Speaking fluency and study abroad: what factors are related to fluency development?

Karen Ruth Leonard
University of Iowa

Copyright 2015 Karen Ruth Leonard

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/1676

Recommended Citation
https://ir.uiowa.edu/etd/1676. https://doi.org/10.17077/etd.5ku9rrs7

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the First and Second Language Acquisition Commons
SPEAKING FLUENCY AND STUDY ABROAD: WHAT FACTORS ARE RELATED TO FLUENCY DEVELOPMENT?

by

Karen Ruth Leonard

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Second Language Acquisition in the Graduate College of The University of Iowa

May 2015

Thesis Supervisor: Assistant Professor Christine Shea
Associate Professor Judith Liskin-Gasparro
CERTIFICATE OF APPROVAL

____________________________

PH.D. THESIS

_________________

This is to certify that the Ph.D. thesis of

Karen Ruth Leonard

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Second Language Acquisition at the May 2015 graduation.

Thesis Committee: __________________________________________

Christine Shea, Thesis Supervisor

________________________________________

Judith Liskin-Gasparro, Thesis Supervisor

________________________________________

James Maxey

________________________________________

Lia Plakans

________________________________________

Thomas Farmer
ACKNOWLEDGEMENTS

As I reflect on the experience of completing my dissertation, I am grateful for the help of many individuals. First, I would like to thank my co-directors, Dr. Christine Shea and Dr. Judith Liskin-Gasparro. Christine, thank you so much for your guidance throughout the dissertation process. I appreciate the ways in which you have both encouraged me and pushed me to keep improving my work. Judy, thank you for your detailed feedback as well as all of the time you have spent mentoring me since I entered the FLARE program.

I would also like to thank the rest of my committee members, Dr. James Maxey, Dr. Lia Plakans, and Dr. Thomas Farmer. Dr. Maxey, your teaching has helped me gain a much better understanding of statistics. Dr. Plakans, thank you for your input on improving the methodology of the study. Thomas, I appreciate your help in thinking through the relationship between my study and models of language processing.

I am also grateful for the assistance of fellow graduate students in FLARE and in Spanish. Tania Leal, thank you for reading and recording the sentences for one of the processing tasks. Brian Olovson, Raychel Vasseur, Christian Supiot Pérez, Fernando Castro Ortíz, Iván Parra García, Marisol Fonseca, and Stacey Alex, thank you for your help in piloting the speaking tasks. Brian and Raychel, thanks also for your advice related to life in Buenos Aires.

Additionally, I greatly appreciate the collaboration of the International Programs staff at the universities where I conducted my research. Thank you for
passing along to students the invitation to participate in the study as well as providing space to meet with participants.

Finally, I would like to thank the students who participated in the study. I am grateful for your commitment to complete the study in the midst of all the exciting and challenging opportunities of a semester abroad. Though I cannot acknowledge you by name, I am very thankful for each one of you!
ABSTRACT

This study explores the development of second language (L2) fluency during a semester abroad and its relationship to the development of grammar, vocabulary, and language processing speed. It also considers the influence of individual participants’ first language (L1) and pre-study abroad (SA) L2 fluency on the development of fluency during study abroad. Additionally, the study examines issues in the measurement of fluency, focusing on questions related to measuring pauses in L2 speech.

Thirty-nine undergraduate students (L1 English) studying in Buenos Aires, Argentina, completed a pretest consisting of speaking tasks in English and Spanish, Spanish grammar and vocabulary tests, a picture-naming task, and a measure of sentence processing speed. Approximately three months later, near the end of their time abroad, they completed a posttest consisting of the same tasks, with the exception of the speaking tasks in English. