging aspects of being abroad.

The present study looks at L2 fluency in terms of flow and ease of speech. More specifically, it uses the Segalowitz (2010) cognitive science approach to fluency as an "observable characteristic of real-time speech behaviour…that reflects the developmental history and current state of the speaker… the mechanisms a speaker has developed…through socially contextualized, communicative activities" (p.6-7). Cognitive science is an interdisciplinary field that attempts to understand the mind from the perspective of disciplines such as psychology, brain sciences and social sciences. Second language fluency from a cognitive perspective, then, takes into account more than the semantics and syntax of speech production; rather, it reflects the cognitive state and flow, or level of automaticity Segalowitz (2010) a learner has access to as they speak in their L2. A full definition of *cognitive fluency* is discussed in relation to speech processing models in the next section, 2.3.1.

In the present study, speaking fluency will not be operationalized or assessed in the context of communicative interaction, lexical diversity, or speech, but rather in the ease and flow of one's L2 speech and the improvement of speech over time, in a SA context.

Second language teachers seek to understand how learners can work to experience this smooth, easy, flowing speech in their L2, or even develop speech that is perceived as more native-like. The effect of a particular learning context on their oral fluency development is also of interest to language teachers, and to learners themselves, especially those who choose to study abroad to enhance their language skills. As discussed in the previous section on SA research, participants of SA programs and their teachers often assume this context to be superior to an AH context, while this is not always the case. Most importantly, understanding appropriate measures and benchmarks aids fluency assessment and is important for pedagogy, as understanding and measuring students' progress is paramount to their L2 speaking success. As Tavakoli and Hunter (2018) suggest in a study of 84 British ESL teachers, teachers often use techniques for promoting fluency that are a "mismatch to what fluency research recommends"; the researchers advocate for a narrower definition of fluency and a push to adopt fluency practice techniques in the ESL classroom.

Chambers (1997), in her review of *fluency* definitions, argues that the field lacks a precise and comprehensive definition, and oral fluency should be defined specifically in SLA studies. Due to the multiple ideas that comprise the idea of *fluency*, SLA researchers explore and measure this construct in a variety of ways (see section 2.2.2), resulting in a wide range of studies on the topic with methodological differences in both the definition and operationalization of L2 fluency. Oral fluency continues to be an important and not fully understood aspect of L2 learner development, both in classroom settings and during stays abroad, decades after the concept was first explored. In light of the wide range of studies in SLA literature presented in the sections that follow, to move forward in L2 fluency research that is directly comparable to other studies, researchers must agree on both what oral fluency is, and how to measure it.

2.3.3 *How to measure utterance fluency*

SLA researchers have operationalized *utterance fluency* in a variety of ways, though recent literature is coming to a more of a consensus as to how this concept should be measured, following Segalowitz's (2010) categorization. Looking at the *breakdown, speed, and repair fluencies* separately, the following sub-sections discuss the importance of each type of utterance fluency in the overall speech act; that is, the underlying speech processes that occur in the L2, and the role each type of fluency plays in Segalowitz's model from formulation to overt speech. It also discusses ways in which researchers have chosen to measure participants' utterance fluency. Each section discusses ways in which the type of fluency may be operationalized and measured, and a final section discusses the most common utterance fluency measurements in current SLA and SA literature in the L2 fluency domain.

The process of speaking fluency is highly complex, and recent research has begun to look at more L2 fluency measures in studies that take place abroad and, in more detail. Wright and Tavakoli (2016), in assessing the last decade of speaking fluency work in the SLA context, suggest that in oral fluency research there is an urgent need for research into the "challenges [that] cognitive factors impose on learners' L2 speech development and social interaction" (p. 74), which has scarcely been investigated in the SA context.

2.3.3.1 *Speed fluency*

Speed fluency measures the rate at which a speaker speaks, adjusted for the length of their speech. While *speech rate* is an indicator of fluency simply measuring syllables/speaking time, *articulation rate* expresses the speech rate excluding all pauses. That is to say, the *articulation rate* is an adjusted speech rate that eliminates any individual differences in pausing behaviour. The measure of *speech rate* was modified to use *syllables per minute* (Nation, 1989) to incorporate a more accurate reflection of rate, as word lengths vary. Time spent speaking (phonation time) without pauses is used to account for individual differences in speech length.

Articulation rate, the resulting measure, shows the efficiency of the underlying cognitive system. Those who speak faster, more fluidly, may process linguistic information more efficiently in the underlying system. Most SA research has found speed fluency measures, especially speech rate and mean length of runs (how much one can speak without pausing) to increase after an SA experience (e.g., Llanes & Muñoz, 2009; Mora & Valls-Ferrer, 2012; O'Brien et al., 2007; Segalowitz & Freed, 2004; Serrano, Llanes, & Tragant, 2011; Serrano, Tragant, & Llanes, 2012; Towell et al., 1996). Speed fluency measures such as speech rate also correlate strongly with native speakers' fluency ratings of learners' L2 speech (Kormos & Dénes, 2004).

2.3.3.2 *Breakdown fluency*

Breakdown fluency refers to both pausing behaviour, and repetitions, restarts or repairs – any hesitations that cause a breakdown in speech. Early studies of breakdown fluency focused only on temporal measures to measure L2 speech. In addition to the speed fluency measure of speech rate, the number of words spoken per minute, two main measures originally used in fluency research are a measure of speed through *speech rate*; *silent pauses* a measure of the hesitations a speaker makes; and mean length of runs (henceforth, MLR), the average time spent speaking between pauses. Additional temporal variables have been introduced in the literature to provide a complete picture of speaking fluency. These measures attempt to assess how fluently a learner processes the preverbal message in the conceptualizer, encode the message and articulate it.

Most studies look at pausing behaviour when assessing L2 oral fluency changes. L2 learners may pause differently in their L1 than they do in their L2, depending on the efficiency of the underlying speech processing mechanisms at play. Pauses in L1 behaviour may occur when a speaker is searching for a thought, for dramatic effect, or as a part of an individual's personal speaking style. Pausing in the L2, however, is often indicative of hesitation due to a lack of oral proficiency - that is, a lack of fluidity in the conceptual planning, or a hiccup in one of the encoding stages, before the individual’s utterance becomes speech.

2.3.3.3 *Types of pauses*

Hesitations (pauses) are usually categorized into two main types – silent pauses (periods of no sound), and filled pauses, which are sounds that are not words in the language, such as *um, uh,* or other fillers. Silent pausing is often measured simply by adding up the silent time that occurs during a speaker’s total speaking time, and dividing it by total time to acquire a ratio (De Jong, 2013). Filled pauses are usually counted manually by a listener, and also divided by time resulting in a ratio value. Filled pauses can also refer to lexically filled pauses which may not be seen as disruptive to the speech process (e.g. well, you know, etc.).

Riggenbach (1991) suggests three categories of filled pauses:

> *a) nonlexical ("uh," "urn")* - fillers that are not recognized as words
>
> *b) sound stretches* - vowel elongations of .3 seconds or greater
>
> *c) lexical ("y'know," "I mean")* - fillers that are recognized as words but
>
> in context contribute little or no semantic information
>
> Riggenbach, 1991, p.426

Early research on filled pauses (Schachter, Christenfeld, Ravina, & Bilous, 1991) investigated filled pauses in native speaker university lecturers and found “no indication that filled pauses serve any particular function" (p.2) in the speech production system. The study finds that filled pauses occur with "astonishing frequency" (p. 3), at an average rate of 1.33 "umms" per minute when not gesturing, and 3.68 when gesturing. Hence, an L2 learner's use of filled pauses may not necessarily be attributed to a search for the correct lexical or grammatical term in the L2, but rather be indicative of L1 pausing behaviour.

Nation (2007) accounts for the speaker's use of "umm", "err" or other fillers as filled pauses. Some researchers such as (Tavakoli, 2011) categorize the pauses into native and non-native sounds, the native sounds being more "fluent". For example, "umm" or "uh" would be English fillers, whereas "ben" or "euh" would be French fillers. Native speaker sounds may be perceived as more fluent than non-native speaker filled pause sounds. While a non-lexical filler may indicate a block in the underlying cognitive speech process (in the mental lexicon, for example), a *lexical* filler may actually be similar to native-like speech patterns and not indicative of disfluency.

2.3.3.4 *Pause frequency, length and distribution*

In addition to the *type* of pause that occurs in L2 speech, it is often helpful to look at several other features of the pause. Specifically, the frequency of pausing in a speech sample, the duration of the pause, and the distribution of pauses (e.g. within clause vs. at clause boundaries) may all show evidence of difficulty in the underlying speech processing mechanisms to which Segalowitz (2010) refers in his adaptation of Levelt's (1999) model. Pauses at clause boundaries occur frequently in the speech of native speakers and are much less disruptive than clause-internal pauses (e.g. Pawley & Syder, 2000), which reflects inefficiencies in the cognitive system that produces overt speech. For listeners of L2 speech, within-clause pauses appear to affect a listener’s perception of fluency more than end-clause pauses in L2 speech (Kahng, 2018; Saito et al., 2018; Suzuki & Kormos, 2020).

Although it is accepted that speed fluency generally improves after a stay abroad, SA research examining breakdown fluency does not yield conclusive results. Some studies find phonation-time ratio (e.g., Towell, 2002), pause frequency or mean pause duration to improve (e.g., Du, 2013, Huensch & Tracy-Ventura, 2017; Kim et al., 2015; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004), but others have not. For example, Llanes and Muñoz (2009) did not find improvement in pause frequency or length after SA, or Towell (2002) did not find mean pause length to improve. Some studies measured pause types together, while others only looked at silent pausing and not filled pauses, and nearly all of them did not take into account pause distribution, which has only recently become a methodological consideration in L2 fluency research especially in the SA context.

A high pause frequency, or pause rate, shows evidence of a lack of fluidity in L2 speech. Likely, there is a lack of fluidity at the fluency vulnerability points (*fs* in Segalowitz’s (2010) model), such as grammatical or phonological encoding of the message. With high pause frequency, especially when a participant’s pausing behaviour in L1 speech is controlled for, there is clearly a point in the speech production process that slows down the delivery of the overt message. Low pause frequency, then, would indicate more *flow and ease* in the underlying cognitive system; participants who pause less by definition have more fluid speech, and are more likely to have higher cognitive efficiency in their speech process.

Similarly, *pause duration* plays an important role; a longer pause (controlled for L1 speaking behaviour) would indicate more of a delay in the underlying system. Finally, the location of the pause can help explain a participants’ level of fluency. Pausing at clause boundaries may be seen as more “native-like”, or a more natural way to pause. Pausing within a clause, however, is less natural and more likely to be a symptom of inefficiency, or lower cognitive fluency in the underlying speech processes. Taken together, *pause duration and distribution* contribute to an understanding of pausing behaviour as much as pause rate does.

De Jong (2016), for example, found that pause duration and pause distribution significantly contribute to breakdown fluency. De Jong (2016) studied 74 Turkish and English speakers and found that L2 pause distribution was significantly different for those with a higher L2 proficiency; they paused more at Analysis of Speech Unit (ASU), boundaries than within the ASU (speech units). Overall, L2 speakers paused more within ASU than at the ASU boundaries. Interestingly, there was no significant differences between silent or filled pauses in this respect. L2 learners paused more often, longer, and between clauses more, perhaps because unlike L1 speakers, they often had to formulate their conceptual planning mid-utterance in their inefficient system.

Tavakoli (2011), too, found that though L2 learners have longer and more frequent pauses than native speakers, the “distinctive feature” of L2 speech in 40 learners of English is their tendency to pause mid-clause. Pausing at clause boundaries, however, was equal or less to L1 speakers, and like L1 speakers they also rarely paused in formulaic sequences, perhaps because these phrases were more automatized in the underlying cognitive system.

2.3.3.5 *Repair fluency*

Another quantitative measure used in fluency research is repairs or restarts per speaking time that a learner makes in his/her own speech. This definition follows Lyster and Ranta (1997) who coined the term *repair* to mean an adjustment in a learners’ own speech to correct for a mistake they realize they make, as they make it. Repair fluency includes restarts and repairs divided by phonation time. This includes number of repetitions, (repeated words), restarts (beginning a phrase again) and repairs (self-correction while speaking) (Lyster & Ranta, 1997). Restarts are false starts where the speaker stops and starts the utterance again. Repetitions occur when the speaker repeats a word, or syllable.

Referring back to Segalowitz's model (see Figure 3 – L2 Speech Production Model (Encoding) Segalowitz 2010 p.9) a repair occurs because of the operation of the monitoring system; this process, including reformulation, occurs in the *formulator*, as the speaker produces new overt speech. A repair, much like a pause, may indicate a lack of efficiency in the underlying speech process. As the speaker is re-attempting their utterance, or perhaps repeating it while they perform lexical search, the lack of efficiency likely arises from the mental lexicon feeding into grammatical or phonetic encoding. The latter, as well as the point of articulation comprising of *fs*, Segalowitz (2010) areas of fluency vulnerability, affects the number of repairs or restarts the L2 speaker experiences. Studies in the SA context of repair fluency thus far do not have consistent findings, with some reporting decreased disfluencies after SA (e.g., Mora & Valls-Ferrer, 2012) and others finding no improvement in repair fluency (e.g., Huensch & Tracy-Ventura, 2017).

2.3.3.6 *Measures of utterance fluency*

De Jong (2013) reviews temporal measures of fluency progress over time. Following Tavakoli and Skehan (2005), she discusses *breakdown, speed and repair* components of fluency, she provides the following proposed measures of fluency. The list below includes a new, more accurate method of calculation. In most cases, this refers to the division of the measure by phonation time (time spent speaking without pauses), rather than syllables, as this is arguably a more accurate measure of length of speech (De Jong, 2013). As De Jong et al. (2015) note, frequency measures of fluency such as silent pauses, filled pauses, repetitions and repairs are "theoretically less confounded" (p.230) if divided by phonation time, rather than total duration. In this way, time spent pausing is eliminated when comparing an individual's performance.

Fluency measures categorized as speed, breakdown and repair fluency measures (from De Jong (2013), following Segalowitz (2010)) are listed below:

Speed Fluency:

Speech Rate: syllables divided per total time

Articulation Rate: syllables divided by phonation time (speaking time not including pauses)

Phonation Time Ratio: phonation time divided by total time

Breakdown Fluency:

Silent Pauses rate: number of silent pauses divided by total time

Mean Pause Duration: total length of silent pauses divided by total number of silent pauses

*Filled pauses rate:*number of filled pauses divided by total time

Repair Fluency:

Repairs:

Restarts and repairs divided by total time

As several reviews of oral fluency in SLA literature have noted, studies are neither consistent in which measures they choose, nor in how they operationalize them; reviews of studies in the oral fluency domain call for more consistently chosen measures to better compare studies (Tullock & Ortega, 2017; Huensch & Tracy-Ventura, 2017;Yang, 2016). Llanes (2011), too, notes that oral fluency measurement in SA literature is often inconsistent, making it difficult to compare studies and assess the effect of the SA context on L2 fluency development. However, in terms of measuring *utterance fluency* there does seem to be a trend towards collecting L1 data for comparison and control purposes, and a trend towards conducting more detailed analysis of pauses in terms of location and duration, rather than solely number (Leonard & Shea, 2017; De Jong, 2015). Furthermore, a growing body of studies over the last couple of years has begun to look at the relationship between utterance fluency and select cognitive fluency factors.

The SA context provides increased input and opportunities for output, making it a context in which L2 oral fluency development could thrive. With respect to cognitive fluency, being given the opportunity to be constantly exposed to the L2 and the opportunity for continual L2 speaking practice could increase automatization. As we know from Segalowitz's (2010) model, before overt speech is created, speakers must go through the process of both macroplanning (not language specific) and microplanning (language specific); the SA context could foster more efficient microplanning, resulting in more cognitive fluency. With respect to utterance fluency (speed, breakdown and repair), disfluencies in speech could be reduced after a stay in the SA context as participants become less affected by *Fs,* "fluency vulnerability points" as each stage of their preverbal speech develops. For example, grammatical encoding (*f3)* and phonetic encoding (*f5)* may be less likely to be affected as participants gain more and more exposure to the L2 spoken by native speakers in their environment. Furthermore, *f7*, self-perception could potentially change, as participants in SA programs may gain more confidence in their linguistic abilities while abroad. Knowledge of the internal and external world, and exposure to more L2 vocabulary in the SA context may allow the speech production process to happen more smoothly.

In the section that follows, details on the oral fluency literature thus far that has investigated *utterance, cognitive and perceived fluency* in the field of SLA are presented. For a discussion of L2 oral fluency development in the SA context specifically, see section 2.4.3.

2.4 Oral fluency development in SLA

To understand L2 fluency development in a SA context, it is important to understand what we know about L2 oral fluency development overall, which primarily takes place in other contexts, such as the foreign language classroom, or general interactions in the daily life of immigrant adults. L2 fluency research started to take off with Lennon's (1990) aforementioned case study of four learners of English studying abroad who experienced linguistic gains. The field has expanded over the last three decades to include studies on both Segalowitz's (2010) classification of *broader* and *narrower* senses of L2 fluency, as L2 researchers attempt to understand both the processes underlying fluency and optimal conditions for L2 fluency development. While measures of speed fluency are included in most fluency studies, studies choose different aspects of breakdown fluency and repair fluency to measure and report on, often reporting only on a handful of measures. Additionally, though speech elicitation tasks are generally done through story-like narratives, studies have changed over time as to how they approach speech elicitation. A detailed explanation of commonly used task types used to measure utterance fluency can be found in the discussion preceding the explanation of the present study's methodology (Section 4.1.1). The subsections of this chapter below first discuss the role of the L1 in understanding L2 fluency, and then discuss L2 oral fluency development in two contexts – learners at home, and learners abroad. This distinction is useful as learners at home generally have less L2 input and less opportunity for L2 practice. Furthermore, the research in L2 oral fluency is more highly developed *outside* the SA context; SA research tends to study fluency as one of many components, rather than its focus.

2.4.1 *L1 role in L2 fluency*

To understand L2 fluency well, we must take L1 fluency into consideration, as L2 fluency research has shown that L1 speaking style affects L2 speech (De Jong, 2013; 2014). Thus far, few fluency studies have taken into account the multilingualism of their participants (Tullock & Ortega, 2017), regardless of whether they are studies of L2 learners at home, or studying abroad. Fewer still have investigated the impact of the L1 on the L2 or adjusted for any differences in their measurement especially in an SA context, creating a gap this dissertation intends to begin to fill.

Segalowitz (2010) cautions that L1 fluency data is necessary to control for individual differences in speaking style, which is likely to transfer over to the L2. If one generally speaks more slowly than average in their L1, for example, this is unlikely to change in their L2. De Jong, Grouenhout, Schoonen, & Hulstijn (2015), in a study of Turkish learners of Dutch, find it useful to report measures of L2 proficiency corrected for L1 behaviour – L2-specific measures. Fluency, they conclude, also includes "personality and speaking style", and for most measures except silent pause duration, L2-specific fluency measures predicted proficiency.

Overall, when L2 fluency is included in studies, there is usually a positive correlation between the L2 and L1 when examining speech rate and silent pauses (Daewele & Towell, 2005; De Jong 2013; Derwing et al, 2009). For example, Kahng (2014) studies utterance fluency in L1 Korean, L2 English speakers and finds silent pause rate within clauses relates to L2 proficiency. L1 mean syllable duration (ASD), Mean length of runs (MLR) and silent pause rate were most strongly associated with L2 oral fluency, showing empirical evidence of the link between L1 and L2 fluency. Groenhout, Schoonen, Hulstijn, and De Jong (2013), too, find positive correlations between L1 and L2 fluency measures in a study of English and Turkish native speakers learning Dutch; the more fluent an individual was on one measure in their L1 (e.g. silent pauses), the more they were on that measure in their L2. This was found to be true on most measures, but a negative correlation was found between silent pause duration and repetitions, indicating those who pause less in their L1 may fill this time with decreased repair fluency (increased repetitions) in their L2. Importantly, “frequency measures such as the number of filled pauses, number of repetitions and number of repairs cluster together” (Groenhout et al., 2013, p. 233); those who use more filled pauses also repair their speech more. L1 measures are not seen to be significantly related to L2 proficiency (vocabulary) scores, though all L1 fluency measures could predict (to some extent) L2 fluency. When using L1 fluency to predict L2 proficiency, Groenhout et. al (2013) found no significant relationship between pausing behaviour and L2 proficiency; L1 average syllable duration best predicted L2 proficiency. Derwing et. al (2009), however, look at NS judgements of fluency and only find correlations of L1 and L2 fluency measures at the initial stages of their two-year longitudinal study of Slavic learners of Mandarin, indicating that the relationship between L1 and L2 may not always be straightforward.

Despite methodological difficulties that may arise when controlling for the L1, doing so allows researchers to obtain L2-specific measures of fluency that are not biased by L1 speaking styles and can thus better predict speaking proficiency (De Jong et al., 2015) and perceived fluency (Kahng, 2020). Not only has L1 performance been largely ignored in L2 fluency research, but there are almost no studies that address it within the SA context. Parts of L2 speech that indicate individual speaking styles (e.g., de Jong et al., 2013) need to be differentiated from those that could possibly help researchers identify a certain level of L2 speaking proficiency or progress in L2 acquisition. As Segalowitz (2016) suggests, temporal features speaking fluency and their underlying cognitive processes are likely to reflect fluency phenomena that are mainly L2-specific. For example, L2 speakers are more likely to pause within clauses rather than at clause boundaries or at syntactic boundaries as has been shown in studies assessing L2 utterance fluency (De Jong, 2016; Kahng, 2014; Suzuki & Kormos, 2020). Furthermore, we know that lexical retrieval is slower in the L2, switching in the L2 takes longer (higher switch costs) and that participants find it more difficult to inhibit the L1 than the L2. (Costa, Santesteban, & Ivanova, 2006), which clearly affects the speed and temporal measures involved in utterance fluency.

Huensch and Tracy-Ventura (2017) caution that when taking into consideration the L1 effect on L2, cross-linguistic differences should be studied. For example, Spanish speakers generally speak faster than English speakers, and German speakers slower than French, as English and German have a "more complex syllable structure" (Huensch & Tracy-Ventura, 2017, p. 759). Syntactic complexity, too, can elicit hesitations and produce less fluent speech; there is an even greater effect of hesitations on fluency in the L1 than the L2 (Sadri, Mirdamadi & De Jong, 2014).

De Jong et al. (2015) looked at whether L2 fluency measures can accurately predict L2 proficiency (measured by vocabulary knowledge) for Turkish learners of English. Filled pauses and repetitions were strongly positively correlated, and silent pauses and filled pauses were strongly negatively correlated. All fluency measures were found to be significantly correlated to the equivalent measure in the L1. Using regression analysis, L1 pause duration and speech rate (speed) were found to accurately predict the same measures in L2 speech. When looking at the interaction between language and fluency measures, the researchers find "the relation between L1 and L2 is not different for native speakers of Turkish or native speakers of English" (p.12). Regardless of language, L1 fluency is a good predictor of L2 fluency. Some studies look at adjusting or correcting the L2 measure for L1 fluency; they create an L2-specific measure of fluency based on the L1 speaker's speaking style.

Huensch and Tracy-Ventura (2017) examine the relationship between L1 fluency and L2 fluency English L1 learners of French and Spanish after 5 months' residence abroad. English learners of Spanish only had three measures of fluency that significantly correlated in L1 and L2 at both pre-test and post-test – average syllable duration, silent pause duration between ASUs, and silent pauses per speaking time. English learners of French, however, had two measures of fluency correlate between L1 and L2 at pre-test (average syllable duration and silent pauses per speaking time), and *five measures* (all except repairs per speaking time and silent pause duration) correlate at post-test, indicating that L1 and L2 speech had more similar speech patterns after a residence abroad, but that perhaps pausing and correction behaviour continues to vary in L1 and L2.

A recent study on Turkish learners of English investigated the relationship between L1 and L2 fluency, interestingly finding that breakdown and repair measures were correlated, but speech rate and articulation rates were not. (Duran-Karaoz & Tavakoli, 2020) Specifically, Duran et. al. (2020) found that end clause filled pauses were similar in participants' L1 and L2, but surprisingly found statistically significant differences in L1 and L2 mid-clause pauses, articulation rate, and speech rate, indicating that participants speak differently in an L1 than L2. However, this research has not been done on other languages, and the authors make the important consideration that their participants have not studied abroad or used their L2 for professional purposes, indicating less language contact than other fluency studies considering the role of L1 on L2. Kahng (2020) also found that L1 silent pause duration and filled pause frequency predict the corresponding L2 utterance fluency measures. Mid-clause pauses and mean syllable duration were predicted by L2 cognitive fluency measures (based on lexical retrieval tasks), in combination with the L1 measure. Again, this research was on bilinguals not L2 learners, De Jong and Mora (2017) provide an example of research that does look at L2 learners. They find L1 fluency measures predict L2 spontaneous speech in Spanish learners of English Interestingly, very little research with L1 and L2 speech data has been done in the SA context.

Taken together, the limited research done thus far on the role of L1 on L2 oral fluency generally shows that L2 speakers speak in a similar way – that is, have similar speed and pausing patterns – than they do when speaking in their L2. L1 speech has often been found to predict L2 speech patterns. Thus, individual differences in a learner's L1 are important to take into consideration when evaluating L2 oral fluency.

Within the SA context, the role of L1 has been explored much less than in L2 fluency literature in general, even though we know L1 fluency is a good predictor of L2 and a participant's L1 speech characteristics may influence their L2. In other words, since the majority of studies exploring this look at fluent bilinguals, late bilinguals, or immigrants, we do not have a clear picture of what oral fluency progress would look like abroad if fluency measures were adjusted for L1 speech.

2.4.2 *Fluency development in L2 learners at home*

Fluency studies of L2 learners in a domestic context are reviewed in this section, as learners at home are usually thought to have less input, and less language contact with the L2. SA studies on L2 fluency often include a group of at-home learners who show a more modest development. Importantly, more in-depth research has been done in SLA field on utterance fluency development in contexts that were *not* abroad, and therefore this research provides an appropriate starting point to delve into a more in-depth analysis. The literature reviewed in this section primarily focuses on utterance fluency development. L2 cognitive fluency studies are relatively rare in SLA research, but have been increasing in number since Segalowitz's (2010) and Segalowitz (2016) appeal for studies that focus on the underlying cognitive and social conditions that influence utterance fluency. Cognitive-based studies on L2 fluency that currently exist study a wide range of different cognitive abilities and individual differences, making comparisons between studies difficult. Most studies do include a measure of working memory or lexical access when examining cognitive capacity and fluency gains, but there is little consistency between studies making them difficult to compare. There is an increase of researchers conducting studies that explore the underlying psychological processes behind speech production, and cognitive individual differences that explore this; key findings of these studies are presented below.

L2 fluency research tends to focus, unsurprisingly, on measuring utterance fluency – the *speed, breakdown, and repair* characteristics of L2 speech that directly affect its *fluidity or flow*. Prior to Tavakoli and Skehen's (2005) categorization, most studies focused on *speed* and *breakdown* fluency, specifically speech rate and pausing behaviour changes, though some studies now incorporate *repair fluency.* What we know about L2 utterance fluency development is that speech rate and pausing behaviour generally improves over time; L2 learners speak faster over time, pausing behaviour is seen to be stable or to have improved (e.g. De Jong & Perfetti, 2011; Skehen, Foster & Shum, 2016; Kormos & Denes, 2004; many others).

With respect to *speed and breakdown fluency* measures, studies find that repetition and practice of a particular topic increases fluency and keeps pausing behaviour stable over time. De Jong and Perfetti (2011), for example, argue that proceduralization of linguistic knowledge "represents a change in the underlying cognitive mechanisms" (p. 533) that impacts fluency when participants practice repetition. In their study of 47 mixed L1 ESL students in the United States, they tested the effectiveness of Nation's (1989) 4/3/2 task for developing fluency. Dutch learners of English in a "fluency training" group are asked to speak on a certain topic for 4 minutes with one partner, and then switch partners and have 3 minutes to talk about the same topic, and then 2 minutes with a third partner. A second group of participants performed a similar task, except they spoke about three *different* topics. Results of the pre- and post-tests, four weeks apart show that mean length of runs (MLR) significantly increased, and phonation-time ratio (speaking time without pauses/total time) stayed stable, but only for the training group.

In recent years, there has been a shift towards looking at more detail in fluency measurement, especially with regards to *pause distribution,* rather than only location. Skehan, Foster and Shum (2016) make an important theoretical contribution to the literature, as they focus their analysis with respect to pause distribution (within or between clauses), rather than categorizing measures in terms of speed, breakdown and repairs. Curiously, they also find that for non-native speakers, greater lexical diversity is not associated with greater fluency (in terms of less pausing within and between clauses). However, for NS counterparts, having greater lexical diversity in their speech means pausing more at clause boundaries. The researchers conclude that this may be due to L2 learners having a smaller mental lexicon to draw from, and purposely working within the constraints they know they have in their L2.

De Jong (2016), too, emphasizes the importance of pause location and looks at AS-units (ASUs) to divide utterances, rather than clauses following Foster, Tonkyn, and Wigglesworth, (2000). An ASU is a unit of meaning which incorporates at least one clause. In a study of 52 L2 speakers of Dutch and 18 native speakers, she finds that L2 speakers pause more at within ASUs and L1 speakers pause more at ASU boundaries on 8 different narrative speaking tasks. Pauses used for macroplanning in Levelt’s (1999) model are at the clause boundary. De Jong (2016), finds between-clause pausing to have a similar location (before low-frequency words) and duration in both L1 and L2 speech; duration of L1 and L2 pausing within clauses was significantly different, emphasizing the need to examine pause location and duration.

Narrative tasks are consistently used in the L2 fluency literature, particularly in picture and cartoon format to elicit speech production. One well-known example is the “Picnic Story” (Heaton, 1966) cartoon-strip based task in which children walk to a picnic and are surprised to find their dog has jumped into the picnic basket and accompanied them, eating their food along the way. Huensch and Tracy-Ventura (2017) used a similar story called the Cat Story (Dominguez, Tracy-Ventura, Arche, Mitchell and Myles, 2013), and a story called The Brothers Story (Langley, 2006) at post-test to investigate British English learners of Spanish and French studying abroad for the final year of their degree. Fluency data was calculated using PRAAT and data was transcribed and input into CLAN for lexical information. In addition to the study on learners abroad discussed further with similar studies in 2.4.2, the researchers also looked at fluency measures of native speakers of English, Spanish and French to see if there was a significant difference in pausing behaviour. Specifically, they looked at seven fluency measures suggested by De Jong (2013) - average syllable duration, average silent pause rate (between clause and within clause), silent pause rate, filled pause rate, repetition rate, and repair rate. Results indicated that speed and breakdown fluency differ by language. Mean syllable duration, filled pause rate, repetitions and repairs were found to be significantly different between the three languages, leading the researchers to emphasize the importance of understanding cross-linguistic differences when comparing studies. L2 oral fluency development is also sometimes looked at in terms of the CAF framework – in relation to complexity (lexical diversity) and accuracy (Housen & Kuiken, 2009).

Some research, too, cautions that the chosen type of narrative task may have an effect on oral fluency results, as the task type can affect the complexity of structures produced; more background detail provided can result in more complex structures (Tavakoli & Parveneh, 2008). Skehan et al. (2016) use a narrative re-telling task for 28 Cantonese and Mandarin learners of English, but change the tightness of structure in the topic, rather than the topic itself. Participants watch 5 to 7-minute Mr. Bean videos ranging from loose structure to tight structure (sequence of events) and must re-tell the story in their L2. With respect to task structure, they find "more structured tasks…are associated with less pausing at clause boundaries" (p.107). Pausing less may be due to less resources in the speech system allocated to macroplanning due to the tight structure. Perhaps a loose structure is more indicative of an L2 learner's speech in daily conversation.

Taken together, we can see that utterance fluency – specifically speech rate and MLR – generally increase over time for L2 learners, whereas breakdown fluency measures often decrease, indicating positive gains in L2 oral fluency development. Task type can also have an effect on breakdown fluency measures, and most task types are monologic tasks that are easy to administer, but reflect less authentic speech (Préfontaine & Kormos, 2015; Tavakoli, 2016). Thus, the task structure should be carefully chosen when we design fluency tasks. The following section presents an overview of our understanding of L2 oral fluency development within the SA context itself.

2.4.3 *Fluency development in L2 learners abroad*

Oral fluency development in L2 learners abroad has been explored in less depth (using fewer measures and rarely using L1 data) than oral fluency development in domestic programs, and this subsection aims to discuss the main outcomes of L2 oral fluency research done within this context.

While task types are similar in most L2 fluency studies (speech elicitation, narrative tasks usually based on visual prompts or story starters), measures used in L2 fluency in SA studies are usually fewer in number and less thoroughly investigated than those used in domains focusing on L2 fluency in other contexts, as SA fluency students tend to have a wider focus. In fact, the majority of studies conducted in the SA context use a handful of measures, such as speech rate and number of filled or unfilled pauses, often calculated in ways not used in most L2 fluency studies outside the context (Kim, 2015), such as counting words per minute or manually counting mean length of runs or words per minute (Moehle & Raupach, 1983). While manual counting may be reliable, time and resource limitations may mean these studies can only investigate a handful of measures. Most research prior to Tavakoli and Skehen's (2005) classification of *speed, breakdown and repair* fluency focused on speech rate and pausing behaviour.

In general, learners abroad experience gains in L2 oral fluency (utterance fluency, primarily speed) by the end of their stay, speaking faster, for a longer duration, and with less hesitations than at pre-test. However, the results with respect to breakdown fluency in the SA context are not as clear.

For example, in Towell's (1996) four-year longitudinal study of 12 university level English learners of French, participants told a story using a comic-strip based narrative as a speech elicitation test. Average speech rate and pausing were not significantly different before and after a SA experience, but a change in the MLR between pauses was. MLR were significantly different from pre-test and post-test, indicating that participants could speak for longer stretches of time without pausing after a stay abroad. Segalowitz and Freed's (2004) seminal study on L2 oral fluency development abroad examined 40 American learners of Spanish studying abroad highlights the benefit of the SA context. Students in the SA context significantly improved their speech rate, mean length of runs longest fluent run and pausing rate. Those studying domestically did not improve on these measures, illustrating the benefit of the SA context for L2 oral fluency gains. Pause frequency and duration, have at times been found to improve (e.g., Du, 2013, Huensch & Tracy-Ventura, 2017; Kim et al., 2015; Mora & Valls-Ferrer, 2012), but others have only found that stays abroad resulted in significant gains in speech rate (Serrano, Tragant & Llanes, 2012; Llanes & Munoz, 2009).

Self-perception of oral fluency in SA has also been studied, with studies finding that L2 learners abroad perceive themselves to have a significantly higher level of oral ability at the end of their stay than the beginning (Allen & Herron, 2003; Allen, Dristas & Mills, 2006). It is important to note that perception of “oral ability” may include proficiency and pronunciation, not only fluency. SA studies on perceived fluency (by raters) also vary in their findings. Some have found that NS perceived fluency ratings correlate with temporal fluency measures of MLR (e.g. Préfontaine, 2012) and speech rate and pausing measures (Kahng, 2018; Saito et al., 2018; Suzuki & Kormos, 2020). while others find no relationship (Towell, 2002; Llanes & Muñoz, 2009).

In Freed’s (2000) study of 29 English L1 undergraduate students, half on a SA programme in France and 14 at home had NS judges rate speech samples on a fluency scale of 1 “not at all fluent” to 7 “extremely fluent”. Importantly, and unexpectedly, there was no significant difference between the SA and AH groups except when high proficiency learners were removed from the groups, indicating NS ratings may be more helpful in assessing fluency of lower-proficiency learners. NS raters in Freed (2000) were not given a definition of “fluency” on which to base their judgments; an interview conducted indicates that NS raters have varying definitions of fluency, and rated the scores based on varying characteristics, such as accent, grammar, vocabulary, idiomaticity and speech rate.

In terms of the SA effect on decreasing L2 disfluencies, Keppie, Lindberg, and Thomason (2016) note that though some studies find improvement in rates of disfluencies like repairs and repetitions (Mora & Valls-Ferrer, 2012; “almost an equal number of studies” (p.46) can be found in which learners do *not* see any significant improvements in their disfluencies abroad. D’Amico’s (2012) study of American learners of Spanish abroad is one example in which changes in disfluency did not reach significance, but there was still a trend towards less filled pauses at post-test. Kim (2015), however, looked at American learners of Chinese abroad and found that *unfilled* and *filled* pauses *increased* at post-test, despite other measures of L1 fluency. Participants did see significant improvements on speech rate and repairs, however.   Keppie et al. (2016)study of six L1 English learners of French on a 6-month stay abroad program using an interview method found that *filled pauses, silent pauses and repairs* significantly decreased. Participants had less self-corrections and significantly less disfluencies than their AH colleagues when speaking in informal small-group conversations, aligning with an earlier research finding that SA participants show more significant changes in fluency in informal rather than formal conversations (Diaz-Campos, 2004).

A line of research looks at spoken language development, using the Oral Proficiency Interview (OPI) tool and generally finds that participants improve their score after a stay abroad. (Moehle & Raupauch, 1983; Veguez, 1984; O'Conner, 1988; Liskan-Gasparro, 1984; Magnan, 1986). However, as Pinar (2016) points out, the OPI results are not directly comparable with studies that look at temporal measures of fluency, which is the focus of this dissertation.

Another interesting aspect in the literature is the concept of fluency training abroad. Tavakoli, Campbell, & McCormack (2016) find that L2 learners benefit from *fluency training* even over as short a period of time as 4 weeks. Participants were trained twice a week on fluency awareness, fluency improvement tips using discourse markers, and given opportunities to practice speaking. The researchers follow De Jong, Steinel, Florijn, Schoonen, and Hulstijn (2012)'s advice in measuring pause location in addition to speech rate and MLR and find that the group exposed to training were able to produce longer MLR and less between-clause pauses (at clause boundaries) than the control group, pointing to the importance of having L2 learners become aware of their fluency levels and train for fluency improvement. Kirsner, Dunn, and Hird (2003) stress the importance of measuring pause duration and mean length of runs (MLR) and studied 3-minute L1 and L2 speech samples of Japanese learners of English in Perth, Australia asked to talk about their favourite holiday. They found longer MLR and significant negative correlations between L2 pause duration and self-reported hours of training and practice in English.

When comparing contexts and examining fluency development in At-home students (AH), domestic immersion students (IM) and study abroad (SA) students the literature generally finds participants speak faster and (silent) pause less after a stay abroad (e.g. Segalowitz & Freed, 2004), but are inconsistent in their findings of other fluency measures (see Lopez-Serrano, 2010). Collentine and Freed (2004), for example found both IM and SA groups improved on fluency measures, contrary to expectations, the domestic immersion group (IM) improved *more* than the SA group, casting doubt on SA as the superior context. Both domestic immersion and SA contexts can lead to significant improvement (Collentine & Freed, 2004; Serrano et al., 2011). While fluency studies have compared learning contexts in the past, as Tullock & Ortega (2017) note, this tendency may be “gradually falling to disuse” (p. 2) as the research moves towards focusing on the benefits of each context separately, rather than determining an optimal context. This is in line with Perez-Vidal’s (2015) observation that participants’ language learning experiences in different contexts build up over time.

The length of stay is another component of interest in the stay abroad fluency literature. Short stays abroad have also been shown to allow L2 learners to significantly improve on oral fluency measures. For example, Llanes and Muñoz (2009) studied Spanish/Catalan child learners of English on 3-4 week stays abroad and found significant improvements in oral fluency measures. They found a decreased filled pause rate, increased speech rate and articulation rate, and longer MLRs, even after a short time abroad. Izmaylova-Culpepper and Olovson (2017) find that L1 American English learners of Spanish participating on an 8-week SA program experience significant gains in speed and accuracy by the end of their stay. Interestingly, participants were allowed to choose their own personal story to narrate, by choosing an image from a pile of cards with prompts one that closely related to an event they experienced. They were then asked to re-tell the same story at the end of their stay. Importantly, the study had a low number of participants (4), and all participants lived with a host family during their stay. Studies on longer stays, too, generally show that participants experience fluency gains in terms of utterance fluency, especially in MLR, speech rate, and pausing (hesitation) behaviour (e.g. Llanes, 2011).

Mora and Valls-Ferrer (2012) investigate the Complexity, Accuracy, Fluency (CAF) relationship at three time periods, pre-test, test after instruction at home, and a test after their SA experience. Participants were found to speak significantly faster; they also produced longer speech runs and they had less pauses, which were shorter at time 3. This provides strong evidence for the positive impact of a SA period on the development of oral fluency. Although there were moderate gains in accuracy, it was on utterance fluency measures L2 learners made the most progress abroad. Mora and Valls-Ferrer's (2012) discovery that gains in fluency were "not at the expense of learners producing less accurate or complex language; as their speech contained fewer errors per ASU their vocabulary became slightly richer and their syntax did not become less complex" (p.621).

Study abroad context effects do not necessarily produce linguistic gains. In fact, (Tragant et al., 2017) find "no effect of context" (p.558) when comparing domestic immersion summer camp programs with summer programs of similar length in language schools; although both groups significantly improved, participation in the program did have a positive effect. Recent research also finds significant improvement for short (3-to-4 week) domestic immersion programs for school-aged children (Tragant et al., 2017). Other studies, however, show that even on short, intensive programs such as a 4 week English for Academic Purposes program in the UK participants show increased speed and repair fluency compared to a control group (Tavakoli et al., 2016). In fact, they find that “tailor-made training aimed at improving fluency can have short term positive effects” (2016, p 466). Interestingly, participants were trained to have more awareness of their own fluency by listening to and identifying disfluent non-native speaker speech and were given strategies and opportunities to increase their fluency in class (such as completing picture description tasks). One study found that learning context was less of a predictor of fluency gains than length of time an L2 is studied prior to studying abroad (Diaz-Campos, 2004).

More recent studies do investigate several measures of speed and breakdown fluency, and sometimes repair fluency using De Jong and Wempe’s (2015) PRAAT script (e.g. Préfontaine & Kormos, 2009; Llanes & Muñoz 2009; Derwing et al., 2009). The majority of studies report basic fluency measures such as speech rate, articulation rate and MLR and silent pause rate (generally finding an increase in speech rate and varied results on other measures), but do not investigate details such as pause frequency vs. duration, filled pauses, pause location, or compare to L1 data. There are some exceptions to this as linguists have begun to find improvements in pause location (within or at clause boundaries) (e.g. Mora & Valls-Ferrer, 2012) with one study even finding those with more L2 classroom experience before studying abroad paused more at “juncture” pauses (at clause boundaries), rather than within clauses (Keppie et., al, 2016). Only one known study in SA literature shows a significant *increase* in filled and unfilled pauses at post-test (Kim et al., 2015), in American learners of Mandarin Chinese abroad. This result is puzzling, as other measures of L2 utterance fluency are aligned with most findings in the literature (speech rate increases and mean pause duration decreases). Given individual differences in speaking style, and a plethora of experience factors at play in the SA environment, there are several reasons this could have occurred, though the study does not provide a potential explanation.

Taken together, these studies show that L2 oral fluency development does generally increase after a SA experience, particularly showing participants speak faster after a SA experience. However, other measures of utterance fluency that improve, and the effect of other variables – such as complexity, accuracy, proficiency and learner experience – vary from study to study, and from participant to participant, confirming the need for more depth in analysis of fluency in SA studies, and more studies on individual differences. The present study works towards filling that gap.

To summarize, what we know about L2 fluency in the SA context is much less than we do about L2 oral fluency studies in general, as SA studies often investigate a variety of skills, rather than focusing on fluency measures in depth. Not only do studies vary in what linguistic gains they focus on, but they differ in how they capture information to measure linguistic performance (Llanes, 2011). Over 40% of studies examine learners of English, with Spanish and French as the two next common languages; few investigate other languages (Yang, 2016). In general, it is true that some temporal measures such as speech rate usually increase after a SA period (Tullock & Ortega, 2017). Most studies do find *some* improvement in linguistic gains in fluency over the course of a SA, though evidence has not been consistent and the measures to assess improvement are sparse, compared to the general literature on L2 fluency (which assesses contexts such as immersion classrooms for late bilinguals, daily interactions for immigrant adults, or EAL classrooms for immigrant children or refugees). In general, the SA context *does* appear to be conducive to fluency gains, though not always, and not for all participants

As mentioned previously, SA studies often have a wider focus than L2 fluency – looking primarily at experience factors abroad (Grey et.al, 2015), or vocabulary development (Briggs, 2016), for example, the research they conduct on L2 oral fluency development uses only a handful of measures. Typically, speech rate, mean length of runs, and a couple silent and filled pausing measures – sometimes taking into consideration the pause location – are used. Understandably, when the primary focus of SA studies is a construct other than oral fluency – vocabulary development, motivation and confidence, reading and writing development, for example – only the most used L2 utterance fluency measures in the literature are explored. This dissertation seeks to add to the growing body of literature doing more in-depth analysis on speed, breakdown and repairs in L2 fluency abroad, more in line with L2 oral fluency literature for participants not on a SA experience (see 2.3.1 for details).

2.5 Cognitive individual differences in SLA

Individual differences in SLA have been researched from a variety of perspectives in different bodies of literature, including cognitive psychology, applied linguistics, and social psychology, with each discipline evaluating different aspects of language acquisition (Segalowitz, 1997). Differences in age, aptitude and motivation have been studied, finding that adults can in some cases be as successful as younger learners in L2 acquisition, and that language aptitude tests can predict linguistic gains. As Segalowitz (1997) points out, though, it is difficult to isolate one individual difference factor (such as age) from the “multitude of social, affective, and other experiential factors confounded with it” (p.88). Robinson (2001) provides a framework for examining cognitive differences in SLA, based on aptitude testing and evaluating the cognitive resources of attention and memory. In general, we know that individual differences in “age and aptitude interact in predicting language learning” (Robinson, 2001, p. 381), though it should be noted that language learning in these cases has been measured primarily using grammaticality judgment tests.

A small, but growing body of literature is beginning to look at cognitive individual differences and their effect on L2 oral fluency development. Specific research in the study abroad context (Grey et al., 2015; Marijuan & Sanz, 2018; Pérez-Vidal, 2014; Sanz, 2014) cautions against generalizing the effect of context without looking at individual differences in participants, and of participants’ personal experiences in the SA context (as discussed in Section 2.2.3). While individual differences can occur in a plethora of categories, for the purposes of this section, we will focus on the implications of *cognitive* individual differences on SLA. More specifically, this section examines two main areas investigated in SLA literature – attention and memory. A discussion on inhibitory control (the individual difference explored in the present study) follows in the next section, 2.6. As literature on cognitive individual differences within the SA context is limited, the sections below use examples of studies in both AH and SA contexts where available.

Some research is beginning to use the cognitive bases of fluency to the Levelt’s (1989; 1999) Speech Processing Model. We can see that individual differences in *cognitive* fluency have the potential to greatly influence speech production. Individual differences in both utterance and perceived fluency are important to consider when assessing differences in oral fluency, but it is cognitive fluency that controls the processing of the message and the speed at which it reaches the *articulator*. Every utterance of speech is processed through grammatical, morphological, and phonological encoding before it becomes overt speech. In fact, in the 2016 IRAL Special Issue on Fluency, Segalowitz (2016) argues that future research must focus on "L2-specific modes of cognition that might underlie L2 utterance fluency" (p.5). Thus, it is imperative that research go beyond the measurement of changes in utterance fluency, and even beyond the extensive research on contextual variables, motivational and affective factors; understanding the relationship between cognitive and utterance fluency may be crucial in understanding why some learners experience greater gains than others, both in the AH and SA context.

Cognitive individual differences (i.e.: individual differences in cognitive ability such as memory, attention and inhibition), then, can be seen as components of the concept of *cognitive fluency* to which Segalowitz (2010) refers (Granena et.al, 2016). To be "cognitively fluent", processing must move smoothly, with ease, throughout the speech production process. Sub-conscious processes must also flow with ease – whether they relate to conceptualizing, planning, retrieving lexical or grammatical structures, or articulating one's message. Segalowitz's (2007) concept of *automaticity* occurs not only in the formulator in the 2010 model, but in all of them. If each of these areas functions in a fluid way, L2 speech is more likely to exhibit more fluent characteristics (in terms of speed, breakdown or repair fluency).

Thus, L2 oral fluency may be one area of SLA where individual differences in cognitive ability become apparent; more *cognitively fluent* speakers (those with more efficient processing in the aforementioned areas) are likely to have more fluent sounding speech. They would, by definition, have more fluidity in processes such as memory, attention, choosing from lexical resources, and inhibiting extra or conflicting information. If all of these areas of Segalowitz's

(2010) model work in a relatively superior way in *Participant X* compared to *Participant Y*, for example, *X* would be able to go from conceptualizing a message to articulating it with much more ease than *Y* would. While this may not be apparent in receptive tasks such as listening or reading, in L2 speech the difference would be noticeable to both the speaker and the listener. Compared to other areas of processing, such as L2 phonological processing, individual differences in cognitive ability may have a greater impact on L2 oral speaking performance as they contribute to the efficiency of the speech system. Also, as seen in the literature discussed to this point, L2 utterance fluency often strongly correlates with perceived fluency. To be perceived as more *fluent*, one needs both smooth utterance fluency (less pauses, corrections) and smooth cognitive processing in the underlying system. Cognitive individual differences (differences in ability such a memory, attention and inhibition) lead to cognitive fluency (or a lack thereof). While L2 oral fluency may be the area of L2 performance, L2 *cognitive fluency* may also have a great impact (such as in phonological processing) on L2 speech itself.

Cognitive research in SLA to date spans a wide range of topics– such as attention (noticing – see (Ellis, 2002)), working memory e.g.(Grey et al., 2015; Sunderman & Kroll, 2009), aptitude, motivation (e.g. (Hessel, 2017), phonological memory (O'Brien et al., 2007) - and, inhibitory control e.g. (Schmidt, 1992; Service, 2007). Both second language learners and fluent bilinguals must manage their multiple languages within a complex cognitive system that consists of many systems that are not language related.

In his methodological review on first language (L1) speech production, and discussion on its uses in L2 research, Crookes (2008) refers to the L2 learner's production system as "a very incomplete apparatus…accessible to consciousness, rather than automatic" (p.117). *Macro-planning* in speech production, of course, is not automatic and this stage of the Speech

Production Model may vary from learner to learner. For example, the extent to which *micro-planning* is needed in order to develop a pre-verbal message depends on the L2 resources available to the learner to form the pre-verbal message in the planning stage. Clearly, depending on an L2 learner's exposure to the L2, L2 ability level, external knowledge or even cognitive capacity, different resources will be available. In a forward-looking review of SA literature and sub-areas of study (Marijuan & Sanz, 2018) suggest that studies that include cognitive variables are sorely lacking in the SA field and future research needs to further investigate the plethora of cognitive variables at play when learning and speaking an L2.

Research has shown the profound impact that a bilingual mental lexicon has on speech production in fluent bilinguals, affecting L2 accuracy in speech production. (e.g. Colome & Miozzi, 2010; Costa & Sansteban, 2003; Paradis (1985); Runnqvist, Strijkers, Xavier Alario and Costa, 2012). The impact of mentally switching from one language to another, known in the literature as "switching-cost" (Rayner & Ellis, 2007, p.59) can impact different aspects of speech production, such as the vocabulary that speakers can retrieve from their lexicon. In L2 learners, then, this cross-linguistic influence may impact the speed and ease with which they speak, playing a role in oral fluency development. Inhibitory control of the L1 and L2– the ability to suppress the other language's competing information when speaking - and its relationship to oral fluency is an under-researched area of cognitive development in L2 learners.

There has been "growing suspicion" (Collentine & Freed, 2004, p. 154) that SLA theories lacked a strong cognitive basis to explain linguistic behaviour. Surprisingly, nearly two decades later few SA studies (and even fewer SA studies measuring oral fluency) take a deep look into cognitive variation in their participants and its effect on L2 performance. Few studies focus on cognitive differences in the SA context, and even fewer look at speaking proficiency or oral fluency development in relation to cognitive differences.

When studying oral fluency development of L2 learners, it is important to consider cognitive processes that learners must undertake to speak in their L2; without exploring this in L2 fluency research, we miss a large component of what makes each individual's speaking experience unique and may affect their development. Factors such as attention (noticing), working memory, and other cognitive factors influence how speech is processed, automatized. De Jong et al. (2012) is one study that does investigate this construct in relation to utterance fluency. Cognitive fluency of L1 Dutch speakers is measured using a picture naming task to evaluate lexical selection speed. Results show that the mean duration of a silent pause is not a good indicator of cognitive fluency. The syllable duration measure was the one that could best be explained by lexical selection speed accounting for 50% of the variation in utterance fluency.

Segalowitz and Freed (2004) were one of the first to incorporate cognitive fluency measures into L2 oral fluency development studies, working with both attention and working memory and finding a positive relationship of attention and WM ability to L2 oral fluency. However, relatively few studies have explored other aspects of cognitive fluency development as they relate to L2 oral fluency development, especially abroad.

Other areas of cognitive development have been less explored in SLA research. While working memory has been explored in several studies, Wen et al. (2017) advocate for the inclusion of other aptitude tests measuring automatization speed, implicit learning, or control over a language system. Few studies have explored L2 oral fluency development and cognitive abilities and fewer still in the SA context. Granena (2018)is one such example that looks at a battery of eight widely accepted cognitive tests for aptitude LLAMA (Meara, 2005) and HI-LAB

(Linck et al., 2013) in 135 learners of Spanish in a classroom context in the United States. Results showed that implicit memory ability could predict L2 speed fluency and implicit learning ability could predict lexical diversity in L2 speech when participants completed a picture-description task from the popular B1 *New English File* textbook. Interestingly, Saito, Tierney, and Sun (2019) looked at L2 pronunciation in 48 Chinese learners of English studying in the UK and found that sensitivity to speech signals (an aptitude test) predicted L2 pronunciation attainment.

The following subsections take a brief look at cognitive fluency studies other than those related to *inhibition* in SLA – namely, attention and working memory. Section 2.6 then specifically looks at inhibitory control as a construct, and discusses studies in this area, especially as they relate to L2 oral fluency development.

2.5.1 *Attention*

Attention is one area of cognitive development that has been investigated in SLA and L2 learning development research. *Perceptual* attention is defined as automatic behaviour – cues that we pick up on without paying specific attention, and *focal* attention is defined as to some degree voluntary executive control function (Robinson, 2017). As Schmidt (1990) notes, *noticing,* or attention plays a role in developing awareness of linguistic features. Individual differences in the latter may change the efficiency with which one can conceptualize, formulate and articulate a message in the L2. SLA studies on attention show that breakdowns in performance may be caused by limits on attentional resources. For example, Skehan (1998) shows that L2 learners have greater oral fluency when given a short amount of time to plan their speech vs. the no-planning condition. Hulstijn (2001) also discusses allocation of attentional resources as a key component of fluent production; despite negative connotations associated with behaviourist theories of repetition and drills, to reach automaticity of vocabulary knowledge (and thus speech), frequent reactivation is needed. We also know that at time, no significant difference has been found when comparing to AH on L1 or L2 visual word recognition, a measure of attention (Segalowitz & Hulstijn, 2005). As Kormos (2015) points out, the role of attention in bilingual speech production is even more complex than in monolingual speech production. If the cognitive demands of a task are high, as in the case of speaking an L2, tradeoffs between accuracy, fluency, and complexity occur.

Segalowitz (2007) points to the importance of two factors in speech production – *access fluidity* (in terms of access to the mental lexicon) and *attention control* – focuses and re-focuses attention in real-time. He advocates for the use of measures of access fluidity and attention control to assess learner experience in specific situations, as these can be obtained only through "extensive exposure and practice with the target language" (p.184). In fact, as a cognitive individual difference, *attention* must work not only in the cognitive function of noticing, but also in the metalinguistic awareness that comes when processing input and integrating new knowledge (Kormos, 2013).

Torres and Sanz (2015) look at cognitive differences in attention in Spanish heritage bilinguals and late immersion bilinguals through an Attentional Network Task (ANT), including a Flankers task, where participants must ignore competing information about directional arrows (non-verbal). While the Flankers task itself tests inhibition (the ability to suppress one piece of information while expressing another), the ANT is a more comprehensive test that includes both flankers type trials with attentional cue types, and is classified as a task measuring attention. Reaction time measures showed that heritage bilinguals and late immersion participants had no significant difference in their attentional capacity. The results suggest that there is no advantage for heritage bilinguals compared to late immersion bilinguals in cognitive capacity in terms of attention.

Segalowitz and Freed's (2004) seminal study was one of the first to investigate individual differences in L2 oral fluency development related to attention in the SA context. Using a lexical access test, they investigated the relationship between cognitive fluency of SA and AH learners and L2 oral fluency gains and found a significant relationship with one measure – pause-free runs, referred to as "Filler-free". This indicates that performance on cognitive fluency tests has been shown to relate to L2 oral fluency gains in the SA context. Tests specifically related to attention, however, did not show a relationship to L2 oral fluency gains.

2.5.2 *Working memory*

Another area of cognitive development sometimes explored in SLA literature is working memory. Working memory is defined as the ability to temporarily store and manipulate information necessary for complex cognitive tasks (Baddeley, Lewis, & Vallar, 2013) and is explored more than other areas of cognitive capacity in SLA. Kormos (2013) asserts that working memory is not only at play in input processing, but throughout the speech production process – in noticing, integrating new linguistic information, and in automatization leading to processing speed.

Whether working memory (WM) capacity is directly related to linguistic gains in L2 learners is unclear, both within and outside of the SA context. We do know that phonological memory (a specific component of working memory defined as the “short-term maintenance of speech sounds” - see Perrachione, Ghosh, Ostrovskaya, Gabrieli, and Kovelman (2017) - can predict oral fluency development in participants studying abroad; some find those with a higher working memory capacity had higher fluency or linguistic gains (O’Brien et al., 2007; Sunderman & Kroll, 2009); contrarily, other studies indicate that working memory capacity has no significant impact on linguistic gains in lexical or morphosyntactic judgement tests (Grey et al., 2015). O’Brien et al. (2007) look at phonological memory in adults through the recognition of non-words and find increased phonological memory predicts L2 oral fluency gains. Grey (2015), however, found no relationship between individual differences in working memory and linguistic gains while abroad.

Wen, Biedro’n, and Skehan (2017), in their research on SLA and aptitude, suggest that working memory is a promising avenue on which to continue investigating individual differences and their impact on SLA development, but few studies have focused on L2 oral fluency, and fewer still on the SA context. Sunderman and Kroll (2009) is one example that does look at 48 learners of Spanish abroad. In an effort to understand why SA learners often come back with “seemingly no change” (p.79) in their L2 ability, they investigate individual differences in cognitive capacity and its relationship to vocabulary growth, comprehension and accuracy. Those who performed better on a working memory (reading span) test did not fare any better than their lower-performing counterparts on aural comprehension tests; the SA environment itself allowed for comprehension development. However, when it came to *production* ability, those with higher WM capacity did experience more fluency gains. The researchers caution that the effect of cognitive differences in WM may be task dependent. Inhibitory control ability, or other types of cognitive capacity were not measured.

Grey (2015), too, examine the cognitive capacity of 26 advanced L1 English learners of Spanish on a 5-week intensive stay abroad program. Using an L1 reading-span tasks to assess working memory, they find that individual differences in working memory capacity are not significantly related to L2 accuracy in grammatical judgment. Grammatical ability improved over the 5-week stay abroad regardless of working memory capacity. Learners of Spanish completed sentence span working memory tests along with grammar and lexical decision tests. However, there were no correlations between working memory and any linguistic gains. Puri'c, Vuksanovi'c, and Chondrogianni (2017) find that 58 Serbian children in English language immersion programs have a cognitive advantage on executive control activities, including a verbal working memory (WM) test compared to a control, but their study does not look at fluency gains. However, Rezai and Okhovat (2016) do find that working memory capacity plays a significant role in L2 oral fluency development in L1 Persian learners of English. Participants were asked to speak on four topics– two simple, and two more cognitively demanding tasks - and experience gains on silent pause rate and pause duration. As Juffs and Harrington (2011) point out, while WM capacity measured by WM tests could be an explanatory variable for individual differences in L2 speech production, there are other explanatory variables that could have a stronger impact. Higher WM, however, especially in an SA context, could allow the learner to "marshall resources [more] effectively" (Juffs & Harrington, 2011, p.156).

Clearly, the results as to the extent to which cognitive capacities affect L2 linguistic gains (in both AH and SA contexts) are mixed and have not yet been fully explored in the literature. Further research is needed to discover which individual differences in cognitive capacity consistently affect L2 oral fluency development.

What we know about individual cognitive differences in the SA context is considerably little. Clearly, few studies look at L2 acquisition from a cognitive perspective, though this body of research is growing and is starting to include L2 oral fluency development. SLA studies on L2 learners both at home and abroad are beginning to follow Segalowitz (2016)'s call for more studies that investigate the cognitive fluency and underlying speech processes that lie behind differences in utterance fluency, but the relationship between inhibitory control and linguistic development (especially oral fluency, and especially abroad) remains largely unknown.

In particular, we are interested in factors that may inhibit the components where the fluency vulnerability points are present in the bilingual speech production model, such as morpho-phonological or grammatical encoding. This inhibition may in turn restrict one's speaking fluency, and fluency development. For example, an increased cognitive capacity to restrict interference from one language when speaking another has the potential to increase the *flow*, or *fluency* of one's macroplanning, microplanning, and consequently path from the preverbal message to overt speech. The next section explores *inhibitory control*, the cognitive individual difference measure discussed and analyzed in the present study. After a description of the construct of *inhibitory control* and a presentation of recent studies in the field, inhibitory control is discussed with respect to L2 learner's oral fluency development while abroad.

Leonard and Shea (2017) is one of very few SA studies that takes cognitive individual differences into account when investigating L2 fluency abroad. They look at CAF progress of 39 L1 English learners of Spanish studying abroad for 3 months in Argentina and also find participants experience gains on oral fluency measures, accuracy and complexity (in terms of lexical diversity). However, individual differences were found to affect performance. Specifically, those who performed better on picture-naming and sentence-verification tasks *(measures of cognitive fluency)* at pre-test were found to experience greater utterance fluency gains (speech rate, articulation rate and repair rate) at post-test.

2.6 What is inhibitory control and why is it important in L2 speech?

Inhibitory control is defined as the ability to suppress one action while performing another (Baus, Costa, & Carreiras, 2013; Misra, Guo, Bobb, & Kroll, 2012). *Inhibition* itself is an executive process that is automatic and is one of the underlying mechanisms of L2 speech. In his seminal paper on the bilingual mental lexicon (Green, 1998) suggests a now widely accepted *inhibitory control* model of bilingual speech production. Green's (1998) bilingual speech production model uses a *conceptualizer* that is independent of language. A supervisory attentional system responsible for allocating resources when processing is not automatic, comes into play "where automatic control is insufficient" (1998, p.69). The model posits that a lexical stimulus in the L2 evokes several responses within a speaker and the speaker must *inhibit* other lexical information (such as the name of the word in the L1) to correctly produce the L2 word.

This occurs in Segalowitz's model using Levelt's original concept of the *conceptualizer* – the part of the speech production model where the preverbal message is formed. The particular pieces of information that need to be controlled and inhibited depend on an individual's speech goal (e.g. saying a word in one of their languages). The Supervisory Attentional System (SAS) monitors the language production and "organizes and maintains non-linguistic representations for mapping" (Green, 1998, p.77). The more efficiently the SAS can do so, and the more separate the system can keep linguistic information from different languages, the more easily one can limit interference from one language when speaking the other.

It is this idea of *inhibitory control* ability that may shed light on why hesitations and other disfluencies are more prevalent for some individuals than others. If the cognitive functions underlying speed production itself are not fluent, overt speech may contain disfluencies. The fact that an individual's inhibitory control ability may change over a period of time makes it an interesting cognitive concept to explore with respect to its effect on an individual's L2 language performance. (Lev-Ari & Peperkamp, 2013)

2.6.1 *Inhibitory control in SLA research*

While some studies find a positive relationship between better inhibitory control ability and L2 speaking ability using lexical retrieval picture naming tasks (Costa & Santesteban, 2004, 2004; Linck, Hoshino, & Kroll, 2008; Pivneva, Palmer, & Titone, 2012), lower switching costs for fluent bilinguals than learners (Costa & Santesteban, 2004), or a positive relationship between phonological memory and inhibitory control of L2 learners (Darcy, Mora, & Daidone, 2016) few have explored this relationship for students learning an L2 in a study abroad context. In fact, the majority of this research focuses on *fluent bilinguals,* or multilinguals and not on L2 learners.

In addition to the limited number of inhibitory control studies investigating L2 learners, it is important to note that when relating inhibitory control ability to L2 speaking performance, few studies measure speaking ability in terms of oral fluency when examining cognitive capacity. Instead, most studies focus on vocabulary (proficiency) growth or reading comprehension.

2.6.1.1 The role of inhibitory control in language learning

Inhibitory control is an executive function concept that is not widely investigated in SLA literature, as the focus on cognitive ability and linguistic gains often falls on working memory and attention, as discussed in section 2.5. Some research on inhibitory control ability has been done on L1 learners in the classroom in Montreal, Canada, primarily in studies on English reading and mathematics. For example, Aite et al., (2016) examined 10 year old native English speakers' ability to correctly adopt a third person perspective in a body/figure transformation task which requires them to inhibit the first person perspective. Those with higher performance on the Stroop Task had higher performance on the perspective task. Recent research on pseudowords to investigate the discrimination of the often confused, reversible "b" and "d" letters in L1 child learners of English has shown that inhibitory control is required to successfully make this distinction (Foisy, Ahr, Masson, Houdé, & Borst, 2017).

To date, the research on inhibitory control ability of L2 learners has varied greatly in its focus. Pivneva et al., (2012) use a map task in both monologue and dialogue form in 42 L1 and L2 bilingual English-French and French-English speakers, and also have participants complete a battery of inhibition tasks – nonverbal and verbal Simon tasks, Stroop tasks and an Antisaccade task (measuring brain reflexes). Both Simon and Stroop tests ask the participant to focus on certain information while providing competing information simultaneously. For example, a colour Stroop test presents a word, such as "green" in a different colour font (such as yellow), asking the participant to identify the colour (not the word). In a Simon Task participants must quickly and accurately decide if an arrow is pointing right (clicking a YES/NO option), with half the trials appearing congruently (e.g. right arrow, right side of the screen) and the other half incongruently (e.g. right arrow, left side of the screen). The researchers have native speakers rate the speech in terms of clarity of instruction, speaker fluency and speaker native-ness. They defined perceived speech fluency as "whether people spoke in a fluid or halting way" (Pivneva et al., 2012, p. 10). They found that speaker fluency ratings were unaffected by proficiency scores (as measured by a LEAP-Q proficiency test), but speaker nativeness was. Neither speaker fluency nor nativeness (as perceived by native speaker raters) was affected by participants' inhibitory control scores on Simon or Number Stroop tasks. However, differences in temporal measures of *utterance fluency* and their relationship to inhibitory control were not explored.

Colomé (2001) conducts a Letter (Phoneme) Decision study to test phonological activation of a translation word in the non-target language. Catalan-Spanish bilingual participants were prompted with a letter, and then a picture, and asked if the letter was contained in the Catalan word for the picture. Participants indicated "yes" or "no" to the on their keypads to the question *Is the letter in the Catalan word for the picture?*, but with three conditions present – Yes (corresponded with the onset of the first syllable of the Catalan noun), No (corresponded with the onset of the first syllable of the Spanish noun), and No (corresponded to a phoneme not in the Catalan or Spanish noun). For example, *T for taula (Catalan), M for mesa (Spanish) and F (neither).* To balance out the trials there were an equal number of No and Yes responses. Reaction times were measured to compare rejection of a phoneme included in Spanish with the rejection of a phoneme not included in either Catalan or Spanish. Seeing a letter corresponding to the Spanish word would activate Spanish, thus it was expected that participants would take longer (higher reaction times) to reject the non-target phoneme in Catalan. Higher reaction times would be expected if both the Spanish and Catalan phonemes were activated, as it would take participants longer to reject the incorrect phoneme. The study showed both languages were activated. Participants took longer to reject a phoneme that was present in the Spanish than when the phoneme was not in either the Spanish or Catalan word, suggesting Spanish was activated in performing a task exclusively in Catalan with Catalan words. In terms of inhibition, the study shows that participants (fluent bilinguals) must inhibit one language even when completely immersed in and using their other language (Colomé, 2001).

Lev-Ari and Peperkamp (2013, 2014), in a study on VOT, also investigated individual differences in inhibition, hypothesizing that English-French bilinguals residing in France who performed a retrieval-induced-inhibition task would have more influence of their L2 on the L1 if they had lower inhibition skills. Their hypothesis was proven, indicating that those who performed better on inhibition tasks could better control activation of both languages. Similar studies have not yet been conducted with L2 learners.

Some research has also been done with phoneme rejection studies to investigate the extent to which both languages are activated, thus investigating cognitive fluency through the study of lexical access. Catalan-Spanish bilingual participants are asked to accept or reject a phoneme based on whether it (presented as a letter, or as a sound) was present in their less dominant language; results showed naming latencies were shorter for cognate words (which only have meaning for bilinguals), which was in line with the researcher's predictions (e.g.Costa, Caramazza, & Sebastian-Galles, 2000)). In a second experiment on production, results show that bilingual participants have faster reaction times when naming cognate words than non-cognates, while this difference does not exist for Spanish monolinguals.

Coulby, Poulton and Clayards (2017) study two groups (older and younger) of 23 participants in Canada who complete a Simon task (a measure of inhibitory control) and a visual paradigm task done through eye-tracking. In the latter, they must correctly identify a picture associated with a word on the screen; words are minimal pairs and a "competitor image" is included in the four images of which the participant must choose one. Results showed that there was no significant effect of age on inhibitory control ability. However, as a group (older and younger learners), participants who had higher reaction time scores on the Simon task – *had worse inhibitory control* – fixated more on the competitor image before choosing the correct option, indicating "they have more difficulty with uncertainty… [they are] more distracted by competitors" (Coulby, Poulton & Clayards, 2017, p.76). While the study is not on L2 learners, the researchers conclude that "general inhibitory skill…is important for how well competitors can be ignored" (Coulby et al., 2017). Though the participants in this study were in a domestic (at-home) context, for SA learners this could indicate a key difference in the L2 learning experience. If better inhibitors are less distracted by competing lexical information, they may in turn have more efficient processes in the cognitive speech system, from the macroplanning stage to lexical encoding. Thus, it is possible that these "better inhibitors" would have more fluent speech, though this was not investigated in the study. Linck and Weiss (2015), however, in studying L1 American learners of Spanish studying in an AH classroom environment found that WM capacity was related to higher proficiency, but inhibitory control was not.

Taken together, the research done on inhibitory control in language learning so far has shown us that better performance on inhibitory control tests is related to faster identification and rejection of competing linguistic information. Research specifically on L2 oral fluency measures has not been conducted.

2.6.1.2 Inhibitory control and study abroad

Few studies have looked at the effect of individual differences in inhibitory control in the SA context, limiting our knowledge of the effect of inhibitory control in L2 learners abroad.

Linck, Kroll, and Sunderman (2009) investigate inhibitory control abilities in adult L2 learners of Spanish abroad in Spain and their AH counterparts using a Simon task, a reading-span task (a working memory task in which participants read sentences aloud and are required to memorize end of sentence words), and a verbal-fluency task in which participants were asked to produce exemplars from a particular semantic category. They also a completed a comprehension task (a translation recognition in which they had to identify if an L2 translation was correct). Results showed that on both production and comprehension tasks SA learners significantly outperformed their AH counterparts. However, SA learners also experienced *reduced L1 access*; that is, they produced less exemplars of semantic categories than their AH counterparts and also showed no sensitivity to lexical-neighbour distractors in the production task, while AH participants did. The researchers suggest that due to decreased L1 use while abroad, L2 learners used their "L2 mental set" when performing tasks and relied less on their L1, supressing it. L1 inhibition in L2 learners, then, is facilitated by the immersion context. Interestingly, in a delayed post-test upon returning home SA learners retained their lack of sensitivity to lexical-neighbour distractors, but experienced "a rebound at retest" (p.1513) and in fact produced *more* exemplars in semantic categories than their AH counterparts did.

Linck et al. (2008) study different groups of L2 learners to investigate whether there are positive cognitive consequences of the immersion environment for bilinguals, too. In particular, they study L1 American English learners of Spanish in four groups – -immersion (study abroad), post-study abroad, classroom learners and a monolingual control group - to see if living in L2 immersion environment or having greater L2 proficiency are related to having better inhibitory control. They tested participants on inhibitory control using a Simon Task and a (L1 and L2 word naming) test, and on working memory using language and numerical processing tests.

Results showed that even late bilinguals (L2 learners) performed better on inhibitory control tests than monolinguals. Interestingly, and contrary to other studies, results showed AH participants to have significantly *better* results (higher inhibitory control ability/lower reaction times) on the Simon test compared to the SA participants. More proficient L2 learners abroad did not show a difference in inhibitory control ability compared to AH learners. The researchers suggest that being immersed in an L2 might "initially induce a cost to inhibitory control…[until] the bilingual has more experience controlling their two languages within the L2 environment" (Linck et al., 2009, p.367). While there was no relationship between self-reported proficiency and inhibitory control, when using a translation-recognition task (an online measure of linguistic ability), more proficient SA and AH learners were found to have *worse* inhibitory control. As the researchers note, improving inhibitory control is "not simply a matter of acquiring greater L2 proficiency" (p. 369). If we refer back to Segalowitz's (2010) model, lexical knowledge (a measure of proficiency) is only one component of the encoding necessary to move a message from the conceptualizer to articulation. To improve the underlying cognitive process of inhibition, and therefore cognitive fluency, it is likely that more than just lexical improvement is necessary. SLA research to date, especially in the SA context, has only begun to explore this construct.

LaBrozzi (2012), in his eye-tracking study finds that inhibitory control scores based on a Simon task had no relationship to linguistic processing ability (morphological cues) for either group. These results paint a different picture for inhibitory control compared to other cognitive capacities. For example, Sunderman and Kroll (2009) in their investigation of working memory find that once learners reach a certain threshold of WM capacity, they experience more accurate production in the L2. Labrozzi (2017), however, in another eye-tracking study provides evidence to support more native-like development of visual cue (image) recognition when given a close competitor visual cue; the study also found that inhibitory control measured using a Simon task predicts more distraction by competitor items on an eye-tracking test. Importantly, they look at another variable, age, and find “generally inhibitory skill along with older age is important for how well competitors can be ignored” (Labrozzi, 2017, p.83).

Taken together, the research on inhibitory control thus far indicates that, at least for fluent bilinguals, inhibitory control affects L2 production. Those with increased cognitive capacity for inhibition - “inhibitory skill” - are able to better ignore lexical competitors in their L2, and the immersion context may facilitate inhibition of the L1. More research is necessary to investigate whether inhibitory control (especially in L2 learners, and especially abroad), has a positive effect on L2 speech production.

2.6.2 *Language switching and inhibition*

When L2 learners are abroad in a multi-lingual context, they may be required to switch from their L1 to their L2. SLA researchers have also investigated *inhibitory control* in terms of language switching – suppressing one type of information (the L1) while executing the other (L2). If L2 learners can smoothly switch from one language to another, in terms of Segalowitz’s (2010) adaptation of the *blueprint of the speaker* model, they have efficient underlying processes that are not negatively impacted by competing linguistic information. When an idea is conceptualized, encoded using grammatical, morphological and lexical encoding, and articulated, and the speaker is forced to switch languages, they must fluidly re-conceptualize, encode and articulate.

The present study looks at language switching ability of learners abroad, to determine if the SA context positively affects this ability, and to investigate the impact switching from one language to the other has on L2 fluency.

2.6.2.1 *Language Switching in Bilinguals*

Much of the early research on *language switching* in bilinguals as an inhibitory control phenomenon has been done through L1-L2 or L2-L1 reading activities, where participants are asked to read a text which switches languages mid-sentence. Meuter and Allport (1999) were the first to look at *switching costs* in speaking, in a number-naming task in which participants were cued to switch languages by a change in colour in the screen. Participants were found to incur greater switching costs (reaction time delays) when switching into their L1. As greater inhibition is required on the L1 “because it is stronger and receives more activation relative to the L2” (Olson, 2016, p. 728), speakers experience greater switching costs.

Olson (2016)investigates the impact of language context on language switching costs in fluent bilinguals completing a picture naming task comprised of Snodgass and Vanderwart standardized pictures. In this study, “context” refers to whether L1 (Spanish)-dominant Spanish-English bilinguals speak in an English speaking or bilingual (English/Spanish speaking) environment. Participants either had to name 95% of the words in their L1, 95% of the words in their L2, or 50% each, in the particular set of pictures. Participants experienced greater switching costs when switching into their dominant language. Interestingly, however, *no significant switching costs were found* when participants operated in a bilingual context. That is to say, when the task involved 50% Spanish and 50% English items, and participants were required to switch into either one language or the other an equal number of times, switching costs were not significant. When considering SA participants, it is important to remember that they are a) likely operating in a bilingual context, but b) not fluent bilinguals and may therefore behave differently.

One inhibitory control study of 56 L1 English learners of French (L2) and Spanish (L3) finds a direct link between inhibitory control and language switching as participants show "asymmetrical switching costs…with smaller switching costs in the L3 relative to the more dominant L1 or L2" (Linck, Schwieter, & Sunderman, 2011, p. 656). Participants completed the commonly used Simon Task. Switching ability was evaluated using a verbal fluency picture naming task in which participants had to name the picture on their computer screen in the language associated with a colour (blue for L1, red for L2, etc.); the faster one was able to switch and correctly identify the next word in another language, the lower the switching cost. Reaction times scores on the Simon task in this study predicted the speed at which bilinguals were able to switch *into* their L2 or L3. Thus, inhibitory control ability on a non-verbal inhibition test in this study was found to be able to predict language switching costs. Other studies in which participants complete picture naming tasks with language switches also find inhibitory control to be strongly related to language switching ability in multilinguals both in children (e.g. Deepthi & Nataraja, 2014); and adults (de Bruin, Roelofs, Dijkstra, & Fitzpatrick, 2014), even in unbalanced multilinguals who do not use one of their languages as often or have an "abandoned" language (Philipp & Koch, 2009).

Language switching studies that look at individual differences in cognitive capacity are limited, especially in SLA research, and especially in the SA context; we do not know how L2 learners switch languages abroad and if this improves over time. One study investigating the impact of a switch in L2 learners found that speech after a forced, unexpected switch is more heavily accented (Goldrick, Runnqvist, & Costa, 2014), but the majority of language switching literature focuses on fluent bilinguals (see Declerck & Philipp (2015) for a review). Understanding how language switching works in L2 learners, and how it relates to individual differences in cognitive capacity (specifically inhibitory control) would help us better understand the underlying cognitive process of L2 speech production.

2.7 Study abroad experience factors and L2 fluency

Our knowledge of the impact on SA experience factors – such as length of stay, type of residences, type of classes taken and language use - on L2 oral fluency development is limited, as most studies that analyze individuals' experiences in the SA environment focus either on the sociolinguistic aspect of the SA experience or focus on linguistic gains other than oral fluency (see 2.2.1 for a discussion of linguistic gains in SA other than oral fluency). Those that do include fluency as a linguistic gain tend to look at one or two measures (such as speech rate, or lexical diversity), without considering a full range of oral fluency measures. Experience factors, for the purpose of this study, are defined as elements of the stay abroad experience that differ among individuals and may influence their language environment. The Language Contact Profile is one example of a questionnaire that primarily asks for demographic and experience factor information. Information collected includes previous experience with the L2, living situation while abroad, languages spoken with people in the same residence, courses taken in the L2, and days per week or hours per day spent speaking, writing and listening in the L2 in a variety of circumstances (Freed et al., 2004). At times, these experience factors are related to linguistic gains, but the variables analyzed and the outcomes of the studies are not consistent. Out of classroom contact has been found to have a positive relationship with intercultural sensitivity

(Martinsen, 2011), and language contact has not always been found to be related to vocabulary gain (e.g. Briggs, 2016).

Many studies do not analyze linguistic gains and rather examine the individuals' experience itself using self-reported questionnaire and interview data. Dewey, Ring, Gradner & Belnap (2013), for example, look at L1 English learners of Arabic's ability to make native-speaker friends while abroad and the factors they feel facilitated or inhibited these social connections. The strongest factor influencing the growth of social networks and related increased L2 use was the program's institution of an L2 language use requirement.

Literature that has looked at the impact of SA experience factors on L2 oral fluency development has found that more language use and language contact contributed to a difference in more and less fluent groups of participants abroad and at home (Freed, Segalowitz, et al., 2004). Some studies find a relationship between oral fluency gains in mean length of runs (amount of time spent speaking without pausing) and amount of out of classroom contact in L2 writing, but not reading (Freed, Segalowitz, et al., 2004), or between perceived oral fluency gains and total out of classroom contact with the L2 (Cigliana & Serrano, 2016), but others show us that gains in oral fluency are not necessarily a function of out-of-class contact hours or extra-curricular activities in the target language (Segalowitz & Freed, 2004). Hernández (2010) shows that participants who study abroad (and have more L2 language contact and use than participants at home) experience gains in oral proficiency when tested using SOPI, a standardized test that measures spoken narrative and persuasive ability. However, the study uses an overall test score rated by professional teachers and does not look at L2 oral fluency measures in particular. In a recent study where the researchers do examine L2 oral fluency progress, Duran-Karaoz and Tavakoli (2020), find scores on the Oxford oral proficiency test predict participants' mid-clause pauses in their L2, and scores on the Elicit Imitation Task (another proficiency test) predict speech rate, repairs and mid-clause silent pause.

Most SA studies that examine SA experience factors use the Oral Proficiency Interview (OPI), journals or logs, or questionnaires to gather information on SA experience factors. Individual differences in L2 oral fluency gains are investigated in terms of the amount of L2 input and output participants gain abroad. Surprisingly, studies such as Segalowitz and Freed (2004) find no significant relationship between L2 oral fluency gains and extra-curricular activities, out-of-classroom L2 language contact hours, or even the presence of a host-family. Host-family interactions are often described to be “banal”, repetitive interactions that do not require participants to hold a conversation at length (Segalowitz & Freed, 2004, p.193). Martinsen (2011), too, finds that participants with more complex interactions in their language contact – attending church groups or having formal conversations, for example – experience more linguistic gains.

What we know from previous studies on language contact in the SA context is that there is, in general, much room for improvement in SA program design. Researchers have made suggestions to maximize students’ linguistic gains while abroad. Some research finds that participants on SA programs are limited in the type of oral interactions they have and advocate for more varied opportunities in out of class contact facilitated by programs (Briggs, 2016; Kinginger, 2011; Dewey et al., 2014).

2.8 Gaps in the literature

As is clear from the above review of the literature, and in several meta-analysis studies that have been conducted e.g. (Yang, 2016) there are several gaps in SA literature with respect to the study of individual differences in L2 oral fluency development. Although many L2 learners continue to expect that the (assumed) increased opportunity of input and output in the SA context will lead to L2 oral fluency gains while abroad, there is still not enough scientific evidence in the literature to support SA as a superior context for linguistic gains, or even L2 oral fluency gains, with the exception of speed fluency, which generally improves. Context effects of the SA environment, especially with respect to its role in L2 speaking fluency development or other linguistic gains, are unclear despite two decades of research on the topic, and despite the continual expectation from students and educators that SA or immersion contexts are optimal learning environments.

While some studies show that L2 learners improve on fluency measures, others show no significant differences on L2 speaking performance or proficiency. Individuals, too, seem to fare differently in the SA context – some thrive, and others do not, but we do not have a clear answer as to which individual differences affect performance the most. Most SA studies show an improvement on measures of L2 fluency, though the focus of studies in the SA context is often on other variables, with fluency being only one component, often measured by one or two measures. Most importantly, there is no conclusive evidence on which the optimal context (or combination of contexts) is for L2 fluency development– or perhaps, at which point L2 learners reach a proficiency threshold necessary before they can benefit from the SA context. As the number of students who study abroad increases every year, filling some of these gaps would be an important step in moving forward in SA literature.

First and foremost, within L2 oral fluency literature itself in the SA context, only a few studies go into depth on pausing and disfluency behaviour. Early studies measure pause frequency in general, and do not take into account where the pause occurs (between or within a clause), meaning that overall, we have little idea of *how* participants pause abroad and how this relates to their personal speaking style; in other words, how they pause in their L1 Taking both pause distribution and location in addition to pause type into account when looking at oral fluency gains, and using L1 data, could add to our understanding of L2 oral fluency development abroad. Specifically, the literature has begun to look at within clause vs. between clause (or within/between an ASU) silent and filled pauses, and other disfluencies in L2 speech. Adding these measures may allow for a more comprehensive understanding of how and L2 speech is disfluent in the SA context, and how this speech progresses after a stay abroad. Marijuan and Sanz (2018) advocate for L2 oral fluency studies to control for inter-subject differences in L1 fluency. Controlling for L1 behaviour in speech when assessing L2 speech production and development in the SA context is an important gap in the literature that has only begun to be researched.

Second, we have little understanding of *cognitive fluency* in the oral fluency literature. Segalowitz's (2010) model of speech production puts emphasis on the underlying cognitive processes that affect L2 speech development. The impact of cognitive individual differences, especially in the SA context and especially with L2 learners (as opposed to fluent bilinguals) is an area that still needs to be explored. While we have some knowledge of the role of attention, working memory, and other cognitive individual differences impact L2 oral fluency, we know surprisingly little about how these differences play out in L2 learners studying abroad. Given the growing need for research into cognitive components in SLA areas (Marijuan & Sanz, 2018), investigating one or more individual cognitive differences can contribute to our understanding of how L2 oral fluency develops abroad.

Specifically, inhibitory control is one cognitive factor that is under-researched, particularly in a stay abroad context. In most studies, only generic and non-linguistic, icon (arrow) or colour-based tests of inhibitory control such as the Simon or Stroop are used; perhaps a more linguistic or lexical inhibitory task would have a stronger relationship to speaking fluency and performance. The majority of studies on inhibitory control investigate fluent bilinguals or multilinguals, not L2 learners. As is clear in Segalowitz's (2010) model, we cannot fully understand L2 oral fluency development without considering cognitive processes such as inhibition. The majority of the work done so far on cognitive capacity of L2 learners abroad has focused on working memory, but inhibition has not been researched as much.

Language switching as an inhibitory control measure, and its relationship to other inhibitory control measures, and to oral fluency development in the SA context is also under-researched in L2 learners. Understanding how quickly L2 learners can switch from one language to another, and which individual differences help them do so will enhance our understanding of L2 speech production.

Third, while there is a plethora of research on individual differences in the SA experience, researchers have not yet discovered which SA experience factors are crucial in order to improve one's L2 speaking ability abroad. As the focus of the present study is L2 oral fluency gains, we will look at the relationship between experience factors (such as self-reported language use, classroom hours, and self-reported motivation levels) have on L2 fluency gains. Research thus far has shown us that language use and language contact are typically (but not always) related to vocabulary gain and gains in oral proficiency. Little research has been done on the relationship between L2 oral fluency gains and experience factors such as language contact and living situation. Further studies are needed to create a clearer picture of how and why L2 development in the SA experience differs so greatly for different individuals, and how these differences impact L2 oral fluency development.

As is clear in the review of the literature presented up to this point, and in meta-analysis studies on SA e.g. (Yang, 2016), the SA experience is not uniform in its nature. Context effects of the SA environment generally show that studying abroad plays a positive role in L2 speaking fluency development, but individual differences in the learner experience affect linguistic outcomes. Individuals fare differently in the SA context – some thrive, and others do not, but we do not have a clear answer as to which individual differences affect performance the most. Working to fill this gap is an important step in moving forward in SA literature. Practical implications of this research for SA programs are also a major interest in the field.

The present study, detailed in the chapters that follow, adds evidence to start to fill some of these gaps. Given the need for a focus on underlying cognitive factors when looking at L2 speaking ability, the study adopts Segalowitz's (2010) speech production model as a basis for the investigation. The study investigates *utterance fluency,* the impact of *cognitive* individual differences, and L2 experience factors, to gain more understanding on the L2 learner's experience abroad and factors that affect L2 oral fluency development.

3 Aims and research questions

The present study seeks to contribute to SLA literature by examining L2 fluency development in detail, controlling for differences in L1 speech. Importantly, the study looks at the individual cognitive difference of inhibitory control, an area that is understudied in the SA context, as explained in section 2.8. Individual differences (in both cognitive abilities and contextual experience factors) may play a significant role in an individual's oral fluency performance over the course of a stay abroad. Following Segalowitz's (2010) model of L2 fluency, this study is unique in that it measures both *utterance fluency* and *cognitive fluency* and examines their relationship.

3.1 Aims of the present study

The present study has three main objectives:

1. The first aim of this study is to assess the effects of a short SA period on L2 oral fluency development, and to determine the extent to which the SA experience affects L2 oral fluency. To do this, present study investigates whether *oral fluency* (speaking fluency) performance in US English learners of Spanish improves after a three-month SA period in Barcelona (in terms of speed, breakdown, and repair fluency). It adds to the literature by a) taking into consideration differences in their L1 speech patterns, such as speed and pause rates, and b) investigating pause frequency within and between clauses and ASUs (idea units).

Within L2 oral fluency literature in the SA context, only a few studies go into depth on pausing and disfluency behaviour. SA studies tend to focus on speed fluency without considering many measures of breakdown fluency, and address pausing behaviour as a whole, regardless of where the pause occurs. Taking pause location into account could add to our understanding of L2 oral fluency development. Specifically, studies have begun to look at within-clause vs. between-clause silent and filled pauses. Many SA studies do not take into consideration participants' L1 data. Controlling for L1 behaviour in speech when assessing L2 speech production is an important gap in the literature that has only begun to be researched (Kahng, 2020; Segalowitz, 2016). SA studies that control for inter-subject differences in L1 fluency could contribute to our understanding of L2 oral fluency development abroad.

2. Second, this study aims to determine whether there is a relationship between individual differences in inhibitory control and L2 oral fluency gains in an SA context. There is a lack of fluency studies that include measures of *cognitive fluency* in oral fluency literature, especially in the SA context and especially with L2 learners (as opposed to fluent bilinguals). Segalowitz's (2010) model of speech production puts emphasis on the underlying cognitive processes that affect L2 speech development. Most of the work done so far on cognitive capacity of L2 learners abroad has focused on working memory, but inhibition has not been researched as much. While some literature looks at inhibitory control ability and L2 speaking ability e.g. (Costa & Santesteban, 2004; Pivneva et al., 2012), few studies have explored this relationship for students learning an L2 in a SA context (Linck et al., 2011). Referring back to Segalowitz's (2010) model, inhibition is important as the L2 speaker moves from having an idea in the *conceptualizer and formulator*, to articulating it as overt speech. Before becoming speech, the message must pass through grammatical and phonological encoding, and L2 learners must inhibit competing

information in their languages to select from their L2 mental lexicon; stronger cognitive ability to inhibit could increase the efficiency of the underlying cognitive processes and therefore produce more fluent L2 speech.

Previous SA studies have not dealt with language switching as a component of inhibitory control. This study investigates whether the ability to better suppress an L1 (through better inhibition skills measured through performance on inhibitory control tests) is related to L2 oral fluency gains abroad.

Given the growing need for research into cognitive components in SLA areas (Marijuan & Sanz, 2018), research that investigates individual cognitive differences and its relationship to oral fluency can greatly contribute to the literature.

3. The third aim of this study is to determine the extent to which individual differences in SA experience factors affect L2 oral fluency gains over the SA period. Experience factors include amount of self-reported language use and related language use factors such as living situation and hours of language classes taken.

The present study, detailed in the chapters that follow, adds evidence to start to fill some of these gaps. Given the need for a focus on underlying cognitive factors when looking at L2 speaking ability, the study adopts Segalowitz's (2010) speech production model as a basis for the investigation. The study investigates *utterance fluency, cognitive fluency* and L2 experience factors, to gain more understanding on the L2 learner's experience abroad and factors that affect L2 oral fluency development. While there is a plethora of SA research that includes some experience factors as variables, we do not yet know which SA experience factors are to improve one's L2 speaking ability abroad.

3.2 Research questions

The sub-sections below present the three research questions and hypotheses for the present study, arising from the research objectives outlined in the previous section.

Research question 1: Does L2 Spanish oral fluency improve after 3-month stay abroad in Barcelona?

The first research question will examine L2 Spanish oral fluency ability in terms of *breakdown, speed,* and *repair* fluency. Specific measures chosen to examine this research question, and a justification for those measures are detailed in Section 4.0.

Hypothesis 1

The first hypothesis is that participants' L2 oral fluency will increase after a 3-month stay abroad. Our hypothesis is based on the fact that most previous research has found an increase in some measures of oral fluency (usually at least speed fluency) after spending a few months in an environment in which they have more L2 input and are given more opportunity for L2 output. In terms of *speed* fluency, participants are expected to speak faster in the L2 at post-test. *Breakdown* fluency measures are also expected to improve, as participants are expected to have less disfluencies at post-test. For example, pausing behaviour is expected to change, both in terms of number of pauses and pause location. First, overall pause rates are expected to improve (decrease) after a SA period. Second, participants are expected to begin pausing more between clauses and ASUs (and less within clauses and ASUs), as this is more similar to native-like speech, as has been shown in recent fluency studies (e.g. De Jong et al, 2013).

Disfluencies that contribute to breakdown (e.g. filled pause rate) and repair fluency (repair rate) are expected to decrease at post-test. At post-test, participants should have longer mean lengths of run (MLR); that is, they will speak for longer stretches of time without pausing. They will also have less elongations, repetitions and repairs in their speech, effectively making *gains* on these measures at post-test by having less of these disfluencies in their speech.

L1 behaviour has been shown to be indicative of personal speaking style (e.g. De Jong et al., 2013) and may influence participants' L2 production. Therefore, L2 oral fluency measures will be adjusted to control for the L1 where possible.

Research question 2: Does inhibitory control ability relate to L2 oral fluency gains in the SA context?

To investigate this question, we will look at whether individuals who perform better on inhibitory control tests have greater L2 oral fluency gains in the SA context. An individual's inhibitory control ability may have consequences for their L2 speech production. We know that the effect of inhibitory control on L2 speech in fluent bilinguals becomes more pronounced with more exposure (Lev-Ari & Peperkamp, 2013). However, this has not been tested in the SA context.

Hypothesis 2

The second hypothesis is that inhibitory control is strongly related to L2 fluency gains in the SA context. "Better inhibitors" (better performers on inhibitory control tests) will be more successful at suppressing the L1 when in an L2 context, and therefore experience more fluency gains. Less interference from the L1 would lead to increased fluidity in the underlying cognitive system – the conceptualizer and the articulator - and result in more fluent speech. Additional language exposure and speaking opportunities present in the SA environment may lead to L2 oral fluency gains (Hypothesis 1). If a participant has higher inhibitory control ability and therefore less L1 interference, this would allow them to take greater advantage of L2 speaking experiences during their stay abroad, resulting in larger L2 fluency gains.

Research question 3: To what extent are L2 experience related factors (such as self-reported language use, Spanish classes taken and living situation) related to fluency gains after a 3-month stay abroad?

Although participants spend the same amount of time in the L2 environment, several factors individual to their experience may influence their use of the L2 in the environment, and therefore affect changes in their oral fluency development. Experience related factors such as self-reported language use, living situation, hours of Spanish class taken are of interest as they vary among participants.

Hypothesis 3

The third hypothesis predicts that participants who engage more with their L2 environment (use the L2 more, live with locals, etc.) will experience more gains in L2 oral fluency on *speed*, *breakdown* and *repair* measures than those who do not. They are predicted to have positive gains in speed fluency (e.g. higher speech rate), and negative gains on breakdown and repair measures (e.g. lower pause rate and repair rate). Specifically, participants with a higher percentage of self-reported L2 use are expected to be more fluent by the end of their stay than the beginning; they will experience more fluency gains than those who report speaking Spanish less. Participants in a more multilingual living situation that fosters language use (host family, or Spanish monolingual housemates) are also expected to experience more fluency gains than those living in other situations, due their more immersion-like experience (Cundick, 2007; Baker-Smemoe et. al, 2012; Saito et al., 2018).

4 Methodology

As the theoretical background for this study has been established, this chapter reports the methodology used to answer the three research questions. The research design is discussed in 4.1, which stems from the literature review in the previous section. Section 4.2 provides information on participants. In 4.3, each of the tasks is explained in the context of their relationship to *oral fluency* development (used interchangeably in the literature and in this dissertation with *speaking fluency*), *inhibitory control* or the *study abroad experience,* with respect to the research question they address. Details are given about the task purpose, stimuli used, and task procedure. Section 4.4 explains how the data was analyzed.

4.1 Study design

This section explains the research design of the present study. Details on participants are presented, followed by an explanation of the tasks and procedures used.

The schema diagram below, *Figure 1,* summarizes the entire research design:

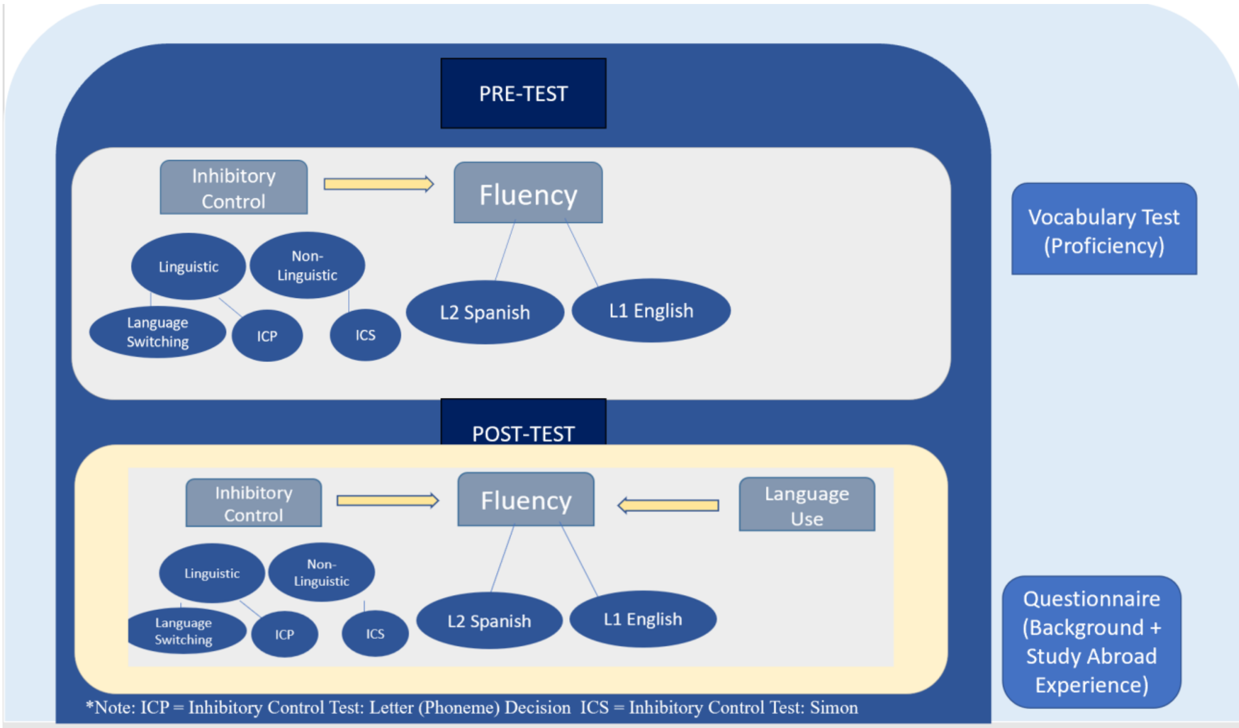


Figure 5 – Research Design Schema

A three-month long, non-intensive program was chosen as the specific context as most SA students choose this length of program, and the numbers of US SA students going on these experiences increases yearly (Murphy, 2017). While the purpose of their stay abroad may not be only for language learning (and could include cultural experience or travel), most SA participants do expect to improve their L2 oral skills while abroad (Isabelli, 2004)

As shown in Figure 5, the present study is set up as a pre-test/post-test design to investigate L2 learners' oral fluency development during a stay abroad, taking into consideration individual differences in inhibitory control, language learning background, language use, and proficiency.

The three facets of the study are shown in the central portion of the schema - oral fluency development, inhibitory control ability and experience factors; the schema shows them within the SA context. Individual differences in general (e.g. cognitive skills leading to *cognitive fluency* as defined by (Segalowitz, 2010)), as well as individual differences related to the SA experience itself both affect the main focus of investigation in the study – oral fluency development. Specifically, the relationship between inhibitory control ability and oral fluency development, and L2 fluency improvement after controlling for individual differences in L1 fluency are of interest. Perceived fluency is out of the scope of this study. Several studies have shown utterance fluency to be strongly related to perceived fluency (e.g. Kormos & Denes, 2009; 2016), thus we made a decision to focus on utterance fluency for the present study. In terms of RQ#2, we decided not to investigate other cognitive individual differences such as attention or working memory for two reasons; one, it was felt inhibitory control was the most under-researched area especially in the SA context, and two, to avoid test fatigue, because the battery of tests to investigate our research questions resulted in a 45-minute session and a 10-minute at-home questionnaire at pre-test and post-test. As participants were not highly motivated to participate (despite payment), test fatigue (and its effect on the data) was a major concern.

The present study measures oral fluency using a story-based task in both the L1 and L2 at pre-test and post-test. Inhibitory control is measured using three different tasks – a standardized inhibitory control task (the Simon), a more linguistic based decision task (Letter (Phoneme) Decision), and a language switching ability narrative task. Although previous research has shown that an individual's inhibitory control remains constant (Linck et al., 2008), data was collected at both time points, in case this differed for SA participants. Becoming more exposed to their L2 in the SA context, and using the L1 less than usual, could allow them to improve their ability to suppress L1 interference (Leonard, 2017). These three tasks were chosen to gain a broad picture of inhibitory control ability and thus, participants' underlying cognitive fluency. Many studies in the literature use both linguistic and non-linguistic measures when assessing inhibitory control Linck et al. (2008), as a linguistic measure may more closely simulate the inhibition of language. Language switching is used as a measure of inhibitory control as the ability to switch between languages shows the participants' ability to suppress one language (and its competing information) while speaking the other. Participants performed three inhibitory control tasks in total – two linguistic (language switching and letter (phoneme) decision task (see 4.3.5), and one non-linguistic, which is widely used (Simon task, see 4.3.4). Although the tests all measure inhibitory control, having both non-linguistic and linguistic measures allows us to analyze inhibitory control using reaction times in a task that pertains more specifically to language and may be more relevant to linguistic development.

Referring back to the schema, the two boxes on the right-hand side display individual differences in learners' background that they bring to the learning context. Specific details as to an L2 learner's individual prior experience learning Spanish and living situation in Spain may be experience factors that affect their L2 oral fluency development. These variables help investigate RQ #3, as they show how learners' individual language learning experiences in the SA context differ.

4.2 Participants

Participants for this study were 52 students (8 male and 44 female, aged 19 to 21) from American universities who came to Barcelona either from September to December 2015 or January to April 2016. All participants were of American nationality (one had dual Greek-American and one had Mexican-American nationality). The Mexican-American participant was not a heritage speaker of Spanish and began learning Spanish at age 11 (self-reported), though one member of his immediate family was a native Spanish speaker. Two other participants had one native Spanish speaker in their immediate family, but both did not begin learning the language until primary school at age 6. None of these three participants reported speaking Spanish to family during their stay. In fact, having a native Spanish speaker family member is not expected to greatly influence the results of this study, as these three participants reported speaking Spanish only 5%, 10% and 30% of the time respectively at post-test. One additional participant did report learning Spanish from the age of 1 but did not report any native Spanish family members. One possible explanation could be a caregiver outside the family who spoke Spanish. This participant reported speaking English 85%-90% of the time while in Barcelona This one participant was more fluent than the average participant, which may account for his high language use at post-test.

The majority of participants did not speak a language other than English or Spanish during their stay, with the exception of five participants speaking to family members or flatmates: one spoke Turkish (25% at post-test), one spoke Persian (5% at post-test), one spoke Lithuanian (5% at post-test), one spoke Finnish (2% at post-test), and one spoke Polish (10% at post-test). With respect to living situation while they were abroad, 10 participants had a two-week homestay experience, after which they lived in a shared apartment with peers, 36 participants lived in a shared apartment for the whole stay, 5 participants lived in halls of residence, and 1 participant lived alone.

All participants had taken intermediate level Spanish classes at their university before arriving and were taking at least one of their courses in Spanish during their exchange program. Many participants reported starting to learn Spanish at a young age in primary school. Table 1 shows the numbers of participants by gender, and average (mean) values for age, Spanish class and proficiency (the LEXTALE-ESP text, a measure of vocabulary size often used as a proxy for proficiency).

The LEXTALE-ESP test was designed so that some words should be known by low-proficiency learners, others by high-proficiency learners, and others only likely to be known to native speakers, like the word for *birdseed.* The test is a yes/no vocabulary test which includes vocabulary words from different frequency bands, as well as non-words. The test has a word-nonword ratio of 2:1 (Izura, Cuetos, & Brysbaert, 2014), where participants are asked only to indicate which words they are certain are correct Spanish words (-2 points for incorrect answers ,+1 for correct). The maximum possible score on the test is 60 (if a participant were to mark all the Spanish words as words, and not incorrectly mark any of the 30 nonwords as words). The minimum (worst) possible score on the test would be -60, if a participant were to incorrectly identify every one of the 30 non-words as words, and not identify a single real word as a word. For an example of possible scores, when the test was developed, the mean score of the participants tested was 11.9, with a SD = 17.9 and a range of -16 to 58 (see Izura et al., 2014).

The LEXTALE-ESP pre-test results for our participants showed an average score of 8.8 and a mean score of 7.3, which was substantially lower than expected from participants deemed to be at an "intermediate" level by their universities. Izuru et.al., (2014), authors of the LEXTALE-ESP test, show 12 as a "low proficiency" score and "54-60" as near-native or native. The test in his original English version (LEXTALE) has been shown to be a useful estimate of English proficiency and a more reliable predictor than self-ratings (Lemhöfer & Broersma, 2012). Descriptive statistics for the participants in this study show vocabulary test scores increased from pre-test to post-test by an average of 1.53 with a SD of 7.3 at pre-test and 8.6 at post-test.

Students were attending one of several university or language school exchange programs in Barcelona: CEA Language School, Universitat Autonoma de Barcelona (UAB), University of

Barcelona (UB) or University Pompeu Fabra (UPF). Due to the difficulty in motivating participants to join the study, we were not able to get participants who were all in the same academic program or university. All students except one lived in shared dorms or apartments with other American students while one student lived in an apartment alone; students chose these options before their arrival based on their personal preference. Ten students were required to spend the first two weeks of their stay in a homestay family before moving to student dorms, as a mandatory part of their specific program. All participants were paid 20€ for their participation, upon completion of the study.

Extensive recruitment was done in an attempt to find the largest number of willing participants. Recruitment involved attending information sessions for newly arrived SA students at several universities in Barcelona to ask if they would be willing to participate. Those who signed up at these sessions were then contacted via WhatsApp message and email. Students were also recruited from a SA language school in the city centre (of Barcelona) through meetings with the program managers. The educational institution agreed to encourage participation and Spanish teachers agreed to award an extra 10% on the participants' homework mark if they participated.

As each participant's experience abroad varied, and they were exposed to different levels of language use, questionnaire data was collected at the end of their stay. When analyzing the data and taking into consideration the results, it is necessary to consider data from the background questionnaire. Data presented in the table below reflects the varying experiences of the 52 participants in the study. Although not all of this information is not data directly related to the research questions presented and was collected both near the beginning and end of the participants stay, it does have the potential to help in interpreting the results in section 5.0.

Table 1 - Participants' Study Abroad Demographic and Experience Factor Information (n=52)

Demographic or experience factor	**Mean**	**SD**	**Min**	**Max**
Age	20	1	19	21
LEXTALE-ESP Pre-Test (-60 – 60)	8.8	7.3	-4	32
LEXTALE-ESP Post-Test (-60 – 60)	10.4	8.6	-6	38
Age of Beginning Spanish instruction (years)	11	4.29	1	20
Spanish taken prior to SA (months)	49	24.7	8	120
Class in Spanish (hours/week)	6.9	4.41	2	16
Self-Reported Language Use (%)	21.9	16.4	0	50
Comfort Level Speaking Spanish (1-10)	6.40	1.71	2	9
Desire to Reach Native Speaker Level (1-10)	5.62	1.44	2	7
Avoid switching to L1 (1-10)	3.73	1.87	1	7

4.2.1 *Excluded participants*

Some participants were excluded from some tasks due to missing data, or because their data showed they were an extreme outlier. All 52 participants successfully completed the Letter (Phoneme) Decision Test, and 49 completed the fluency tasks at both pre-test and post-test. The participants per task are shown in Table 2. Participants with a participant number beginning with "p" indicates they were first semester students. A limitation of the study design (as discussed in section 6.4 on limitations of this study) is that the switch point chosen was fixed. Unfortunately, the fixed point chosen for the first semester was too short and most participants did not have enough pre-switch data to analyze; thus, this data was used as a pilot. The language switching task was used as a pilot in first semester because the chosen length before the switch was too short and most participants did not have enough pre-switch data to analyze. In other words, they did not have enough time to speak in the first language they were asked to speak in (L2 for half the participants, L1 for the other half, counterbalanced at post-test), before being asked to switch to the other language. The task in its final form was run in second semester, with 35 participants. Table 1 below shows the original and final number of participants for each task. Participants were excluded if they had missing data at pre or post-test.

Table 2 - Participants per task

	Oral Fluency	Language Switching	Simon Test	Letter (Phoneme) Decision Test	Vocabulary Test	Questionnaire
#Participants completing task	52	38	52	52	52	52
#Participants final	49	35	49	52	52	52

4.3 Tasks and measures

This section explains the general procedure of the study and details the 5 main tasks used. Due to the number of tasks, for ease of reading this section is organized by task and measures used in each task are explained in the final subsection of each task.

4.3.1 *General procedure*

Participants were involved in two sessions of approximately 45 minutes (pre-test and post-test), plus a 10-minute online questionnaire completed on Google Forms before arriving at the post-test session. If participants had not completed the questionnaire upon arrival, they were asked to complete it before leaving the data collection session. Sessions were held in quiet classrooms at the University of Barcelona. For the first semester, quiet classrooms at CEA language school were used. A maximum of 3 participants were in the classroom at one time, on opposite ends of the classroom, completing tasks individually. Computer tasks were completed with noise-cancelling headphones for inhibitory control tasks to ensure maximum concentration. One researcher was in each classroom.

In the study there were 6 tasks: oral narrative tasks (one L2 and one L1 story), non-linguistic cognitive inhibitory control (Simon Task), linguistic cognitive inhibitory control task (Letter (Phoneme) Decision Task), language switching tasks (one L1-L2 switch and one L2-L1 switch), a vocabulary test (LEXTALE-ESP) and a background questionnaire. All tasks were piloted on 8 participants to ensure instructions were clear and the timing of the tasks went well. Instructions were in English to ensure all students could understand them, as their background with Spanish was varied and the universities categorized their level as, broadly, “intermediate”.

4.3.2 *Speech elicitation tasks (English and Spanish)*

Before presenting the speech elicitation tasks and procedure used for L1 and L2, tasks commonly used in the literature are described in the section below.

4.3.2.1 *Commonly used speech elicitation tasks for fluency measurement*

One of the most well-known tools for assessing L2 oral fluency is a speech elicitation task. Oral fluency tasks to measure utterance fluency have been shown to be effective when they are communicative, decision making tasks of approximately 60 seconds (Groenhout et al., 2013). Other studies recommend narrative tasks (explained below), while some even look at a variety of speaker tasks using different registers and communicative situations (De Jong et al., 2013). In most tasks, participants speak for one to three minutes, and few are tasks with dialogue or unstructured communication; rather, they are narrative tasks, asking participants to tell a story either from a picture prompt, or a personal experience. Using oral narratives based on picture materials helps ensure methodological consistency between learners in terms of the lexical and grammatical resources necessary to complete the tasks. Only Tavakoli (2016) has investigated which whether monologic, dialogic, oral narrative or interview speech elicitation tasks are more sensitive at detecting linguistic development during SA programs, finding no difference in speaking fluency resulting from different task types.

Nation (1989) used a different approach to investigate fluency changes before and after practice in a pedagogic setting he calls the *4/3/2 Task*, where students speak about a topic for a four-minute time limit, then speak about the same topic for three minutes, then two minutes. In his case study of eight learners of English, Nation (1989) found most learners increased their speech rate, although the percentage increase ranged from 1% to 67%. This method was used in a classroom setting.

In the SA context, however, most studies used a picture description or story-like format, even if the objective of the study is to assess dimensions other than fluency, such as comprehensibility (Bergeron & Trofimovich, 2017; Crowther et al., 2015, 2018) or complexity, accuracy and lexical diversity (Tavakoli & Foster, 2011). For example, in SA tasks participants are often asked to describe a story sequence. One commonly used task is The Picnic Story (Heaton, Longman & London, 1966)**,** which tells the story of a young girl and her brother who go on a picnic their mother has packed for them. Their puppy jumps into the picnic basket and to the children's surprise, eats the food before they can sit down to a picnic. A similar task The Suitcase Story task (Derwing, Munro, Thomson, & Rossiter, 2009b), in which a man and a woman bump into each other while turning the corner on the street, dropping their luggage as they fall is also commonly used . Their luggage is similar, and the two swap suitcases but only realize this as the return home with the wrong suitcase. These tasks are usually chosen due to their simplicity and the ease with which participants could see the course of events in the story. The present study uses both above tasks in a speech elicitation task when assessing *language switching* abilities, due to the tight storyline that can be easily interpreted from the multi-framed picture.

Studies have shown that the tighter a storyline is, the more accurate student's speech is; multiple framed stories also elicit more complex syntax than a single picture description task (Tavakoli & Foster, 2008). Some studies advocate creating various tasks that are either simple or complex, formal or informal, and descriptive, narrative, or argumentative to show participants' performance in a wide range of situations (De Jong et al., 2012). The present study follows the latter method for speech elicitation tasks in both the L1 and L2, to explore measures of utterance fluency.

4.3.2.2 *Task purpose and description*

The purpose of the oral narrative task used in the present study to assess participants' utterance fluency at pre-test and post- test. Speaking tasks were inspired by fluency tasks done in De Jong (2013). Tasks were adapted in an attempt to present speaking situations that would feel relevant to participants. Participants completed two speaking tasks – one in Spanish and one in English, both at pre-test and post-test.

4.3.2.3 *Stimuli and materials*

The stimuli and materials chosen for this task were created based on consensus in the literature that story-based tasks provide opportunities to assess participants' fluency in an L2. Tasks were created that would seem personally relevant to the college-aged participants and chose scenarios related to college life, such as a decision as to where a college party should be held. Oral narrative tasks were similar, but not exactly the same, for participants at pre-test and post-test. A major advantage of choosing different, but similar speech elicitation tasks at post-test is avoiding task habituation; participants cannot simply repeat what they said at pre-test. Tavakoli et al. (2016), for example, calls for studies to "avoid a practice effect" (p. 454). In each situation, participants were given background information, and required to make a simple decision and justify it to an audience they would find personally relevant (student council, a group of their peers, etc.).

It is important to acknowledge that using different tasks at pre- and post-test could influence the type and amount of oral production output as speech elicited at Time 1 would not be directly comparable to the content of speech elicited at Time 2. Both L1 and L2 data was collected.

Four tasks were used in total (two for each language), and the design was counterbalanced. Participants who received Version 1 at Time 1 received Version 2 at Time 2, and those who received Version 2 at Time 1, received Version 1 at Time 2. Statistical analyses showed there were no significant differences in fluency with regards to task type; participants who completed Version 1 were not significantly more or less fluent than those who completed Version 2. Participants first completed the Spanish task, and then completed the English task. Data was recorded by the researcher using a Marantz PMD660 recorder hooked up to a Shure SM58 microphone in which participants were speaking. All task materials for speech elicitation tasks can be found in Appendix A.

Participants complete tasks such as whether to hold a graduation even on campus or off-campus) and convince other people of their point of view. Task instructions were in English and each task was accompanied by visual cues (royalty-free stock photography) for clarity. Task materials can be found in Appendix A.

The four scenarios used for speech elicitation tasks, inspired by De Jong (2014), are listed below:

Oral Fluency Version 1:

Task 1: Spanish Task: Town Hall Meeting

Students were asked to imagine they were speaking at the local town hall meeting with regards to their opinion of the location for the new cinema (on campus or at a different location). An arrow indicated their "preferred location" and they were asked to convince the board of reasons for this; some suggestions were given in the prompt so participants had an equal, fair opportunity to think of ideas.

Task 2: English Task: Students' Union Meeting

Students were asked to imagine they were at a students' union meeting where the president is discussing two options for holding a charity fundraising event – at a campus bar, or an off-site location. Students are asked to explain their "preferred location", which is on campus, and told that the off-campus event could be less accessible

Oral Fluency Version 2:

Task 1: Spanish task: Graduation Party

Students were asked to imagine they were planning for their university graduation party, and had were discussing two locations: a campus bar need the university, or a restaurant off-campus that could hold more people.

Students were asked to explain their "preferred location", which is on campus, and told that the off-campus event could be less accessible for students. They were instructed to thank the president for the presentation, explain their preferred option, and justify their choice.

Task 2: English Task: Planning A Party

Students were asked to imagine they were discussing the location of an upcoming party with other students. Two choices were presented – on campus, in a small apartment, or off-campus in a large house. Again, students were presented with a couple reasons for each location (small apartment would be a cozy environment, off-campus house would be difficult to access, etc.) They were told that they knew the neighbours of the big house and would prefer that location, and to thank the current speaker and explain and justify their choice.

An example of an oral fluency task (Version 1, Task 1) is presented in Figure 6 below. The complete tasks with all versions can be found in Appendix A.

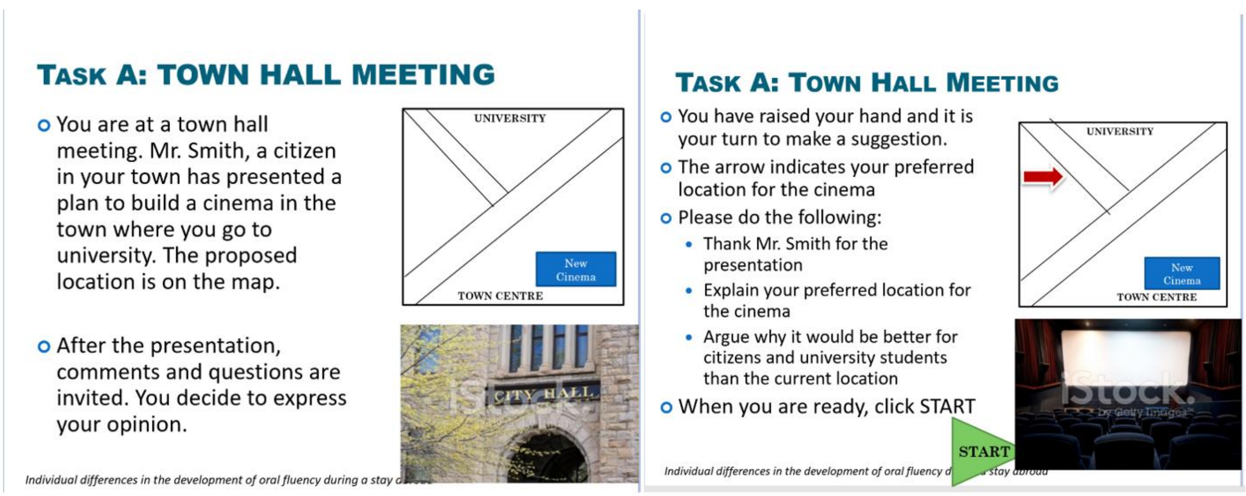


Figure 6 - Speech Elicitation Task 1 Version 1

4.3.2.4 *Procedure*

The instructions were to read the instructions carefully, make a decision on what they would recommend to the audience (depending on the specific prompt), and make their recommendation with a justification.

Before participants began, the researcher took them through PowerPoint slides with the instructions and asked if they had any questions. Participants were shown a photo and short description of a situation in which they had to make a simple decision and justify it. Instructions were given as to the context of the situation, and the student's "role" in the situation. English was chosen as the language of instruction as the students were not fluent bilinguals. Students were instructed that they would be given two tasks – one in which they needed to speak for approximately one minute in Spanish, and the other in which they needed to speak for approximately one minute in English. Students were given time to read the instructions on the screen, and 30 seconds to prepare to speak. They were allowed to jot down notes about the task on a scrap piece of paper if they wished, during the 30 second preparation time. As seen in various studies in the literature (e.g. Chang, 2010; Huesch & Ventura 2017), participants are often given planning time of around 30 seconds to gather their thoughts and familiarize themselves with the topic before speaking. They were also shown a green status bar to indicate 30 seconds of preparation time before the task began and were thus familiar with the status bar. They were instructed to use the status bar as a guideline but continue their final thoughts even if the status bar ran out of time, to encourage fluent speech. They were then asked to speak for approximately one minute, and to click the "Done" button if they finished early and had no more to say. The green status bar was clearly visible across the screen that counted down the time from 60 seconds, so students had an idea of how much they had spoken.

4.3.2.5 *Sample Transcript (L1 and L2)*

A sample transcript of the Spanish (L2) and English (L1) speech elicitation (narrative) tasks for one participant is provided below to help the reader comprehend the language level of the participants, understand how their L1 speech compares to their L2 speech, and see examples of disfluencies. The data for all participants was not transcribed. This example is for illustrative purposes and indicates in *italics* the disfluencies found. These are the disfluencies heard during analysis of the recordings and marked in PRAAT before analysis.

L2 Spanish Task (Pre) Participant 20

"**Ehh** *(filled pause)*…Gracias Sr. Smith por mostrarnos la presentación. (*silent pause).* **Ehh** *(filled pause)*…es importante que **uh** *(filled pause)* tenemos un cine en nuestra ciudad. Pero para mi **uh** *(filled pause)* yo prefiero que el cine va a estar **uhh** *(filled pause)*…**eh***(filled pause)*…no **en..en** *(repetition)*…… **uh** *(filled pause)* en la calle que hemos dicho…pero la calle…**uhh** *(filled pause)*..a lado de la universidad. Creo que es mejor para los estudiantes universitarios **uh** *(filled pause)* si el cine es más cerca de la universidad y si los estudiantes pueden…**uh***(filled pause)* … *(silent pause)*..ver **la… el cine** *(repair)*..…**ehh** *(filled pause)*....**cuando es.…cuando están** *(repair)* en la univerdad porque…**uh** *(filled pause)*.. la parte de la calle **queeee** *(elongation)*..…**uh** *(filled pause)***..…**que**…***(silent pause)* hemos **hab-hablado** (*repetition)* es **más oscuro…más oscura** *(repair)* **queee** *(elongation)*lo que es **eh** *(filled pause)* a lado de la universidad."

L2 English Task (Pre) Participant 20

"Okay, I'd like to say thank you to the President for..**um***(filled pause)* …presenting her ideas **um**. However, I'm in favour of having the charity fundraiser **uh** *(filled pause)* on campus at the campus bar as opposed to an off-campus site like a restaurant for dinner **umm** *(filled pause)* just because **umm** all the students would have easier access to the on-campus bar as opposed to the off-campus dinner site, **um** *(filled pause).* It's a more relaxed environment because it's a place we frequent often as students as at the same time **uh** *(filled pause)* y'know the extra revenue can go back to our campus bar and help something that's local to us as opposed to a more random off-campus site and everyone can have easier access and transportation on-campus as so many more students live on campus these days as opposed to in the past"

4.3.3 *Inhibitory control task- language switching (English and Spanish)*

4.3.3.1 *Task purpose and description*

Participants performed two short speech tasks (different to the fluency tasks) to assess their language switching abilities. Language switching is a measure of inhibitory control (since switching "requires active inhibition or suppression of the stronger competitor language" (Meuter & Allport, 1999). Activation has been shown to be related to proficiency; high activation on bilingual picture-naming tasks was shown in higher proficiency learners. Shifting to the L1 has also been shown to require more time to overcome inhibition (Calabria, 2012; Costa & Santesban, 2004; Costa, Santesban & Ivanova, 2006). Research has also been done in read aloud tasks, showing that previously activated non-target phonemes may affect phonetic processing of the target phonemes (Antoniou et al., 2011). Some research has been done into the cross-linguistic influence in the pronunciation of phonemes after a switch at the beginning, middle, and end of a sentence read aloud, finding that all groups had interference from Spanish, regardless of whether the language they switched into was their L1 or L2 (Bullock & Toribio, 2009). However, what we know so far about the suppression of one language (inhibition) when speaking in the context of prompted or natural speech, (not sentences read aloud), is a topic that has not been thoroughly tested in SLA research. Furthermore, the effects of language switching using neither read aloud tasks nor spontaneous speech has been tested in the SA context. Previous research has primarily investigated fluent bilinguals or immigrants in an L2 speaking context.

In the present study, SA participants completed both story tasks at pre-test and post-test and were asked to switch into one of their languages at a certain time point. The purpose of this task was to investigate additional measure of inhibitory control to examine the relationship between inhibitory control and fluency gains. In most studies, only non-linguistic tests of inhibitory control such as the Simon or Flankers tasks are used. The Stroop task is one exception sometimes used in the literature, in which participants are presented with colour names and colours and are required to suppress one piece of information while communicating the other. A linguistic and lexical inhibitory task related to vocabulary (where the stimuli are words and images, not numbers, shapes, or colours) was chosen as it was anticipated to have a stronger relationship to speaking fluency and performance.

Furthermore, ease of switching (the time it took participants to successfully switch from one language to another) was of interest. “Switching costs” moving from one language to another (L1 to L2, and L2 to L1) may change after a time period with greater than usual exposure to the L2 context. Meuter and Allport (1999), for example, show that L1-dominant bilinguals have more difficulty *switching into* their L1, as it requires more effort to inhibit the L1. In the present study, too, it may become easier for students to switch into their L2 after they have resided in the L2 context for three months. Also, it may become more difficult for them to switch into the L1, if they truly have been constantly inhibiting the L1 during their stay. There is evidence in the literature that bilinguals lose access to their L1 when using their L2 in an L2-dominant environment (Linck, Kroll & Sunderman, 2009b), and this task was created to gain more insight into this process in L2 learners in the SA context.

Piloting was especially crucial for the tasks we designed (language switching and Letter (Phoneme) Decision). Although for the language switching task, the 8 pilot participants were able to switch languages after 7 seconds, when this was performed on 14 participants in Fall 2015, this was not enough time for our participants to speak in the first language, before switching into the second language, rendering the tests unusable. Many participants only uttered a few words or stalled. Thus, the decision was taken to have participants in the second semester switch at 27 seconds (the average half-way point for pilot participants), allowing for enough data to compare L1 with L2. This may have happened despite piloting because pilot participants were not in the same population as our main participants; they were graduate students and researchers, not SA participants. Therefore, only 35 participants (those in the Winter 2016 term) completed the task.

4.3.3.2 *Stimuli and materials*

Both tasks used were in the format of a framed cartoon story, in which an action occurred in each frame, making the story easier to tell by guiding the flow of events. The first task used was the Picnic Story task (Heaton, Longman & London, 1966), which tells the story of a young girl and her brother who go on a picnic their mother has packed for them. Their puppy jumps into the picnic basket and to the children's surprise, eats the food before they can sit down to a picnic. The second task used was *The Suitcase Story* task originally used in (Derwing et al., 2009b), in which a man and a woman bump into each other while turning the corner on the street, dropping their luggage as the fall. Their luggage is similar, and the two swap suitcases but only realize this as the return home with the wrong suitcase. These tasks were chosen due to their simplicity, previous successful use in SLA literature (e.g.O'Brien, 2014), and the ease at which participants could see the course of events in the story. Computers were used to indicate a prompt for switching using a coloured PowerPoint slide with the words English or Spanish that changed colour and the word at the beep.

4.3.3.3 *Procedure*

Before beginning, to encourage fluency in this task participants were given a list of key words they may have wanted to use and their English to Spanish translations (e.g. *hill* = "colina"). This was done since the task was not to measure baseline fluency (as the speech elicitation tasks did) but focused on fluency around a switch. Participants could keep these pieces of paper in front of them as they spoke, to take away the stress of speech planning, and retain the focus of the activity on the language switch, and the ability to switch into another language smoothly. Participants were also given a chance to ask the researcher for any other necessary translations before beginning, but no participants had further questions. The list of vocabulary items that participants were provided with can be found in Appendix B.

Participants performed one task beginning in Spanish and then switching into English, and the other task beginning in English and switching into Spanish. They were asked to speak in one language, and then asked to switch to a second language after 27 seconds (27 seconds was chosen as during pilot test the average total speech time was 54 seconds, so the halfway point to switch at would be 27 seconds). Participants were instructed to continue speaking until they had finished telling the story. They were not told at exactly which point they would have to switch, so that the forced switch came as a surprise and was as natural as possible. The switch was prompted both by a beep, and a screen in front of them changing from displaying the language in capital letters that say ENGLISH (blue screen) to SPANISH (green screen). The language switch was indicated with colour and text so that it was obvious when participants needed to switch.

One language switching task was completed in English, and a second, similar task, in Spanish, so a baseline measure of L1 speech could be collected. Tasks were counterbalanced according to language. Half of the participants had Version 1, with Spanish first, and the others had Version 2, with English first) and no significant differences in speech rate or pausing were found as to participants who completed Version 1 or 2. Those who completed Version 1 at pre-test completed Version 2 at post-test to avoid task habituation. For example, those who did the Suitcase Story from English to Spanish at pre-test, completed the Suitcase story from Spanish to English at post-test.

Participants were recorded doing the language switching tasks using a Marantz recorder and microphone, a in a quiet space in a classroom. Participants were at far ends of the classroom and were not able to hear one another speak; only one student performed the task at a time, although up to 2 other students were in the classroom completing computer-based tasks while wearing noise-cancelling headphones.

Participants had more than enough time to complete their story telling within 60 seconds; only three students exceeded the 60 second time limit, and those who did wrapped up their story once they saw the status bar depleted. Once participants had finished both English and Spanish tasks, the speaking portion of the testing was finished. Participants were instructed to stop speaking the first language in the task as soon as they heard the beep, and to switch to the second language as quickly as possible.

Participant 1 (Version 1)

Pre-test: Picnic-Story for L1-L2 switch (Eng/Sp),

Suitcase story for L2-L1 switch (Sp/Eng)

Post-test: Suitcase story for L1-L2 switch (Eng/Sp)

Picnic-Story for L2-L1 switch (Sp/Eng)

Participant 2 (Version 2)

Pre-test: Suitcase story for L1-L2 switch (Eng/Sp)

Picnic-Story for L2-L1 switch (Sp/Eng)

Post-test: Picnic-Story for L1-L2 switch (Eng/Sp)

Suitcase story for L2-L1 switch (Sp/Eng)

4.3.3.4 *Measures used in analysis*

The main measures taken from the language switching task are those of the basic fluency measures described in section 4.3.2.4. Language Switching files were pre-processed in the same way as the fluency files, as they contained speech. While the speed of the "switch" itself (switching time, switching costs) is to be considered a measure of inhibitory control, fluency data can still be taken from the speech before and after the switch, in analyzing the impact of the switch on L1 and L2 speech. Therefore, files were pre-processed alongside the fluency speech files detailed below.

As the portion of speech in each language is only a maximum of 30 seconds and not long enough to look at disfluency measures in detail, basic speed and breakdown fluency measures will be used (*speech rate*, *articulation rate, MLR, silent pause frequency*, *filled pause frequency)*. In particular, fluency after switching *into* each language will be compared to fluency before the switch, for both L1 and L2, as can be seen in the diagram below.

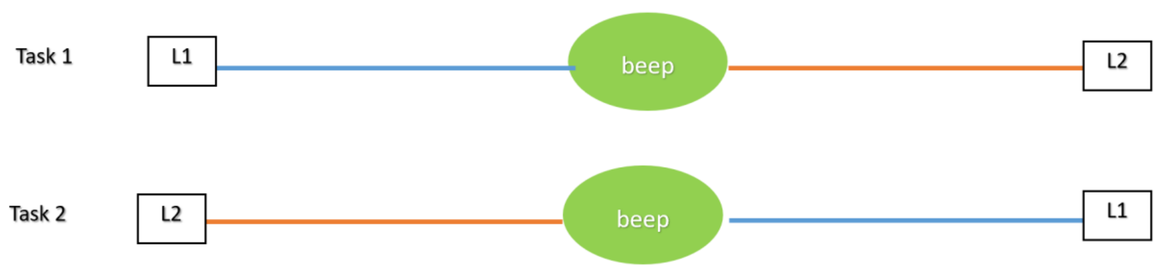


Figure 7 - Language Switching Task Diagram

The effect of the switch on fluency measures will be investigated both when switching into L1 and when switching into L2, since when speaking either language, the other must be suppressed. It is this inhibitory control mechanism and measure of *cognitive fluency* that we are investigating. Fluency *before the switch* in each language is used as a baseline measure of fluency. This baseline is more representative of the task and the forced switch than the fluency measures taken in the *oral fluency* tasks. The speed of switching into Spanish vs. switching into English will also be measured, calculating a Switch Score (the speed at which one switches into Spanish relative to English).

Measures taken from the language switching task were the same as the fluency measures taken for L1 and L2 speech in the speech elicitation tasks. Each measure was taken before and after the switch. For example, participants would have the speed (speech rate measure) of their L1 speech before a switch, L1 after a switch, L2 speech before a switch and L2 after a switch.

Importantly, a novel measure of inhibitory control was taken to show *the time it took a participant to switch after being asked to.* This measure is be called *overflow* and is calculated as the time at which the beep occurred (B) and the time at which the participant *actually started speaking the second language* (S). The *overflow* measure is a measure of language switching ability and will be used when investigating relationships between inhibitory control and fluency variables. In an additional analysis, relationships between language switching results, other inhibitory control results (Simon Test and Letter (Phoneme) Decision tests) and oral fluency results will be investigated. As the beep was planned to occur after a time limit of speaking (after 27 seconds), and thus occur at the same time for everyone, where the beep occurred syntactically of course, could not be controlled for. It is possible that some participants were mid-clause when the beep occurred, while others were between clauses or sentences; the latter may have made the switch more fluent and allowed participants to switch more quickly into the other language. This is a limitation of the present study design.

While fluency data was collected on fluency before and after the switch, a decision was taken to use the Overflow measure as the main measure in analysis, and discard fluency measures, given the short length of speech samples participants produced before and after the task. The measures were originally taken in the "before switch" portion of the speech and "after switch" portion (separate measures), for each language. However, due to the very short length of speech before and after the switch, these measures were later not used, as they did not present a solid picture of fluency before and after a switch. We decided to focus on the time taken to switch itself, using the overflow measure to measure the speed of the switch, looking at the relationship between fluency measures.

The list below shows the measures used in analysis discussed above. Switch Score and Switch Gain were used in preliminary analysis to examine the differences between pre and post-test switching, and Overflow was used as the main measure from this task.

Switch Score:
The speed at which a participant switches into Spanish relative to switching into English
(<Switching time English into Spanish> - <Switching time Spanish into English>)

Switch Gain

The difference in switching time from pre- to post-test (gain)

(Switch Score Post-Test- Switch Score Pre-Test)

Overflow (s)

The time it took a participant to switch after being asked to
(Switch Time) – (Beep Time)

4.3.4 *Inhibitory control task (Simon)*

4.3.4.1 *Task purpose and description*

To investigate *cognitive fluency*, and more specifically individual differences in inhibitory control ability, we used a Simon Task. This task is commonly used in the psycholinguistics literature. Some studies use arrow or shape based variants of the test (e.g. Colby et al., 2017) or even number based versions ((Philipp & Koch, 2009), while others use a Stroop test which uses mismatching colours and colour-words as conflicting sources of information (Linck et al., 2008). See Section 2.6.1 of the literature review for further examples.

4.3.4.2 *Stimuli and materials*

A standard Simon Task was completed, which contained 80 trials of right and left arrows, half of which were congruent and half of which were not. Congruent trials had the arrow on the same side of the screen as the direction it was pointing, while in incongruent trials the arrow appeared on the opposite side of the screen to which it was pointing, as shown in Figure 7.

Twenty each of Right Congruent, Left Congruent, Right Incongruent and Left Incongruent trials were presented. The Simon task contained 9 practice trials. Practice trials were presented to familiarize participants with the task before beginning.

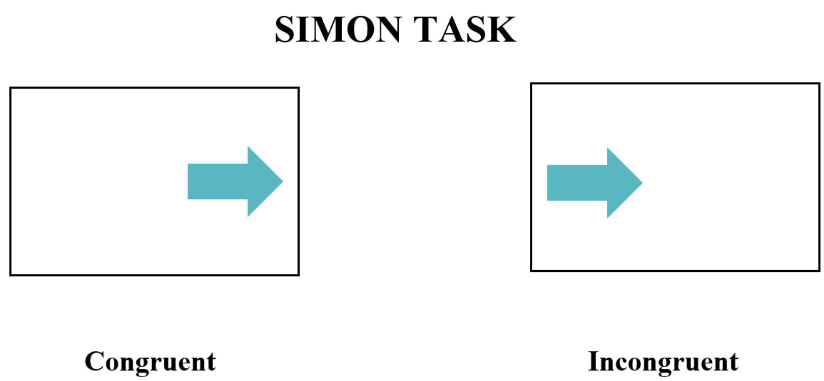


Figure 8 - Simon Task Diagram

4.3.4.3 *Procedure*

The task was administered using DMDX software (Forster & Forster, 2003), and participants were given feedback as to whether they had identified the correct answer. Then, the participants continued to the real test.

The actual test contained 80 items. Participants were asked to push buttons with their fingers “right” or “left” *as quickly and accurately as possible*, not compromising accuracy for speed. Participants were able to follow the instructions and complete the task without any problems. Participants had to click a key labelled “right” or a key labelled “left” to indicate in which direction the arrow was pointing, irrespective of which side of the screen the arrow was pointing. DMDX recorded reaction time in milliseconds and accuracy (incorrect or correct response) data for each trial.

4.3.4.4 *Measures used in analysis*

For the Simon Task, data was screened for values falling outside +/- 2.5 standard deviations from the individual means. Scores were computed from each individual’s reaction time data *on correct trials only*. Incorrect trials were excluded from the analysis, and inhibition scores were computed as the difference between the mean reaction times (RTs) on incongruent trials and congruent trials is used (Linck et al., 2011). A small difference would mean participants have higher inhibitory control ability, as reaction times are shorter, and participants are able to more easily inhibit the information they must suppress. A large difference would mean a lesser inhibitory ability, as there is a greater difference between the way in which one inhibits competing information as compared to congruent, or similar, information. The score, which we refer to as the *ICS Score (Inhibitory Control Simon score)* is one measure of individual differences in inhibitory control used in analysis. Correlation to other inhibitory control data and to oral fluency measures is investigated.

4.3.5 *Inhibitory control task (Letter (Phoneme) Decision)*

4.3.5.1 *Task purpose and description*

The purpose of the third inhibitory control task was to use a more linguistic measure of inhibitory control. While the Simon task is a commonly used measure, it measures reaction times in general (reactions to arrows icons) and is not a task that activates words in a participants' L1 or L2, providing a domain-general, rather than domain-specific measure of inhibition. The present task was inspired by Colomé's (2001) task in which Catalan-dominant Spanish bilingual participants were prompted with a letter, and then a picture, and asked if the letter was in the Catalan word for the picture. Reaction times were measured to compare rejection of a Spanish phoneme in the Catalan word for a particular picture with the rejection of a phoneme absent from the word in both languages. Colomé (2001) found slower reaction times for Spanish phoneme than the absent phoneme, indicating that a participants' less dominant language influenced their processing of the Catalan word.

We wanted to use this test for English learners of Spanish who were not equally dominant in their languages to determine how fast they could reject competing information in their L2 when presented with a similar task. This is a measure of inhibitory control as participants' must suppress linguistic information from one language (English, their L1) while responding based on their L2.

4.3.5.2 *Stimuli and materials*

In this task, participants were shown a consonant, and then an image, and asked if the letter was present in the Spanish word for the item. While participants were only asked to respond “Yes” or “No” as to whether the letter was in the Spanish word, three conditions were present – “Yes, (the letter is in the Spanish word) ”, “No, (the letter is in the English word”, and “No, the letter is in Neither ”.

The figure below shows an example of the three conditions presented. “M” is in the Spanish word for table*,* *mesa*, “T” is in the English word, *table*, and “F” does not appear in either the English or Spanish word.

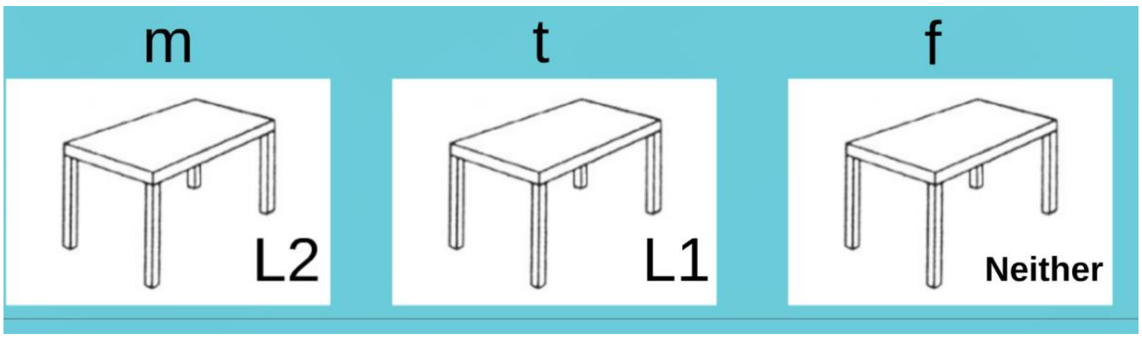


Figure 9 - Picture Selection Task Diagram

Target items were carefully selected so that they contained equal numbers of trials of each condition. Three trial items of each type were also chosen for participants to complete before the task began. This task provides a measure of inhibition as participants must suppress information in their L1 while selecting the correct response using information about their L2. In order to correctly select that yes, “m” is in the Spanish word for table, participants must inhibit the “t” from table in their L1. Target items were chosen to satisfy the following conditions, as defined in Colomé (2001)

- Target words should have 2 or more syllables
- Target words should be non-cognate words
- Target words should start with a consonant

- There should be no more than 3 “YES” answers or 3 “NO” answers in a row

To satisfy the latter condition, a fourth type of category was created “YES, the letter is in both”, where the correct answer was “Yes”, because the letter existed in both L1 and L2.

Items were presented in 3 versions. Each version was a randomized order of 20 blocks of 6 items each - 3 fillers and 3 target words). No significant difference was found between the results of participants completing any of the 3 versions of the test (randomization of blocks).

A list of all items used can be found in Appendix D).

4.3.5.3 *Procedure*

Crucially, for the Letter (Phoneme) Decision test, participants first started with a Lexical Knowledge test, including the target words in the Letter (Phoneme) Decision test. In the Lexical Knowledge test they had to indicate whether they knew the Spanish word for the item in the picture by clicking “Yes” or “No”. The picture was presented first, then the Spanish word for the item in the picture was presented, and then participants were asked “Did you know that word?” and asked to answer honestly, yes or no. Participants went through every word presented in the actual test, so they were familiar with the words and could activate them from their lexicon. With all words already activated in the lexicon in the Lexical Knowledge portion of the task, we could ensure that participants had had equally recent access to the lexical representations of the words appearing in the text. This also ensured that participants associated the correct word with the correct image. Yes and No answers were recorded in case any words needed to be eliminated due to the majority of participants not knowing a word.

This was a modification to the original task in Colomé (2001) because our participants were not fluent bilinguals; they might not know all the words in their L2, Spanish language. Words chosen were high frequency nouns, non-cognates, and words that could easily be depicted in images. Furthermore, images were pictures from a database commonly used in the literature (Snodgrass & Vanderwart, 1980) of standardized pictures. However, images associated to words in this database can sometimes be unclear. We wanted to ensure that all participants understood the images in the same way.

Next, participants completed the main activity, which began with a 9-item trial period with feedback. Participants were then given a chance to ask the researcher questions and clarify any unclear instructions before beginning the main Letter (Phoneme) Decision Task. In the task, they were asked was -*"Is the letter___ in the Spanish word for this item?.* They were instructed to click right for yes, and left for no. To make it clear, "YES" and "NO" letters were taped onto the appropriate keys on the keyboard. Participants were instructed to answer *as accurately and quickly as possible*, not compromising speed for accuracy.

4.3.5.4 *Measures used in analysis*

Pre-processing for Task 4 (Letter (Phoneme) Decision) involved screening the data for reaction times within +/- 2.5 standard deviations from the mean and calculating participants' scores on the tests. These scores (called ICP) were used in all further analysis.

Recall the example of the Letter (Phoneme) Decision Task trials involving an image of a table, and three conditions – the letter *m* appearing (for the Spanish word, *mesa), t* for *table* and *f*- in neither. In this task "Incongruent" trials correspond to rejecting L1 items (the "t" condition in the example and the "Congruent" trials correspond to rejecting letters in neither the L1 or L2 (the "f" condition in the example). These are congruent because we expect that rejecting the "Neither" condition is less cognitively demanding than rejecting a phoneme that occurs in the L1 as in the incongruent trials. A case in which the letter is neither in the L1 nor L2 word for the item would be the easiest to reject because neither would have been likely to have been activated upon viewing the image of a table, because the languages are not in conflict with the decision.

An individual's ICP score was created in a similar manner to the ICS score in the Simon Task. An inhibitory control score was created using reaction times for correct trials only for each individual. The score was created as the difference between the mean reaction times (RTs) on incongruent trials and congruent trials (on correct trials only). Correct trials within +/- 2.5 standard deviations of the mean RTs were used for the score.

Therefore, the calculated score is the RT of the L1 condition minus the baseline, the RT of the Neither condition). Rejection is taken as a measure because it is only there that inhibition would play a role; a "yes" condition in either Spanish would not require inhibiting one of the languages.

4.3.6 *Questionnaire*

4.3.6.1 *Task purpose and description*

Participants completed a questionnaire to gain background information and ascertain participants' exposure to the L2 and their L2 language use during their SA experience. Continual extra input over the length of their stay could affect both utterance fluency development and the efficiency of the underlying cognitive system, key components in this study. The questionnaire pertaining to the participants' experience while abroad was administered at post-test, as post-test data provides a better evaluation of the whole SA experience.

The third research question examined the extent to which L2 experience factors (such as language use, out of classroom contact, and living situation) were related to L2 oral fluency gains. To answer the third research question, we looked at questionnaire data taken at the end of the participants' stay abroad period – self-reported language use data, background information and other conditions of the stay such as living situation and exposure to Spanish during class time The questionnaire was adapted from the Language Contact Profile (LCP), but shortened due to the length of the other tasks in this study, to avoid participant test fatigue. Open-ended questions, and questions to assess participants' perception of their fluency improvement were added at post-test. A copy of the questionnaire can be found in Appendix E.

Research primarily focused on experience factors may go so far as to collect weekly data in language logs (e.g. Martinsen et al, 2010), while some focus on questionnaires that measure a single construct with respect to the study abroad experience, such as intercultural sensitivity development (Anderson, 2004).

Due to concern over participant fatigue after having to complete an intense 45-minute session, we designed the questionnaire to be about 10 minutes long. Although the questionnaires were completed before arriving at the session by most participants (all except those who did not follow the instructions and completed the questionnaire before leaving the session), the questionnaire length was kept short to keep participants motivated and engaged, while leaving room for some open-ended responses. A standard questionnaire such as the Language Contact

Profile (LCP) (Freed et al., 2004) for example, is seventeen questions long, but allows little space for open-ended answers. Following Cigliana and Serrano (2016) and Briggs (2015), we gathered data on language contact in terms of number of hours spent in the L2 classroom. Cigliana and Serrano (2016) had participants self-report changes in their reading, writing, speaking, grammar and pronunciation skills, but for the present study focused on L2 oral fluency, only speaking skills were of interest.

4.3.6.2 *Stimuli and materials*

The list below summarizes the data sections and content within each section of the questionnaire.

Demographic Questions 1-11: Age, Gender, Nationality, Spanish Heritage

Experience with Spanish (Questions 12-13): Years of learning Spanish prior to arrival, age of beginning Spanish instruction

Experience Studying in Spain (Questions 14-17): Hours of Spanish class (and other classes in Spanish) taken while abroad, course/program taken

Experience Living in Spain (Questions 19-25):

Living Situation (apartment, homestay, etc.) and nationality of roommates

Daily language use (for all communication) of L1, L2 and other (% - totally up to 100%)

Language Use at “home” to others in living quarters (the person, which language was spoken and % - e.g. Sofia, Canadian, 100% English)

Extracurricular Activities (number of activities, description)

Spanish Learning Perception - Open ended:

Most effective Spanish learning methods abroad, experience with Catalan language

Motivation and Additional Language Use Questions

Rating scale: Level of comfort speaking Spanish, importance of reaching native speaker level, resistance to speaking English

Open ended: Perception of the effect of SA on Spanish language learning

Three questions about their attitudes and opinions towards Spanish were added. Participants were asked to respond using a scale ranging from 1-10 how comfortable they felt speaking Spanish, how much they felt inspired to reach a goal to speak like a native speaker, and how likely they would be to *resist switching* back into English if the person they were having a conversation with switched to English. This was asked to assess the participants' motivation and willingness to communicate, and willingness to adapt, which may have contributed to L2 oral fluency development.

4.3.6.3 *Procedure*

Most participants completed the background questionnaire before they arrived at the post-test using a Google Forms link, which they were provided by email. Those who had not finished the questionnaire completed it after finishing the other tasks in the study. Data was collected via Google Forms Responses in Google Sheets.

4.3.6.4 *Measures used in analysis*

General background information was coded nominally, while the main portion of the questionnaire was divided into the sections as described in 4.6.3.2. Following Cigliana and Serrano (2016), analysis as a whole group was done to show an overall picture of the SA experience, followed by comments on individual differences in the qualitative analysis.

Variables in questionnaire sections other than demographic information were coded into high and low levels relative to the rest of the group. For example, overall self-reported L2 use was coded into high L2 use >30% vs. low L2 use < 30%). Living situation data was also coded into categories – high, medium, low L2 use - with values relative to the group (e.g. L2 use with the 5 people participants spent most time with). Values in the questionnaire were used in correlation analysis to see the relationship between specific experience factors and speaking fluency. Correlational analysis was also used to look at the relationship between motivational factors (e.g. comfort level of L2 use, desire for native-likeness, resistance to speaking English) and fluency development.

Open ended questions were analyzed qualitatively to understand trends in students' views on positive aspects and challenges they faced during their SA experience. The questions asked participants to indicate aspects that helped them learn Spanish while in Spain, how they personally felt the SA experience affected their Spanish, and their experience with the Catalan language. Answers to open-ended questions were coded using the comment type criteria below:

<ins>Positive:</ins> A positive reflection on the experience, mentions benefits of the stay.

<ins>Mostly Positive:</ins> Mentions benefits of the stay and at least one negative aspect of the stay.

Negative: A generally negative reflection on the experience, mentions disadvantages or points of discontent with the stay abroad experience.

Mostly Negative: Mentions disadvantages of the stay and at least one positive aspect of the stay.

The researcher rated the comments, and comments were then independently rated by another graduate student in SLA. Inter-rater reliability tests (percentage agreement) showed rater agreement to be 90.3%.

4.4 Data Analysis

All analyses were carried out using SPSS v. 25. This sub-section first discusses excluded participants and then the normality of the distribution.

4.4.1 *Pre-processing of speech data*

Prior to analysis, speech data (all fluency recordings for Task 1) was pre-processed in PRAAT. Using PRAAT software, speech files from both the Fluency and Language Switching tasks were cut to ensure dead space before and after speech was eliminated. The files were then de-noised to eliminate any background noise and ensure the file would be readable by the software. The file was then normalized for peak and mean amplitude using PRAAT's functions; this function "maximizes the audibility of each selected sound without distorting it" (Praat Toolkit, 2019)

Using De Jong and Wempe's (2013) *Praat V2 Oral Fluency* script (the standard used in oral fluency literature – see De Jong, 2011), syllables, silent pauses, articulation rate, phonation rate (percentage of time spent speaking), measures were calculated automatically. These markings were then visible in the software with a depiction of the sound waves (see Appendix F for images and description of the PRAAT tool).

Next, three tiers were added to each file in PRAAT – one to mark disfluency measures, one to mark clause boundaries, and one to mark Analysis of Speech Unit (ASU) boundaries. ASU, similar to T-units in writing research, are defined as "single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000). An ASU may contain one or more clauses; a clause contains at a minimum, one verb.

All 104 speech samples for oral fluency tasks, and 208 files for language switching tasks were listened to in PRAAT, and manually marked for clause and ASU boundaries. Then, each file was marked for each disfluency heard in the middle of the disfluency. The disfluencies were marked as point on a tier, according to their type (*f – filled pause, e – elongations, n – repetition and r- repair)*. Using a combination of De Jong and Wempe (2009)'s PRAAT script for fluency measures, and further PRAAT scripts written by the researchers, the data was pulled from PRAAT into SPSS 25.0 to run statistical analysis. Each file went through the described process twice to ensure accuracy. Intra-rater reliability for a small sample (5%) of the data was taken and was 87%.

The data was not transcribed as transcription is not necessary for most speaking fluency measures; the focus of the present study is not semantics, lexical diversity, or linguistic errors made in speech. Several L2 oral fluency studies in the literature focus on temporal measures and thus do not transcribe participant's speech data, (see Mehrang & Rahimpour, 2010; Sadri Mirdamadi & De Jong, 2014; Préfontaine & Kormos, 2015).

4.4.1.1 *Fluency measures used in analysis*

To measure L2 oral fluency, measures of speed, breakdown and repair fluency were taken, following De Jong (2013; 2011). A decision was taken to mark both clauses and ASUs (idea units) in speech samples, as most studies choose one or the other and we wanted to investigate whether looking at ASUs was the same as looking at clauses in terms of pausing data.

In terms of speed fluency, speech rate, articulation rate, phonation-time ratio and MLR were calculated. (For more detail on commonly used measures, see 2.3.2). Fluency measures used are summarized in the table at the end of this section. Regarding breakdown fluency, one of the main objectives of this study was to look at pausing behaviour in terms of both quantity and location (often referred to in the literature as *pause frequency* and *pause distribution)*, following (Segalowitz, 2016). The measures above will be looked at in terms of whether the disfluency falls within or at a clause boundary, and an AS-unit (ASU) boundary. While the T-unit is commonly used for analysis written data, the ASU is the equivalent standard measure for looking at speech behaviour (Foster et al., 2000). It is an *idea unit*, and defined as "a single speaker's utterance consisting of a clause or subclauses unit, and any subordinate clauses associated with either"(Foster et al., 2000, p. 365). Analyzing whether students pause within or between an idea unit shows their ability to speak more fluently; native speakers are more likely to pause between ideas than within an idea unit (Segalowitz, 2016; Tavakoli, 2011).

In terms of repair fluency, some previous research categorizes repairs as *disfluencies* (e.g. (Kormos & Dénes, 2004), while others split this measurement into repairs and restarts (Lyster & Ranta, 1997). Again, the literature diverges with respect to exact measures, but by quantifying restarts and repairs as *disfluencies* researchers attempt to capture a deeper understanding of fluent speech.

Elongations were calculated as an additional measure of disfluency in the present study. Elongations are defined as instances when participants stretched out a word to stall, in effect creating a filled pause. *Disfluency measures* will be referred to in this dissertation as those utterances that impair speech, namely *repetitions, repairs, elongations* and *filled pauses.*

Below is a summary of the measures used to evaluate a participant's fluency from their speech samples. Filled pauses are utterances that are not speech such as "umm, uhh", etc., whereas silent pauses are the absence of speech or disfluencies. A time of silence in speech of at least 2.5 milliseconds. Each measure of pausing for silent pauses and filled pauses was calculated as four separate measures:

1) as a within clause or between clause disfluency
2) within ASU or between ASU disfluency

Fluency Measures Used In The Present Study (following De Jong, 2013; Segalowitz, 2010):

<u>Speed fluency measures:</u>

1) *Speech Rate:*

syllables divided per total time

2) *Articulation Rate:*

syllables divided by phonation time (speaking time not including pauses)

Breakdown fluency measures:

1) *Mean Length of Runs*

Average time spent speaking between pauses

2) *Silent pause rate:*

number of silent pauses **divided by duration**

Silent pause rate (pruned):

number of silent pauses **divided by phonation time (speaking time without pauses)**

3) *Phonation Time Ratio:*

phonation time **divided by total time**
(percentage of time spent speaking, not pausing)

4) *Within clause silent pause rate* (frequency)

5) *Between clause silent pause rate* (frequency)

6) *Within ASU silent pause rate* (frequency)

7) *Between ASU silent pause rate* (frequency)
8) *Mean Pause Duration*:

total length of silent pauses divided by total number of silent pauses

9) *Duration of within clause silent pauses*

10) *Duration of between clause silent pauses*
11) *Duration of within ASU silent pauses*
12) *Duration of between ASU silent pauses*
13) *Filled pauses rate:*
(filled pauses/duration)

Filled pauses pruned:

number of filled pauses **divided by phonation time**

14) *Within clause filled pause rate* (frequency)

15) *Between clause filled pause rate* (frequency)

16) *Within ASU filled pause rate* (frequency)

17) *Between ASU filled pause rate* (frequency)

18) *Elongations rate*

(number of elongations divided by duration)

19) *Total silent pause duration*
20) *Total disfluency rate*
21) *Total pause rate (frequency)* – pauses (silent and filled) and elongations combined

Repair fluency measures:

1) *Repair rate*

Restarts and repairs (divided by duration)

Repair rate (pruned) (Restarts and repairs divided by phonation time)

2) *Repetition rate*

(repetitions/duration)

Repetition rate (pruned)

Repetitions divided by phonation time)

Duration of speech - (overall time in seconds)

All speed, breakdown and repair fluency measures were computed as rates (divided by total duration of speech in seconds to adjust for varying amounts of speaking time in seconds) allowing measures to be compared across participants. Fluency gains (pre-test to post-test differences) were calculated for speed, breakdown and repair measures to use in analysis to answer the research questions. Relative gains were calculated to express the T1 to T2 difference in terms of percentage, rather than the raw gain (the number itself). For example, a participant with a speech rate of 20 at T1 and 25 at T2 would have a raw gain of "5" from pre- to post-test; the value 5/20 would be used as the relative gain value. Relative gains were used as this is an individual differences study, and we wanted to look at participants' individual progress from pre- to post-test.

4.4.1.2 *Pause duration and location*

The length of a silent pause for this study is 0.25 seconds or greater, a widely accepted criterion (De Jong, 2013; 2011).

Pause location is an aspect of pausing that is understudied in SA literature. A pause's location can be determined not only as to whether a pause falls within or between a clause, but whether it falls between or within an ASU, as defined by (Foster et al., 2000). Tavakoli (2011) investigates pausing patterns in 40 NSs and 40 learners of English and finds that on a variety of story speech elicitation tasks, L2 learners distinctively "pause frequently in the middle of clauses rather than at the end" (p.1) than native speakers. Pause location was chosen for analysis in the present study to allow for deeper insight into fluency behaviour.

4.4.1.3 *L1 Adjusted Measures*

As is noted extensively in the literature, L1 fluency and pausing behaviour affects L2 fluency (e.g. De Jong et al, 2012), and L2 fluency should take the L1 into consideration, adjusting measures to create an L2-specific measure of fluency for individuals (see Segalowitz, 2016; Kahng, 2020). An L2-specific measure takes into account the differences that individuals have in their L1 speech; unadjusted measures assume that all speakers can be compared without considering their L1 behaviour.

As an individual's L1 speaking style can influence their L2 speaking style including pausing behaviour(De Jong et al., 2015), we decided to control for individual differences in L1 fluency by calculating an adjusted measure. As there was no significant difference between L1 fluency at pre-test or post-test for most measures in our dataset, pre-test data was chosen to use in calculations for L1 adjusted measures. Pre-test data was chosen as it was taken near the beginning of participants' stay abroad and would therefore not have been influenced by their exposure to their L2 in the study abroad environment.

To calculate an L1-adjusted measure of fluency, the following formula was used, to indicate the proportion of the L2 measure using L1 as a reference:

L1 Adjusted rate= L2 rate /L1 rate

For example, if a participant's L1 Adjusted speech rate was 0.8, it would indicate that their L2 speech rate was 80% of their speech rate in the L1. If a participant had a speech rate of 3.2 in the L1, and 2.6 in the L2, their adjusted rate would be 0.8125. A decision was taken to only use adjusted measures for speed and breakdown values where all participants had a value (no null values). For instance, a participant may have had 0 repairs in their L1, while another participant had 0 elongations, and yet another had 0 repetitions or filled pauses. Therefore, L1-adjusted oral fluency gains scores were calculated where all participants had a value (for example, speech rate and silent pause rate).

To gain a global understanding of disfluency behaviour across participants, a composite measure - *disfluency count* - was calculated, summing disfluency measures. The *disfluency count rate* was chosen to allow us to better compare participants, as not all individuals are disfluent in the same way. For example, two participants who spoke for the same length of time, but Participant X may have had 2 repetitions, 3 filled pauses, and 1 elongation in their speech while Participant Y may have had 5 filled pauses and 1 repetition. These two participants would have the same disfluency count rate, indicating that they were equally disfluent but there were differences in their individual way of being disfluent.

4.4.2 *Summary of all measures used in analysis*

To adequately answer the three research questions presented in section 3.2, many measures were needed. With respect to investigating whether Spanish L2 oral fluency changes after a 3 month stay abroad (RQ#1), and the portion inhibitory control (RQ#2) please refer to 4.4.1.1, which lists all 25 measures. All measures used in the analysis, listed by research question and instrument used can be seen in below.

RQ1: Oral Fluency *Does L2 oral fluency improve after a 3 month stay abroad?*	Speech Elicitation Task; speech production sound files	Please see 4.4.1, which presents all 25 fluency measures by speed, breakdown and repair. These measures were used in SPSS to test the difference between pre and post test, and in correlational analysis with inhibitory control measures and questionnaire measures. Speech rate, silent pause rate, MLR (measures where all participants had a value in L1 and L2 and an adjusted measure could be calculated)
RQ2: Inhibitory Control *Does inhibitory control ability relate to L2 oral fluency gains in the SA context?*	Language Switching Task; speech production sound files	Please see 4.4.1, which presents all 25 fluency measures, and the explanation in 4.3.3 as to why these were later not used for this task. Overflow – The time taken between the “beep” and switching into the target language. Overflow was used in correlational analysis with fluency measures. Overflow was also used to test pre and post-test differences after the stay abroad.

		L1 Adjusted Measures & L1 Adjusted Gains
	Simon Task	ICS - Inhibitory control score based on reaction times for the Simon task (used in correlational analysis with fluency gains) In a preliminary analysis, the difference between ICS scores at pre-test and post-test was also analyzed.
	Letter (Phoneme) Decision Task	ICP - Inhibitory control score The ICP score is used in correlational analysis to investigate the relationships between this task and the Simon Task, language switching task and fluency tasks. In a preliminary analysis, the difference between ICP scores at pre-test and post-test was also analyzed.
RQ3: Experience factors	Post-test questionnaire	Spanish class hours taken Total hours of class taken in Spanish % of native Spanish speakers in living situation % of native English speakers in living situation % of time spent communicating in English % of time spent communicating in Spanish % of time spent communicating in any other language Number of extracurricular activities in which Spanish is used Comfort level speaking Spanish Desire to speak at a native-speaker level Willingness to resist switching to English Experience with Catalan (open ended)

4.4.3 *Normality of the distribution*

As the sample size for this data set are greater than 30, tests for normal distributions can generally be used (Field, 2005). Due to the large number of variables, all normality tests discussed in this section have been placed in Appendix H for ease of reading. With respect to fluency data, the sample size is 49 after removing outliers. Fluency data was normally distributed at T1 and T2. Visually inspecting histograms and boxplots also confirmed histograms showed a normal distribution for most fluency measures. Furthermore, a Shapiro-Wilk test showed fluency data were normally distributed at T1 and T2, with the exception of measures of repair fluency (repetitions, repairs) and elongations. So, parametric tests were used in all subsequent fluency analysis, with non-parametric tests calculated for repair fluency measures. The Shapiro-Wilk test was chosen over the Kolmogorov-Smirnov test due to small sample sizes (Pallant, 2016). L1-Adjusted fluency scores later used in the analysis for measures where this was possible were also checked for normality, and it was found that these measures were normally distributed.

For the Simon test, to check the normality of the distribution the distributions were examined using histograms and box-and-whisker plots before removing outliers. The three participants discussed in the previous section as excluded participants. A Shapiro-Wilk test also confirmed that normality assumptions were not violated ($W=0.936$, $p=0.06$) for inhibitory control scores from the Simon test (ICS scores).

For the Letter (Phoneme) Decision test, however, the responses were not normally distributed. A Shapiro-Wilk test of normality showed the PICS scores were not normally distributed, indicating the need for non-parametric tests when assessing the Letter (Phoneme)

Decision Task (W=0.936, p=.010). With respect to the Letter (Phoneme) Decision test, there were three versions (to randomize blocks for participants). We checked version for the Letter (Phoneme) Decision tests using the Kruskal Wallis test and no significant differences were found at pre or post test between the 3 versions (Mean_Neither = (5.689, p=0.058) Mean_L1 (5.434, p=0.061), Mean_L2 (0.654, p=0.721). All versions of the test were equally valid.

5 Results

Chapter 5 of this dissertation presents the data analysis and statistical analysis of the tasks involved in this study, in order to answer the three research questions. Results concerning each research question are presented in a sub-section of this chapter. 5.1 presents the results of the data analysis to answer RQ1, examining L2 Oral Fluency behaviour from the beginning to the end of the SA period. The section is further split by *speed (5.1.1), breakdown (5.1.2)* and *repair (5.1.3)* fluency. Due to the number of variables and for clarity of reading, descriptive statistics are presented in each relevant sub-section, followed by t-tests reporting fluency gains from pre-test to post-test on all measures. 5.2 presents the descriptive statistics of all three inhibitory control tests in relevant subsections–non-linguistic (*5.2.1* Simon Task), linguistic (*5.2.2* Letter (Phoneme) Decision Task and *5.2.3* Language Switching and correlation data among the tests. Section 5.2.4 presents the relationship between inhibitory control and fluency, showing results that answer RQ2. Section 5.3 goes on to present results pertaining to language use and other questionnaire data and its relationship to fluency gains, thus answering RQ3. An alpha level of 0.05 is used as a significance criterion in all tests.

5.1 L2 Oral Fluency Behaviour

Section 5.1 presents the results of the first research question.

RQ1: Does L2 Spanish oral fluency improve after a 3-month stay abroad in Barcelona?

To answer the first research question, descriptive statistics for pre-test and post-data data were examined, before ANOVA and paired sample *t*-tests were conducted on the various fluency measures. All fluency measures were calculated as ratio measures (over speaking time in seconds) to account for differences in length of speech samples.

The following sections look first at the descriptive statistics of L2 oral fluency behaviour, followed by inferential statistics. The data is presented in three separate tables containing measures of *speed, breakdown and repair* fluency, respectively, for ease of reading. As described in detail in the Methodology chapter, L1 data was collected in this study. As there is a moderate to strong relationship between L1 and L2 fluency or most measures (see Appendix I for correlation tables), we opted to use L1 Adjusted measures throughout the analysis. Measures were adjusted (L1/L2) where possible (where all participants had a value for the variable), thus resulting in proportion of the L2 variable (e.g. speech rate) using L1 as a reference. This was not possible for filled pausing and repair fluency measures as they did not all have a value in the L1, but was possible for speed fluency measures. All measures are ratio measures (divided by duration of speech in seconds) to allow for differences in length of speech samples. As discussed in the Methodology chapter, relative L2 fluency gains (adjusted for the L1 where possible) are used for further analysis.

5.1.1 *Speed fluency*

Descriptive statistics for speed fluency can be seen in Table 3 below. Each measure is shown with its raw value and ratio measure in the L1 and L1 at pre-test and post-test. Raw gains and relative gains (pre-post/pre) are shown below the raw data. As explained in the Methodology chapter, raw gains refer to a post-test to pre-test differences, while relative gains adjust this difference for the pre-test value, which differs per participant. Raw and relative gains are both based on adjusted measures are shown in the tables, though only relative gains are used in all further analysis. As all participants had values for speed fluency measures, they were adjusted for the L1, as discussed in the Methodology chapter. For fluency measures used in analysis (ie: speech rate and articulation rate), L1-adjusted gains information (based on L1-adjusted measures) is also shown in the table.

In terms of speed fluency, descriptive statistics suggest that on average participants spoke more slowly at post-test compared to pre-test. On average, even after adjusting for L1 speech, participants spoke slightly fewer syllables, but did have a higher speech duration, meaning they spoke for a slightly longer time (47.4 seconds in Spanish at pre-test and 50.2 seconds at post-test). Relative measures were calculated as the raw gains divided by the pre-test raw value, that is, (T2-T1)/T1, showing participants' gain relative to their own performance at pre-test. However, contrary to expectations, the mean participants' speech rate *decreased* slightly from 3.2 syllables/sec to 2.9. Articulation rate, too, (speech rate excluding silent pauses) decreased, from 4.3 to 3.9, contrary to expectations that participants would speak *faster* and *longer* in their L2 after a stay abroad in which they theoretically had more L2 exposure, input and output. Relative gains were calculated using the L1 as a baseline. Relative gains, from pre-test to post-test were negative, for both speech rate (-0.03) and articulation rate (-0.04), indicating that after a stay abroad participants had slower speech and articulation rates in the L2, even after taking into consideration participants' L1 speech. As Figure 11 indicates, the standard error bars show variation in the data from the mean, which suggests that our participants did not all exhibit speed fluency performance similar to their peers.

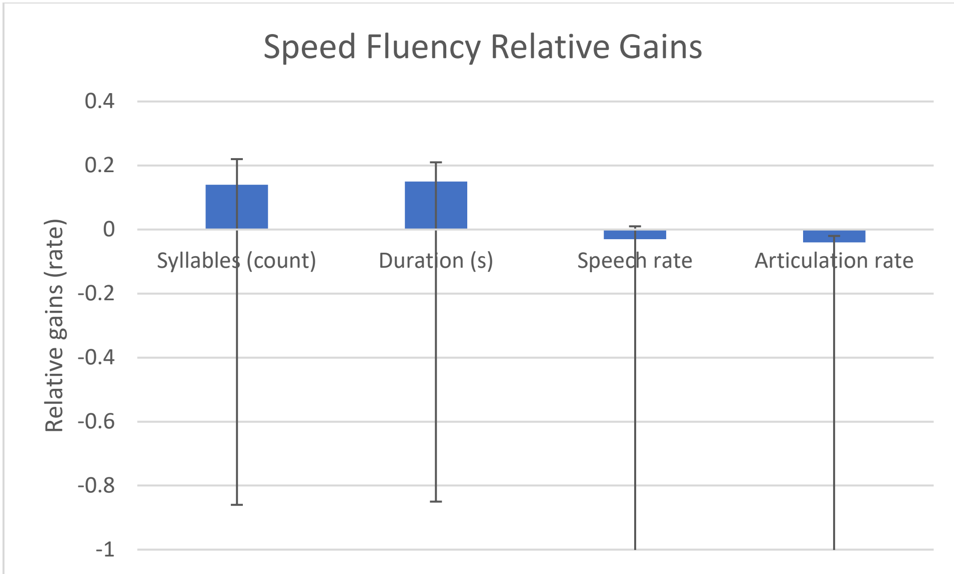


Figure 10 - Speed Fluency Relative Gains (n=49)

As can be seen in Figure 10, while syllables and duration increased on average, relative gains and speech rate and articulation decreased. Standard error refers to the standard deviation of a sampling distribution, and depicts the variability from the mean (Field, 2020) As the data for these relative gains are spread out, we can see there are a wide range of (positive and negative) relative gains our participants had on speed fluency measures. On average, though, there were negative relative gains in speech rate and articulation rate, which implies that participants did not improve their L2 oral fluency on these measures. Relative gains were calculated for L2 measures only, as they were relative to L1 measures used as a baseline, as explained in section 4.4.1.3 of the Methodology chapter.

Table 3 - Speed Fluency (Raw Counts, Rates & L2 Gains)

Syllable count								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	160.3	62.2	82	321	159.3	54.5	71	286
Postest	171.5	47.7	63	277	153.1	49.6	46	209
Gain (Post-Pre)	11.2	-14.5	-19	-44	-6.2	-4.9	-25	-77
L2 Relative Gain (Raw Gain/Pre)	-0.04							
Duration								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	48.3	18.3	19.4	90.1	47.4	17.6	18	89
Postest	46	12.3	13.3	64.8	3	0.6	1.9	4.2
Gain (Post-Pre)	-2.3	-6	-6.1	-25.3	-0.1	-0.1	0.3	-0.6
L2 Relative Gain (Raw Gain/Pre)	0.06							
Speech Rate								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	3.5	1	1.7	4.8	3.1	0.7	1.6	4.8
Postest	3.8	0.5	2.8	4.7	3.0	0.6	1.9	4.2
Gain (Post-Pre)	0.3	-0.5	1.1	-0.1	-0.1	-0.1	0.3	-0.6
L2 Relative Gain (Raw Gain/Pre)	-0.1	0.3	-0.06	0.03				
L1 Adjusted Gain	-0.03	0.29	-0.63	0.59				
	L1				L2			
Articulation Rate	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	4.5	0.7	3.1	5.3	4.2	0.6	2.2	5.3
Postest	4.5	0.4	3.6	5.1	3.9	0.5	3.2	5.1
Gain (Post-Pre)	0	-0.3	0.5	-0.2	-0.3	-0.1	1	-0.2
L2 Relative Gain (Raw Gain/Pre)	-0.05	0.02	-0.3	0.5				
L1 Adjusted Gain	-0.04	0.17	-0.4	0.5				

5.1.1.1 *Speed fluency from pre- to post-test*

To test the first research question and determine whether there was a significant change in speed fluency from pre- to post-test, paired sample t-tests were run, using measures adjusted for L1 speech. No significant difference was found for the length of the speech sample, in either syllables $t(49)=0.186$, $p=.854$ or duration $t(49)=-1.12$, $p=.270$); these measures are relevant to calculate speech rate. Interestingly and unexpectedly, the results suggest participants spoke significantly *slower* at post-test $t(49)=2.305$, $p=.026$ and had a significantly *slower* articulation rate $t(49)=2.55$, $p=.029$. While this is an unexpected result, as speed fluency normally increases after a SA experience, participants may have been producing speech more carefully, attempting make less mistakes, slowing down their speech rate. No other significant differences in speed fluency were found. Results of all speed fluency t-tests can be found in Appendix J. To summarize, of the three speed fluency measures, and one duration measure, only two were significantly different at post-test, but suggested a decrease, rather than the expected improvement. Section 5.1.2 below presents the results of different aspects of breakdown fluency.

5.1.2 *Breakdown fluency*

Descriptive statistics were run on all breakdown fluency measures and are reported below in Table 4. The table presents descriptive information for pre-test data and post-test data (L1, L2, and adjusted measures where applicable) in the first two sections of the chart followed by L2 gains data in the third section. Raw counts are provided for informational purposes (e.g. number of silent pauses) followed by ratio measures (e.g. silent pause rate (number of silent pauses/ duration)), which were used in all analyses. Both frequency and duration measures are reported where applicable. Raw gains on adjusted measures are reported for informational purposes; however, relative gains adjusted for the L1 (where possible to calculate) are used in all analysis. The mean relative gain is shown in the table. See the Methodology chapter for a complete list and definition of measures.

Table 4 - Breakdown Fluency (Raw Counts, Rates and L2 Gains - Silent Pause Measures) n = 49

* L2 Raw Gains: (Post Mean Gain-Pre Mean Gain) *L2 Mean Relative Gains (L1 Adjusted where applicable)

	Silent pauses (count)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	16.78	10.06	4	46	**20.45**	**9.29**	7	44	1.2	0.3
Postest	13.63	5.63	2	26	**20.47**	**7.91**	2	35	1.5	0.4
L2 Raw Gains	0.02									
L2 Mean Rel Gains	0.001									

	Silent pause rate (frequency)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.37	0.1	0.2	0.7	0.43	0.1	0.2	0.7	1.24	0.42
Postest	0.32	0.1	0.2	0.6	0.4	0.1	0.1	0.5	1.30	0.47
L2 Raw Gains	1.22									
L2 Mean Rel Gains	0.21									

	Silent pause rate pruned (frequency)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.5	0.21	0.2	1.0	0.6	0.18	0.2	0.9	1.39	0.64
Postest	0.39	0.13	0.2	0.7	0.55	0.21	0.2	1.0	1.56	0.79
L2 Raw Gains	0.17									
L2 Mean Rel Gains	0.47									

	Mean pause time									
	L1				L2			L2/L1 Adj		
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.37	0.11	0.2	0.7	0.62	0.19	0.4	1.2	1.84	0.89
Postest	0.59	0.15	0.4	1.3	0.59	0.15	0.5	1.2	2.04	1.3
L2 Raw Gains	0.2									
L2 Mean Rel Gains	0.38									

	Silent pause duration (total)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	9.96	8.56	1.6	40.9	13.03	6.93	3.2	34	1.3	0.4
Postest	6.88	3.43	0.9	14.1	12.36	6.73	0.9	39.1	0.9	0.2
L2 Raw Gains	-0.40									
L2 Mean Rel Gains	-0.31									

	Silent pause duration rate									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	9.96	8.56	0.1	0.5	0.27	0.09	0.1	0.5	1.54	0.88
Postest	0.16	0.06	0.1	0.3	0.24	+0.1	0.1	0.5	1.59	0.58
L2 Raw Gains	0.05									
L2 Mean Rel Gains	0.68									

	Mean length of runs (MLR)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*(SD)*
Pretest	0.11	0.06	0.4	0.03	0.15	0.05	0.05	0.3	1.59	0.86
Postest	0.09	0.03	0.03	0.2	0.14	0.06	0.04	0.3	1.82	1.03
L2 Raw Gains	0.23									
L2 Mean Rel Gains	0.72									

	Phonation time ratio									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.79	0.1	0.5	0.9	0.73	0.09	0.4	0.8	0.95	0.18
Postest	0.84	0.06	0.7	0.9	0.76	0.1	0.5	0.9	0.91	0.12
L2 Raw Gains	-0.04									
L2 Mean Rel Gains	-0.01									

	Within clause silent pauses (count)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	13.55	9.74	2	43	16.45	8.19	2	38	2.9	0.4
Postest	10.06	5.29	1	21	16.88	7.08	3	30	1.58	0.9
L2 Raw Gains	0.43									
L2 Mean Rel Gains	0.03									

	Within clause silent pause rate (freq)									
	L1				L2			L2/L1 Adj		
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.29	0.12	0.09	0.06	0.35	0.11	0.06	0.56	1.44	0.92
Postest	0.23	0.01	0.04	0.45	0.33	+0.1	0.14	0.58	1.71	0.93
L2 Raw Gains	0.27									
L2 Mean Rel Gains	0.58									

	Within clause silent pause dur rate									
	L1				L2			L2/L1 Adj		
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	8.16	6..38	0.99	38	10.39	6.64	0.77	38	1.91	1.56
Postest	4.85	2.96	0.23	11.2	4.85	2.96	0.91	34	2.22	1.63
L2 Raw Gains	0.31									
L2 Mean Rel Gains	1.22									

	Between clause silent pauses (count)								
	L1				L2				
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max*	Not adjusted for L1 (not all had values)
Pretest	3.49	2.17	0	11	4.49	3.0	0	14	
Postest	3.98	2.61	0	13	3.8	2.06	0	8	
L2 Raw Gains	-0.3								
L2 Mean Rel Gains	-0.23								

	Between clause silent pause freq								
	L1				L2				
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	0.09	0.05	0	0.2	0.1	0.06	0	0.3	Not adjusted for L1 (not all had values)
Postest	2.03	1.48	0	0.2	0.08	0.04	0	0.2	
L2 Raw Gains	-0.28								
L2 Mean Rel Gains	-0.2								

	Between clause silent pause dur								
	L1				L2				L2/L1 Adj
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	1.8	1.19	0.5	4	2.64	2.17	0	12	Not adjusted for L1 (not all had values)
Postest	2.03	1.48	0	7	2.31	1.58	0	7	
L2 Raw Gains	-0.33								
L2 Mean Rel Gains	-0.1								

	Between clause silent pause dur rate								L2/L1 Adj
	L1				L2				
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M (SD)*
Pretest	0.04	0.03	0	0.1	0.06	0.05	0	0.3	Not adjusted for L1 (not all had values)
Postest	0.05	0.03	0	0.1	1.45	1.75	0	0.89	
L2 Raw Gains	-0.17								
L2 Mean Rel Gains	1.3								

	Within ASU silent pauses (count)								
	L1				L2				L2/L1 Adj
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	13.9	9.74	3	43	17.43	8.57	4	39	1.3
Postest	10.59+	5.15	1	21	17.37+	7.49	2	31	-0.003
L2 Raw Gains	1.6								
L2 Mean Rel Gains	-0.01								

	Within ASU silent pauses (freq)									
	L1				L2				L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.31	0.12	0.1	0.6	0.37	0.11	0.1	0.6	.1.3	-0.003
Postest	0.34	0.1	0.1	0.6	0.25	0.1	0.1	0.5		
L2 Raw Gains	-0.02									
L2 Mean Rel Gains	-0.42									

	Within ASU silent pauses (dur)									
	L1				L2				L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	8.29	7.37	1.2	38.1	10.93	6.87	1.6	31	0.23	0.1
Postest	5.13	2.92	0.3	11.2	10.31	5.98	0.9	34	0.9	0.1
L2 Raw Gains	-0.6									
L2 Mean Rel Gains	-0.06									

	Within ASU silent pauses (freq)									
	L1				L2				L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.31	0.12	0.1	0.6	0.37	0.11	0.1	0.6	1.42	0.79
Postest	0.34	0.1	0.1	0.6	0.25	0.1	0.1	0.5	1.60	0.84
L2 Raw Gains	-0.02									
L2 Mean Rel Gains	-0.42									

	Within ASU silent pauses (dur)									
	L1				L2				L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	8.29	7.37	1.2	38.1	10.93	6.87	1.6	31		
Postest	5.13	2.92	0.3	11.2	10.31	5.98	0.9	34		
L2 Raw Gains	-0.6									
L2 Mean Rel Gains	-0.06									

	Between clause silent pauses (count)								
	L1				L2				L2/L1 Adj
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	3.49	2.17	0	11	4.49	3.0	0	14	Not adjusted for L1 (not all had values)
Postest	3.98	2.61	0	13	3.8	2.06	0	8	
L2 Raw Gains	-0.3								
L2 Mean Rel Gains	-0.23								

	Between clause silent pause freq								
	L1				L2			L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M (SD)*
Pretest	0.09	0.05	0	0.2	0.1	0.06	0	0.3	Not adjusted for L1 (not all had values)
Postest	2.03	1.48	0	0.2	0.08	0.04	0	0.2	
L2 Raw Gains	-0.28								
L2 Mean Rel Gains	-0.2								

	Between clause silent pause dur								
	L1				L2				L2/L1 Adj
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	1.8	1.19	0.5	4	2.64	2.17	0	12	Not adjusted for L1 (not all had values)
Postest	2.03	1.48	0	7	2.31	1.58	0	7	
L2 Raw Gains	-0.33								
L2 Mean Rel Gains	-0.1								

	Between clause silent pause dur rate								L2/L1 Adj
	L1				L2				
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M (SD)*
Pretest	0.04	0.03	0	0.1	0.06	0.05	0	0.3	Not adjusted for L1 (not all had values)
Postest	0.05	0.03	0	0.1	1.45	1.75	0	0.89	
L2 Raw Gains	-0.17								
L2 Mean Rel Gains	1.3								

	Within ASU silent pauses (count)									
	L1				L2				L2/L1 Adj	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	13.9	9.74	3	43	17.43	8.57	4	39	1.3	0.3
Postest	10.59+	5.15	1	21	17.37+	7.49	2	31	-0.003	0.001
L2 Raw Gains	1.6									
L2 Mean Rel Gains	-0.01									

	Within ASU silent pauses (duration rate)									
	L1				L2				L2/L1 (Adj)	
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*
Pretest	0.17	0.11	0.01	0.5	0.23	0.1	0.01	0.6	1.9	0.49
Postest	0.12	0.06	0.01	0.3	0.12	0.06	0.01	0.4	2.06	1.54
L2 Raw Gains	-0.03									
L2 Mean Rel Gains	-0.93									

	Between ASU silent pause (count)								
	L1				L2				L2/L1 (Adj)
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M (SD)*
Pretest	3.14	2.27	0	10	3.51	2.62	0	9	Not adjusted for L1 (not all had values)
Postest	3.45	2.23	0	13	3.31	2.16	0	8	
L2 Raw Gains	0.94								
L2 Mean Rel Gains	0.27								

	Between ASU silent pause freq								
	L1				L2				
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	0.07	0.05	0	0.2	0.08	0.05	0	0.2	Not adjusted for L1 (not all had values)
Postest	0.08	0.05	0	0.2	0.07	0.05	0	0.3	
L2 Raw Gains	-0.1								
L2 Mean Rel Gains	-0.13								

	Between ASU silent pause duration								
	L1				L2				
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	1.67	1.23	0	5.1	2.1	2.05	0	12	Not adjusted for L1 (not all had values)
Postest	1.76	1.33	0	7.5	2.04	1.6	0	6.2	
L2 Raw Gains	-0.06								
L2 Mean Rel Gains	-0.03								

	Between ASU silent pause dur								
	L1				L2				L2/L1 Adj
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	*M (SD)*
Pretest	0.04	0.03	0	0.1	0.05	0.03	0	0.3	Not adjusted for L1 (not all had values)
Postest	0.04	0.03	0	0.1	0.04	0.03	0	0.2	
L2 Raw Gains	-0.1								
L2 Mean Rel Gains	-0.2								

First, this section will look at descriptive statistics of overall breakdown fluency measures, followed by data specific to silent pause location and duration, filled pause location, and elongations. Descriptive statistics on silent pausing data in Table 4 above show that participants, on average, produced approximately the same number of silent pauses at pre and post-test (from 20.45 to 20.47), and this number decreased when examining ratio measures (0.43 to 0.40). However, after using L1 data to adjust for their L1 pausing patterns, the data shows that participants *increased* the number of pauses in their L2 from pre-test to post-test. Participants were hypothesized to decrease the number of silent pauses (a major disfluency) at the end of their stay abroad, one indicator of speaking more fluently. Contrary to expectations, adjusted pause rate (number of pauses/duration) increased slightly from 1.2 to 1.3. When using raw measures, mean silent pause duration (the average length of a pause) decreased from 0.62 to 0.59 milliseconds, but after adjusting for L1 speech, mean pause time was shown to increase. The total time spent silent pausing also increased once adjusted for the L1, pointing to the fact that it may be important to make these adjustments, as they more accurately reflect one's personal speaking style (De Jong, 2014) and may change the results. Silent pause duration as a ratio measure (over duration) is also shown to increase. Phonation rate (time spent speaking excluding silent pauses) decreased slightly from pre- to post-test, after adjusting for the L1, and mean length of runs (MLR) – the average length of a speech stream without (silent) pausing - in the L2 decreased slightly from pre-test to post-test as a raw measure. However, after adjusting for the length of runs L1, we can see that participants' MLR *increased* from pre-test to post-test, indicating that at post-test they were able to speak for longer stretches of time without (silent) pausing. This result is more in line with expectations and implies that participants could speak more fluently (at least one this one breakdown fluency measure) after their stay abroad. As shown in Figure 12, on average, participants had positive gains in L2 silent pause rate, mean pause time, silent pause duration and mean length of runs.

Taken together, the silent pausing data indicates that participants spent *less* time speaking (and more time silent pausing) and had slightly longer pauses at the end of their stay than the beginning. These findings are generally contrary to expectations of a SA experience, but importantly, by the end of their stay abroad participants were, as expected, able to speak for longer stretches of time without pausing (MLR) which could be interpreted as a gain in fluency. The descriptive data also shows the importance of adjusting for the L1 where possible.

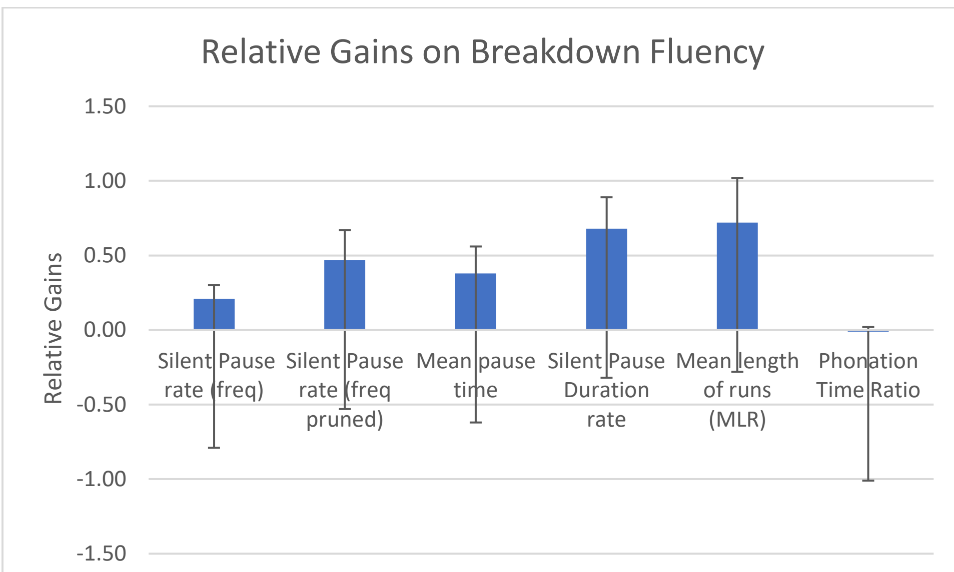


Figure 11 - Relative Gains on Breakdown Fluency – Silent Pauses (n=49)

Silent pauses were examined in terms of their location within or between clause and within and between ASUs, as explained in the Methodology chapter. At first glance, the number of silent pauses appear to remain the same (increasing slightly with an average of 16.45 to 16.88

at post-test). After adjusting for L1 speech, silent pause frequency within clauses increases at post-test; there is a relative gain of 0.58. Within clause silent pause rate, however, shows the silent pause frequency within clauses decreases slightly from 0.35 to 0.33, with a small standard deviation. Descriptive statistics suggest total duration of within clause silent pauses, after adjusting for the L1, also increases, showing a relative gain of 1.22, and between clause silent pauses decrease at post-test. This is contrary to expectations that participants would pause *less* within clauses and more *between* clauses at the end of their stay, thus pausing in a more natural or native-like manner. However, it is possible that participants do not “trade-off” between clause pausing for within-clause pausing. Rather, participants may have a personal speaking style that influences the location of their pauses. Figure 11 shows relative gains in silent pausing measures.

Descriptive statistics for between clause data show that between clause silent pause frequency *decreases* on average at post-test with a relative gain of -0.28. After adjusting for L1, there is a negative mean relative gain between clause silent pauses (-0.20), with a high standard deviation, indicating that pausing location behaviour varied widely among participants. With respect to duration of between clause silent pauses, the data also varied widely, with large standard deviations from the mean. On average, the duration of between clause pauses decreased slightly. After adjusting for L1 pauses, however, there was a positive relative gain (1.3), indicating that the average time spent silent pausing between clauses increased. Importantly, there is a very high standard deviation, indicating that individual participants varied a lot in the amount they paused.

ASUs (units of meaning consisting of clauses) were examined in addition to clause data in this study following Foster, Tonkyn and Wigglesworth’s (2000) definition of the term. More fluent participants would theoretically pause less within a unit of meaning, and more between a unit of meaning. Descriptive statistics for within ASU silent pausing data show that the frequency of within ASU silent pauses decreased at post-test, with a relative gain of -0.42. The duration of within ASU pauses also decreased with a relative gain of -0.06 (-0.93 as a ratio measure over total duration). Between ASU silent pauses decreased slightly at post-test. Participants had a relative gain of -0.01. The duration of between ASU silent pauses also decreased showing a relative gain of -0.03 (-0.20 as a ratio measure over total duration). This may indicate that participants may not have been making a tradeoff of between ASU pauses to within ASUs pauses. Rather, they experienced a decrease in pauses both within and between ASUs. What stands out is that when examining ASU data, participants did have a decrease in pauses at post-test. However, if examining only *clause* data, it appears that participants did not have any negative gains (reduction) in pausing, pointing to the fact that examining units of meaning, rather than only clauses, can help us understand more about pausing behaviour. Figure 12 shows relative gains in silent pause location measures.

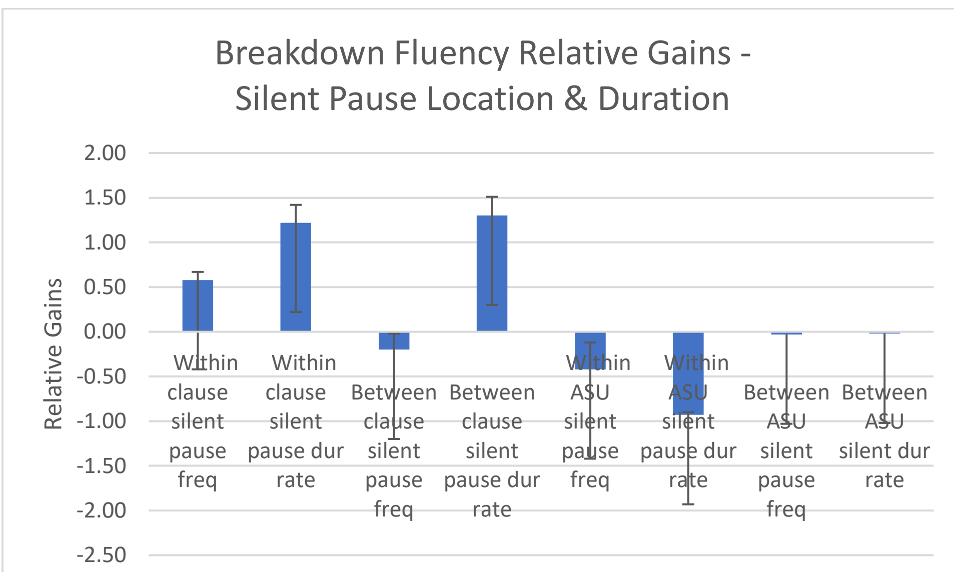


Figure 12 - Breakdown Fluency Relative Gains -Silent Pause Location & Duration (n=49)

Table 5- Breakdown Fluency (Raw Counts, Rates and L2 Gains - Filled Pause Measures n=49

* L2 Raw Gains: Post Mean Gain-Pre Mean Gain *L2 Mean Relative Gains (L1 Adjusted where applicable).

	Filled Pauses (count)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	4.43	3.67	0	14	5.33	4.39	0	20
Postest	3.39	2.96	0	13	6.86	4.84	0	21
L2 Raw Gains	1.53							
L2 Mean Rel Gains	0.29							

	Filled Pauses (freq)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.1	0.06	0	0.3	0.11	0.07	0	0.3
Postest	0.08	0.07	0	0.3	6.86	4.84	0	0.3
L2 Raw Gains	0.02							
L2 Mean Rel Gains	1.8							

	Filled Pauses (freq pruned)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	1.04	1.01	0	4.2	0.16	0.11	0	0.4
Postest	0.77	0.69	0	2.9	0.18	0.12	0	0.5
L2 Raw Gains	0.02							
L2 Mean Rel Gains	1.3							

	Within clause filled pauses (count)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	3.63	3.91	0	14	4.41	3.66	0	15
Postest	4.41	3.66	0	13	5.71	4.18	0	17
L2 Raw Gains	1.3							
L2 Mean Rel Gains	0.29							

	Within clause filled pauses (freq)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.08	0.06	0	0.3	0.09	0.07	0	0.3
Postest	0.06	0.06	0	0.3	0.11	0.07	0	0.3
L2 Raw Gains	0.02							
L2 Mean Rel Gains	0.22							

	Within ASU filled pauses (count)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	3.67	3.92	0	14	4.61	3.99	0	18
Postest	2.63	2.56	0	12	5.84	4.28	0	17
L2 Raw Gains	1.23							
L2 Mean Rel Gains	0.27							

	Within ASU filled pauses (freq)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.08	0.07	0	0.3	0.1	0.07	0	0.3
Postest	0.06	0.06	0	0.2	0.11	0.08	0	0.3
L2 Raw Gains	0.01							
L2 Mean Rel Gains	0.09							

	Between clause filled pauses (count)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.8	0.93	0	4	0.92	1.34	0	6
Postest	0.86	1.14	0	4	1.14	1.37	0	6
L2 Raw Gains	0.22							
L2 Mean Rel Gains	0.002							

	Between clause filled pauses (freq)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.022	0.027	0	0.1	0.020	0.026	0	0.1
Postest	0.022	0.029	0	0.1	0.023	0.027	0	0.1
L2 Raw Gains	0.003							
L2 Mean Rel Gains	0.15							

	Between ASU filled pauses (count)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	3.14	2.27	0	10	3.51	2.62	0	9
Postest	3.45	2.23	0	13	3.31	2.16	0	8
L2 Raw Gains	-0.2							
L2 Mean Rel Gains	-0.06							

	Between ASU filled pauses (freq)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.020	0.03	0	0.1	0.016	0.02	0	0.1
Postest	0.019	0.03	0	0.1	0.021+	0.03	0	0.1
L2 Raw Gains	0.01							
L2 Mean Rel Gains	0.31							

	Elongations (count)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.41	0.64	0	2	0.3	0.6	0	2
Postest	0.2	0.04	0	3	0.33	0.59	0	3
L2 Raw Gains	-0.08							
L2 Mean Rel Gains	-0.2							

	Elongations (freq)							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.010	0.02	0	0.011	0.012	0.02	0	0.1
Postest	0.009	0.02	0	0.010	0.008	0.02	0	0.1
L2 Raw Gains	-0.002							
L2 Mean Rel Gains	-0.25							

	Total pauses (silent+filled) count								
	L1				L2				L2/L1 Adj
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*	
Pretest	21.2	12.81	4	60	25.78	12.51	8	61	1.2
	17.02	7.12	4	31	27.33	11.14	2	50	1.1
L2 Raw Gains	1.52								
L2 Mean Rel Gains	0.06								

	Total pauses (silent+filled) freq							
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.48	0.15	0.2	0.9	0.536	0.14	0.2	0.9
Postest	0.4	0.12	0.1	0.6	0.542	0.16	0.2	0.9
L2 Raw Gains	-0.01							
L2 Mean Rel Gains	0.01							

Filled pauses were also examined in terms of their location within or between clauses, and within and between ASUs. Filled pause duration was not examined. As not all participants had filled pauses or elongations in both their L1 and L2, these disfluencies could not be adjusted for L1 data and non-adjusted ratio measures were used in all further analysis.

Descriptive statistics on filled pauses in Table 5 above show that filled pause frequency (on average) *increased* at post-test. There was a relative gain of 1.8 on filled pause frequency (1.3 when considering the pruned measure excluding silent pauses). When examining the location of filled pauses, we see positive relative gains in within clause frequency (0.29), within ASU frequency (0.09), between clause frequency (0.15) and between ASU frequency (0.31). It was expected that participants would *decrease* filled pause frequency overall after a stay abroad, especially within clauses and ASUs. Descriptive statistics on elongations show that there is a relative gain of =0.25, indicating that participants exhibited less elongations at post-test than pre-test, in line with expectations. Figure 13 shows relative gains in filled pausing measures.

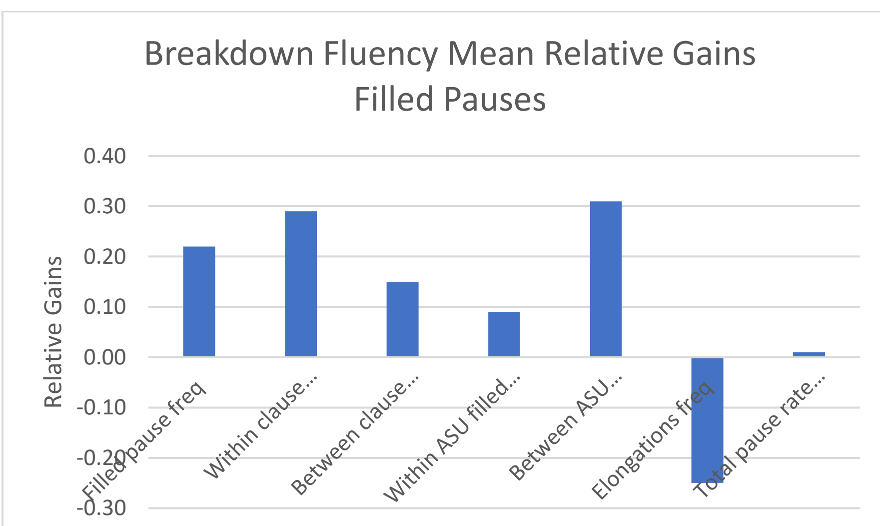


Figure 13 -Breakdown Fluency Relative Gains -Filled Pause Location & Duration (n=49)

A composite measure, *total pause rate (frequency)* was also calculated to consider silent, filled pause and elongation frequencies together, as speakers may not be disfluent in the same way. Adjusting for the L1, (as all participants, of course, paused in some way), we see that total pausing frequency *increased* slightly (relative gain of 0.01) at post-test, contrary to expectations.

Taking silent and filled pausing data together as a rate of total pausing disfluency, we can see that, contrary to expectations, participants paused *more* and for a longer time at post-test than a pre-test. In sum, descriptive statistics on silent pause and filled pause location indicate that participants paused more within and between clauses, but slightly less within ASUs than at pre-test. As standard deviations are high for some measures, it is clear that at least for within-clause and between-clause silent and filled pausing measures, the average may not be representative of the population, showing a wide range of individual differences in pausing patterns.

Elongations, however, decreased at post-test (relative gain of -0.25), indicating that participants elongated their words as a long, filled pause less at post-test than pre-test. This could potentially indicate that they were able to access L2 words more easily and did not have the need to "stall" by dragging out the previous word while they searched for the next word in their mental lexicon.

5.1.2.1 *Breakdown fluency from pre- to post-test*

To examine if there were statistically significant changes in L2 oral fluency from pre- to post-test and answer the first research question, paired sample t-tests were used. Results in this subsection are presented in the same order descriptive statistics were presented above - general breakdown fluency measures, silent pausing data (within and between clause & ASUs), filled pausing data (within and between clause & ASUs), and finally the composite breakdown fluency measure.

Paired sample t-tests were conducted to examine this difference, and found no significant differences in L1-adjusted measures of silent pause rate ($t(48) = -1.057$, $p = .296$), silent pauses per speaking time ($t(48) = -1.117$, $p = .270$), mean pause time($t(48) = -0.854$, $p = 0.397$), silent pause duration, ($t(48) = -0.390$, $p = .698$) mean length of runs ($t(48) = -1.158$, $p = .252$) or phonation time ratio ($t(48) = -1.057$, $p = .296$). The data indicates that participants did not significantly change their (silent) L2 pausing behaviour after a stay abroad.

With respect to silent pause duration and location measures, paired sample t-tests found no significant differences from pre- to post-test in most measures. Due to the large number of measures, non-significant results tables can be found in Appendix J, and significant results are reported within the text. Between-clause silent pause rate was significantly less at post-test than pre-test, after adjusting for the L1 ($t(48) = 2.038$, $p = .047$), indicating that participants paused less between clauses at post-test. The duration of between clause pauses also decreased, but not significantly ($t(48) = 1.517$, $p = .136$). Within ASU pause rate, within ASU duration, between ASU pause rate and between ASU duration did not significantly differ from pre-test to post-test.

When examining filled pause duration measures, paired sample t-tests found no significant differences in filled pause rate, pruned filled pause rate, within clause or between clause pause rates, or between or within ASU pause rates. No significant difference was found in elongations rate from pre-test to post-test.

The most surprising aspect of the data was a paired sample t-test on the composite *total pausing* measure (silent + filled pauses) found significant differences between pre- and post-test pausing measures. Participants significantly *increased* their total pause rate at post-test ($t(48) = -2.035$, $p = .047$), contrary to expectations.

5.1.3 *Repair fluency*

In terms of repair fluency, descriptive statistics indicate that participants showed some changes on repair fluency measures, indicating changes in this aspect of fluency. Descriptive statistics for repair fluency are reported below. As we can see from Table 6, standard deviations indicate a wide variance in the data in both repetition and repair behaviour in L2 speech.

On average, both *repair rate* and *repetition rate* increased slightly from pre- to post-test, contrary to expectations. This also occurred when repair rates were taken over speaking time without silent pauses (pruned rates). Relative gains (using pre-test values as a baseline) in both *repetition rate* and *repair rate* were positive, indicating an increased incidence of repetitions and repairs at post-test. L1-adjusted measures were not possible for repair fluency, as not all participants had a value in both L1 and L2 for these measures.

5.1.3.1 *Repair fluency from pre- to post-test*

To investigate the changes from pre- to post-test with respect to repair fluency, inferential statistics were used. As explained in section 4.4., repair fluency measures were not normally distributed, so non-parametric tests were used. Wilcoxon Signed ranks test showed that there was no significant difference from pre- to post-test for *repair rate* ($z = -0.743$, $p = .458$) or *repetition rate* ($z = -1.385$, $p = .166$). However, when examining *pruned* rates (over speaking time not including pauses), while there was no significant difference in *repair rate (pruned)* (z = -0.857, p= .391), there was a significant difference in *repetition rate (pruned)* (z = -1.903, p= .057). Participants had significantly more repetitions in their speech at post-test than pre-test.

Participants were expected to decrease their rates of disfluency, thus hesitating less in their speech at post-test. However, it is possible that an increased repetition and repair rate at the end of their stay abroad could indicate more awareness of their mistakes and effort to correct them. It should also be noted that the duration of the speech samples is small (approximately one minute), thus many participants did not repair their speech at all in this sort of a sample. The following section considers repairs alongside other disfluencies, in a composite measure.

.

Table 6 - Repair Fluency Descriptive Statistics (n=49)

Repairs								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.4	0.7	0	2	0.3	0.6	0	2
Postest	0	0	0	0.04	0.6	0.9	0	2
Gain (Post-Pre)	-0.4	-0.7	0	-1.96	0.3	0.3	0	0
L2 Relative Gain (Raw Gain/Pre)	0.01							
Not adjusted for L1								
Repair rate (frequency)								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.007	0.014	0.052	0.1	0.007	0.014	0.052	0.1
Postest	0	0	0	0	0.009	0.014	0	0.036
Gain (Post-Pre)	-0.007	-0.014	-0.052	-0.1	0.002	0	-0.052	-0.064
L2 Relative Gain (Raw Gain/Pre)	0.29							
Repair rate (frequency) pruned								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.009	0.019	0	0.064	0	0	0	0
Postest	0	0	0	0.004	0.012	0.018	0	0
Gain (Post-Pre)	-0.009	-0.019	0	-0.06	0.012	0.018	0	0
L2 Relative Gain (Raw Gain/Pre)	0.33							
Repetitions								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.267	0.653	0	3	0.481	0.803	0	2
Postest	0.231	0.574	0	2	0.889	1.289	0	3
Gain (Post-Pre)	-0.036	-0.079	0	-1	0.408	0.486	0	1
L2 Relative Gain (Raw Gain/Pre)	0.85							
Not adjusted for L1								
Repetitions rate (frequency)								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.01	0.012	0	0.06	0.009	0.017	0	0.045
Postest	0.009	0.009	0	0.003	0.016	0.02	0	0.058
Gain (Post-Pre)	-0.001	-0.003	0	-0.057	0.007	0.003	0	0.013
L2 Relative Gain (Raw Gain/Pre)	0.78							
Repetitions rate (frequency) pruned								
	L1				L2			
	M	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Pretest	0.007	0.019	0	0.09	0.009	0.018	0	0
Postest	0	0	0	0	0.022	0.027	0	0
Gain (Post-Pre)	-0.007	-0.019	0	-0.09	0.013	0.009	0	0
L2 Relative Gain (Raw Gain/Pre)	1.44							

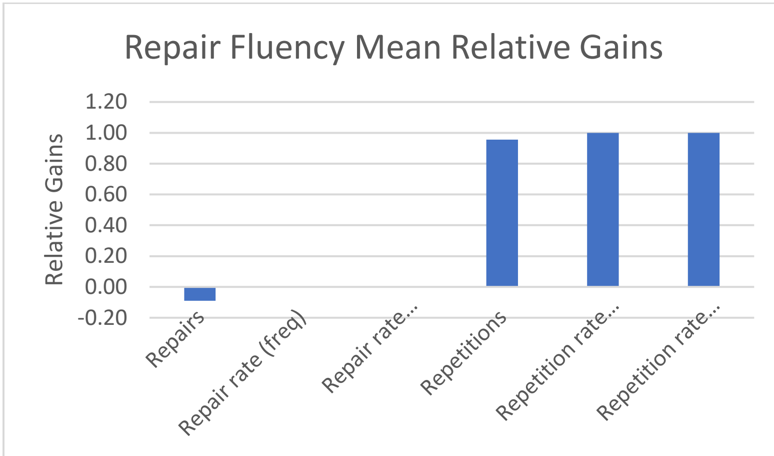


Figure 14 - Repair Fluency Mean Relative Gains (n=49)

While examining the relationship between fluency gains measures themselves was not a main aim of this study, it is important to note some fluency gains measures were highly correlated with each other, as expected. For speed fluency, gains in speech rate and articulation rate were strongly and significantly correlated $r(49) = 0.72$, $p=.01$). For breakdown fluency, silent pause rate gains were strongly negatively correlated with gains in MLR $r(49) = -0.86$, $p=.02$), indicating that those who gained more in length of runs, paused significantly less. Within clause pause rate gains and within ASU pause rate gains were strongly, positively and significantly correlated $r(49) = 0.92$, $p=.01$), as were between clause pause rate gains and between ASU pause rate gains $r(49) = 0.85$, $p=.01$). Repair fluency gains were not significantly correlated with speed or breakdown fluency measures. Correlation tables for fluency gains measures can be found in Appendix J.

5.1.4 *Composite disfluency measure*

Taken separately, changes in disfluency measures show participants, on average, only show an expected gain on some measures of fluency, namely a decrease between clause silent pauses. They also showed an unexpected significant change on two measures – an increase in pruned repair rate, and a decrease in speech rate. As there are individual differences in how speakers speak, even when taking their L1 data into consideration (De Jong, 2016), participants could be "disfluent" in a variety of ways. Some participants have a personal preference for pausing with "umms" and "ahhs", while others tend to remain silent when unsure how to continue a stream of speech. Thus, the decision was taken to calculate a composite measure – *total disfluency rate*, taking into consideration silent and filled pauses, elongations, repairs and repetitions. As explained in the Methodology section, to calculate the *total disfluency rate*, all pausing, filled pause, elongation and repairs disfluency counts were added together and taken as a rate over speech rate. As all participants were disfluent in some way, an L1-adjusted rate was used in all further analysis.

Table 7 - Total Disfluency Measures (n=49)

		Mean	SD	Min	Max
Total disfluency (count)	L1 Pre	26.5	16.49	4	74
	L1 Post	36.55	15.81	5	44
	Gain (Post-Pre)	10.05	-0.68	1	-30
	L2 Pre	32.82	17.33	9	84
	L2 Post	36.55	15.81	2	75
	Gain (Post-Pre)	3.73	-1.52	-7	-9
	Adjusted Relative Gains (L2/L1)	0.1			
Total disfluency rate (freq)	L1 Pre	0.61	0.21	0.2	1.1
	L1 Post	0.49	0.17	0.2	0.9
	Gain (Post-Pre)	0.72	0.23		
	L2 Pre	0.69	0.21	0.3	1.2
	L2 Post			0.2	1.3
	Gain (Post-Pre)	-0.69	-0.21	-0.1	0.1
	Adjusted Relative Gains (L2/L1)	0.75			

As shown in Table 7 above, the raw number of total disfluencies participants exhibited *increased* from pre-test to post-test, contrary to expectations, meaning participants were actually more disfluent on this measure at post-test. The mean total number of disfluencies increased by 3.73, with a relative gain of 0.1. The disfluency rate (ratio measure) increased from 0.69 to 0.72, showing a relative gain of 0.75 after adjusting for the L1. Participants may be disfluent in different ways (e.g. repairs, repetitions, filled or silent pauses, etc), but at post-test they were more disfluent than at pre-test. A paired samples t-test showed that the L1-adjusted total disfluency rate *significantly increased* at the end of the stay t (48) = -2.115, p = .040). This result is counterintuitive, as we would expect that after a stay abroad, participants would have less disfluencies in L2 speech.

5.1.5 *Summary of results*

In summary, the results for the analysis for research question 1 show that there was little change in participants' L2 oral fluency behaviour at the end of their 3-month stay abroad in Barcelona. Speed fluency data shows that participants spoke significantly *slower* at the end of their stay, while breakdown fluency results indicate, surprisingly, that participants exhibited *more* silent and filled pauses at the end of their day. While between clause silent pauses decreased significantly, other measures of silent pause location or duration, filled pause location, or elongations showed no significant change from pre-test to post-test. A composite disfluency measure showed participants, on average, were *less* fluent in their L2 at post-test, contrary to expectations. Out of a total 1 measure of duration, 3 speed fluency measures, 22 breakdown fluency measures, and four repair fluency measures, and 2 composite measures, only 7 measures reached significance, suggesting that participants' L2 oral fluency changed very little over the course of their SA experience.

5.2 Inhibitory Control

The second research question examined the relationship between participants' inhibitory control ability and L2 fluency to investigate whether individual differences on this cognitive skill played a role in L2 oral fluency development.

RQ#2: Does inhibitory control ability relate to L2 oral fluency gains in the SA context?

To answer the second research question, we looked at domain-general inhibitory control tests and linguistic inhibitory control tests, which use linguistic stimuli rather than arrows. As detailed in the Methodology chapter, reaction time scores were calculated for both the Simon Test "ICS Score" and the Letter (Phoneme) Decision Test "ICP Score". Scores were calculated considering congruency of the items (e.g. right and left arrows) with participant responses, as is standard for Simon Test data. Inhibitory control tests were taken at both pre- and post-test to see if the SA experience (and L2 input and L2 use they would theoretically have during the SA) affected inhibitory control ability. As inhibitory control tests were significantly correlated from pre- to post-test, and there was no significant difference between the scores, pre-test data was chosen for use in tests with other measures. L2 oral fluency gains measures (used to answer the research question) have all been adjusted for the L1 where possible.

5.2.1 *Non-linguistic inhibitory control*

The ICS score refers to the reaction time score from the Simon Task (for correct answers, within +/- 2.5 standard deviations), while the ICP is the corresponding score for the Letter (Phoneme) Decision Task. Descriptive statistics in Table 8 below indicate that on average, inhibitory control ability improved slightly (participants had lower reaction times). The standard deviation indicates high variability in the data, and correlational data shows a weak, but significant positive correlation between pre-test and post-test scores (r=.354, p=.01).

Table 8 - Simon Test Results (n=52)

	Mean	SD	Min	Max
Pre-Test ICS Mean (ms)	41.42	53.9	-198.2	167.2
Post-Test ICS	33.76	29.27	-30.3	110.6

Participants had no significant difference in their Simon Task scores from pre-test to post-test ($t(51) = 1.074$, $p = 0.288$) and pre-test results were weakly correlated with post-test results, indicating that better performers on this test at pre-test are slightly more likely to be better performers at post-test.

5.2.2 *Linguistic inhibitory control*

To examine linguistic inhibitory control, we examined the Letter (Phoneme) Decision and Language Switching tasks, the results of which are detailed in the sections below. For the Letter (Phoneme) Decision test, as explained in the Methodology section, a Lexical Knowledge test was done to ensure participants knew the words associated with each picture. There were no words that were overwhelmingly "unknown" by the majority of the participants. The word "boca" was removed from results because from the researchers' observations and participants' comments during the test, the picture was easily confused with the word "labios" and participants were unsure which word the picture referred to. Then, words that were unknown (individually) for each participant were removed before conducting the analysis to ensure the calculated score used words participants knew.

5.2.2.1 *Letter (Phoneme) Decision Task*

Linguistic inhibitory control as measured by the Letter (Phoneme) Decision Task did not change from pre-test to post-test. While inhibitory control ability may be unlikely to change over the course of a few months (Linck et. al, 2011), it is interesting to note that the mean reaction time score on the Letter (Phoneme) Decision test did decrease at post-test, indicating slightly better inhibitory control performance at post-test. However, given the high standard deviations, this must be interpreted with caution. As mentioned in the Data Analysis (4.5) section, Letter

(Phoneme) Decision data was not normally distributed, so a Wilcoxon Sign Rank test was used. No significant difference was found between pre- and post-test ICP scores, (z(51)= -.483, p= .629), as can be seen in Table 9 below. A Spearman Rank correlation test was then conducted to look at the relationship between pre-test and post-test scores and no significant correlation was found, (rs= -0.9, p=.51). Participants who performed well on this test at pre-test were not necessarily likely to perform well on the test at post-test. There was no significant difference between pre-test and post-test ICP scores as shown using a Wilcoxon Rank test z(51) = -.483, $p = .629$. L2 oral fluency gains were not related to ICP scores.

Table 9 - Letter (Phoneme) Decision Test Results (n=52)

	Mean	SD	Min	Max
Pre-Test ICP	6.92	128.9	-415.57	290.08
Post-Test ICP	11.52	120.61	-403.36	282.00

5.2.3 *Language switching as inhibitory control*

As explained in the Methodology section, in the language switching task participants were asked to tell the story in a comic frame in one language, and then switch to the other when prompted. This task provided a linguistic measure of inhibitory control testing learners' skill in switching from L1 English to L2 Spanish and vice-versa. As with other inhibitory control tasks, prior to examining the main research question (the relationship of language switching as a measure of inhibition to L2 oral fluency gains), some preliminary analyses were done to look at whether changes in inhibitory control and switching behaviour from pre- to post-test. This section presents the results of both these preliminary analyses, and the subsequent analysis done on the Overflow measure (the measure chosen to relate to L2 fluency gains).

Two types of *overflow* were measured for participants at pre-test, and post-test: the time between the “beep” and the actual start time of the “switch” (the participant speaking the other language) for switching *into* the L1, and for switching *into* the L2. This was measured to see how long participants took to switch into each language after being prompted to (and therefore suppress their first language), and whether this measure of inhibitory control changed from pre- to post-test. This also allowed us to calculate a measure of switching into L2 relative to switching into L1. Descriptive changes in the “overflow” measures (the time it took in seconds for a participant to switch after being prompted) are detailed below.

Table 10 - Descriptive Statistics for Overflow Measures

	Mean (s)	SD (s)	Min	Max
Spanish:				
Pre Overflow into Spanish	2.66	1.91	.80	9.01
Post Overflow into Spanish	1.98	1.43	0.1	7.02
Overflow into Spanish Relative Gain (Pre-test to Post-test)	0.09	1.16	-1.0	5.0
English:				
Pre Overflow into English	1.64	0.66	.58	2.95
Post Overflow into English	1.6	0.78	0.1	4.01
Overflow into English Relative Gain (Pre-test to Post-test)	0.03	0.91	-0.60	3.20

Mean overflow time was higher when switching into Spanish (Overflow into Spanish) than English (Overflow into English) at both pre- and post-test. After a forced switch from L1 to L2, participants took more time before beginning the target language (switched more slowly) than when doing so from L2 to L1, indicating inhibitory control was lower when switching into the L2. As the participants in this study were low to intermediate proficiency in their L2 and not fluent bilinguals, this was expected. Some research has shown that switching from L2-L1 can be more difficult after at time abroad, but these tests have been done with fluent bilinguals in immersion contexts (Linck et al., 2009), not with low to intermediate proficiency learners. Overflow time for switching *into* Spanish decreased from pre-test to post-test indicating participants improved on average on this inhibitory control measure (switched faster). The corresponding measure for the L1 (Overflow into English) remained approximately the same for from pre- to post-test, indicating that participants' inhibitory control ability when switching into English was not affected after a 3-month stay abroad. Overflow Gain (the difference between post-test and pre-test Overflow, without taking L1 switching as a baseline) was also calculated for switching both Spanish and English. A relative gain was calculated, considering the pre-test value (Relative Gain = (Post-test – Pre-test)/Pre-test) for each participant. The mean values of this score are shown in Table 10 and indicate that on average participants were faster at post-test for switching into English (relative gain 0.03), and substantially faster for Spanish (relative gain 0.09). The standard deviations are high, however, indicating great variability in the data. Some individuals switched more slowly at post-test than pre-test.

Results of paired samples t-tests showed no significant differences in overflow time for switching into English from pre-test to post-test $t(34) = 2.09, p = 0.836 > .05$. Overflow time for switching into Spanish, however, was significantly different from pre-test to post-test, indicating participants switched significantly faster into Spanish at post-test $t(34) = 2.09$, $p = .05$.

The results showed that participants switched significantly faster into English than Spanish ($t(34) = 2.991$, $p=.005$), a difference that did not reach significance at post-test ($t(34) = 1.429$, $p=.162$). This is interesting as it is possible that participants were beginning to gain the ability to switch slightly faster between their languages at post-test, minimizing the difference in switching capabilities from one language to another.

A Switch Gain measure was calculated to assess whether a participant was able to switch more quickly at post-test than pre-test, after having been in the SA environment for three months. Participants' individual changes in overflow behaviour were calculated using a measure called *SwitchScore (<English into Spanish> - <Spanish into English>)*. We used switching into English as a baseline, which is why switching into English is subtracted from switching into Spanish. A baseline of English was used as participants were dominant English speakers and switching into the L1 would have less of a cognitive load than switching to an L2. *SwitchGain* was then obtained by subtracting the pre- from the post-test Switch Scores *(PostSwitchScore - PreSwitchScore)*. Therefore, a lower *Switch Gain* shows improvement at post test (ie: faster switching into Spanish). Table 11 below shows the descriptive statistics and results of paired-sample t-tests.

Table 11 - Descriptive Statistics & Switch Score & Switch Gain Measures (n=35)

	Mean	SD	Min	Max
Pre SwitchScore (s)	1.02	2.02	-1.56	7.69
Post SwitchScore (s)	0.38	1.55	-2.75	6.06
Switch Gain (s)	-0.65	2.24	-6.55	4.75

Descriptive statistics for SwitchScore and SwitchGain measures indicate a lower mean SwitchScore at post-test. However, the standard deviation shows large variability in the sample, indicating great individual differences on this measure. As the standard deviations were large, a non-parametric Wilcoxon Signed Ranks test was used as an alternative to the paired sample t-test. There was no statistically significant difference between Switch Scores at pre-test and post-test, indicating no difference in the speed at which participants switched into Spanish vs. English ($z(34) = -1.523$, $p = .128$). As can be seen from the graphs below, some participants improved (switched faster at post-test) while other participants did not, or in some cases switched *slower* at post-test. Figure 15 illustrates changes in SwitchScore from pre to post test for each individual participant.

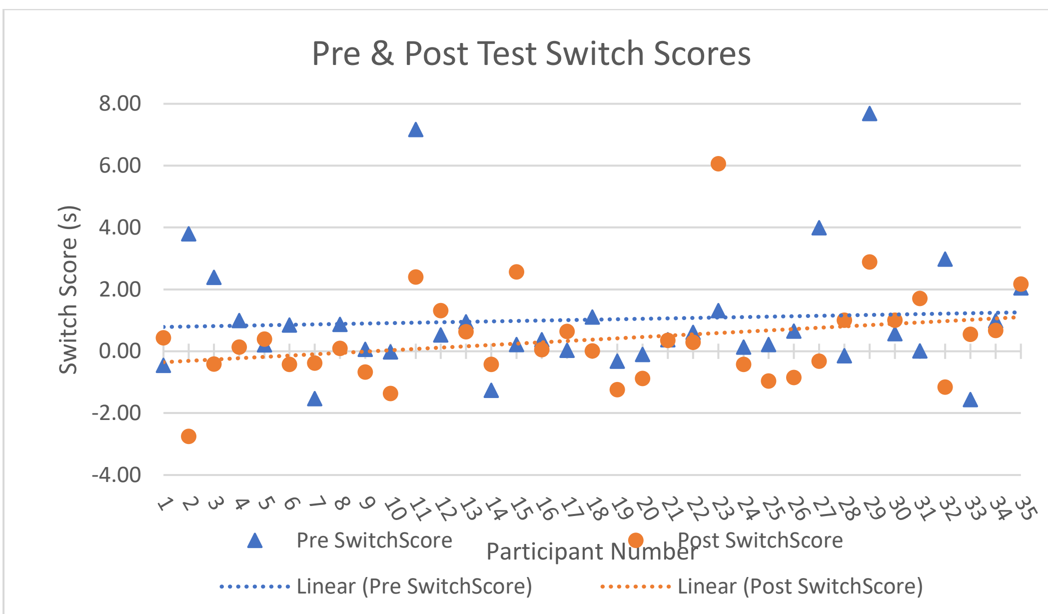


Figure 15 - Pre & Post Test Switch Scores

As we can see from Figure 15, there are large individual differences in pre and post Switch Scores. Some participants switch into Spanish at approximately the same speed at post-test. Others have a difference of up to 6 seconds. Switching from the L1 to the L2 was more cognitively demanding for some participants, who stalled or paused for several seconds before being able to continue and start speaking the L2. The line of best fit shows that the majority of participants were able to switch to the L2 in less than one second, and slightly faster at post-test.

Figure 16 below illustrates the individual differences in the Switch Gain score Post Switch Score minus Pre Switch Score). Participants varied greatly in their switching speed at the end of their stay.

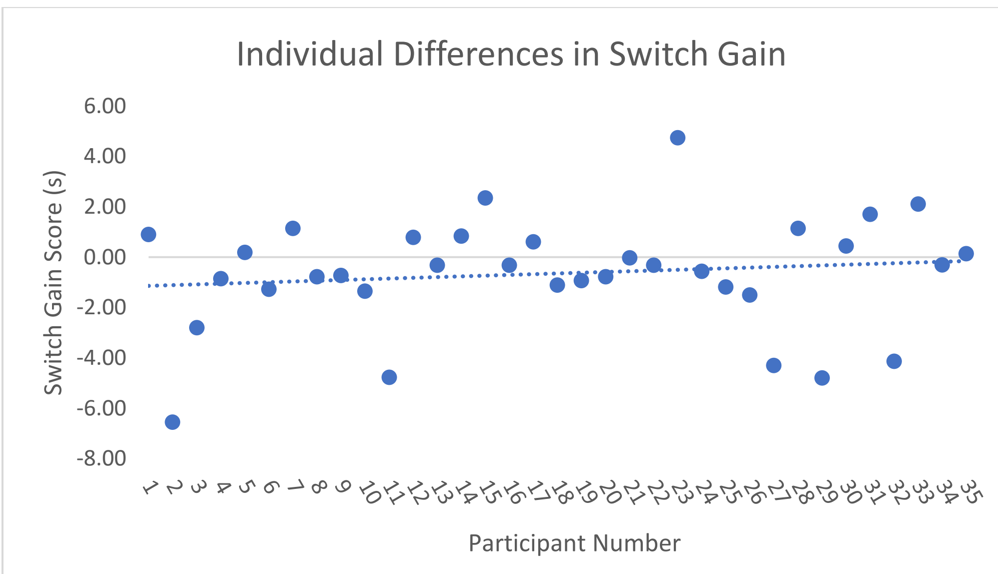


Figure 16 - Individual Differences in Switch Gain

ICS, PICS and Overflow Gains measures were also not significantly correlated, indicating that those who performed better on one type of inhibitory control test did not necessarily perform better on another type.

5.2.4 *Inhibitory control and fluency gains*

To examine the relationship between inhibitory control and overall fluency gains, Pearson correlations were run for speed, breakdown and repair fluency gains measures. The tables in the subsections that follow show correlations between inhibitory control measures and fluency gains. As the PICS score was not normally distributed, non-parametric correlations were used for this measure.

5.2.4.1 *IC and Speed Fluency Gains*

Pearson correlation tests show no significant correlations between the ICS score (Simon test score) and speed fluency gains. Those who performed better on the Simon test (and therefore exhibited more inhibitory control) did not necessarily experience more speed fluency gains. Spearman rank non-parametric tests confirm that the PICS score was also not significantly correlated to speed fluency gains.

Table 12 – Pearson Correlations of Inhibitory Control Measures and Speed Fluency Gains (n=49)

		1	2	3	4	5	6	7	8
1	ICSPre								
2	Syllable Gain	.08							
3	Duration Gain	.05	.86**						
4	Speech Rate Gain	.02	.43**	.11					
5	Articulation Rate Gain	-.02	.46**	.27	.72**				
6	SwitchGain	.07	.08	.09	.15	-.21			
7	OverflowSPGain	.1	.15	.12	.23	-.21	.92**		
8	OverflowENGain	.05	.11	.05	.08	.03	-.44**	-.05	

*Correlation significant at the 0.05 level ** Correlation significant at the 0.01 level

Table 13 – Spearman Rank correlations of PICS Inhibitory Control Measure and Speed Fluency Gains (n=49)

		1	2	3	4	5	6	7	8
1	PICS								
2	Syllable Gain	0.13							
3	Duration Gain	0.13	.84**						
4	Speech Rate Gain	0.01	.44**	0.10					
5	Articulation Rate Gain	0.08	.52**	0.26	.76**				
6	SwitchGain	0.19	-0.07	-0.06	0.10	-0.05			
7	OverflowSPGain	0.12	0.13	0.07	0.27	0.01	.86**		
8	OverflowENGain	0.02	0.28	0.27	0.12	0.10	-.45**	-0.05	

*Correlation significant at the 0.05 level   ** Correlation significant at the 0.01 level

To test the influence of inhibitory control variables on the fluency gain of speech rate, a linear regression was run (F=0.663, R^2=0.077, p=.62). The regression analysis indicated that only 7.7% of speech rate gain could be explained by individuals' inhibitory control test scores, and this result did not reach statistical significance.

5.2.4.2 *IC and Breakdown Fluency Gains*

To examine the relationship between breakdown fluency relative gains and inhibitory control measures, Pearson Correlations were conducted. Spearman Rank correlations were also conducted as the PICS measures (from the Letter (Phoneme) Decision task) was not normally distributed. Due to the large number of measures, the correlations tables below show breakdown fluency measures in three groups - general measures (Tables 14 and 15), silent pause location and duration measures (Tables 16 and 17) and filled pause location measures (Tables 18 and 19).

No significant correlations were found between ICS and breakdown relative fluency gains measures. However, there was a significant medium negative correlation between the PICS inhibitory control score and between ASU filled pauses *rs(49)*= -0.47, *p*= .03. Those who had a higher score on the Letter (Phoneme) Decision inhibitory control test had a lower gain in between ASU filled pauses (thus indicating speech with less of this type of disfluency at post-test). Surprisingly, no other measure of breakdown fluency was found to be significantly correlated to ICS, PICS or Overflow Gains measures.

Multiple regression analysis showed that the inhibitory controls measures (ICS, PICS and Overflow into Spanish) were not significant predictors of breakdown fluency measures. Inhibitory control test scores explained only 11.9% (F=1.09, R^2=0.011, p=.39) of the variance in pause rate gain, but this result did not reach statistical significance.

Table 14 - Pearson Correlations of Inhibitory Control Measures and Breakdown Fluency Gains (n=49)

	ICS	
	r	p-value
Phonation rate gain	.04	.08
MLR gain	-.15	.32
Silent pause rate gain	-.07	.64
Silent pause rate pruned gain	-.12	.04
Mean pause time gain	-.26	.07
Silent pause duration gain	-.03	.85
SwitchGain	.07	.68
OverflowSPGain	.10	.55
OverflowENGain	.05	.78

Table 15 - Spearman Rank Correlations of Inhibitory Control Measures and General Breakdown Fluency Gains (n=49)

	PICS	
	rs	*p-value*
Phonation Rate gain	.07	0.65
MLR Gain	-.13	.37
Silent Pause Rate gain	-.18	.21
Silent Pause Rate Pruned gain	-.15	.30
Mean pause time gain	-.06	.71
Silent Pause Duration gain	.03	.84
SwitchGain	.19	.29
OverflowSPGain	.12	.49
OverflowENGain	.02	.09

Table 16 - Pearson Correlations of Inhibitory Control Measures and Silent Pause Breakdown Fluency Gains (n=49)

	ICS	
	r	p-value
Within clause silent pause gain	.06	.67
Between clause silent pause gain	-.28	.07
Within ASU silent pause gain	.05	.72
Between ASU silent pause gain	.04	.81
Duration of within clause pauses gain	-.03	.83
Duration of between clause pauses gain	-.14	.38
Duration of within ASU pauses gain	-.07	.65
Duration of between clause pauses gain	-.26	.09
SwitchGain	.07	.68
OverflowSPGain	.10	.55
OverflowENGain	.05	.78

Table 17 - Spearman Rank Correlations of Inhibitory Control Measures and Silent Pause Breakdown Fluency Gains (n=49)

	PICS	
	rs	*p-value*
Within clause silent pause gain	-.09	.54
Between clause silent pause gain	.17	.28
Within ASU silent pause gain	.06	.69
Between ASU silent pause gain	-.15	.34
Duration of within clause pauses gain	-.04	.79
Duration of between clause pauses gain	.17	.28
Duration of within ASU pauses gain	.08	.60
Duration of between clause pauses gain	-.17	.28
SwitchGain	.19	.29
OverflowSPGain	.12	.49
OverflowENGain	.02	.90

Table 18 - Pearson Correlations of Inhibitory Control Measures and Filled Pause Breakdown Fluency Gains (n=49)

	ICS	
	r	*p-value*
Filled pause rate gain	.09	.55
Filled pause rate gain pruned	.08	.60
Within clause filled pause rate gain	.12	.44
Between clause filled pause rate gain	-.11	.62
Within ASU filled pause rate gain	.18	.25
Between ASU filled pause rate gain	-.61**	0.01
	.	
Elongations Rate Gain	.50*	.05
SwitchGain	.15	.40
OverflowSPGain	.17	.32
OverflowENGain	.01	.97
Total pause rate gain	-0.05	0.74

Table 19 - Spearman Rank Correlations of Inhibitory Control Measures and Filled Pause Breakdown Fluency Gains (n=49)

	PICS	
PICS	*rs*	*p-value*
Filled pause rate gain	-.04	.8
Filled pause rate gain pruned	-.07	.65
Within clause filled pause rate gain	-.09	.59
Between clause filled pause rate gain	-.05	.81
Within ASU filled pause rate gain	.1	.52
Between ASU filled pause rate gain	-.47*	.03
Elongations Rate gain	.54*	.03
SwitchGain	.19	.29
OverflowSPGain	.12	.49
OverflowENGain	.02	.9
Total pause rate gain	03	.98

5.2.4.3 *IC and Repair Fluency and Total Disfluency Gains*

When examining the relationship between *repair fluency* gains and inhibitory control, we can see that there are also no significant correlations to inhibitory control measures. Those who perform better on inhibitory control tests (PICS and ICS) do not necessarily experience positive gains in fluency (negative gains in disfluencies). Multiple regression analysis also showed no significant predictors of repair fluency gains.

With respect to total disfluency, inhibitory control test scores could only explain 4.1% of the variance in total disfluency gains for our participants, and this number did not reach statistical significance ($F=0.468$, $R^2=0.041$, $p=.71$), indicating that IC test scores are not a good predictor of speech disfluency. The *total disfluency* measure (which was created, as explained in the Methodology section, by adding the count of all disfluencies and taking it as a rate) can be seen in the table below. When examining the relationship between *total disfluency* gains and inhibitory control, we see that there are no significant correlations to inhibitory control measures.

Table 20 - Pearson Correlations of Inhibitory Control Measures and Repair Fluency Gains (n=49)

	ICS	
	r	*p-value*
Total Disfluency gain	0.01	0.97
Repair Rate gain	0.16	0.58
Repetition Rate gain	-0.10	0.66
SwitchGain	0.07	0.68
OverflowSPGain	0.10	0.55
OverflowENGain	0.05	0.78

Table 21 -Spearman rank correlations of PICS Inhibitory Control Measure and Repair Fluency Gains (n=49)

	PICS	
	r	*p-value*
Total Disfluency gain	0.15	0.30
Repair Rate gain	0.21	0.46
Repetition Rate gain	0.12	0.59
SwitchGain	0.19	0.29
OverflowSPGain	0.12	0.49
OverflowENGain	0.02	0.90

5.2.5 *Summary of results*

Overall, the results of this study pertaining to the second research question suggest that performance on inhibitory control tests (Simon Task, Letter (Phoneme) Decision Task, and Language Switching Overflow scores) does not relate to fluency gains after a three month stay abroad. With respect to language switching examined in the preliminary analysis, as expected for low to intermediate proficiency L2 learners, participants switched faster into English than they did into Spanish. Overflow into Spanish was not significantly related to inhibitory control measures. What is interesting, though, is that participants were able to switch significantly faster into Spanish at the end of their stay than the beginning, which may be one indicator of increased inhibitory control (though this was not true for other inhibitory control measures). With respect to the relationship with inhibitory control and fluency gains, decrease in one disfluency measure – between ASU filled pauses – is significantly related to performance on the PICS test, indicating that participants with higher performance on linguistic inhibitory control tests are likely to experience less disfluencies (a negative gain) in terms of filled pauses between ASUs. Inhibitory control ability as measured by linguistic and non-linguistic measures, could not significantly explain the variation in fluency gains scores.

5.3 Language use and study abroad experience

The sub-sections below look at the relationship between fluency gains and experience factors (self-reported language use, classroom instruction factors, extra-curricular activities, and other language use factors). A subset of participants who had high L2 language use throughout their stay are also analyzed, as it is possible that their higher level of interaction in the L2 could affect their L2 fluency gains. Finally, qualitative information from participant questionnaire data is provided to gain more insight on language use during the stay abroad.

5.3.1 *Self-reported language use percentage*

Self-reported language use percentage was examined as a measure of language use. Other language contact factors more indirectly related to language use (such as living situation) are discussed in the sub-sections that follow. On the questionnaire, participants were asked to estimate their total percentage of time spent communicating (speaking, sending text messages, etc.) in their L2 (Spanish), L1 (English) and any other languages, and estimate this as a percentage. Descriptive statistics were run on this measure to examine the distribution of the data.

Table 22 - Self Reported L2 Use

	% L2 Use
Range (Min-Max)	0-75
Mean	22.4
SD	15.9
Median	20

As can be seen in Table 22, participants an average of 22.4% of the time speaking Spanish. Most participants spoke English the rest of the time, with a select few speaking another language with their families up to 10% of the time. However, the mean of Spanish speaking time does not represent the data well, as the standard deviation is nearly as large as the mean. Only 6 (of 52) participants reported speaking Spanish more than 30% of the time.

5.3.1.1 *Self-reported language use percentage and fluency gains*

Pearson correlation analyses were conducted to examine the relationship between self-reported language use percentage and speed, breakdown and repair fluency gains measures. L1-adjusted gains were always used in analysis for measures where adjustment was possible. With respect to speed fluency, no significant correlations were found between self-reported pre or post-test language use percentage and any of the speed fluency gains measures (speech rate and articulation rate), indicating that there is no relationship between this language use measure and speed fluency. With respect to breakdown fluency, four significant relationships were found. Phonation time ratio (the time spent speaking without pausing) was weakly, but significantly positively correlated with self-reported language use, suggesting those who reported using the L2 more could speak longer without pausing (r= 0.30, p=.037). A gain in between ASU silent pause rate was also weakly but positively correlated with post-test self-reported language use (r= 0.43, p=.005), implying those who reported speaking the L2 more may have had more pauses between ASUs, potentially indicating more fluent pausing behaviour, as pauses were between meaning units and not mid-units. Interestingly, there was a weak, but significant positive correlation between pre-test self reported language use and gains in the duration of between clause silent pauses (r= 0.31, p=.047), but a weak but significant negative correlation between gains in the duration of within ASU silent pauses (r= -0.28, p=.046). These results point to the possibility that participants who used the L2 more had longer between clause silent pauses, but shorter within ASU pauses. The duration of within clause silent pauses also showed a negative correlation at with self-reported language use that was approaching significance (r= -0.25, p=.055), suggesting that those who use the L2 more while abroad may have shorter silent pauses within clauses, and therefore be somewhat more fluent with respect to breakdown fluency compared to their peers who use the L2 less. Correlation tables for self-reported language use results can be found in Appendix K.

Together, these results show that those who self-reported speaking the L2 more experienced longer between clause silent pauses, shorter within clause pauses and shorter within ASU silent pauses. It is possible that there was a trade-off between pausing between clauses and within clauses, since participants who spoke more Spanish spent more time pausing between clauses and less within clauses. However, these results must be interpreted with caution because, as we know, individuals have differences in their personal speaking styles (De Jong, 2013), so less of one type of pause and more of another on average may not necessarily indicate a trade-off. Nonetheless, the relationship does suggest that those who used the L2 more while abroad were able to exhibit less disfluencies in pausing behaviour, especially with respect to duration. Of a total 22 breakdown fluency gains measures, only the 3 reported above had a significant relationship.

No significant relationship was found between any of the four repair fluency gains measures and self-reported language use percentage, conveying the idea that the rate at which participants repeated (stalled) in their speech or self-corrected their speak was not related to the amount of Spanish spoken during the stay abroad.

To further explore the impact of self-reported language use percentage on the variables for which there was a significant correlation, a regression analysis was run. Self-reported language use was used in all analysis as stated at the beginning of this section. The regression analysis suggests that post-test language use explained 9% of the variance in phonation rate gains (F=4.63, R^2 = 0.089, p=.04).

A regression analysis on between ASU silent pause rate found that post-test self-reported language use explained 8.4% of the variance in between ASU silent pause rate gains. As the regression is significant (F=8.892, R^2 = 0.084, p=.05), post-test self-reported language use % can be interpreted as a good predictor of between ASU silent pause rate. A variable with a significant F value can be interpreted as a good predictor of the independent variable, even if there is one variable in the model (Field, 2005). Self-reported language use percentage also explained 18% of the variance in duration of between clause silent pause rate gain (F=3.676, R^2 = 0.18, p=.05), and explained 1.6% of the variance in between ASU silent pause rate gains, but self-reported language use percentage was not a good predictor of the latter (F=0.762, R^2 = p=.39).

Fluency gains for a small subset of "high language use" participants who self-reported communicating in Spanish more than 30% of the time (speaking/texting/writing) were also examined, as we thought perhaps those who claimed to speak more Spanish would experience higher gains. As mentioned at the beginning of this chapter, participants' average language use was 26.1%, so even though 30% may seem low for a study abroad context, these participants were consistently above their peers' average Spanish speaking time. There were 19 high language users and 30 low language users. T-tests conducted between these two groups indicated there was no significant difference on any measures of fluency between the two groups, although there was trend towards significance on one measure – between ASU silent pausing measures $t(19,30) = 1.87$, $p=.07$. This result suggests participants with higher language use may potentially (silent) pause less between ASUs than their lower language use peer group. The result implies being slightly more fluent, albeit only on one (of 22) measures of breakdown fluency; no difference was found in speed or repair fluency gains. Results tables for self-reported language use can be found in Appendix K.

5.3.2 *Classroom instruction hours and fluency gains*

Another SA experience factor that potentially affects language contact is one's exposure to and use of the L2 in the second language classroom. In particular, exposure to Spanish in the classroom was examined through a question on the questionnaire asking participants to state the number hours they spent in Spanish class (including other university classes taught in Spanish). Though taking Spanish classes was a requirement to participate in the present study, participants varied with respect to the number of hours of class required to be taken by their home universities. Descriptive statistics for hours of Spanish class showed that class hours ranged from 2 hours to 16 hours per week. This includes both Spanish language classes, and content classes taught in Spanish. It was assumed that participants spoke in Spanish during class hours, or at least gained exposure to the L2, which is why we took it into consideration as an L2 use variable. The average number of class hours in Spanish was 6.6, though the standard deviation was quite high (4.3), indicating that the mean was not representative of the sample as a whole.

Table 23 - Spanish weekly classroom contact hours

n=49	Mean	Std Dev	Median	Min	Max
number of Spanish classroom contact hours	6.6	4.3	4	2	16

Pearson correlation analysis between number of classroom hours and fluency gains variables was conducted to examine the relationships between Spanish classroom contact hours and fluency gains. No significant correlations were found in all measures of speed and breakdown fluency (26 measures in total). However, with respect to the four repair fluency measures, we found that the gain in the number of repetitions was significantly and moderately correlated with Spanish classroom contact hours (r=-0.512, p=.015). This data illustrates that those who took more hours of Spanish class had a lower gain in repetitions rate (and therefore less repetitions in their speech at the end of their stay). Perhaps those taking more Spanish class contact hours had more speaking practice and were thus able to speak with less repetitions. To summarize, of the over 30 fluency gains measures, only one was significantly correlated with classroom instruction hours, suggesting that classroom instructions hours may not have a large impact on speaking fluency gains in the SA context, at least for our participants. Correlation tables for all questionnaire data and fluency gains can be found in Appendix K.

5.3.3 *Other language use factors and fluency gains*

Using the self-reported questionnaire data, the relationships between living situation and other language use factors such as extra-curricular activities done in Spanish and fluency gains were analyzed. Ten participants stayed in a homestay for 2 weeks (a part of their specific SA program), and then moved to a shared apartment, while the majority (35) lived in shared apartments with mostly American and some L1 Spanish speaking roommates. Three participants lived in dorms, while one participant lived in an apartment alone.

To examine how living situation was related to fluency gains, living situations were categorized as *most immersive* (those with a 2 week homestay), *somewhat immersive* (shared apartments), *somewhat isolated* (halls of residence) *and most isolated* (apartment alone). A t-test was conducted to see if there were differences between the Homestay group and participants who did not have a Homestay experience with respect to most L2 fluency gains measures. Only one fluency measure – within clause silent pause rate gains – was significantly different between the groups (t (39,10)= p=0.04. Homestay participants had significantly less gains in within clause silent pause rate, indicating they were *more* fluent on this breakdown fluency measure at the end of their stay than their peers. However, this result should be interpreted with caution as the sample size numbers are different (10 vs. 39), which is not idea for statistical analysis of two independent groups. Descriptive statistics and t-test results for all non-significant measures for Homestay measures can be found in Appendix K.

Extracurricular activities taken in Spanish were another way in which individuals differed on their SA experiences and could potentially have more L2 language use during their stay. At the end of their stay, participants were asked to indicate whether or not they participated in an extra-curricular activity in which they spoke Spanish and to identify the activity. Twenty-six of the 49 participants had at least one extracurricular activity. Only two participants reported having two activities in Spanish. Activities were diverse, including church groups, gym memberships, volunteer organizations, language exchange groups and others. A Pearson correlation analysis was conducted to examine the relationship between fluency gains and the presence of extracurricular activities. No significant relationships between speed and repair fluency and extracurricular activities were found. Interestingly, only one measure of breakdown fluency significantly and moderately correlated with extracurricular activities. There was a significant, positive and moderate correlation between ASU filled pauses and extracurricular activity level (r=0.442, p=.039), indicating that those with more activities had a higher filled pause rate between ASUs. This is an unexpected result, as those with more language use opportunities were expected to pause less (both with respect to silent and filled pauses) than those who did not. Descriptive statistics and t-test results for all non-significant measures for extra-curricular activity measures can be found in Appendix K.

5.3.4 *Other study abroad experience data*

Finally, we wanted to examine how study abroad experience factors other than language use affected fluency gains. At the end of the questionnaire at post-test, participants were asked to provide additional information on their experience in terms of what aspects of their stay abroad had been helpful in improving their Spanish, their exposure to the Catalan language in Barcelona their confidence level having spent 3 months abroad, and how helpful their stay abroad was overall in improving their Spanish level. As described in the Methodology section, we asked these questions to gain more insight on factors that may have influenced language use, and participants' personal experiences over the course of their stay (with respect to confidence level, for example), and therefore may have an impact on oral fluency results.

The table below shows descriptive data for questions where participants were asked to answer the question on a scale of 1 (low) to 10 (high).

Table 24 – Descriptive Statistics for Self-Reported Confidence Level

	Mean	SD	Min	Max
Comfort level speaking Spanish	5.6	1.71	2	7
Desire to speak like a native speaker	3.7	1.44	1	7
Willingness to resist switching to English	6.4	1.87	2	9

At the end of their stay, participants were somewhat comfortable speaking Spanish (5.6 on a scale of 1-10). Although few had a desire to eventually speak as a native speaker would (3.7 on the Likert scale), interestingly, participants were somewhat willing to resist switching to English when, after having spoken in Spanish, their conversation partner (realizing the participant was not a native speaker, or was having difficulty communicating), switched languages (6.4 on the Likert scale). Given the range of minimum and maximum scores, it is clear that there was great individual variation on these measures.

Despite their lack of desire to speak as a native speaker would, most participants *did* feel that their study abroad experience helped them improve their Spanish. In open-ended questionnaire questions, several noted this improvement was not as much as expected. This is despite data that shows that there were no dramatic improvements in speech fluency over the three-month period. This data suggests that while a SA experience may not necessarily improve a participant's fluency with respect to speed, breakdown and repair measures, it still has the potential to increase one's confidence in speaking the L2.

Within the comments, participants indicated whether they had perceived gains in their own L2 fluency and perceived an increase in confidence in their L2 speaking skills. Fifty percent (50%) of participants commented that they felt more confident in speaking Spanish at the end of their stay. A greater percentage (67%) indicated that they perceived their L2 oral fluency to have increased.

Overall, students had positive comments about their SA experience. Comments were categorized as positive, mostly positive, negative, mostly negative. For example, 39/52 (75%) of comments in response to the open-ended question *"How do you think your study abroad experience has affected your level of Spanish so far? Please provide details"* were positive or mostly positive and only 25% of comments were negative or mostly negative. Inter-rater reliability (percentage agreement) for the categorization of comments was 90.3% For a descriptive of how comments were categorized, please refer to section 4.3.6.4 of the Methodology chapter. Many participants commented that although SA programs were useful, they did not give them the opportunity to speak that they had hoped for. Two participants commented that they perceived their L2 listening skills to have improved, though the majority referred to a perceived improvement in L2 speaking skills.

Percentage of Participant Comments on the Effect of the SA experience on Spanish Level (Participant ID number)

Positive (57.6% percent of all responses)	*"I think study abroad has really improved my level of Spanish even though I speak English the majority of my day" (20)* *"The immersion of my study abroad experience makes it that much easier to speak." (21)* *"It has definitely improved" (27)* *"I learned a ton of new vocab" (62)* *"I believe my Spanish has improved at least 20%" (P006)* *"It helped me with less formal lingo" (P016)* *"My level of Spanish has increased dramatically and my Spanish confidence in speaking has as well – practicing daily helps a lot"(50)*
Mostly Positive (17.3 % percent of all responses)	*"It has helped but not as much as I anticipated. It would have been more beneficial to have classes with speakers native to Catalan/Spanish instead of 100% native English speakers" (26)* *"Somewhat…vocabulary and grammar"(43)* *"More second nature…but not as much as I'd hoped" (47)* *"**Study abroad provides all the resources for you to improve your Spanish, but it´s up to you to use the resources and push yourself out of your comfort zone to try to speak more fluently**." (53)* *"It has helped me a good bit, but it has also made me nervous because I don't want to mess up" (61)* *"My Spanish class has helped a little…but I got most of my practice talking to cab drivers"(P011)*
Mostly Negative (11.5% percent of all responses)	*"I believe my listening/understanding skills have improved greatly but unfortunately I can tell that my speaking skills have declined due to a lack of opportunities for practicing**. Coming here, I assumed our program would have offered plenty of opportunities to practice in class, but there are virtually none. It took me too long to realize that practicing speaking was my own responsibility"** (27)* *"Not as much as I had hoped" (33)* *"Only slightly better" (36)* *"Not very well, still, but I can get my point across" (40)* *"**It´s made me a less confident speaker. Before coming I used to think my Spanish was at a decent level, but I have been corrected so many times in my speaking by natives that I have started to avoid speaking it all together."** (66)* *"It has helped me a little, but I wish I spoke it more with locals" (68)* *"As the group translator I got lot of practice. No one else really spoke Spanish" (P001)*
Negative (13.5 % percent of all responses)	*"Just made me more comfortable using it, but it did not improve that much" (55)* *"It just exposed me to Spanish culture…**you kinda have to struggle**" (58)* *"**Have not used it at all**…can't understand but I've learned a few words" (67)*

As can be seen from the examples in the Table 35 (above), those whose comments were completely positive (17.3%) mentioned the immersion experience, daily practice, and informational, incidental vocabulary gain as reasons why they enjoyed their stay. Over half of the comments (57.6%) were mostly positive, indicating that participants enjoyed their stay and believed they benefitted from a SA experience especially in terms of speaking skills and vocabulary knowledge– at times overestimating the amount by which their Spanish fluency improved.

Those who found their experience mostly (but not entirely) positive, additionally commented on important information that SA participants should fully understand before embarking on their semester abroad. They acknowledged that being around English native speakers both in and out of class hindered their learning opportunities, and that real-life practice was both useful and limited (for example, to cab drivers and others who spoke no English). Some acknowledged that even in an L2 SA environment, it was truly up to the individual learner to step out of their comfort zone and make a deliberate effort to learn the language. Others cited their own fear of failure or nervousness at making mistakes as tainting their immersion experience. This data suggests that participants embarking upon SA programs may need to be more prepared for the reality of the SA experience. Organizations and universities than run SA programs may want to help participants set realistic expectations and establish a baseline level of confidence in L2 speaking before arriving, or a strategy to manage a fear of failure before their experience.

Participants in the present study who found their experience mostly negative commented on interesting and valuable areas of improvement for SA experiences, specifically in terms of its isolationist nature. For example, one participant comments that they became a *less* confident speaker by the end of their stay, reducing their language use due to anxiety and frustration at being corrected too often by native speakers. Another noted that they had wrongly assumed that taking Spanish classes and living in Barcelona would provide them with sufficient opportunities for language practice. Yet another explained that being a “group translator” was the only way he spoke Spanish, indicating that participants often clustered in groups of native English speakers relying on one individual to do the speaking. Participants also noted that speaking an L2 was a “struggle” and some commented on learning more about the culture than the language, pointing to an important consideration for SA programs.

Those who only had negative comments when asked about their SA experience and its effect on their speaking skills (13.5%) focused on key ideas – avoidance of speaking the language, and the frustration at having to struggle to speak and be understood.

To further explore themes in participants’ comments about their stay, text-mining analysis was done using Monkey Cloud word cloud generator, which integrates artificial intelligence and text mining software to bring out semantic themes in qualitative questionnaire data. Words that are larger or more prominent in the word cloud have been mentioned more frequently in the data set. Figure 19 below shows the prominent words that appeared in the open-ended questionnaire question regarding the participants’ impressions of their SA experience and its effect on their level of Spanish.

Figure 17- Effect of SA on Spanish level (text-mining analysis)

The terms "conversation", "native speaker", "confidence" and "confident speaker" are prominent. "Native speaker teacher", "native speaker friend" and similar terms referring to contact with those speaking the L2 also frequently appear. This data suggests that while a 3-month stay abroad may provide limited fluency gains (as seen in the first research question), participants do perceive it to be a valuable experience in which they gain contact with native speakers and have some opportunity (usually daily) to speak the language, get "lots of practice" and improve their confidence levels. Some did, however, notice a "lack of opportunities" compared to their expectations. Interestingly, vocabulary came up frequently in the open-ended questionnaire comments, as participants perceived their L2 vocabulary to have increased. This was especially noted in terms of "local sayings", expressions, new words and less formal vocabulary.

Participants were also asked which aspect of their stay they found to be the most helpful in helping them learn Spanish during their SA experience. Reasons were coded into 5 categories based on participant responses – speaking with native speakers, classes in Spanish, the immersion environment itself (being exposed to the language), living with locals and becoming comfortable with making mistakes. If a participant answered more than one of these within their open-ended response, it was counted in both categories. Figure 20 shows the number of times each reason was mentioned in response to the open-ended question. Native speaker interactions (primarily at restaurants and with taxi drivers) were most often listed. Spanish class taught by native speaker teachers was also cited as one of the most helpful aspects, and 12 of the 77 comments (15%) cited the immersion environment and being exposed to as a whole as beneficial.

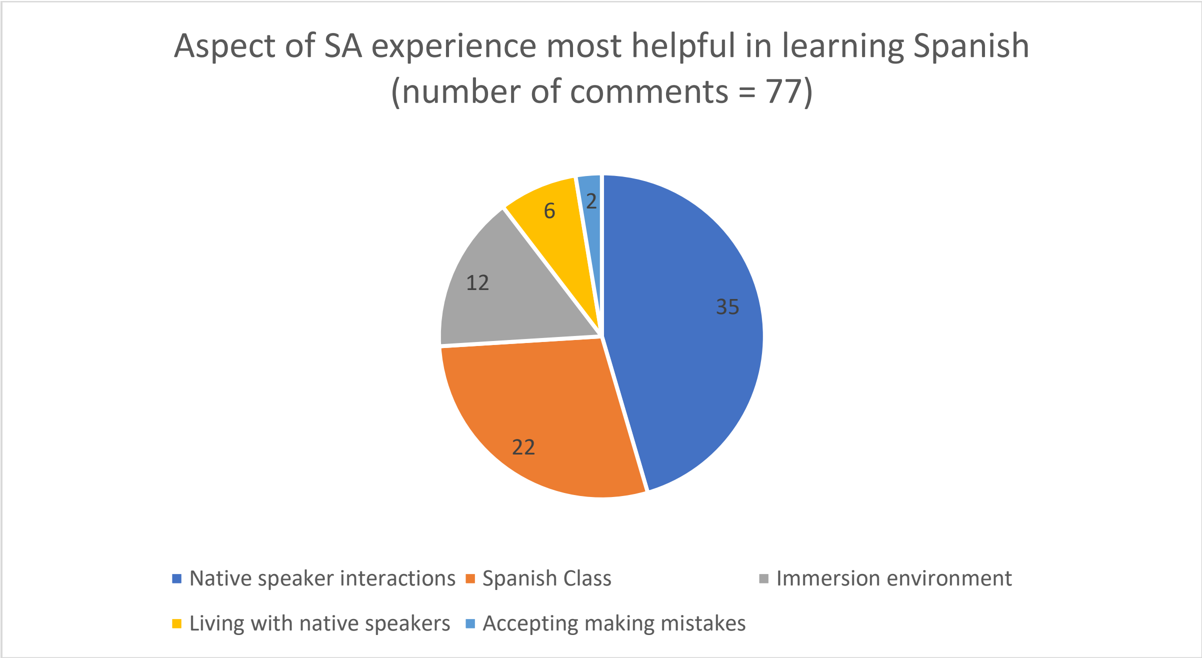


Figure 18 - Most helpful aspect of SA experience in learning Spanish (self-reported)

Figure 21 below shows the word cloud data for the question in which participants were asked about the most helpful aspect of their stay. Though caution must be used when interpreting word cloud data (as the words such as “Spanish class” could have a positive or negative context), the data below was taken from questions were participants were asked to name the *helpful* aspects of their stay. Clearly, participants found Spanish classes and speaking with locals to be the top aspects that helped their Spanish level during their stay. Conversations with cab drivers, host parents and native speaker friends were among the most mentioned aspects. Interestingly, extracurricular activities such as church groups were also cited as being helpful in L2 Spanish improvement.

Figure 19 - Perceived aspects of SA experience that helped L2 Spanish learning

In short, the word cloud data for both open-ended questions suggest that participants have a positive perception of their stay abroad, see the benefit of interacting with native speakers and perceive they speak more fluent Spanish at the end of their stay. Given the few fluency measures on which participants actually improved, this data is surprising. Taken together with the comments analysis done in this section, we can see that the data imply that SA experience may not be what participants expect; SA program organizers can benefit from this data and better prepare students for the challenges they are likely to face abroad.

5.3.5 *Summary of results*

To answer the first part of the research question, regarding how experience factors such as language use relates to fluency gains, it is clear that, although participants came to Barcelona to study Spanish in some cases with Spanish language university-level content courses, they did not use their L2 as much as expected while abroad. Experience factors analyzed included self-reported language use, classroom instruction hours, extra-curricular activities and living situation. Most fluency gains were not related to self-reported language use or other experience factors. In fact, we can see that of the many fluency measures (3 speed fluency measures, 22 breakdown fluency and 4 repair fluency measures), only 4 measures were significantly related to self-reported language use, one measure was significantly related to classroom instruction hours, and one was significantly related to living situation. The results imply that these specific experiential factors themselves do not have a large impact on participants L2 oral fluency gains while abroad.

Data from the questionnaire implies that participants had little motivation to speak their L2, to become as "fluent" as a native speaker, though they did tend to be willing to resist switching to English if their conversation partner attempted to do so. Despite the majority of the group believing that their Spanish had improved and stating that they were now more confident speakers of Spanish, the fluency data in the previous research question has shown that their speaking fluency did not significantly improve on most measures. A few students communicated that their SA program provided much less opportunity to speak Spanish (even in class) than they had hoped for, though overall, they cited Spanish class and daily native speaker interactions (such as taking a taxi) as important aspects contributing to their perceived L2 speaking improvement.

Although neither classroom contact nor extracurricular (out-of-class contact) were significantly related to L2 oral fluency gains, participants did have the impression that both classes and extra-curricular activities were valuable components of their stay that affected their L2 learning.

6 Discussion

This chapter discusses the results reported in the previous chapter, 5.0. The discussion is divided in three sub-sections by research question. A summary of the results is presented at the beginning of each sub-section. The objective of the present chapter is to use the theoretical frameworks presented in the literature review, and reference previous studies in SLA literature to discuss answers to the research questions presented in chapter 3.0, as found through the present study, and to discuss possible explanations for the (mostly unexpected) results. Section 6.4 discusses limitations of the study and section 6.5 discussion pedagogical implications of the present study, and ideas for future research.

In the present study, speaking fluency was examined using speech elicitation tasks at pre-test and post-test to help understand how the SA experience impacts L2 oral fluency behaviour for American learners of Spanish. The findings in the present study are quite different to the majority of SA literature and present some interesting perspectives to consider regarding language learning in the context of SLA. As few L2 oral fluency studies thus far have adjusted for participants' individual differences in the L1, especially in the SA context, the present study provides a unique contribution to the literature. With respect to analyzing a dimension of cognitive fluency – inhibitory control – the present study provides a start to incorporating this dimension in L2 oral fluency research abroad that needs to be researched further.

Combined with SA experience factor data, the findings in the present study may help us understand individual differences in the development of L2 oral fluency for learners abroad and also provide important considerations for the development of future SA programs aimed at fostering L2 oral fluency.

6.1 L2 Fluency

In the present study, speaking fluency was examined using speech elicitation tasks at pre-test and post-test to help understand how the SA experience impacts L2 oral fluency behaviour for American learners of Spanish. The findings in the present study are different to the majority of SA literature and present some interesting perspectives to consider regarding language learning in the context of SLA, especially for Americans studying Spanish abroad. While most SA studies that include L2 fluency measures cover a wider range of topics and limit their analysis to an average of 3 to 4 L2 fluency measures (Tullock & Ortega, 2017) – usually one or two of each of *speed*, *breakdown* and *repair* fluency, the present study attempted a more in-depth and broader analysis. While the intent of the present study was to bring this in-depth analysis done in L2 fluency studies in other contexts to the SA context, given the results of the study we must question to what extent a plethora of measures is necessary. As detailed in the subsections below, very few significant differences were found from pre-test to post-test on the measures tested. However, it is true that measuring fluency using both frequency and duration, and location measures (within and between clauses or ASUs) did bring some insight to the data, so we would recommend following this growing trend in L2 oral fluency research within the SA context. It would also be interesting to see if a wider range of measures provides more conclusive data when larger monologic speech samples (longer than one minute), or dialogic speech samples are used.

In the sub-sections below, the results are discussed in terms of speed, breakdown and repair fluency, Segalowitz's (2010) adaptation of the model and concepts of *cognitive* and *utterance* fluency, and in relation to what is already known in study abroad literature. Possible explanations for the lack of fluency gains found on most measures (across all three types of fluency) are then discussed.

6.1.1 *Speed Fluency*

Results (detailed in Section 5.1) for the first research question showed that participants performed unexpectedly on most measures in terms of speed fluency. When interpreting the descriptive statistics, we can see participants gained on the duration of speech indicating that they did speak for a longer time during the post-test, as expected. They may have felt more confident or were trying more to express themselves more at the end of their stay, or they may indeed have had more disfluencies in making more effort to communicate their point.

Surprisingly, on average our participants spoke *significantly slower* in their L2 at post-test, even after adjusting for the L1, which was indicated by a decrease in both speech rate and articulation rate from pre- to post-test, a highly unexpected result. Relative gains from pre-test to post-test, even after adjusting for the L1, were negative, suggesting that participants did not improve on this measure (speak faster) in their L2 at the post-test. Similar results have not been found in the literature as participants are usually found to improve on temporal fluency measures, especially speech rate. Adjusting for L1 speech did affect the outcome of some pausing measures, but in speed fluency this was not the case.

Given both seminal SA fluency studies and more recent studies that show participants significantly improve on speed fluency measures after a stay abroad (Freed & Segalowitz, 2004; D'Amico (2012), Izmaylova--Culpepper & Olovson (2017)) our hypothesis was that speed fluency measures - speech rate and articulation rate– would improve after at stay abroad. Despite expectations to the contrary, participants became *less* "fluent" in terms of speed fluency as a result of their 3-month stay, a highly unexpected finding. One possible explanation for this result is that participants could have been attempting to express themselves using more vocabulary, or using grammatical or sentence structure aspects more correctly, slowing down their speech rate, though a lexical and grammatical analysis of participants' speech was beyond the scope of this study. The use of only one type of speaking task may also have affected results. Du (2013), for example, found in his study on American learners of Mandarin Chinese studying abroad, that those who told a narrative story spoke faster at post-test than those who were asked to defend or explain a position, as they were in the present study. Perhaps three months abroad in is not enough time for intermediate learners of Spanish (some of whom were taking only 2 hours of Spanish class per week, for example) to have significantly faster speech. Du (2013), for example, found that those who committed to a language pledge (to not speaking their L1) experienced significant gains in speech rate, while others did not. (For a full discussion on experience factors and fluency gains in the present study, see 6.3).

Importantly, SA studies, and even those in foreign language housing settings (e.g. Martinsen et al. (2011), sometimes show that some individuals experience gains, whereas others do not. While Martinsen et al., (2011) looked at oral proficiency as a whole using the OPI test, not specifically L2 oral fluency gains, it is important to note that individuals differ in their progress abroad.

It is possible that the slower speech rate at post-test (and a lack of relative gains) was due to participants' *attempting* to speak more and express themselves more clearly, but not being able to fluently get their message across at a more fluent speed. A lack of automaticity would mean speech processes may be slow, but at post-test participants may have been more confident trying to use the language, whereas at pre-test they simply paused or stalled rather than expressing themselves. Speech rate is certainly one of the most commonly used measures of fluency in SA literature and *usually* increases after a stay abroad. However, as Tullock & Ortega (2017) rightly note in their scoping review, it is one of few measures of fluency that is consistently measured and taken by itself, is an imprecise measure of fluency improvement. The present study did not find results consistent with the literature in speech rate, either. Perhaps using a more dialogic task type or one that elicits more speech would provide a more accurate measure of speech rate.

6.1.2 *Breakdown Fluency*

Studies in the SA literature on L2 fluency generally show positive gains in breakdown fluency measures, though they vary greatly in the measures they use to assess this construct. As shown in 5.1.2, participants in the present study experienced gains on some breakdown fluency measures, but not others. Importantly, adjusting for participants' L1 data where applicable (where all participants had a value), affected the results and in some cases, revealed gains that were not present before adjustment. MLR, once adjusted for L1 speech did show positive gains, indicating participants could speak for longer stretches without pausing after their stay.

In the literature, MLR is a breakdown fluency measure that is often seen to improve, as are silent and filled pauses. Llanes and Muñoz (2009), for example, in addition to speed fluency gains, found improvements in breakdown fluency – participants had longer MLRs even after a short stay abroad. Our participants, too, were able to speak for a longer duration, and had longer runs (speech without pauses) at post-test. The positive gain in mean length of runs (MLR) indicates though the SA environment may have provided potential boost in their confidence as they could sustain more time speaking without pausing or disfluency at the end of their stay than they did at the beginning.

With respect to silent pause frequency, however, after adjusting for the L1 we saw that participants had *gains* in silent pauses as well as mean pause duration, indicating they paused more and for a longer time at post-test. This is in stark contrast with the majority of studies in the literature over the last decade, which show SA participants do usually improve in their breakdown fluency when considering hesitation measures e.g (Freed & Segalowitz, 2004; (D'Amico, 2012; N Segalowitz et al., 2004; Tavakoli et al., 2016; Tragant et al., 2017).

When looking at silent pausing location and duration, we had expected within clause pausing to decrease, and between clause pausing to remain the same or increase which would have shown a trade-off between pausing in a more disfluent way (within clauses or AS-units, units of meaning) to a more fluent way (between clauses or AS-units). This did not occur for most measures. When one pauses *between* a clause or ASU, it is less disfluent than when one pauses *within* a clause, as the disfluency falls between two separable pieces of speech, rather than in the middle of a meaning unit, breaking the flow of the speech. In fact, contrary to expectations, participants in our study had relative gains in within-clause silent pause rate and within-clause pause duration. Between-clause frequency and duration both showed negative relative gains, indicating participants paused less, and for less time, between clauses. This can be interpreted as an improvement in breakdown fluency, as participants did indeed pause less, though not for all types of pauses.

However, our participants' performance with respect to within and between ASU silent pausing data does align more closely with the literature. De Jong et al, (2016), for example, found that L2 learners of Dutch paused less (had less silent and filled pauses) *within* ASUs in L1 speech than they did in their L2, but between ASU they paused about the same in both languages. Importantly, De Jong et. al.,'s (2016) study found that proficiency level mediated both frequency and duration of within-ASU pauses in the L2. Our participants were of a lower intermediate level, and we did not thoroughly measure their proficiency; this could explain the lack of within-clause improvement (decrease of within-clause pauses), though there were gains in within-ASU fluency. Understanding participants' proficiency level more clearly, we may have been able to see if there was more improvement on within-ASU pauses in different proficiency level groups, as the literature suggests.

For our participants, filled pause rate did not significantly change at post-test. Perhaps participants attempted to more fluently articulate some of the time, falling back on filled pauses when their articulatory skills in the L2 failed them. The literature on speaking style (e.g. De Jong, 2015) has shown that silent or filled pausing may be a personal preference and vary in function with the way speakers speak in their L1. This may show that participants are attempting to speak more, but as they do not have the level or automaticity to continue in a fluid speech stream; it is possible they are trading silent pause time for disfluencies. One possible explanation is that they could be making a choice between using silent or filled pauses or other disfluencies when speaking. The fact that participants showed negative relative gains on elongations, however, may indicate that they stalled less at post-test, and were able to continue their speech stream better than at pre-test. Participants did show a significant difference on the composite total pausing measure, though this was a *gain*, not a decrease in pauses. Other than this aspect, in terms of breakdown fluency, their speech was similar before and after a 3-month stay abroad, indicating little improvement on breakdown fluency measures.

This inconsistency with the literature that shows improvements on pausing measures for intermediate learners on SA programs (e.g. Du et al., 2013; Kim, 2015; Segalowitz & Freed, 2004) could potentially be explained by our participants' lack of practice (L2 use) or motivation, further discussed in 6.3. Examining breakdown fluency from the perspective of Segalowitz's (2010) model, at pre-test the pre-verbal message would in this case be "stuck" in one part of the encoding process in, either unable to draw from the mental lexicon, or unable to encode (and therefore use) a correct grammatical morpheme, or morphological structure. Overwhelmed and susceptible to fluency vulnerability points in the speech system, participants pause. Another possible explanation why our results differ from others is that perhaps at post-test, our participants are only starting to overcome the fluency vulnerability points in the speech system (such as morphological or grammatical encoding) and therefore participants' ability to *attempt* to express themselves. Since the encoding of phonological and morphological system is not fully developed, and their grammatical and mental lexicon are not as complete as in their L1, participants often end up using filled pauses or elongations rather than the utterance they want to use. They do not quite access their desired lexical or morphological item in time to use it in speech, reaching a disfluency point that does not enable them to continue their speech fluently. Taken together, these results could suggest that participants are "on their way" to improving their pausing and disfluency behaviour after a 3-month stay abroad. The data show very little improvement on breakdown fluency measures but do show improvement on some within clause pausing measures. It is possible that SA experience factors discussed in 6.3 (such as language use, and other conditions of their stay) impacted the extent to which breakdown fluency improved. Reflecting on the use of all of these breakdown fluency measures, it appears to be neither efficient nor useful to examine this wide a range of measures. However, we would recommend measuring breakdown fluency frequency and duration, as well as location (within or between ASUs). Given that ASUs and clauses were only significantly different on some measures, it may be more feasible to use only ASUs in future studies (as these represent units of meaning and encompass a large range of speech than a grammatical clause). Total disfluency as a composite measure may be more useful to study than looking at individual types of disfluencies themselves.

6.1.3 *Repair Fluency*

Repair fluency results for the present study were somewhat unexpected as there were increases in both repairs and repetitions at post-test. In the L2 fluency literature, repetitions and repairs are at times found to decrease after a SA experience e.g. (Keppie et al., 2016), Suzuki and Kormos (2019)), though the presence of repairs in L2 speech does not always depend on the proficiency level (Tavakoli 2017). More advanced learners, too, exhibit repair behaviour, as do those who have a personal focus more on accuracy than fluency (Kormos, 1999). Recent research has also shown that repairs may be seen as an indicator of personal speaking patterns rather than a disfluency measure (Duran-Karaoz & Tavakoli, 2020). As most SA oral fluency studies focus on speed and breakdown measures, few comparisons to the previous literature can be made.

Participants in the present study showed a significant *increase* in repairs and repetitions at post-test, indicating participants were more disfluent on this measure, though the increase in repairs did not reach significance. This may suggest that participants are still “stuck” as to what to say after a three-month stay abroad, as they repeat themselves, stalling for time. However, one possible explanation is that after a stay abroad they were more aware of their mistakes and made more of an attempt to correct them. Self-monitoring is a process that is highly relevant in L2 oral fluency development (see Kormos, 2006), and could be responsible for this increase in repairs (and therefore decrease in repair fluency) at the end of a SA experience. In Segalowitz’s (2010) model, this, like the data from breakdown fluency, would correspond to slightly better *cognitive fluency*, producing less hesitations. Before reaching *overt speech*, L2 speech is processed through the conceptualizer, which leads to a preverbal message (which is then phonetically encoded) before it can reach the articulator and convert into a message. If participants have a speech plan but do not have the automaticity to deliver it, they may become stuck in the process, repeating to stall for time, and repairing to correct an error they realize has occurred.

Although composite measures of fluency are rarely used in SA literature with respect to L2 oral fluency measures, we decided to calculate this measure since individuals differ in how they express their disfluency (silent or filled pauses, elongations, repairs, etc.). Following the L2 oral fluency literature, repairs and disfluencies are counted in the overall speech rate, and repair rates are taken from the count of repairs over total speech. Our results show that, after adjusting for L1 disfluencies, participants were *significantly more disfluent* at post-test. This finding was unexpected and suggests that participants, though speaking more, are experiencing no improvement after a stay abroad with respect to total disfluency.

Taken together, the data on speed, breakdown and repair fluency was generally not aligned with expectations in that we expected that the majority of these measures found to improve in the literature, would also improve for our participants. When analyzing the present study's results in terms of SLA speech production theory, looking at Segalowitz's (2010) Model of the L2 speaker) the results are not in line with the expectation that participants would show a tendency towards gaining more *automaticity* during their SA experience. While it may be unrealistic to expect drastic changes in automaticity after only 3 months abroad, we did expect, in line with the literature, that speed and articulation rates would have improved with increased L2 usage and opportunities for practice available in the SA environment. This would have meant high automaticity in speech processes, and smoother processing within the conceptualizer and articulator, thus improving speech and articulation rates - but this was not our case. SA experience factors discussed in 6.3 may help explain this puzzling result.

According to Segalowitz's (2010) Model of the L2 speaker, to produce *overt speech*, participants generate a pre-verbal message created after macro and microplanning in the *conceptualizer*, which is then encoded with grammatical, morphological, phonetic, and lexical information before proceeding through the *articulator* and then producing the message. In a study abroad context, it was expected that participants would gain lexical, grammatical, morphological knowledge and speaking practice over the course of their three months in Barcelona. This knowledge, and extended practice using it, facilitating these types of encoding of the preverbal message and aiding the L2 speech production thereby producing more "fluent" speech.

Segalowitz (2010) identifies seven areas of potential disfluency at different stages in his L2 speech production model, labelled as {f} areas (See *Appendix G - Segalowitz's L2 speech model*

(2010)). There are several reasons a SA experience could help overcome these potential hurdles that slow down L2 speech and cause the learner to produce disfluencies. Through increased L2 input in the SA environment, speakers may be able to have more success in conceptual preparation and forming the preverbal message{f1} (not simply thinking of a message, but how to "build correct construal information" (Segalowitz, 2010 p.12) to reflect their viewpoint in the L2. Given participation in Spanish language classes and activities, participants in a SA experience may gain grammatical {f2}and lexical {f3} knowledge that allows them to more efficiently encode the message, but experience disfluencies at these points. If participants have improved their vocabulary knowledge over the course of their stay abroad (an aspect beyond the scope of this study), they may have less of a tendency to experience disfluencies – like repairs or repetitions – as a result of lexical access from the mental lexicon. In the present study, vocabulary was only measured using a basic test, as a proxy for proficiency, but future studies could integrate a more sophisticated measure of both initial proficiency level and vocabulary development throughout the stay. Increased L2 input (and especially native speaker input) while abroad may also allow participants to develop a better understanding of L2 speech as they listen to native speaker speech, aiding in {f4}, morpho-phonetic and {f5} phonetic encoding. Disfluencies in articulation {f6} (execution of speech) arise as a result of the other critical points in the model, which could theoretically be improved during the stay abroad.

Clearly, our participants did not have the L2 oral fluency development experience we expected; rather, they experienced few if any speed fluency gains, and gains on some breakdown measures, and one repair measure. This may have been due to lower L2 input and output during their stay. The remaining critical point at which disfluencies occur{f7}, self-perception will be discussed further in our discussion on language use and SA experience. As there are many individual differences involved in L2 oral fluency development, it would be interested to see the outcomes of studies using a wider variety of measures for other participants, to see if our study is an anomaly, or representative of SA participants.

6.2 Inhibitory Control

The second research question originated from the idea that, as Segalowitz's (2010) model suggests, *cognitive fluency* is crucial in the development of L2 speaking fluency. In choosing to measure inhibitory control, we looked at the ability to suppress one language while speaking the other (an aspect of cognitive fluency) and its relationship to speaking fluency measures and, particularly, fluency gains from pre-test to post-test. Theoretically, if a participant has stronger inhibitory control ability as a cognitive skill, they can more easily suppress conflicting information. As is well-known in SLA literature, all languages an individual knows are activated when they hear and speak language (Green, 1998; Linck, 2008), and one's ability to accurately and quickly activate and select the correct words from their mental lexicon, suppressing that of the other languages and more efficiently encoding the preverbal message, allows them to successfully produce fluent speech.

Fluency gains for participants in the present study were not related to non-linguistic or linguistic measures of inhibitory control. If a participant were immersed in an L2 context where they speak little of their L1 for an extended period, inhibitory control ability itself could theoretically change. As discussed in the Methodology section, this was done as a preliminary analysis prior to answering the second research question about inhibitory control ability and fluency gains while abroad. We expected that 3 months may have been too short of a stay for inhibitory control ability to change, (and perhaps a more immersive context would be needed to do so). However, we did expect the participants' individual ability control scores would relate to fluency gains on several measures. For example, we would have expected a participant with high non-linguistic inhibitory control ability (measured using the Simon Task) to be able to efficiently suppress the other languages in their speech system, resulting in greater cognitive fluently, and thus more fluent speech. Linck et al., (2009), for example, showed that L2 learners of Spanish abroad had low inhibitory control ability (reduced L1 access, measured using a lexical access test in both L1 and L2) after their stay abroad. This does not appear to be the case for our participants.

Linguistic inhibitory control information measured through the Letter (Phoneme) Decision Task was expected to present a more accurate indicator of inhibitory control, as the information required to be suppressed in the test was linguistic, and not simply suppressing information about the direction in which an arrow was pointing, as it was for the Simon task. Participants needed to be able to suppress information about words in their L1 and correctly choose whether a letter occurred in the L2 word to complete the tasks (the task is detailed in section 4.0). However, the ICP score, which measured their reaction times on correct responses, did not appear to be related to fluency gains over the course of the stay abroad – with one exception. A decrease in one disfluency measure – between ASU filled pauses - is related to performance on the PICS test, indicating that those with higher performance on linguistic inhibitory control tests are likely to experience less disfluencies in terms of filled pauses between ASUs.

Possible explanations as to why the ICP score did not relate to oral fluency gains could in fact be that this was not an accurate test to measure linguistic inhibitory control. However, this is difficult to know given the fact that very few of our participants experienced fluency gains, and no measure of inhibitory control was significantly related to these gains. The test design had limitations. Though it was modelled after Colomé (2001), the original test looked at reaction times for Spanish-Catalan bilinguals in their two languages. The Letter (Phoneme) Decision test designed for the present study used the same rules as Colomé (2001) to choose non-cognate words of relatively high frequency and presented words in a random block order, with each block aligning with rules for the block (no more than 3 "yes" answers in a row, for example). Participants needed to decide (as quickly and accurately as possible) whether the letter they were presented with occurred in the L2 for the word for the image they saw (e.g. *m* for *mesa*, when shown *m* and a picture of a table). Participants had to suppress the L1 information of the picture being a "table" to correctly respond yes to "m" and no to "t" for this picture. See Section 4.0 for full details.

Perhaps a different design is needed for L2 learners, as despite being high frequency these words may not have been as solidified in their mental lexicon since L2 learners have less *cognitive fluency* and less *automaticity* in their L2. Unknown words in the task could have been problematic, although care was taken to choose high frequency words, and to eliminate words in the test participants were unfamiliar with (based on self-reported data). Participants went through a training session before beginning the task where they were familiarized with the images and words they represented. Any words they did not know beforehand (self-reported) were then removed for the analysis for each participant, ensuring that participants were only tested on words they knew and were able to, in theory, more easily reject or accept the appropriate word when prompted. Participants who could more easily inhibit this L1 linguistic information were expected to more easily be able to speak fluently as they would be, in a similar way, suppressing

L1 information when encoding and then articulating the preverbal message in the L2 speech system. As this was not the case for our participants, there could be several explanations, discussed below.

As participants' ICP score was not related to any speaking fluency measures, this could either indicate that no relationship exists between linguistic inhibitory control and oral fluency gains or that a different measure is needed to more accurately capture the relationship between fluency and inhibition. Furthermore, a limitation of this test was test fatigue, as it was the second test in a series of tests participants completed in approximately 45 minutes. As it was an intensive task that required a lot of attentional resources (it was more cognitively demanding than the other tasks in the study), it is possible that participants scores were affected by a lack of motivation to fully concentrate on the task. Participants were paid 20 euros for their participation, and given class participation marks, but this may not have been motiving enough to truly focus on the Letter (Phoneme) Decision task, as it required a high level of concentration.

It is interesting to note, though, as our preliminary analysis prior to examining inhibitory control and fluency gains, that there were also no significant differences from pre-test to post-test on the other linguistic measure of inhibitory control we tested, *language switching* (tested through *Overflow*, the time it took to switch after prompted). Participants did take less time on average to switch into Spanish at post-test, indicating there may be *some* effect of inhibitory control in terms of language switching become easier (faster) for participants at the end of their stay, but these results did not produce a statistically significant difference. There was a trend towards significance for switching faster into Spanish at post-test, whereas switching into English remained the same; this may suggest that participants are *almost* able to switch into Spanish better at the end of the stay than the beginning. Given more time, and perhaps a more intensive L2 environment and more practice, participants may have been able to see more improvements in switching speed. Very few studies in the SA literature look at cognitive fluency, though there is an increasing trend to do so (see Segalowitz, 2016). The majority of the studies that exist do not look at inhibitory control. Sunderman & Kroll (2009), for example, looked at working memory and found that participants must reach a threshold of this cognitive skill before the SA experience led to more accurate L2 production. L2 production in that study was measured by a verbal fluency task (producing accurate exemplars of a semantic category), and not using L2 oral fluency measures. To our knowledge, no studies in SA literature have either looked at non-linguistic or linguistic inhibitory control, or the relationship between inhibitory control and its relationship to oral fluency gains in the SA context.

There are several possible reasons that language switching ability may not have changed for our participants over the course of their stay abroad. First and foremost, the questionnaire data (discussed in detail in 6.3) sheds light on the lack of L2 language use during their stay; the majority of participants communicated in Spanish less than 30% of time, with some participants even admitting to only speaking Spanish in class, or not at all. As the process of switching languages stresses the cognitive system and requires participants to inhibit a lot of information (see Green, 1998; Linck et al., 2011, Linck & Weiss, 2015, for example) inhibitory control may not have been the only contributor to a lack of switching ability. Other cognitive factors such as attention and working memory could have played a larger role in the way in which participants switched. Importantly, participants were *beginning* to see progress on the Overflow measure, as descriptive statistics show that the time needed to switch (after being prompted) decreased, though not significantly.

For the overall group, some of the results on the ability to switch *into a* language and fluency (for L1 and L2) were confusing and difficult to interpret. Once again, the importance of studying individual differences in participants should be noted here. Although US learners of Spanish on similar stay abroad programs for the same time period are a seemingly homogenous group, there was a wide variety in the performance of these tasks within the group itself. As language switching in L2 oral fluency is understudied in the literature, and to our knowledge not studied with respect to L2 oral fluency the SA context, the present study provides a starting point for future research on the topic.

6.3 Study Abroad experience factors

In examining our participants' study abroad experience factors and their relationship to fluency gains, we expected that those who self-reported using the L2 more would experience substantially more gains in L2 oral fluency with respect to speed, breakdown and repair measures. Surprisingly, speed fluency was not related to self-reported language use percentage. This result is contrary to most L2 oral fluency studies in the SA context, which generally find participants experience L2 speed fluency gains. As the questionnaire data also shows (discussed further in Section 6.4), participants in this study did not use their L2 much while abroad and were often surprised at how little they were required (or given the opportunity to) use it. Interestingly, measures of speed fluency gains such as MLR or duration did not correlate significantly to self-reported language use. Some studies in the literature do find language use to affect speed fluency gains. Kirsner et al. (2003) for example found those who reported using the L2 more had longer MLR (and less pauses, a breakdown fluency measure) than their peers. Participants in the present study who did use their L2 more than their peers were *still* not necessarily able to improve on most measures of speed fluency by the end of their stay. A 6-

month or longer stay, or one where L2 speech was more strictly enforced as a policy, may have produced different results. Participants may have developed more *automaticity* with more language use input and output, and therefore seen greater gains in speed fluency and decreases in breakdown fluency (e.g. see Garcia-Amaya (2012)).

Breakdown fluency results were more in line with our expectations, we found that those who self-reported using the L2 more had lower pause frequency within clauses and within ASUs. As explained in the Methodology section, both ASUs and clauses were used in the data analysis as there were significant differences measured from a clause and an ASU, on some measures. Repair fluency gains were not related to self-reported language use. Given the literature on repair fluency that indicates an increase in repairs is not necessarily linked to proficiency level (e.g. Kormos, 1999), but the percentage of repairs decreases after a stay abroad (Keppie et. al, 2016) this was a surprising result. Although self-reported language use data may not be as accurate as measures recorded by the researchers, it does give an indication as to how much participants are attempting to use the L2 during their SA experience. Perhaps the most surprising result for self-reported language use percentage was the low percentage itself (an average of 30%, with high variance in the sample and only 6 of 52 participants reaching 30% or higher). High amount of language use (self-reported) correlated with higher fluency gains in phonation time (time spent speaking/total time), indicating that the more one speaks the L2 abroad, the more likely they are to speak without pausing. While this finding may seem self-evident, language use for participants while on a study abroad program is not always emphasized by the program creators or administrators.

Many of the unexpected results in the first two research questions in the present study, especially pertaining to L2 fluency could be explained by the results of the third research question. As SLA researchers, we often assume that the purpose of a stay abroad for participants who are taking L2 language classes is academic - specifically, to improve their L2 language skills. However, given the obvious lack of L2 use present in the majority of participants and the wide range of using Spanish "0 % to 50 %" of the time, even in the same program participants clearly do not have a homogenous experience while abroad. Kinginger (2011) cites "high variability" in the language use of American SA participants due to their lack of preparedness and confidence and finds they are either not "meaningfully engaged in host communities…or lack guidance in interpreting their observations" (2011, p.67). Dekeyser (2010), in a study of American learners of Argentinian Spanish abroad finds participants initially highly motivated, but "ill-equipped" to use the L2 and monitor their progress. Perhaps to improve their L2 fluency over the course of their stay, our participants would have needed not only a higher level, but more practice using the L2 and strategies for monitoring their progress.

Given this shockingly low use of their L2 in an L2 environment on a stay abroad, it is unsurprising that there were few L2 oral fluency gains. Given a longer stay, more input and forced opportunities for output and practice, perhaps our participants would have practised switching languages more and had a more fluent speech after a language switch.

What we know about language contact in the SA context and L2 oral fluency gains is that in general, more language contact does relate to gains in L2 oral fluency. Freed, Segalowitz and Dewey (2004), for example, show that voluntary out of class contact time such as in extracurricular activities was strongly related to gains in the length of runs without pausing (MLR), but surprisingly, not with other temporal measures of fluency. In fact, our participants were found to be *more disfluent* on one measure - between ASU filled pauses – the more they took extracurricular activities. While this result is contradictory to the literature, it is possible that participants were attempting to speak more (as post-test duration was longer) but were unable to speak smoothly and without hesitations. While these results appear surprising due to opportunities for increased L2 input and output abroad, some L2 studies also find no relationship between those who take extracurricular activities or have additional out-of-classroom contact hours and linguistic gains (e.g. Segalowitz & Freed, 2004; Briggs, 2015). In fact, as Briggs (2015) cautions, the type of out-of-class contact is also important; participants who choose not to live with native speakers may miss out on the richest opportunities to use their L2 and develop their vocabulary in the L2.

With respect to classroom hours and L2 oral fluency gains, we know that the results are mixed – participants sometimes experience more linguistic gains when they take more Spanish class, and sometimes do not. Our results, too, show a moderate negative correlation on only one measure - between classroom contact hours and repetitions, indicating that on one aspect, participants had a smoother flow of speech when they had more classroom contact hours; taken more Spanish classes led to some effect on L2 oral fluency gains.

The social aspect of a SA experience is another language contact variable that deserves consideration. Our participants lived with and spent most of their time communicating with their American peers, rather than L2 native speakers. Studies have shown that L2 learners often find requirements to speak daily with locals or opportunities to spend time with native speakers in their own age group highly beneficial to building confidence and speaking skills (Dewey et al., 2013). Baker-Smemoe et al. (2014) surprisingly found that L2 use was not related to lingusitic gains and even different types of L2 use – classroom use, receptive and interactive out-of-class contact – could not explain differences in gainers and non-gainers. The one variable that did predict gains was the *intensity* of social networks and time spent with close friends. This illustrates that in fact, experience factors related to social networks and language use are an important consideration in SA program development.

With respect to our participants, they spent most of their time living and communicating with their English-speaking peers, which some commented they thought was detrimental to their L2 learning. Previous research shows that participants in programs where they are required to speak daily with native speakers of the L2, or spend time with native speaker peers of their age group report the experience is helpful, and also increase their speed and breakdown fluency at the end of their stay(Baker-Smemoe et al., 2014; Dewey et al., 2013). Du et al.'s 2013 study on American learners of Mandarin Chinese also found that some participants who had more out-of-class contact (interacting with Chinese people in extracurricular activities) showed greater gains in speech rate and hesitation free runs. Given more time, and perhaps more classroom instruction or opportunities to engage with native speakers in their L2 and speak *less* of their L1, the participants in the present study might have experienced more L2 fluency gains, and perhaps might have produced longer MLRs and significantly less pauses by the end of their stay.

Furthermore, when assessing *why* US learners of Spanish did not improve as much as expected on speaking fluency measures, it is important to note that there are many motivations for studying abroad. Second language acquisition literature often assumes or even reports that a stay abroad in a country where the L2 is spoken is of an academic nature and for language learning purposes. For students in university, the target population in the present study, the stay abroad could have had many other motivations – a 3-month trip with the opportunity to live alone for the first time, to experience living in a new country and culture, to travel around Europe, or to share an experience with peers or colleagues. A participant whose motivation for coming to Barcelona was to enjoy the city and experience living alone or with friends for the first time, may not have been likely to choose living with a family at the beginning of their stay.

Despite low language use percentages and several students mentioning their disappointment with SA programs in terms of opportunity to practise and speak their L2, most participants believed they improved in their Spanish fluency at post-test (even when they didn't). When asked about the helpfulness of their stay on their L2 language skills, 75% had positive or mostly positive comments. Crucially, though, when asked (at post-test) about their willingness to resist speaking English (specifically to resist switching to English when their conversation partner switched), they on average rated their willingness at 6.4/10, indicating that most would give it a try but were not overly compelled to place themselves in situations where they only spoke Spanish. They also felt relatively uncomfortable speaking Spanish (5.6 on a scale of 10) and had very little desire to eventually speak as a native speaker (3.6/10 on average). As English is a lingua franca and most people in the centre of Barcelona are likely to know some English and therefore take the opportunity to use it with foreigners, rather than requiring that they speak Spanish. Participants were native speakers of English who were relatively uncomfortable speaking their L2, so they, too, may have taken the readily available opportunity to use English to make communication easier, rather than use their not-so-fluent Spanish. However, other studies of Americans abroad (e.g. D'Amico, 2012) do show that participants at least improve on measures of speed fluency, usually at least on speech rate and MLR.

Again, this points to the lack of motivation to speak Spanish or immerse themselves in as much Spanish as possible when English is a readily available alternative. Given this attitude (even at the end of their stay), it is unsurprising that participants did not achieve stellar L2 fluency gains over the course of their stay. As linguists and researchers, we often believe that

SLA must be the main motivation for spending a few months abroad taking language classes, but this may not be the case for many participants. Other studies that have examined language use and fluency gains often do so from the perspective of foreign language anxiety or foreign language enjoyment, finding that even if their motivation to study abroad is to learn their L2, increased anxiety or stress decreases L2 gains in proficiency (e.g. McIntyre, 2017), and that having a higher proficiency before going abroad can decrease anxiety while abroad (Dewey, Belnap & Steffan, 2018).

Qualitative data from open-ended comments reveal an interesting perspective of the language learner. A sample of these comments was provided in Section, 5.3.6, and a few are discussed in detail here. Despite not experiencing fluency gains according to the tests conducted, many participants felt that their oral fluency skills *did* improve, with one participant even estimating an increase of 20% and another claiming their Spanish level "increased dramatically".

Participants' comments that they were able to learn "less formal lingo", "a ton of vocab", and become "much better at understanding", but "wish they had spent more time with locals" perhaps point to an area of research not covered by the present study – *motivation and confidence in language learning.* Though a review of vocabulary acquisition was beyond the scope of this study, participants self-reported being able to acquire vocabulary while abroad, but not to use it to speak more fluently. Vocabulary acquisition may have been receptive, rather than productive. Future studies should look at a more detailed and rigorous method of measuring initial vocabulary or proficiency level as well as L2 vocabulary development during a SA as it relates to L2 oral fluency development.

At the end of the stay, two of our participants commented that they hadn't been prepared for the level of effort needed or the lack of opportunities readily presented to them to use their L2 and to be forced into situations where Spanish was the language spoken. In fact, a third participant even admitted her speaking skills had declined due to not being given (or seeking out) the opportunity to practise, bringing up the question as to whether study abroad programs should impose L2 language use on their participants, or whether this is the participants' responsibility. In the few studies where language pledges or required L2 communication is enforced, participants gain in fluency and report they find the pledge beneficial (Dewey et al., 2013; Garcia-Amaya, 2012)

Sadly, some of our participants even felt *less* confident as a result of their stay and being corrected by native speakers. Previous research has shown discrepancies between SA participants' expectations and perceived improvement, where participants feel they improvement less than expected, especially on speaking and writing skills (see Badsnuber & Ecke, 2009 for example). Furthermore, and perhaps more troubling than the comments above, several participants indicated that they did not feel their stay abroad was full of opportunities to speak the L2. Given that this study took place in the city centre of Barcelona, we cannot ignore the presence of the Catalan language and the exposure our participants had to the language on a daily basis. Had participants been exposed to their L2 (Spanish) in a Spanish monolingual environment, perhaps they would have had more exposure, more input, and therefore taken advantage of more opportunities for output. For example, input from interactions in situations such as banking, retail, transport, getting direction from locals, and participating in extra-curricular activities would have been exclusively in Spanish had the study taken place in a small town outside of Madrid where little English is spoken.

As linguists and educators, we often assume that the stay abroad context is optimal for language learning – or at least, conducive to it. However, several participant comments show that the learners do not feel this way.

See two selected comments below:

*"**Study abroad provides all the resources for you to improve your Spanish, but it´s up to you to use the resources and push yourself out of your comfort zone to try to speak more fluently**." (Participant 53)*

"Unfortunately, I can tell that **my speaking skills have declined** *due to a lack of opportunities for practicing***. *Coming here, I assumed our program would have offered plenty of opportunities to practice in class, but there are virtually none. It took me too long to realize that practicing speaking was my own responsibility"*** *(Participant 27)*

In addition to citing a lack of opportunities to practice and low confidence, some participants admitted to using avoidance techniques, perhaps as a result of low self-esteem caused by over correction of teachers and native speakers. However, this could also be attributed to an individual difference in personality. Some people when faced with an abundance of corrections may see it as a challenging but welcome learning opportunity; others may view it as an overwhelming challenge they are not willing to commit to.

*"**It´s made me a less confident speaker. Before coming I used to think my Spanish was at a decent level, but I have been corrected so many times in my speaking by natives that I have started to avoid speaking it all together."** (Participant 66)*

Participant comments that illustrate overall trends that may contribute to our understanding of L2 fluency development can be found in 5.3.3. Participants may not have been given (or may not have sought out) the opportunity to use their L2 skills outside of class, and those who were less motivated to do so on their own were not "forced" to do so in order to communicate meaning in their messages successfully. Locals, even those of the same age, may see American students studying abroad in the centre of Barcelona as “tourists” and not as foreigners potentially part of their peer group. Furthermore, the influence of the Catalan language on college-aged native speaker locals in Barcelona is an important consideration with respect to the context. Spanish may not be the language they prefer to speak in, making it more difficult to become part of a local, Spanish-speaking peer group. A SA location in a region of Spain where Castilian Spanish is primarily spoken may provide them with better opportunities to speak Spanish. A combination of the environment in Barcelona itself, a lack of motivation on the part of the participants, and a short length of stay could have contributed to the lack of fluency gains we saw at post-test.

6.4 Limitations of the present study and future research

As in all primary research, the study conducted here has several limitations. One limitation of the study is that there was no control group to compare to, so it is difficult to say whether a similar group of participants studying in a domestic, classroom context would have had similar gains or whether the SA students did gain an advantage. Within the SA context itself, despite increased opportunity for input and output, participants may have lacked the motivation necessary to take full advantage of their SA experience. For example, as all participants were returning to their studies in the USA, where Spanish is not necessary to communicate, it is possible that participants were not highly motivated to continue to improve their Spanish.

With respect to the language switching task, there are limitations in this study design. The test was used as a pilot in first semester because the chosen length before the switch was too short and most participants did not have enough pre-switch data to analyze. In other words, they did not have enough time to speak in the first language they were asked to speak in (L2 for half the participants, L1 for the other half, counterbalanced at post-test), before being asked to switch to the other language. The task in its final form was run in second semester, with 35 participants. It may have been better to choose a switch point based on the end of a description of one action, rather than a temporal switching point. However, that design, too, would have had its limitations; switching would then only have been measured for a particular situation based around similar vocabulary. As this was the first time such a task was used, future research designs could test different switch points before determining the optimal switch point.

Additional factors such as attitude or desire to succeed, cognitive skills such as aptitude, working memory or attention, were out of scope of this study. Furthermore, and importantly, asking participants *why* they chose to study abroad, and to honestly answer (or rank) their main goals from the experience would have given us more insight into their experience. Measuring intrinsic or instrumental motivation and relating it to L2 oral fluency gains, may have provided insight into the lack of L2 oral fluency gains for most participants. For example, motivations for studying abroad other than academic reasons for a 19 to 21-year-old young adult could include a desire to experience living away from home or family, a desire to travel with friends, to experience European culture, to enhance future job prospects by gaining "worldliness" which American employers appreciate. Further investigation into these details for both the group as a whole, and the subset of "high language users" who used Spanish more than 30% of the time at pre-test and post-test may reveal why so few participants used the L2 while abroad. Educators, language teachers and study abroad program creators and facilitators could work to ensure their programs are conducive to high L2 language use.

Importantly, the questionnaire reveals that while studying abroad in Barcelona for three months may not be conducive to oral fluency gains, it may give participants the confidence they need to push forward with their L2 studies. Perhaps a delayed post-test after returning home would reveal further confidence to reach a level where they push themselves to speak more or immerse themselves in more Spanish speaking situations. Regan (2005), for example found that Irish learners of French improved on grammatical accuracy and retained this improvement a year after returning from their SA experience.Pérez-Vidal (2014) found similar results on grammatical accuracy for learners of Spanish studying abroad.

Another aspect of the SA program that the present study did not consider (aside from differences in motivation) was individual differences in personality and attitude. In addition to cognitive abilities, previous experience with the L2 and specific conditions of the stay, participants may be affected by their own attitudes towards the L2 language and culture, their personalities, and personal responses in the face of adversity (in this case, speaking an L2 outside of their comfort zone). Research on second language anxiety (see Dewey, 2008) was also beyond the scope of the present study but could provide more insight into the lack of fluency development learners experienced. Previous research has found that as foreign language anxiety increases, language learning and proficiency gains in the SA context decreases (Allan & Herron, 2003; Gardner & McIntyre, 1993). Some research has, however, found that with longer stays foreign language anxiety can decrease (Hessel, 2016). Had we collected data on participants' expectations, motivations and anxiety levels, the present study may have gained insight into possible factors that contributed to the lack of fluency improvement, and especially the decrease in language use. Initial proficiency level was also not tested before departure or upon arrival in Barcelona, other than a vocabulary test which is a very general indicator of proficiency; participants were assumed to have the level of the language class they enrolled in. More accurate proficiency information could have provided more insight into fluency gains (or the lack thereof) while abroad.

Our lack of detailed data on participant motivation is a limitation. Understanding participants' motivations through a more in-depth look at intrinsic and extrinsic motivations could also help explain the unusual results of this study. As a metropolitan and fairly large European city in which participants could get by using only English, Barcelona itself may not have been the best choice for participants if their goal was indeed to immerse themselves in a Spanish-speaking context, or even to improve their speaking fluency. Participants' university classes also took place on campuses close to the centre of Barcelona, where English is more widely spoken than other areas of the city. Collecting more qualitative data (questionnaires and interviews) on participants' motivations and attitudes towards Spanish as an L2, and especially on their L2 experience while abroad would help us better understand changes in L2 speaking fluency behaviour and the impact individual differences have on improvements on these measures. Including participants in different proficiency groups would also be an interesting comparison and help educators and SA program planners understand at what point in a learner's L2 learning trajectory a SA is most beneficial. Working with participants on a year long SA program and collecting data midway through their stay could also help explain L2 fluency changes. A delayed post-test that can be easily completed once participants have returned home would also add to our understanding of the long-term impact of SA fluency gains.

With respect to the measurement of L2 oral fluency itself, one limitation of the present study is the lack of measuring and analyzing complexity and accuracy in relation to fluency development (the CAF construct often analyzed in SA literature). Perhaps data looking at the complexity and accuracy of the speech our participants produced would help explain their lack of L2 oral fluency development.

In terms of inhibitory control measurement, the present study has limitations as although non-linguistic inhibitory control was tested used a standard (Simon) test, linguistic inhibitory control was measured by two tests not yet used in the literature – one, an adapted version of Colomé's (2001) test for fluent bilinguals and the other, a new, language switching task. The present study looks at individual differences in inhibitory control but does not look at multiple other cognitive differences (e.g. attention, working memory) that could affect L2 fluency development. Further research is needed to explore inhibitory control ability in L2 learners and its relationship to L2 oral fluency gains, especially in the SA contexts, as this is an area of cognitive fluency almost completely neglected in the literature.

Finally, though it was not central to the objectives of this dissertation, another limitation of this study is the lack of analysis of sociolinguistic factors that could potentially play a large role in L2 oral fluency development. An analysis of participants' social networks while abroad, for example, could shed light onto the lack of their L2 oral fluency development.

Future research in individual differences in L2 fluency acquisition in a SA context could build upon the present study in several ways. First, to better understand individual difference, further research could expand upon the cognitive variables tested, including tests on working memory, attention, and personality. An addition of a context comparison group – or at least a delayed post-test – would help us understand how L2 learning in a SA context for these participants differences to their AH learning. Adapting data collection (for both cognitive tests and questionnaire data) so it is easier for participants to take, perhaps via an app, or website may also increase participants' motivation and willingness to participate. Given the questionnaire data and its relationship to oral fluency measures, **language use opportunities** (and participant comments/insights both before and after the stay) may be critical to running a successful SA program in which participants end their stay with higher linguistic skills and oral fluency ability than they had at the beginning of their stay.

Pedagogical implications and/or potential solutions for SA programs are discussed with concluding remarks in section 7.0.

7 Conclusion and Pedagogical Implications

As study abroad experiences become more popular year after year, it is crucial that we, as applied linguists and educators, understand how L2 oral fluency develops in this context. Second language acquisition itself is increasingly a globally sought-after skill and many participants study abroad with the goal of improving their L2 speech in mind. The aim of the present study was to understand changes in oral fluency that occur as college-aged American learners of Spanish study an L2 on a SA program. The study looked at individual cognitive differences in inhibitory control ability and SA experience factors (such as language use and contact), and their impact on L2 fluency gains, finding unique and highly unexpected results in some cases.

Participants significantly *decreased* their L2 speech rate at post-test, and only experienced gains on a few breakdown fluency measures after a 3-month stay in Barcelona. They spoke for a longer period of time, and had longer fluent runs, indicating that there was some improvement in their fluency. However, participants were not able to speak with less hesitation that they did when they arrived. In fact, for some measures disfluency significantly increased, suggesting, contrary to most of the literature, that the SA environment did not allow participants to speak with less hesitations.

Inhibitory control ability was not found to relate to L2 oral fluency gains. The Language switching task used was a novel task, looking at the impact of a forced switch on oral fluency, and looking at improvement in (L1-L2 and L2-L1) language switching ability at the beginning and end of the SA experience and provided an additional measure of inhibitory control.

Our participants are a surprising group in terms of the lack of their L2 fluency improvement and particularly in terms of their low L2 use (an average of under 30% with high variability). Although overall participants enjoyed their stay, SA programs could potentially be adapted to maximize L2 oral fluency gains by maximizing opportunities for L2 use and out-of-class contact. Perhaps SA experience programs should include a pre-departure program to moderate participants' expectations, and to give them strategies for optimizing language use opportunities and experiencing optimal L2 fluency gains.

Findings from the present study could be applied in pedagogy in the development of SA programs, as well as potentially in a classroom context. As discussed in 6.6, it is crucial to consider details of the SA program context when designing program outcomes. Participants in the present study spent most of their time with their L1 English speaking peers, gained a little vocabulary and had enough Spanish for basic chores. One participant even cited using Spanish "mostly with cab drivers" and another admitted to being "the translator for the group [in which] nobody else spoke Spanish". To maximize our understanding of a lack of L2 fluency gains, perhaps we need to understand if the participants who enroll in SA programs are highly motivated (and what their motivation for learning the L2 is). Alternatively, or perhaps additionally, the SA program itself could create an environment in which participants are more motivated to take advantage of opportunities to gain more exposure to the L2 and more interaction with native speakers.

Given the mostly positive comments from participants about the effectiveness of a SA experience, the oral fluency data indicating the contrary is concerning. The few who were disappointed in their experience cited anxiety, lack of self-confidence, avoidance strategies and even annoyance and frustration at the lack of opportunity to speak the L2 even in Spanish class, surrounded by L1 speaking peers. As such, the present study points to the importance of expanding on L2 fluency research in a growing part of SA research in including *detailed* qualitative information about participant personality, anxiety, motivations, expectations, and experiences *throughout* their stay to better inform research on L2 oral fluency development. Proficiency level (and a threshold before which to participate in a SA experience) is also an important consideration. Our participants tested as quite low-proficiency whereas most L2 fluency studies finding improvements in the SA context look at upper-intermediate to advanced learners.

The concept of a threshold level that needs to be reached before embarking upon a SA experience (see Perez-Vidal, 2014) may be another important consideration in SA program design. Participants should reach an upper intermediate level before arriving, measured objectively on language tests. As educators, we also need to ensure participants are *not* in isolation with their L1 peer group (yet are not completely isolated in an L2 environment either as this may lead to a state of overwhelm and anxiety). They may need to have a course requirement to speak daily with native speakers or take an enforceable language pledge to create an environment more conducive to language learning. Another idea, following Dewey (2018) would be to have students live in foreign language housing at the home campus first, before embarking on a SA, though it is possible this may not be available in many contexts. Another idea could be to create SA programs which pair together local and exchange students for consistent language practice.

Finally, the results of this study may be an eye-opener for SA language school owners and university departments designing SA experiences (and even potential students planning to embark on a SA experience). Most expect the SA environment to foster L2 use and L2 oral fluency development, but this is clearly not always the case. The location and environment of the SA experience should be carefully chosen, outside of a main city centre or at least not in a neighbourhood where participants can manage their daily lives without speaking the L2 in most situations. SA programs have the opportunity to not only increase the opportunity for L2 contact, but also diversify the *type* of language contact participants have. Ample opportunities should be provided for L2 use with native speaker peers in a similar age cohort, and in a variety of settings with high language use *quality*, as previous research has shown that language use in more complex interactions is related to more linguistic gains.

Participants should undertake a mandatory language pledge or similar L2 use agreement to take advantage of their time abroad. Without this, participants run the risk of isolating themselves further from the language and culture and becoming even more dependent on their L1 as they spend time only with American peers. Barcelona– at least in the city centre - may not be the optimal city to for L2 Spanish learning due to both its accessibility to English speaking tourists, and local youth who primarily prefer to speak a regional language in their peer groups. A monolingual region of Spain where Catalan is not present may be more ideal, although this is speculation based on participant comments on the presence of Catalan and would need to be tested. Furthermore, Barcelona is one of Europe's most well-known cities for tourism and nightlife which, if not carefully leveraged as L2 learning opportunities with native speaker peers, could serve distractors to L2 improvement. In sum, to better foster L2 oral fluency development in the SA context, program creators and educators need to seriously consider and monitor participant behaviour throughout the stay, checking in on participants' expectations and progress.

References

Aite, A., Berthoz, A., Vidal, J., Roëll, M., Zaoui, M., Houdé, O. & Borst, G. (2016). Taking third-person perspective requires inhibitory control: Evidence from a developmental negative priming study. *Child Development*, *87*(6), 1825–1840.

Amuzie, G. L. & Winke, P. (2009). Changes in language learning beliefs as a result of study abroad. *System*, *37*(3), 366–379.

Anderson, R. T. (2004). Phonological acquisition in preschoolers learning a second language via immersion: a longitudinal study. *Clinical Linguistics & Phonetics*, *18*(3), 183–210. doi:10.1080/0269920042000193571

Arvidsson, K., Eyckmans, J., Rosiers, A. & Lundell, F. F. (2018). Self-perceived linguistic progress, target language use and personality development during study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, *3*(1), 144–166.

Baddeley, A., Lewis, V. & Vallar, G. (1984). The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology Exploring the articulatory loop, *36(*2): 233-252

Baker-Smemoe, W., Dewey, D. P., Bown, J. & Martinsen, R. A. (2014). Variables affecting L2 gains during study abroad. *Foreign Language Annals*, *47*(3), 464–486.

Ballinger, S. (2017). Examining peer language use and investment in a distinct North American immersion context. *International Multilingual Research Journal*, *11*(3), 184–198.

Baus, C., Costa, A. & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, *142*(3), 402–9. doi:10.1016/j.actpsy.2013.01.010

Beattie, J., Valls-Ferrer, M. & Pérez-Vidal, C. (2014). Listening performance and onset level in formal instruction and study abroad. *Second Language Acquisition in Study Abroad and Formal Instruction Contexts*, 195–216.

Block, D. (2003). *The Social Turn in Second Language Acquisition.* Georgetown University Press.

Briggs, J. G. (2015). Out-of-class language contact and vocabulary gain in a study abroad context. *System*, *53*, 129–140.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy* (Vol. 129).

Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, *1*(2), 131–151.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535–544.

Chirkov, V. I., Safdar, S., De Guzman, J. & Playford, K. (2008). Further examining the role motivation to study abroad plays in the adaptation of international students in Canada. *International Journal of Intercultural Relations*, *32*(5), 427–440.

Cigliana, K. A. & Serrano, R. (2016). Individual differences in US study abroad students in Barcelona. *Study Abroad Research in Second Language Acquisition and International Education*, *1*(2), 154–185.
Collentine, J. (2009). 13 Study Abroad Research: Findings, Implications, and Future Directions. *The Handbook of Language Teaching*, 218.
Collentine, J. & Freed, B. F. (2004). Learning Context and Its Effects On Second Language Acquistion. *Studies in Second Language Acquisition*, *26*(2), 153-171. doi:10.1017/S0272263104262015
Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language*, *45*(4), 721–736.
Costa, A., Caramazza, A. & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283.
Costa, A. & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511. doi:10.1016/j.jml.2004.02.002
Crookes, G. (2008). Second Language Speech Production Research. *Studies in Second Language Acquisition*, *13*(02), 113. doi:10.1017/S0272263100009918
Cubillos, J. H., Chieffo, L. & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals, 41*(1), 157–186.
Cucchiarini, C., Strik, H. & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989–99.
D'Amico, M. (2012). L2 fluency and willingness to communicate: The impact of short-term study abroad versus at-home study. *US-China Foreign Language*, *10*(10), 1608–1625.
Dalton-Puffer, C. (2008). *Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe*. na.
Darcy, I., Mora, J. C. & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, n/a–n/a. doi:10.1111/lang.12161
De Bot, K. D. E. (1992). A Bilingual Production Model : Levelt's 'Speaking' Model Adapted, *Applied linguistics*, *13*(1), 1–24.
De Bruin, A., Roelofs, A., Dijkstra, T. & Fitzpatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: fMRI evidence from trilingual speakers. *NeuroImage*, *90*, 348–59. doi:10.1016/j.neuroimage.2013.12.049
De Jong, N. (2013). Analysis of fluency. [Conference presentation]. Langsnap Workshop. Southampton, England.
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*, *54*(2), 113–132. doi:10.1515/iral-2016-9993
De Jong, N. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, 1–18.
De Jong, N., Grouenhout, R., Schoonen, R. & Hulstijn, J. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(02), 223–243. doi:10.1017/S0142716413000210

De Jong, N. & Perfetti, C. a. (2011). Fluency Training in the ESL Classroom: An Experimental Study of Fluency Development and Proceduralization. *Language Learning*, *61*(2), 533–568. doi:10.1111/j.1467-9922.2010.00620.x

De Jong, N., Steinel, M. P., Florijn, A., Schoonen, R. & Hulstijn, J. (2012). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(05), 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), 385–390.

Declerck, M. & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin \& Review*, *22*(6), 1630–1645.

Deepthi, M. & Nataraja, N. (2014). Inhibitory control process and lexical robustness in multilinguals. *Journal of the All India Institute of Speech \& Hearing*, *33*.

DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, *43*(1), 80–92.

Derwing, T., Munro, M. J., Thomson, R. I. & Rossiter, M. J. (2009a). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(04), 533–557.

Derwing, T., Munro, M. J., Thomson, R. I. & Rossiter, M. J. (2009b). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(04), 533. doi:10.1017/S0272263109990015

Dewaele, J.-M. & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, *28*(2), 355–365.

Dewaele, J.-M. & MacIntyre, P. (2016). Foreign language enjoyment and foreign language classroom anxiety. *Language Teaching Research*, 22(6), 676–697. https://doi.org/10.1177/1362168817692161

Dewey, D. P., Ring, S., Gardner, D. & Belnap, R. K. (2013). Social network formation and development during study abroad in the Middle East. *System*, *41*(2), 269–282.

Diaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, *26*(2), 249–273.

Dornyei, Z. (2013). *The Psychology of Second Language Acquisition-Oxford Applied Linguistics*. Oxford University Press.

Du, H. (2013). The development of Chinese fluency during study abroad in China. *The Modern Language Journal*, *97*(1), 131–143.

Duran-Karaoz, Z. & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behaviour: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*, 1–25.

Edmonds, Amanda & Gudmestad, Aarnes. (2021). The role of transparency in grammatical gender marking among stay abroad learners of Spanish and French. In M. Howard (Ed.), *Study Abroad and the Second Language Learner: Expectations, Experiences and Development.* Bloomsbury Academic. doi: 10.5040/9781350104228.ch-012.

Eisenstein, M. (2008). Native Reactions to Non-Native Speech: A Review of Empirical Research. *Studies in Second Language Acquisition*, *5*(02), 160. doi:10.1017/S0272263100004836

Ellis, N. C. (2002). Frequency effects in language processing: A Review with Implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2), 143–188. doi:10.1017/S0272263102002024
Ellis, R., Basturkmen, H. & Loewen, S. (2001). Preemptive focus on form in the ESL classroom. *Tesol Quarterly*, *35*(3), 407–432.
Field, A. (2005). Discovering Statistics Using SPSS (Introducing Statistical Methods). Sage Publications, London.
Fillmore, L. W. (1979). Individual Differences in Second Language Acquisition. In *Individual Differences in Language Ability and Language Behavior* (pp. 203–228). Elsevier. doi:10.1016/b978-0-12-255950-1.50017-2
Fillmore, L. W. (1991). When learning a second language means losing the first. *Early Childhood Research Quarterly*, *6*(3), 323–346. doi:10.1016/S0885-2006(05)80059-6
Foisy, L.-M. B., Ahr, E., Masson, S., Houdé, O. & Borst, G. (2017). Is inhibitory control involved in discriminating pseudowords that contain the reversible letters b and d? *Journal of Experimental Child Psychology*, *162*, 259–267.
Foster, P. (2009). Native Speakers and Task Performance: Comparing Effects on Complexity, Fluency, and Lexical Diversity, (December), 866–896.
Foster, P., Tonkyn, A. & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.
Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In *Perspectives on fluency* (pp. 243–265).
Freed, B. F. (1995). *Second language acquisition in a study abroad context*. John Benjamins Publishing.
Freed, B. F., Dewey, D. P., Segalowitz, N. & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, *26*(2), 349–356.
Freed, B. F., Segalowitz, N. & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(02), 275–301. doi:10.1017/S0272263104262064
Garcia-Amaya, L. (2012). Second language fluency and cognition: The study of Spanish second language development in an overseas immersion program and an at-home foreign language classroom. *Bloomington: Indiana University Dissertation*.
Gardner, R. C. & MacIntyre, P. D. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching*, *26*(1), 1–11.
Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Newbury House Publishers.
Goldrick, M., Runnqvist, E. & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, *25*(4), 1031–1036.
Granena, G. (2018). Cognitive aptitudes and L2 speaking proficiency: Links between Llama and Hi-Lab. *Studies in Second Language Acquisition*, (2), 313-336. doi:10.1017/S0272263118000256
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(02), 67. doi:10.1017/S1366728998000133
Grey, S., Cox, J. G., Serafini, E. J. & Sanz, C. (2015). The Role of Individual Differences in the Study Abroad Context: Cognitive Capacity and Language Development During Short-Term Intensive Language Exposure. *The Modern Language Journal*, *99*(1), 137–157.

Groenhout, R., Schoonen, R., Hulstijn, J. & De Jong, N. (2013). Second language fluency : Speaking style or proficiency ? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*, 1–21. doi:10.1017/S0142716413000210

Hassall, T. (2018). Leave-taking in Indonesian during short-term study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, *3*(1), 84–116.

Hernández, T. A. (2010). Promoting speaking proficiency through motivation and interaction: The study abroad and classroom learning contexts. *Foreign Language Annals*, *43*(4), 650–670.

Hessel, G. (2017). A new take on individual differences in L2 proficiency gain during study abroad. *System*, (66), 39–55.

Howard, M; (2005) On the role of context in the development of learner language: Insights from study abroad research. *International Journal of Applied Linguistics*, 147,1-20

Howard, Martin. (2015). On the role of naturalistic and classroom exposure in the acquisition of socio-phonological variation: A longitudinal study of French liaison. *Journal of Applied Linguistics and Professional Practice*. *5*. 159-179. doi:10.1558/japl.v5i2.159.

Huensch, A. & Tracy-Ventura, N. (2017). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, *38*(4), 755–785.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (Cambridge Applied Linguistics, pp. 258-286). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524780.011

Isabelli, C. (2004). Study abroad for advanced foreign language majors: Optimal duration for developing complex structures. *Advanced Foreign Language Learning: A Challenge to College Programs*, 114–130.

Izmaylova--Culpepper, A. & Olovson, B. (2017). Exploring Learner Language Development During Short-Term Study Abroad. *Performance+ Proficiency= Possibilities*, 109.

Izura, C., Cuetos, F. & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1).

Jackson, J. (2016). The language use, attitudes, and motivation of Chinese students prior to a semester-long sojourn in an English-speaking environment. *Study Abroad Research in Second Language Acquisition and International Education*, *1*(1), 4–33.

Juan-Garau, M. (2015). Speaking an additional language: Can study abroad do the trick?. *Language Value*. 46-61. doi: 10.6035/LanguageV.2015.7.4.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854. doi:10.1111/lang.12084

Kahng, J. (n.d.). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, 1–24.

Keppie, C., Lindberg, R. & Thomason, S. (2016). The Benefits of Study Abroad on the Fluency of Learners of French as a Second Language. *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée*, *19*(2), 44–63.

Kim, J., Dewey, D. P., Baker-Smemoe, W., Ring, S., Westover, A. & Eggett, D. L. (2015). L2 development during study abroad in China. *System*, *55*, 123–133.

Kinginger, C. (2011). Enhancing language learning in study abroad. *Annual Review of Applied Linguistics*, *31*, 58–73.
Kinginger, C., Wu, Q., Lee, S.-H. & Tan, D. (2016). The short-term homestay as a context for language learning. *Study Abroad Research in Second Language Acquisition and International Education*, *1*(1), 34–60.
Kirsner, K., Dunn, J. & Hird, K. (2003). Fluency: Time for a paradigm shift. *DiSS.* 13-16.
Klee, C., Lynch, A. & Tarone, E. (1998). *Research and Practice in Immersion Education: Looking Back and Looking Ahead: Selected Conference Proceedings*. Center for Advanced Research on Language Acquisition, University of Minnesota.
Knight, S. M. & Schmidt-Rinehart, B. C. (2010). Exploring conditions to enhance student/host family interaction abroad. *Foreign Language Annals*, *43*(1), 64–71.
Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164. doi:10.1016/j.system.2004.01.001
Krashen, S. (1998). Comprehensible output? *System*, *26*(2), 175–182. doi:10.1016/S0346-251X(98)00002-5
Krashen, S. & Gregg, K. (2012). Issues and Implications The Input Hypothesis: *20*(1), 116–122.

LaBrozzi, R. M. (2012). The role of study abroad and inhibitory control on processing redundant cues. In *Selected proceedings of the 14th Hispanic linguistics symposium* (pp. 228–241).
Lara, A. R. (2014). *Complexity, accuracy and fluency development through study abroad programmes varying in duration*.
Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford University.
Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, *5*(1), 3–18.
Lasagabaster, D. & Sierra, J. M. (2009). Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal*, *1*(2), 4–17.
Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–43. doi:10.3758/s13428-011-0146-0
Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.
Leonard, K. R. & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, *101*(1), 179–193.
Lev-Ari, S. & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*(5), 320–331. doi:10.1016/j.wocn.2013.06.002
Levelt, W. J. (1999). Producing spoken language: A blueprint of the speaker. In *The neurocognition of language* (pp. 83–122). Oxford University Press.
Liddicoat, A. J. (2008). Pedagogical practice for integrating the intercultural in language teaching and learning. *Japanese Studies*, *28*(3), 277–290.
Linck, J. A., Hoshino, N. & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *The Mental Lexicon*, *3*(3), 349–374.
Linck, J. A., Kroll, J. F. & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515.

Linck, J. A., Schwieter, J. W. & Sunderman, G. (2011). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, *15*(03), 651–662. doi:10.1017/S136672891100054X

Linck, J. A. & Weiss, D. J. (2015). Can working memory and inhibitory control predict second language learning in the classroom? *SAGE Open*, *5*(4), 2158244015607352–. doi:10.1177/2158244015607352

Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*, *8*(3), 189–215.

Llanes, À. (2016). The influence of a short stay abroad experience on perceived foreign accent. *Study Abroad Research in Second Language Acquisition and International Education*, *1*(1), 88–106.

Long, M. (1996). The role of the linguistic environment in second language acquisition. *Handbook of Second Language Acquisition*.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals*, *43*(3), 488–503. doi:10.1111/j.1944-9720.2010.01094.x

Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*(1), 37-66. doi:10.1017/S0272263197001034, 37–66.

Macaro, E. (2001). Learning strategies in second and foreign language classrooms. *London: Continuum*, *282*.

Marijuan, S. & Sanz, C. (2018). Expanding boundaries: Current and new directions in study abroad research and practice. *Foreign Language Annals*, *51*(1), 185–204.

Marqués-Pascual, L. (2011). Study Abroad, Previous Language Experience, and Spanish L2 Development. *Foreign Language Annals*, *44*(3), 565–582.

Martinsen, R. A., Baker, W., Bown, J. & Johnson, C. (2011). The Benefits of Living in Foreign Language Housing: The Effect of Language Use and Second-Language Type on Oral Proficiency Gains. *The Modern Language Journal*.

Mehrang, F. & Rahimpour, M. (2010). The impact of task structure and planning conditions on oral performance of EFL learners. *Procedia - Social and Behavioral Sciences*, *2*(2), 3678–3686. doi:10.1016/j.sbspro.2010.03.572

Meuter, R. F. & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40.

Misra, M., Guo, T., Bobb, S. C. & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*(1), 224–237. doi:10.1016/j.jml.2012.05.001

Mora, J. C. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad. *Language Acquisition in Study Abroad and Formal Instruction Contexts*, 167–194.

Mora, J. C. & Valls-Ferrer, M. (2012). Oral Fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*(4), n/a–n/a. doi:10.1002/tesq.34

Morita, N. (2004). Negotiating participation and identity in second language academic communities. *Tesol Quarterly*, *38*(4), 573–603.

Nation, P. (1989). Improving speaking fluency. *System*, *17*(3), 377–384.

Netelenbos, N., Li, F. & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism*, *20*(3), 346–357.
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*(04), 557–581. doi:10.1017/S027226310707043X
O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and Comprehensibility of Native and Nonnative German Speech. *Language Learning*, *64*(4), 715–748. doi:10.1111/lang.12082
OECD. (2012). How many students study abroad and where do they go? *In Education at a Glance 2012: Highlights*, 24–27.
Olson, D. J. (2016). The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics*, *37*(3), 725–756.
Opp-Beckman, L. & Klinghammer, S. J. (2006). *Shaping the Way We Teach English: Successful Practices Around the World: Instructor's Manual*. Office of English Language Programs, Bureau of Educational and Cultural Affairs.
Pérez-Vidal, C. (2014). *Language acquisition in study abroad and formal instruction contexts* (Vol. 13). John Benjamins Publishing Company.
Perrachione, T. K., Ghosh, S. S., Ostrovskaya, I., Gabrieli, J. D. & Kovelman, I. (2017). Phonological working memory for words and nonwords in cerebral cortex. *Journal of Speech, Language, and Hearing Research*, *60*(7), 1959–1979.
Philipp, A. M. & Koch, I. (2009). Inhibition in language switching: what is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(5), 1187–95. doi:10.1037/a0016376
Pinget, a.-F., Bosker, H. R., Quene, H. & De Jong, N. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349–365. doi:10.1177/0265532214526177
Pivneva, I., Palmer, C. & Titone, D. (2012). Inhibitory control and L2 proficiency modulate bilingual language production: Evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology*, *3*, 57.
Pladevall-Ballester, E. & Vallbona, A. (2016). CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills. *System*, *58*, 37–48.
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), 96–112.
Préfontaine, Y. & Kormos, J. (2016). A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 151–169. doi:10.1515/iral-2016-9995
Préfontaine, Y., Kormos, J. & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*, *33*(1), 53–73.
Puri'c, D., Vuksanovi'c, J. & Chondrogianni, V. (2017). Cognitive advantages of immersion education after 1year: Effects of amount of exposure. *Journal of Experimental Child Psychology*, *159*, 296–309.
Ranta, L. & Meckelborg, A. (2013). How much exposure to English do international graduate students really get? Measuring language use in a naturalistic setting. *Canadian Modern Language Review*, *69*(1), 1–33.

Rezai, M. J. & Okhovat, B. (2016). The effect of working memory on EFL learners' oral fluency. *International Journal of English Linguistics*, *6*(5).
Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441. doi:10.1080/01638539109544795
Robinson. (2001). *Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition*. Greenwood Publishing Group.
Roquet, H., Llopis, J. & Pérez-Vidal, C. (2016). Does gender have an impact on the potential benefits learners may achieve in two contexts compared: formal instruction and formal instruction+ content and language integrated learning? *International Journal of Bilingual Education and Bilingualism*, *19*(4), 370–386.
Sadri Mirdamadi, F. & De Jong, N. (2014). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research*, *31*, 105–116. doi:10.1177/0267658314554498
Saito, K., Tierney, A. & Sun, H. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental, prosodic and temporal sensitivity and performance via behavioral and neurophysiological measures. *Bilingualism: Language and Cognition*.
Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcomes. *Language Acquisition in Study Abroad and Formal Instruction Contexts. Amsterdam: John Benjamins*.
Schachter, S., Christenfeld, N., Ravina, B. & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*(3), 362.
Schmidt, R. (1992). Psychological Mechanisms Underlying Second Language Fluency, (March 1991), 357–385.
Segalowitz, N. (1997). Individual differences in second language acquisition. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, 85–112.
Segalowitz, N. (2007). Access Fluidity, Attention Control, and the Acquisition of Fluency in a Second Language, *41*(1), 181–186.
Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 79–95. doi:10.1515/iral-2016-9991
Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N. & D'iaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*, 1–18.
Segalowitz, N. & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*(2), 173–199.
Segalowitz, N. & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. *Handbook of Bilingualism: Psycholinguistic Approaches*, 371–388.
Serrano, R., Llanes, À. & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System*, *39*(2), 133–143. doi:10.1016/j.system.2011.05.002

Serrano, R., Tragant, E. & Llanes, À. (2012). A Longitudinal Analysis of the Effects of One Year Abroad. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *68*(2), 138–163. doi:10.3138/cmlr.68.2.138

Service, E. (2007). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, *45A*(1), 21–50.

Skehan, P., Foster, P. & Shum, S. (2016). Ladders and Snakes in Second Language Fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 97–111. doi:10.1515/iral-2016-9992

Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(2), 174–215. doi:10.1037/0278-7393.6.2.174

Sunderman, G. & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, *30*(1), 79–99.

Suzuki, S. & Kormos, J. (2019). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, 1–25.

Swain, M. (1996). Integrating language and content in immersion classrooms: research perspectives. *Canadian Modern Language Review*, *52*(4), 529–48.

Swain, M. (2000). French immersion research in Canada: Recent contributions to SLA and applied linguistics. *Annual Review of Applied Linguistics*, *20*, 199–212.

Swain, M. & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. Longman, Pearson Education; http://www. longman. co. uk/.

Syder, F. H. & Pawley, A. (2014). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In *Language and communication* (pp. 203–239). Routledge.

Sylvén, L. K. (2004). *Teaching in English or English teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*.

Szyszka, M. (2017). Foreign Language Anxiety in the Context of Foreign Language Oral Performance, Language and Pronunciation Learning Strategies. In *Pronunciation Learning Strategies and Language Anxiety* (pp. 51–85). Springer.

Tavakoli, P. (2011). Pausing patterns: differences between L2 learners and native speakers. *ELT Journal*, *65*(1), 71–79. doi:10.1093/elt/ccq020

Tavakoli, P., Campbell, C. & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*, *50*(2), 447–471.

Tavakoli, P. & Foster, P. (2008). Task Design and Second Language Performance: The Effect of narrative type on learner ouput. *Language Learning*, *58*. 439–473. doi: 10.1111/j.1467-9922.2008.00446.x.

Tavakoli, P. & Hunter, A.-M. (2018). Is fluency being "neglected"in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, *22*(3), 330–349.

Tavakoli, P. & Skehan, P. (2005). Strategic planning, task structure, and performance testing. *Planning and Task Performance in a Second Language*, *239273*.

The chambers dictionary. (1998).

Torres, J. & Sanz, C. (2015). Is there a cognitive advantage for Spanish heritage bilinguals? A first look. *Heritage Language Journal*, *12*(3), 292.

Towell, R., Hawkins, R. & Bazergui, N. (1996). The Development of Fluency in Advanced Learners of French. *Applied Linguistics*, *17*(1), 84–119. doi:10.1093/applin/17.1.84

Tracy-Ventura, N., Dewaele, J.-M., Köylü, Z. & McManus, K. (2016). Personality changes after the "year abroad"? *Study Abroad Research in Second Language Acquisition and International Education*, *1*(1), 107–127.

Tragant, E., Serrano, R. & Llanes, À. (2017). Learning English during the summer: A comparison of two domestic programs for pre-adolescents. *Language Teaching Research*, *21*(5), 546–567.

Trentman, E. (2017). Oral fluency, sociolinguistic competence, and language contact: Arabic learners studying abroad in Egypt. *System*, *69*, 54–64.

Tullock, B. & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*.

University of Alberta Exchange Programs. (n.d.). Retrieved January 15, 2020, from https://www.ualberta.ca/international/go-abroad/returning-to-ualberta/finalizing-credit-transfer

Valls Ferrer, M. & others. (2011). *The development of oral fluency and rhythm during a study abroad period*. Universitat Pompeu Fabra.

Van de Craen, P., Mondt, K., Allain, L. & Gao, Y. (2007). Why and how CLIL works. An outline for a CLIL theory. *Views*, *16*(3), 70–78.

Wen, Z. E., Biedro'n, A. & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*(1), 1–31.

Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT JOURNAL*, *29*(2), 209.

Wright, C. & Tavakoli, P. (2016). New directions and developments in defining, analyzing and measuring L2 speech fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 73–77. doi:10.1515/iral-2016-9990

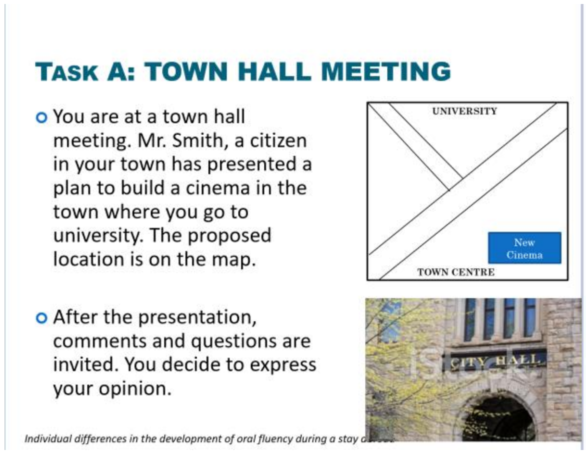
Yager, K. (1998). Learning Spanish in Mexico: The effect of informal contact and student attitudes on language gain. *Hispania*, 898–913.

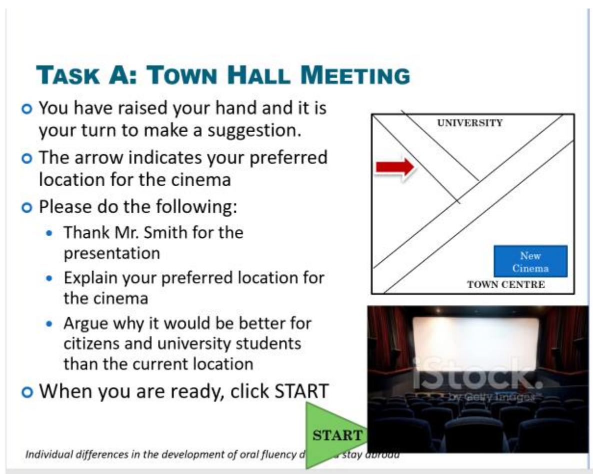
Yang, J.-S. (2016). The effectiveness of study-abroad on second language learning: a meta-analysis. *Canadian Modern Language Review*, *72*(1), 66–94. doi:10.3138/cmlr.2344

Zolta, D. O. (2006). The L2 Motivational Self System, (1959), 9–42.

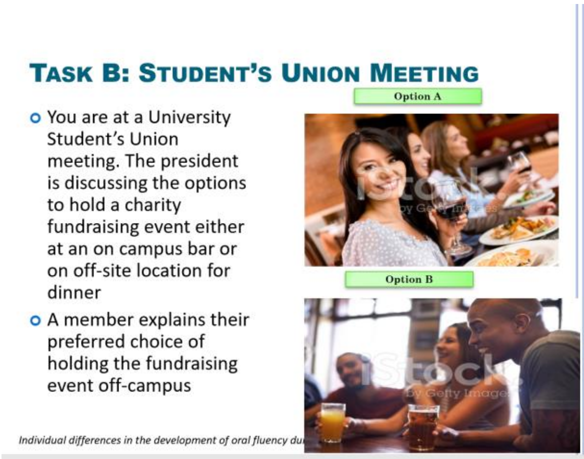
Appendix A - Oral fluency picture prompts

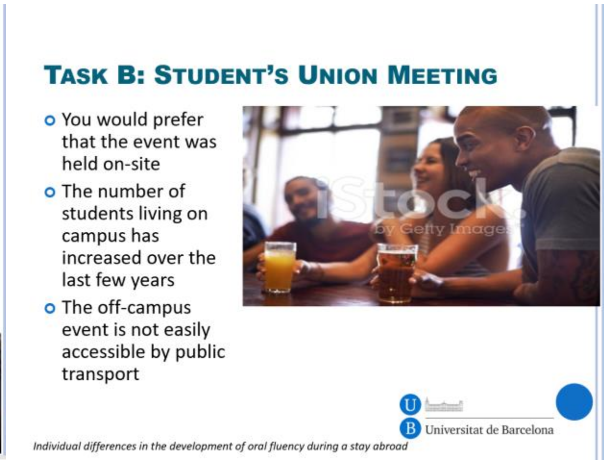
Version 1: Task A

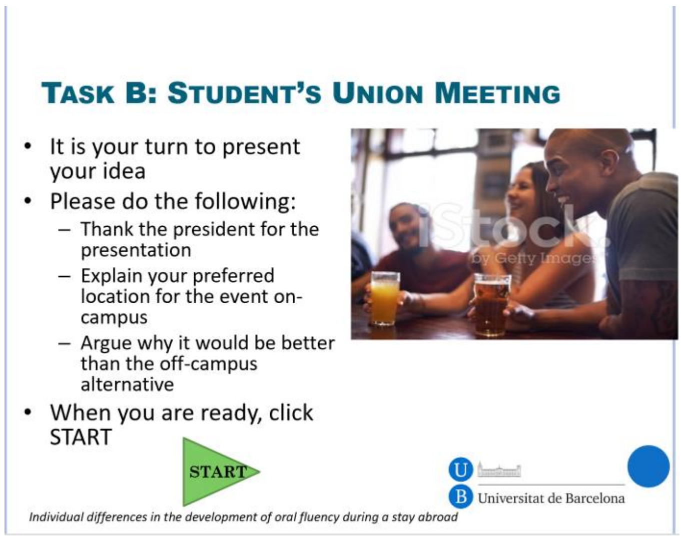




Version 1: Task B







Version 2: Task A

Version 2: Task B

TASK B (VERSION 2): GRADUATION EVENT

You are with a group of friends who are planning an event for their university graduation day and need to decide on a location

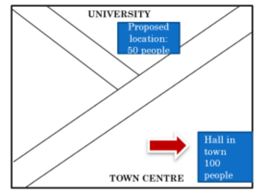
Universitat de Barcelona

TASK B (VERSION 2): GRADUATION EVENT

There are two options for this event. A group member suggests locating the party at a campus bar near the university

The map to the right shows Option A, which one group member suggests. It is an on-campus party at a bar.

You prefer Option B, having dinner at a restaurant that can hold many people



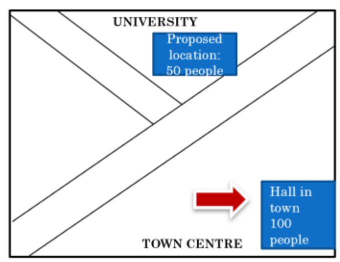
Universitat de Barcelona

Individual differences in the development of oral fluency during a stay abroad

TASK B (VERSION 2): GRADUATION EVENT

It is your turn to speak.

Please do the following
- Thank your friend for the suggestion
- Explain your preferred location for the graduation event, indicated by the red arrow
- Explain why it is a better location that the alternative
-



Universitat de Barcelona

Individual differences in the development of oral fluency during a stay abroad

Appendix B - Language switching task picture prompts and vocabulary lists

Suitcase Story Picture Prompt

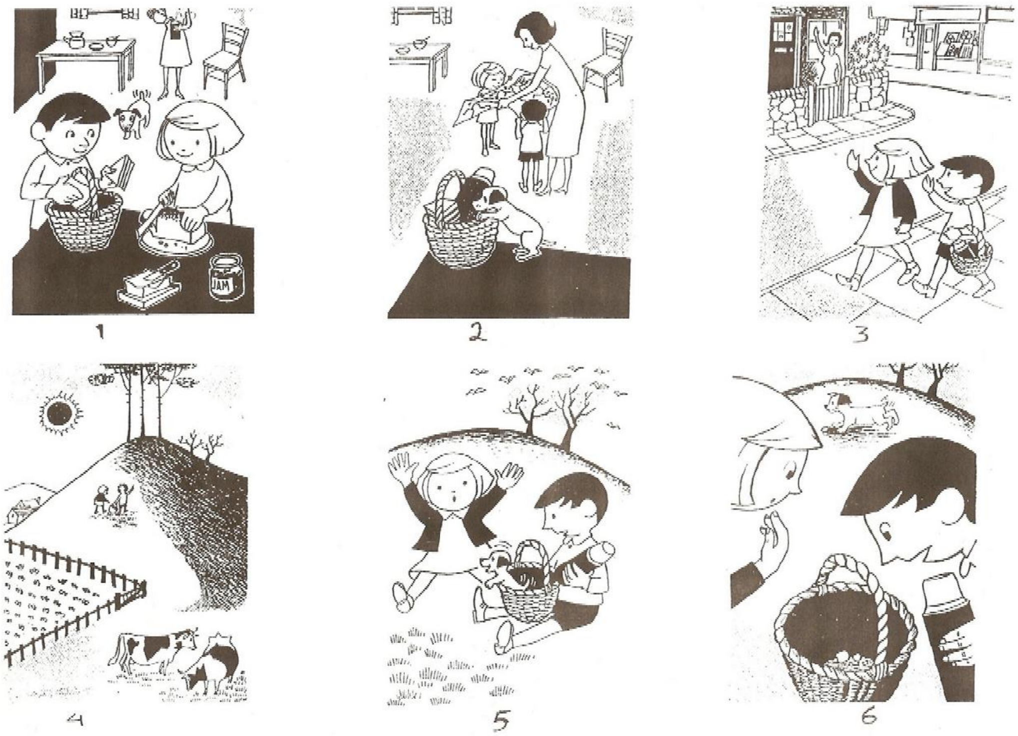
THE SUITCASE STORY

The "Suitcase Story" may be used for research purposes only, provided that the user cites the following source:

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(4), 533-557.

Dog Story Picture Prompt

Oral Picture-Narrative Task: The Dog's Story



Dog-Story Task Vocabulary List For Participants

NOUNS	
English Word	**Spanish Translation**
jam	la mermelada
butter	la mantequilla
basket	la cesta
thermos	el termo
map	el mapa
porch	el porche
sidewalk	la acera
hill	la colina
bird	el pájaro
fence	la valla
grass	la hierba
countryside	el campo
crumbs	las migas

VERBS	
English Word	**Spanish Translation**
To have a picnic	hacer un picnic
Prepare food	preparar la comida
To show	enseñar
To hide	esconder
To wave goodbye	despedir/decir adiós
To be excited	tener ilusión / tener ganas de
To take out	sacar
To jump	saltar
To run away	huir
To realize	darse cuenta

Appendix C - Letter (Phoneme) decision task words

Type	Word	English	YES (L2 only)	NO (L1 only)	NO (Neither)
Practice	casa	house	C	H	S
Practice	mesa	table	M	T	F
Practice	nube	cloud	N	C	R
Target	boca	mouth	B	M	F
Target	bolígrafo	pen	B	P	C
Target	brazo	arm	B	M	T
Target	caballo	horse	C	H	D
Target	conejo	rabbit	C	R	P
Target	dedo	finger	D	F	M
Target	falda	skirt	F	S	H
Target	gato	cat	G	C	D
Target	maleta	suitcase	M	S	H
Target	mano	hand	M	H	B
Target	mosca	fly	M	F	G
Target	**niño**	**boy**	**N**	**Y**	**P**
Target	pájaro	bird	P	B	L
Target	perro	dog	P	D	T
Target	pierna	leg	P	L	T
Target	puerta	door	P	D	F
Target	reloj	clock	R	C	S
Target	taza	cup	T	C	F
Target	tenedor	fork	T	F	L
Target	**toro**	**bull**	T	B	S

Type	Word	English	YES (L2 only)	YES (BOTH)	NO (Neither)
Practice	cometa	kite	O	E	L
Practice	castillo	castle	O	**C**	**R**
Practice	piscina	pool	C	P	**T**
Filler	**bandera**	flag	B	A	T
Filler	barco	boat	C	B	P
Filler	**bolsillo**	pocket	B	O	D
Filler	camión	truck	M	C	P
Filler	cesta	basket	C	E	D
Filler	corazón	heart	C	A	L
Filler	fresa	strawberry	F	A	D
Filler	gafas	glasses	F	G	T
Filler	guante	glove	T	G	S
Filler	lápiz	pencil	Z	L	T
Filler	lechuga	lettuce	G	L	S
Filler	libro	book	L	B	T
Filler	**llave**	key	L	E	R
Filler	luna	moon	L	N	T
Filler	manzana	apple	M	A	R
Filler	pan	bread	P	A	G
Filler	pelota	ball	P	A	D
Filler	ratón	mouse	R	O	P
Filler	silla	chair	S	A	D
Filler	**zanahoria**	carrot	Z	A	S

Appendix D - LEXTALE-ESP vocabulary test

Instructions:

"Hi, this is a test of Spanish vocabulary. You will get 180 sequences of letters that look "Spanish". Only some of them are real words. Please, indicate the words you know (or of which you are convinced they are Spanish words, even though you would not be able to give their precise meaning).

Be careful, however: Errors are penalised. So, there is no point in trying to increase your score by adding tallies to "words" you've never seen before! All you have to do is to tick the box next to the words you know.

If, for instance, in the example below you recognise "sí", "sacapuntas","bien", and "casa", you indicate this as follows:

Estímulo	Palabra?
depiste	
sí	√
coné	
calpar	
joten	
sacapuntas	√

Estímulo	Palabra?
priba	
pelasula	
bien	√
casa	√
lejo	
pretantas	

The results of this test are only useful if you do not use a dictionary and if you work on your own!"

Estímulo	Palabra?
terzo	
pellizcar	
pulmones	
batillón	
zapato	
tergiversar	
pésimo	
cadeña	
hacha	
antar	
cenefa	
asesinato	
helar	
yunque	
regar	
abracer	
floroso	
arsa	
brecedad	
ávido	
capillo	
lacayo	
lampera	
látigo	
bisagra	
secuestro	
acutación	
merodear	
decar	
alardio	

Estímulo	Palabra?
pandilla	
fatacidad	
pauca	
aviso	
rompido	
loro	
granuja	
estornudar	
torpe	
alfombra	
rebuscar	
cadallo	
canela	
cuchara	
jilguero	
martillo	
cartinar	
ladrón	
ganar	
flamida	
candado	
camisa	
vegada	
fomentar	
nevar	
musgo	
tacaño	
plaudir	
besar	
matar	

Estímulo	Palabra?
seda	
flaco	
esposante	
orgulloso	
bizcocho	
hacido	
cabello	
alegre	
engatusar	
temblo	
polvoriento	
pemición	
hervidor	
cintro	
yacer	
atar	
tiburón	
frondoso	
tropaje	
hormiga	
pozo	
empirador	
guante	
escuto	
laúd	
barato	
grodo	
acantilado	
prisa	
clavel	

Estímulo	Palabra?
terzo	
pellizcar	√
pulmones	√
batillón	
zapato	√
tergiversar	√
pésimo	√
cadeña	
hacha	√
antar	
cenefa	√
asesinato	√
helar	√
yunque	√
regar	√
abracer	
floroso	
arsa	
brecedad	
ávido	√
capillo	
lacayo	√
lampera	
látigo	√
bisagra	√
secuestro	√
acutación	
merodear	√
decar	
alardio	

Estímulo	Palabra?
pandilla	√
fatacidad	
pauca	
aviso	√
rompido	
loro	√
granuja	√
estornudar	√
torpe	√
alfombra	√
rebuscar	√
cadallo	
canela	√
cuchara	√
jilguero	√
martillo	√
cartinar	
ladrón	√
ganar	√
flamida	
candado	√
camisa	√
vegada	
fomentar	√
nevar	√
musgo	√
tacaño	√
plaudir	
besar	√
matar	√

Estímulo	Palabra?
seda	√
flaco	√
esposante	
orgulloso	√
bizcocho	√
hacido	
cabello	√
alegre	√
engatusar	√
temblo	
polvoriento	√
pemición	
hervidor	√
cintro	
yacer	√
atar	√
tiburón	√
frondoso	√
tropaje	
hormiga	√
pozo	√
empirador	
guante	√
escuto	
laúd	√
barato	√
grodo	
acantilado	√
prisa	√
clavel	√

Appendix E - Questionnaire

University of Barcelona - Study Abroad Questionnaire

Individual Differences in Oral Fluency after a Stay Abroad

* Required

PERSONAL INFORMATION

1. **Participant Number** *

2. **First and last name** *

3. **Email address** *

4. **Phone/WhatsApp** *

5. **Date of Birth (MM/DD/YY)** *

   Example: December 15, 2012

6. **Gender** *

   Mark only one oval.

   ◯ Female
   ◯ Male

7. **Nationality** *

8. **Are any members of your immediate family of Spanish speaking heritage?** *

   Mark only one oval.

   ◯ Yes
   ◯ No

9. **Date of arrival in Spain (MM/DD/YY)** *

   Example: December 15, 2012

University of Barcelona Study-Abroad End of Semester Questionnaire

PhD Study: Individual Differences in Oral Fluency Development in the Study Abroad Context

* Required

PERSONAL INFORMATION

1. **Participant Number** *

2. **First and last name** *

EXPERIENCE LIVING IN SPAIN

3. **Please list each person you live with, (if any) and their name, age, nationality and the % of the time you spend speaking Spanish, English and Other languages with them** *
   Example: SOFIA, 19, CANADIAN, 60% SPANISH, 40% ENGLISH, 0% OTHER

4. **Please list the FIVE people you have the most contact with in Spain and their name, age, nationality, relationship to you and the % of the time you spend speaking Spanish, English and Other languages with them** *
   Example: GUILLERMO, 30, SPANISH, FRIEND, 100% SPANISH, 0% ENGLISH

5. **On average, what percentage of your day do you spend communicating (speaking, writing, texting, etc.) in Spanish? NOTE: The total % you report of all languages used should add up to 100%. E.g. 50% English, 50% Spanish** *

6. **On average, what percentage of your day do you spend communicating (speaking, writing, texting, etc.) in English? REMEMBER: The total % you report of all languages used should add up to 100%** *

7. **On average, what percentage of your day do you spend communicating (speaking, writing, texting, etc.) in any other language (please specify)? REMEMBER: The total % you report of all languages used should add up to 100%** *

8. **What extracurricular activities are you involved in, where you use Spanish? (e.g. part-time job, church group, sports team)** *

AUDIOVISUAL INPUT - SPAIN

9. **While in Spain, how often have you been exposed to audiovisual material in Spanish (films, television series, etc.)?** *

   Mark only one oval.

   - Never/hardly ever
   - A few times per year
   - Once to twice per month
   - Once per week
   - Several times per week

10. **Was the audiovisual material you watched in Spanish (films, television series, etc) subtitled?**

    If the content was sometimes subtitled, please specify the approximate percentage (%) of time in the 'Other' box

    Mark only one oval.

    - Yes
    - No
    - Other:

11. **What was the language of the subtitles?** *

Mark only one oval.

- English
- Spanish
- Other: ______________________

12. **What did you watch in Spanish? (names of films, television shows, etc.)** *

____

____

____

____

____

13. **Do you think that being exposed to audiovisual input helps you learn Spanish?** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all Helpful	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely Helpful

YOUR STUDY ABROAD EXPERIENCE

14. **How comfortable do you feel speaking Spanish?** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all comfortable	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very comfortable

15. **It is important to me to use (or eventually use) Spanish as a native speaker.** *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all important	◯	◯	◯	◯	◯	◯	◯	Very important

16. **If you are speaking Spanish, and the person with whom you are speaking switches to English, how likely are you to RESIST switching to English?** *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all likely	◯	◯	◯	◯	◯	◯	◯	Very likely

17. **What aspects do you feel have helped you learn Spanish most while in Barcelona? ***

18. **How do you think your study abroad experience has affected your level of Spanish so far? Please provide details. ***

19. **What has your experience been with the Catalan language? ***

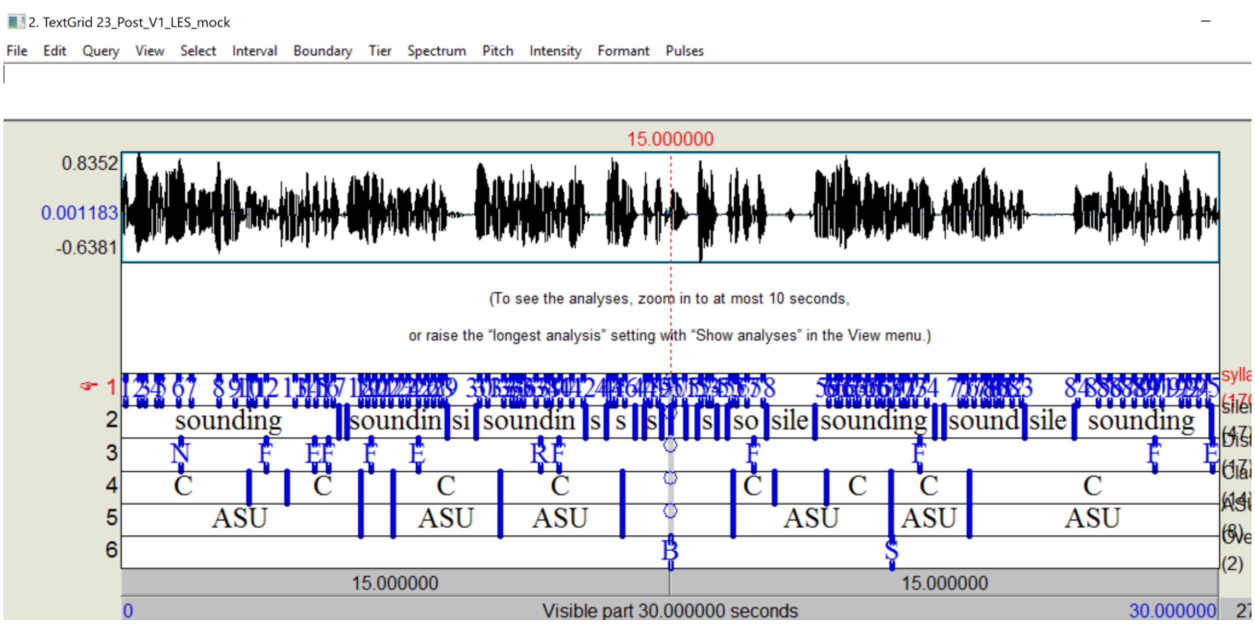
Appendix F - PRAAT

The figure below shows one sample PRAAT Textgrid (Participant 23, Language Switching test English to Spanish).

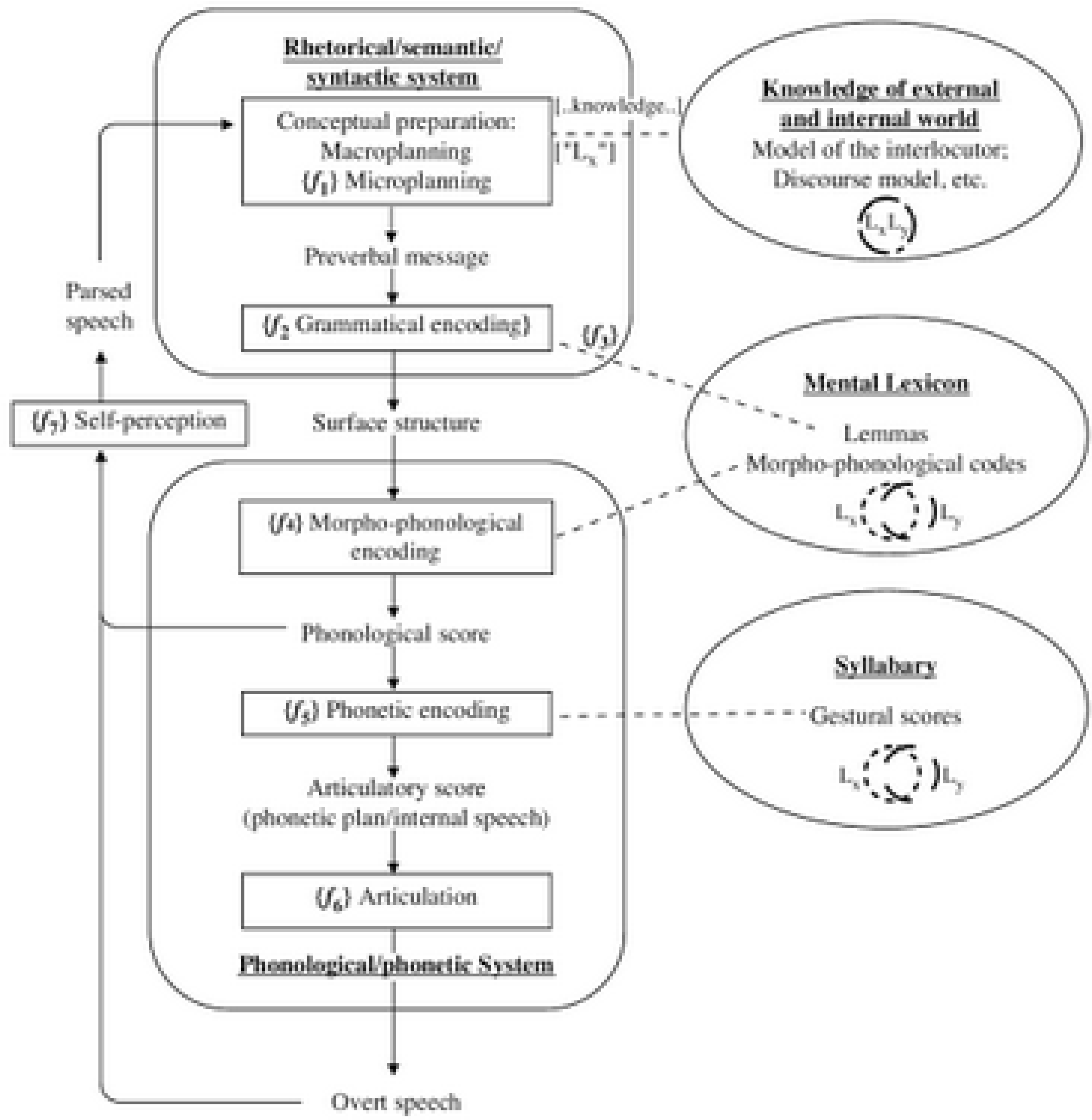
The tiers below show automatic Tiers 1 & 2 added by the script (De Jong and Wempe's (2013) *Praat V2 Oral Fluency* script) for syllables, and "sounding" and "Silent" time. They also show Tiers 3 to 6 added by the researcher:

Tier Number	Description
1	Syllables Count
2	Sounding and Silent (Pausing Time)
3	Disfluencies ("N" = repetitions, "R" = repairs, "E = elongations and "FP" = filled pauses)
4	Clause Boundaries
5	ASU Boundaries
6	"Beep" Location (prompt to switch languages) and "Switch" Location (point where participant started speaking the language they switched into)

Every speech file (for the Speech Elicitation tasks, English and Spanish, and Language Switching tasks, into English and into Spanish) at pre-test and post-test were preprocessed in this way.



Appendix G - Segalowitz's L2 speech model (2010)



Source: Segalowitz (2010) Figure 1.2 on page 9.

Appendix H – Normality tests

Shapiro-Wilk tests were conducted and results are presented below. Boxplots and histograms were also checked visually. If the significance value of the Shapiro-Wilk Test is greater than 0.05, the data is normal. If it is below 0.05, the data significantly deviate from a normal distribution. (Field, 2000).

Simon Task Inhibition Score (ICS)

Table 25 - Shapiro Wilk test for ICS scores

	Statistic	Sig
ICS Pre	0.936	0.055

As the significance is greater than 0.05, the data is normally distributed.

Letter Word (Phoneme) Decision Task Score (PICS)

Table 26 - Shapiro Wilk test for PICS scores

	Statistic	Sig
PICS Pre	0.936	0.011

As the significance is less than 0.05 the data is not normally distributed.

A Kruskal Wallis tests showed all 3 versions of the test were equally valid, as shown in the table below.

Table 27- Kruskal Wallis Test for Significant Differences in Letter (Phoneme) Decision Versions

Null Hypothesis	Statistic	Sig	Decision Regarding Null Hypothesis
Distribution of “Mean Neither Pre” is the same across all versions of VersPre	5.689	0.058	Retain
Distribution of “Mean L1 Pre” is the same across all versions of VersPre	5.434	0.061	Retain
Distribution of “Mean L2 Pre” is the same across all versions of VersPre	0.654	0.721	Retain

Fluency Measures

Shapiro Wilk Test of Normality for L2 Spanish Fluency Measures (n=49)				
Fluency Measure Pre	Pre-test Statistic	Pre-test Sig	Post-test Statistic	Post -test Sig
Speech_rate	.987	.864	.989	.912
Pause_rate	.983	.711	.970	.246
Phonation ratio	.980	.560	.976	.413
MLR	.970	.238	.981	.629
Mean syllable duration	.962	.113	.976	.406
Within clause filled pauses	.960	.092	.954	.055
Between clause filled pause rate	.767	.000	.820	.000
Within clause elongations	.711	.000	.806	.000
Between clause elongations	.262	.000	.	.
Repair rate	.652	.000	.648	.000
Repetitions rate	.632	.000	.769	.000
Within ASU filled pause rate	.956	.064	.952	.046
Between ASU filled pause rate	.737	.000	.705	.000
Within clause silent pause rate	.983	.677	.970	.232
Between clause silent pause rate	.969	.226	.980	.552
Within ASU silent pause rate	.977	.434	.969	.213
Between ASU silent pause rate	.963	.131	.895	.000

Duration of within clause pauses	.975	.390	.966	.166
Duration of between clause pauses	.761	.000	.947	.027
Duration of within ASU pauses	.970	.239	.964	.136
Duration of between ASU pauses	.703	.000	.845	.000
Total filled pause rate	.962	.118	.958	.076
Total filled pauses and elongations	.883	.000	.948	.031
Total filled pauses and elongations rate	.967	.177	.960	.097

As for most measures the significance is greater than 0.05, the data is normally distributed.

Adjusted Fluency Measures

Shapiro Wilk Test of Normality (n=49) for Adjusted Fluency Measures

Fluency Measure (Adjusted)	Pre-test Statistic	Pre-test Sig	Post-test Statistic	Post-test Sig
Speech rate	.995	.983	.756	.023
Articulation rate	.823	.149	.838	.124
Phonation rate	.964	.801	.923	.526
MLR	.908	.470	.838	.125
Silent pause rate	.933	.612	.971	.899
Silent pause rate pruned	.927	.578	.909	.430
Within clause silent pause rate	.835	.180	.911	.441
Between clause silent pause rate	.903	.446	.939	.655
Between ASU silent pause rate	.800	.102	.934	.613
Within ASU silent pause rate	.780	.071	.950	.738
Mean pause time	.933	.613	.842	.134
Duration of within clause pauses	.815	.132	.933	.600
Duration of between clause pauses	.969	.834	.938	.640
Filled pause rate	.858	.255	.704	.007
Within clause filled pause rate	.870	.299	.833	.115
Between clause filled pause rate	.894	.402	.915	.468
Within ASU filled pause rate	.969	.835		

Between ASU filled pause rate	.854	.239	.890	.317
Elongations rate	981	.908	.935	.619
Duration of between clause rate	.729	.012	.816	.081
Duration of within clause rate	.707	.007	.980	.954

As for most measures the significance is greater than 0.05, the data is normally distributed.

Language Switching Fluency Measures

Shapiro Wilk Test of Normality for Language Switching Fluency Measures (n=38)

LES = language switch from English to Spanish LSE = language switch from Spanish to English
"s" – data before the switch "b" – data after the beep

LS Measure Pre	Pre-Test Statistic	Pre-test Sig	Post-test Statistic	Post-test Sig
LES_s_Duration	.856	.001	.889	.004
LSE_b_Duration	.894	.005	.956	.225
LES_s_Between ASU silent pauses	.764	.000	.908	.012
LES_s_Between ASU filled pause_rate	.364	.000	.561	.000
LES_s_Between clause elongations rate	.274	.000	.176	.000
LES_s_Between clause silent pause rate	.360	.000	.638	.000
LES_s_Duration of Between ASU pauses rate	.418	.000	.667	.000
LES_s_Duration of Between clause pauses rate	.416	.000	.638	.000
LES_s_Duration of Within ASU pause rate	.284	.000	.969	.491
LES_s_Duration of Within clause pause rate	.286	.000	.977	.736
LES_s_Repetition Rate	.624	.000	.618	.000
LES_s_Repair Rate	.224	.000	.679	.000
LES_s_Within ASU filled pause rate	.285	.000	.930	.044
LES_s_Within ASU silent pause rate	.283	.000	.925	.032
LES_s_Within clause filled pause rate	.284	.000	.934	.057
LES_s_Within clause silent pause rate	.283	.000	.948	.139
LSE_b_Between ASU filled pause rate	.841	.000	.523	.000
LSE_b_ Between ASU silent pause rate	.969	.491	.974	.639
LSE_b_Between clause elongations rate	.	.	.176	.000

LSE_b_Between clause filled pause	.835	.000	.773	.000
LSE_b_Between clause silent pause rate	.975	.678	.980	.813
LSE_b_Duration of between ASU pause rate	.749	.000	.967	.452
LSE_b_Duration of between clause pause rate	.797	.000	.984	.914
LSE_b_Duration of within ASU pause rate	.927	.036	.958	.257
LSE_b_Duration of within clause pause rate	.920	.024	.954	.196
LSE_b_Repetitions rate	.345	.000	.629	.000
LSE_b_Repair rate	.508	.000	.653	.000
LSE_b_Within ASU filled pauses	.872	.002	.907	.011
LSE_b_Within ASU silent pauses	.972	.563	.975	.654
LSE_b_Within clause elongations rate	.694	.000	.816	.000
LSE_b_Within clause islent pause rate	.953	.193	.969	.503
LSE_s_Between ASU filled pause rate	.545	.000	.705	.000
LSE_s_Between ASU silent pause rate	.959	.268	.928	.040
LSE_s_Between clause elongation rate	.	.	.	.
LSE_s_Between clause silent pause rate	.957	.237	.928	.038
LSE_s_Duration of between ASU pause rate	.926	.034	.944	.108
LSE_s_Duration of between clause pause rate	.946	.122	.948	.135
LSE_s_Duration of within ASU pause rate	.983	.897	.958	.253
LSE_s_ Duration of within clause pause rate	.961	.309	.958	.257
LSE_s_Repetitions rate	.	.	.176	.000
LSE_s_Repair rate	.	.	.176	.000
LSE_s_Within ASU filled pause rate	.815	.000	.764	.000
LSE_s_Within ASU silent pause rate	.974	.646	.976	.000
LSE_s_Within clause filled pause rate	.801	.000	.747	.000
LSE_s_Within clause silent pause rate	.967	.448	.970	.680

As can be seen from the table above, approximately half the variables are normally distributed and half are non-normally distributed, so non-parametric tests will be used.

Language Switching Overflow Measure

Shapiro Wilk Test of Normality for Overflow Measures (n=38)					
Measure	Statistic	Sig	Measure	Statistic	Sig
Pre_LES_overflow	.745	.000	Post_LES_overflow	.877	.001
Pre_LSE_overflow	.955	.165	Post_LSE_overflow	.950	.114
OverflowSPGain	.936	.042	OverflowENGain	.911	.008

The Spanish to English overflow measure was normally distributed at pre-test, but not at post-test. Overflow gain (from pre-test to post-test) was not normally distributed either.

Appendix I – L1 and L2 Oral fluency correlation tables

Speed Fluency Correlations (Pre-test)

		1	2	3	4	5	6	7
1	L1 Speech Rate							
2	L2 Speech Rate	0.174						
3	L2 articulation rate	.292*	.818**					
4	L1 articulation rate	.913**	0.272	**.479****				
5	L2 phonation rate	-0.060	.691**	0.169	-0.137			
6	L1 phonation rate	.885**	0.061	0.069	.641**	0.006		
7	L2 MLR	-0.101	-.808**	-.601**	-0.100	**-.699****	-0.055	
8	L1 MLR	-.857**	-0.110	-0.174	-.721**	0.060	-.897**	0.099
		*Correlation significant at 0.05 ** Correlation significant at 0.01						

Breakdown fluency correlations (Pre-test)

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	L1 total disfluency rate													
2	L2 total disfluency rate	**.293***												
3	L1 MLR	-.604**	0.107											
4	L2 pause rate	-.658**	-0.022	**.919****										
5	L2 pause rate pruned	-.646**	0.025	.973**	.944**									
6	L2 within clause pause rate	-.301*	0.108	.683**	.726**	.648**								
7	L2 between clause pause rate	0.148	-0.264	-0.215	-0.135	-0.189	-0.258							
8	L2 between ASU pause rate	0.135	-0.273	-0.196	-0.123	-0.212	-0.175	.865**						
9	L2 within ASU pause rate	-0.229	0.129	.589**	.610**	.568**	.805**	-0.182	-0.255					
10	L2 meanpausetime	-.652**	-0.213	.557**	.470**	.620**	0.143	0.047	-0.043	0.129				
11	L1 pause rate	.764**	0.264	-.636**	-.691**	**-.673****	-.423**	0.166	0.215	-.364*	-.681**			
12	L1 pause rate pruned	-.550**	-0.118	.663**	.584**	**.673****	.410**	-.437**	-.439**	.406**	.457**	-.712**		
13	L2 MLR	.714**	0.134	-.662**	-.629**	-.685**	-.369**	.462**	.483**	-.349*	-.558**	.844**		
14	L1 within clause pause rate	.600**	0.133	-.646**	-.677**	-.661**	**-.668****	0.278	0.274	-.613**	-.530**	.880**	.787**	
15	L1 between clause pause rate	-0.041	.311*	.359*	.303*	.321*	.524**	-.529**	-.433**	.444**	-0.116	-0.088	-0.278	-.443**

Repair fluency correlations (Pre-test)

	L2 repair rate	L2 repetitions rate	L1 repair rate
L2 repair rate			
L2 repetitions rate	.288*		
L1 repair rate	-0.264	-0.225	
L1 repetitions rate	-0.102	**0.319****	0.208

Appendix J – L2 oral fluency correlations tables

		1	2	3	4	5	6	7
1	ICS							
2	Syllable Gain	.08						
3	Duration Gain	.05	.86**					
4	Speech Rate Gain	.02	.43**	.11				
5	Articulation Rate Gain	-.02	.46**	.27	.72**			
6	SwitchGain	.07	.08	.09	.15	-.21		
7	OverflowSPGain	.1	.15	.12	.23	-.21	.92**	
8	OverflowENGain	.05	.11	.05	.08	.03	-.44**	-.05

		1	2	3	4	5	6	7
1	PICS							
2	Syllable Gain	0.13						
3	Duration Gain	0.13	.84**					
4	Speech Rate Gain	0.01	.44**	0.10				
5	Articulation Rate Gain	0.08	.52**	0.26	.76**			
6	SwitchGain	0.19	-0.07	-0.06	0.10	-0.05		
7	OverflowSPGain	0.12	0.13	0.07	0.27	0.01	.86**	
8	OverflowENGain	0.02	0.28	0.27	0.12	0.10	-.45**	-0.05

		1	2	3	4	5	6	7	8	9	10
1	ICS										
2	Phonation Rate Gain	.04	.05								
3	MLR Gain	-.15	-.07	-.69**							
4	Silent Pause Rate Gain	-.07	-.18	-.73**	.88**						
5	Silent Pause Rate Pruned Gain	-.12	-.1	-.76**	.97**	.93**					
6	Mean pause time gain	-.26	-.03	-.63**	.91**	.77**	.89**				
7	Silent Pause Duration Gain	-.03	-.11	-.69**	.76**	.77**	.75**	.63**			
8	SwitchGain	.07	.15	.28	-.14	-.27	-.24	-.08	-.45**		
9	OverflowSPGain	.1	.17	.33*	-.2	-.31	-.3	-.19	-.48**	.92**	
10	OverflowENGain	.05	.01	.09	-.05	.01	-.04	-.19	.06	-.44**	-.05

			1	2	3	4	5	6	7	8	9	10	11
	ICS												
1	Within clause silent pause gain	*r*	.06										
2	Between clause silent pause gain	*r*	-.28	-.28									
3	Within ASU silent pause gain	*r*	.05	.90**	-.22								
4	Between ASU silent pause gain	*r*	.04	-.18	.48**	-.22							
5	Duration of within clause pauses gain	*r*	-.03	.93**	-.25	.83**	-.18						
6	Duration of between clause pauses gain	*r*	-.14	-.15	.76**	-.11	.56**	-.13					
7	Duration of within ASU pauses gain	*r*	-.07	.83**	-.23	.89**	-.2	.93**	-.11				
8	Duration of between clause pauses gain	*r*	-.26	-.15	.34*	-.14	.71**	-.14	.16	-.14			
9	SwitchGain	*r*	.07	-.17	-.11	-.14	-.28	-.13	-.41*	-.1	.22		
10	OverflowSPGain	*r*	.1	-.3	-.01	-.24	-.1	-.26	-.32	-.21	.23	.92**	
11	OverflowENGain	*r*	.05	-.23	.24	-.15	.44*	-.23	.32	-.19	.	-.44**	-.05

		1	2	3	4	5	6	7	8	9	10	11
1	PICS											
2	Within clause silent pause gain	-.09										
3	Between clause silent pause gain	.17	-.35*									
4	Within ASU silent pause gain	.06	.89**	-.18								
5	Between ASU silent pause gain	-.15	-.24	.73**	-.29							
6	Duration of within clause pauses gain	-.04	.90**	-.39*	.81**	-.24						
7	Duration of between clause pauses gain	.17	-.3	.92**	-.12	.67**	-.28					
8	Duration of within ASU pauses gain	.08	.81**	-.28	.88**	-.31*	.92**	-.2				
9	Duration of between clause pauses gain	-.17	-.18	.79**	-.2	.85**	-.24	.81**	-.31*			
10	SwitchGain	.19	-.2	.05	-.15	-.26	-.21	-.06	-.1	-.13		
11	OverflowSPGain	.12	-.28	.19	-.22	-.07	-.36*	.09	-.24	.1	.86**	
12	OverflowENGain	.02	-.13	.23	.02	.12	-.19	.34	-.07	.39*	-.45**	-.05

		1	2	3	4	5	6	7	8	9	10	11
1	ICS											
2	Filled pause rate gain	.09										
3	Filled pause rate gain pruned	.08	.99**									
4	Within clause filled pause rate gain	.12	.98**	.97**								
5	Between clause filled pause rate gain	-.11	.36	.29	.18							
6	Within ASU filled pause rate gain	.18	.97**	.96**	.98**	.36						
7	Between ASU filled pause rate gain	-.61**	.26	.21	.31	.55**	.08					
8	Elongations Rate Gain	.50*	.4	.42	.48	-.02	.47	.32				
9	SwitchGain	.15	.08	.06	.11	-.2	.04	-.04	.1			
10	OverflowSPGain	.17	.01	-.03	.08	-.1	.05	.	.04	.92**		
11	OverflowENGain	.01	-.19	-.22	-.18	.27	-.04	.13	-.15	-.44**	-.05	
12	Total pause rate gain	-0.05	.30*	.35*	0.25	0.34	0.26	0.23	0.14	-0.06	-0.11	-0.07

		1	2	3	4	5	6	7
1	ICS							
2	Total Disfluency Gain	0.01	0.01	0.16	-0.10	0.07	0.10	0.05
3	Repair Rate Gain	0.16	-0.04					
4	Repetition Rate Gain	-0.10	0.22	-0.29				
5	SwitchGain	0.07	-0.25	-0.06	0.12			
6	OverflowSPGain	0.10	-0.25	0.10	0.21	.92**		
7	OverflowENGain	0.05	0.10	0.32	0.16	-.44**	-0.05	

		1	2	3	4	5	6	7	8	9
1	PICS									
2	Phonation Rate Gain	.07								
3	MLR Gain	-.13	-.86**							
4	Silent Pause Rate Gain	-.18	-.74**	.94**						
5	Silent Pause Rate Pruned Gain	-.15	-.87**	.97**	.96**					
6	Mean pause time gain	-.06	-.68**	.65**	.58**	.64**				
7	Silent Pause Duration Gain	.03	-.78**	.71**	.62**	.71**	.59**			
8	SwitchGain	.19	.21	-.2	-.26	-.25	-.09	-.41*		
9	OverflowSPGain	.12	.32	-.2	-.2	-.25	-.21	-.41*	.86**	
10	OverflowENGain	.02	-.03	.12	.18	.13	-.11	.16	-.45**	-.05

Appendix K –Questionnaire and fluency gains tables (correlations)

Self-reported language use and speed fluency gains

n=49	Mean	SD
Syllables	0.1410	0.57543
Duration	0.1522	0.45431
Speech rate	-0.0345	0.28973
Articulation rate	-0.0352	0.16758

Correlations

		1	2	3	4
1	L2 Use				
2	Syllable gain	0.27			
3	Duration gain	0.16	.857**		
4	Speech rate gain	0.22	.427**	0.11	
5	Articulation rate gain	0.00	.456**	0.27	.724**

*Correlation significant at 0.05 significance level

**Correlation significant at 0.01 significance level

Self-reported language use and breakdown fluency gains

n=49	Mean	SD
Phonation ratio	-0.0111	0.22462
MLR	0.7154	2.13189
Silent pause rate	0.2103	0.65356
Silent pause rate pruned	0.4702	1.37767
Mean pause time	0.3752	1.27669

Pearson correlations

		1	2	3	4	5
1	L2 Use					
2	Phonation time ratio	.299*				
3	MLR	-0.222	-.694**			
4	Silent pause rate	-0.139	-.726**	.876**		
4	Silent pause rate pruned	-0.247	-.755**	.975**	.927**	
5	Mean pause time	-0.186	-.629**	.906**	.767**	.887**

	Mean	SD
Within clause silent pause rate	0.57	1.24
Between clause silent pause rate	0.49	2.18
Within ASU silent pause rate	0.42	0.99
Between ASU silent pause rate	0.07	1.18
Duration of within clause pauses	1.22	2.70
Duration of between clause pauses	1.52	7.1
Duration of within ASU pauses	0.93	2.08
Duration of between ASU pauses	13.78	43.73
Filled pause rate	0.52	1.5
Filled pause rate pruned	0.56	1.69
Within clause filled pause rate	0.36	1.40
Between clause filled ause rate	0.14	1.10
Within ASU filled pause rate	0.36	1.44
Between ASU filled pause rate	-0.01	0.93
Elongations rate	-0.73	0.49
Total pausing rate	0.38	0.91

Pearson Correlations

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	L2 Use																
2	Within clause silent pause rate	-0.174															
3	Between clause silent pause rate	0.290	-0.280														
4	Within ASU silent pause rate	-0.126	.897**	-0.222													
5	Between ASU silent pause rate	.426**	-0.181	.477**	-0.217												
6	Duration of within clause pauses	-0.234	.927**	-0.247	.831**	-0.176											
7	Duration of between clause pauses	.308*	-0.153	.755**	-0.106	.556**	-0.127										
8	Duration of within ASU pauses	-0.240	.829**	-0.230	.892**	-0.200	.925**	-0.113									
9	Duration of between ASU pauses	0.195	-0.146	.342*	-0.138	.712**	-0.139	0.162	-0.138								
10	Filled pause rate	-0.129	.591**	-0.164	.475**	-0.053	.507**	-0.114	.429**	-0.149							
11	Filled pause rate pruned	-0.160	.596**	-0.163	.501**	-0.064	.529**	-0.099	.472**	-0.131	.990**						
12	Within clause filled pause rate	-0.253	.565**	-0.186	.371*	-0.121	.514**	-0.142	.382*	-0.135	.976**	.968**					
13	Between clause filled ause rate	0.095	0.239	0.000	0.180	0.143	0.017	-0.015	0.016	0.048	0.363	0.291	0.177				
14	Within ASU filled pause rate	-0.266	.555**	-0.191	.402**	-0.171	.504**	-0.139	.392*	-0.205	.968**	.960**	.981**	0.359			
15	Elongations rate	-0.096	.705**	-0.018	.685**	0.221	.757**	-0.128	.744**	-0.158	0.398	0.419	0.475	-0.018	0.470	0.316	
16	Total pausing rate	-0.100	.358*	-0.137	.405**	-0.113	.483**	-0.056	.571**	-0.121	.298*	.349*	0.255	0.341	0.263	0.235	0.136

Self-reported language use and repair fluency gains

n=49	Mean	SD
Repair rate	-0.6585	0.48558
Repetitions rate	-0.0323	1.30727

Pearson correlations

	L2 Use	Repair Rate
L2 Use		
Repair rate	-0.83	0.381
Repetitions rate	0.361	-0.294

High language users (>30%L2 use) and speed fluency gains

Descriptive statistics:

High language users (n=8)		
	Mean	SD
Language use	0.5063	0.13
Syllable gain	0.3650	0.98
Duration gain	0.3213	0.58
Speech rate gain	-0.0143	0.18
Articulation rate gain	-0.0627	0.07
Phonation time ratio gain	0.0371	0.13
MLR gain	0.1086	0.33
Silent pause rate gain	0.0225	0.13

Silent pause rate pruned gain	0.0120	0.23
Mean pause time gain	0.2077	0.47
Within clause silent pause rate	0.2566	0.63
Between clause silent pause rate	2.0428	3.93
Within ASU silent pause rate	0.1343	0.63
Between ASU silent pause rate	1.9167	2.47
Duration of within clause pauses	0.2020	0.54
Duration of between clause pauses	9.2329	20.18
Duration of within ASU pauses	0.0729	0.62
Duration of between ASU pauses	61.4284	99.72
Filled pause rate	0.5877	1.90
Filled pause rate pruned	0.6206	2.03
Within clause filled pause rate	-0.2748	0.39
Between clause filled pause rate	0.3538	0.64
Within ASU filled pause rate	-0.4809	0.32
Between ASU filled pause rate	0.5730	0.94
Elongations rate	-0.8044	0.33
Total pausing rate	0.1145	0.27
Repair rate gain	-0.7736	0.45
Repetitions rate gain	1.0890	1.69

Pearson correlations

n=8		1	2	3	4
1	L2 Use				
2	Syllable gain	0.172			
3	Duration gain	-0.226	.853**		
4	Speech rate gain	0.362	0.669	0.243	
5	Articulation rate gain	0.202	.781*	0.504	.865**

High language users (>30%L2 use) and breakdown fluency gains

Pearson correlations

		1	2	3	4	5	6
1	L2 Use						
2	Phonation time ratio gain	0.212					
3	MLR gain	0.154	-.859**				
4	Pause rate gain	0.383	-0.567	.852**			
5	Pause rate pruned gain	0.120	-.860**	.974**	.906**		
6	Meanpause time gain	-.725*	-0.466	0.055	-0.039	0.187	
		0.042	0.244	0.898	0.927	0.657	

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	L2 Use																
2	Within clause silent pause rate	-0.383															
3	Between clause silent pause rate	-0.322	-0.462														
4	Within ASU silent pause rate	-0.269	.886**	-0.281													
5	Between ASU silent pause rate	-0.148	-0.642	0.594	-0.670												
6	Duration of within clause pauses	-0.613	.901**	-0.252	.794*	-0.561											
7	Duration of between clause pauses	-0.147	-0.295	.950*	-0.056	0.497	-0.135										
8	Duration of within ASU pauses	-0.466	.721*	0.351	.898**	-0.348	.820*	0.453									
9	Duration of between ASU pauses	-0.661	-0.371	0.223	-0.501	0.667	-0.129	-0.073	-0.016								
10	Filled pause rate	-0.383	.888*	-0.960	.907*	0.656	.834*	-0.869	.953**	0.060							
11	Filled pause rate pruned	-0.399	.885*	-0.996	.909*	0.904	.839*	-0.990	.960**	0.163	.999**						
12	Within clause filled pause rate	-0.086	-0.539	-0.991	-0.835	0.758	-0.227	-0.933	-0.518	0.103	.899*	0.822					
13	Between clause filled ause rate	-1.000**	1.000**	.[c]	1.000**	.[c]	1.000**	.[c]	1.000**	.[c]	1.000**	1.000**	.[c]				
14	Within ASU filled pause rate	0.418	-0.520	-0.622	-0.434	0.944	-0.522	-0.424	-0.628	-0.457	0.708	0.607	0.831	.[c]			
15	Between ASU filled pause rate	1.000**	-1.000**	.[c]	-1.000**	.[c]	-1.000**	.[c]	-1.000**	.[c]	-1.000**	-1.000**	.[c]	-1.000**			
16	Elongations rate	0.500	-0.145	.[c]	-0.384	.[c]	-0.022	.[c]	-0.412	.[c]	-0.516	-0.529	-1.000**	.[c]	-1.000**	.[c]	
17	Total pausing rate	0.470	-0.059	0.136	0.276	-0.284	-0.020	0.228	0.347	-0.195	0.638	0.661	-0.621	1.000**	-0.346	-1.000**	-0.689

High language users (>30%L2 use) and repair fluency gains

	L2 Use
L2 Use	
Repair rate gain	-0.522
Repetitions rate gain	-0.434

Class in Spanish & Speed Fluency

Pearson correlations

	Class in Spanish	Syllable gain	Duration gain	Speech rate gain
Class in Spanish				
Syllable gain	0.182			
Duration gain	0.134	.857**		
Speech rate gain	0.126	.427**	0.106	
Articulation rate gain	-0.003	.456**	0.266	.724**

Class in Spanish & Breakdown Fluency

Pearson correlations

		1	2	3	4	5
1	Class in Spanish					
2	Phonation time ratio gain	0.133				
3	MLR gain	-0.149	-.694**			
4	Pause rate gain	-0.103	-.726**	.876**		
5	Pause rate pruned gain	-0.139	-.755**	.975**	.927**	
6	Mean pause time gain	-0.105	-.629**	.906**	.767**	.887**

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Class in Spanish																
2	Within clause silent pause rate	-0.139															
3	Between clause silent pause rate	0.209	-0.280														
4	Within ASU silent pause rate	-0.049	.897**	-0.222													
5	Between ASU silent pause rate	0.125	-0.181	.477**	-0.217												
6	Duration of within clause pauses	-0.168	.927**	-0.247	.831**	-0.176											
7	Duration of between clause pauses	-0.068	-0.153	.755**	-0.106	.556**	-0.127										
8	Duration of within ASU pauses	-0.101	.829**	-0.230	.892**	-0.200	.925**	-0.113									
9	Duration of between ASU pauses	0.044	-0.146	.342*	-0.138	.712**	-0.139	0.162	-0.138								
10	Filled pause rate	-0.135	.591**	-0.164	.475**	-0.053	.507**	-0.114	.429**	-0.149							
11	Filled pause rate pruned	-0.130	.596**	-0.163	.501**	-0.064	.529**	-0.099	.472**	-0.131	.990**						
12	Within clause filled pause rate	-0.182	.565**	-0.186	.371*	-0.121	.514**	-0.142	.382*	-0.135	.976**	.968**					
13	Between clause filled pause rate	-0.074	0.239	0.000	0.180	0.143	0.017	-0.015	0.016	0.048	0.363	0.291	0.177				
14	Within ASU filled pause rate	-0.180	.555**	-0.191	.402**	-0.171	.504**	-0.139	.392*	-0.205	.968**	.960**	.981**	0.359			
15	Between ASU filled pause rate	-0.092	0.148	0.155	0.008	.655**	-0.014	0.186	-0.060	0.243	0.255	0.207	0.306	.555**	0.081		
16	Elongations rate	0.075	.705**	-0.018	.685**	0.221	.757**	-0.128	.744**	-0.158	0.398	0.419	0.475	-0.018	0.470	0.316	
17	Total pausing rate	-0.108	.358*	-0.137	.405**	-0.113	.483**	-0.056	.571**	-0.121	.298*	.349*	0.255	0.341	0.263	0.235	0.136

Class in Spanish & Repair Fluency

Spearman rank correlations

	Class in Spanish	Repair rate gain
Class in Spanish		
Repair rate gain	-0.132	
Repetitions rate gain	$.512^{*}$	-0.294

Extracurricular & Speed Fluency

Pearson correlations

	Extracurricular	Syllable gain	Duration gain	Speech rate gain
Extracurricular				
Syllable gain	0.145			
Duration gain	0.090	$.857^{**}$		
Speech rate gain	0.148	$.427^{**}$	0.106	
Articulation rate gain	0.095	$.456^{**}$	0.266	$.724^{**}$

Extracurricular & Breakdown Fluency

Pearson correlations

		1	2	3	4	5
1	Extracurricular					
2	Phonation time gain	0.162				
3	MLR gain	-0.132	-.694**			
4	Pause rate gain	-0.107	-.726**	.876**		
5	Pause rate pruned gain	-0.148	-.755**	.975**	.927**	
6	Mean pause time gain	-0.124	-.629**	.906**	.767**	.887**

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Extracurric																
2	Within clause silent pause rate	-0.181															
3	Between clause silent pause rate	0.162	-0.280														
4	Within ASU silent pause rate	-0.140	.897**	-0.222													
5	Between ASU silent pause rate	.320*	-0.181	.477**	-0.217												
6	Duration of within clause pauses	-0.207	.927**	-0.247	.831**	-0.176											
7	Duration of between clause pauses	0.195	-0.153	.755**	-0.106	.556**	-0.127										
8	Duration of within ASU pauses	-0.152	.829**	-0.230	.892**	-0.200	.925**	-0.113									
9	Duration of between ASU pauses	0.046	-0.146	.342*	-0.138	.712**	-0.139	0.162	-0.138								
10	Filled pause rate	-0.235	.591**	-0.164	.475**	-0.053	.507**	-0.114	.429**	-0.149							
11	Filled pause rate pruned	-0.245	.596**	-0.163	.501**	-0.064	.529**	-0.099	.472**	-0.131	.990**						
12	Within clause filled pause rate	-0.197	.565**	-0.186	.371*	-0.121	.514**	-0.142	.382*	-0.135	.976**	.968**					
13	Between clause filled ause rate	-0.180	0.239	0.000	0.180	0.143	0.017	-0.015	0.016	0.048	0.363	0.291	0.177				
14	Within ASU filled pause rate	-0.251	.555**	-0.191	.402**	-0.171	.504**	-0.139	.392*	-0.205	.968**	.960**	.981**	0.359			
15	Between ASU filled pause rate	.442*	0.148	0.155	0.008	.655**	-0.014	0.186	-0.060	0.243	0.255	0.207	0.306	.555**	0.081		
16	Elongations rate	0.187	.705**	-0.018	.685**	0.221	.757**	-0.128	.744**	-0.158	0.398	0.419	0.475	-0.018	0.470	0.316	
17	Total pausing rate	-0.102	.358*	-0.137	.405**	-0.113	.483**	-0.056	.571**	-0.121	.298*	.349*	0.255	0.341	0.263	0.235	0.136

Extracurricular & Repair Fluency

	Extracurricular	Repair rate gain
Extracurricular		
Repair rate gain	-0.350	
Repetition rate gain	0.235	-0.294

Homestay vs. Non=homestay

Homestay (1.00= Yes)		N	Mean	SD
Speech rate gain	1.00	10	0.04	0.26
	.00	39	-0.05	0.29
Articulation rate gain	1.00	10	0.01	0.11
	.00	39	-0.04	0.17
Phonation rate gain	1.00	10	0.02	0.18
	.00	39	-0.01	0.23
MLR gain	1.00	10	0.13	0.50
	.00	39	0.86	2.35
Silent pause rate gain	1.00	10	0.06	0.28
	.00	39	0.24	0.71
Within clause silent pause rate gain	1.00	10	0.14	0.45
	.00	39	0.68	1.35
Between clause silent pause rate gain	1.00	8	-0.19	0.41
	.00	34	0.65	2.39
Within ASU silent pause rate gain	1.00	10	0.08	0.45
	.00	39	0.51	1.08
Between ASU silent pause rate gain	1.00	9	0.75	1.75
	.00	33	-0.11	0.93
Mean pause time	1.00	10	0.02	0.51
	.00	39	0.46	1.39
Duration within clause pauses	1.00	10	0.22	0.84
	.00	39	1.47	2.95
Duration between clause pause	1.00	8	-0.28	0.50
	.00	34	1.94	7.92
Duration within ASU pauses	1.00	10	0.10	0.77
	.00	39	1.14	2.26
Duration between ASU pauses	1.00	9	8.90	23.98
	.00	35	15.04	47.70
Total pause rate	1.00	10	1.42	0.38
	.00	39	1.17	0.46
FS_totalp_rate_gain	1.00	10	0.06	0.26
	.00	39	0.46	1.00
Repetition gain	1.00	4	0.50	1.91
	.00	12	-0.33	0.77
Repair gain	1.00	5	-0.82	0.40
	.00	10	-0.57	0.52

Independent samples t-test for Homestay participants

	t	p value
Speech rate gain	1.020	0.313
Articulation rate gain	1.021	0.313
Phonation rate gain	0.492	0.625
MLR gain	-0.978	0.333
Pause rate gain	-0.801	0.427
Within clause silent pause rate gain	-2.100	**0.042**
Within ASU silent pause rate gain	-1.211	0.232
Between ASU silent pause rate gain	2.012	0.051
Mean pause time	-0.960	0.342
Duration within clause pauses	-2.306	**0.026**
Duration between clause pauses	-1.627	0.113
Duration within ASU pauses	-2.375	0.022
Duration between ASU pauses	-0.540	0.593
Total pause rate	1.704	0.107
Repetitions gain	1.285	0.220
Repairs gain	-0.906	0.382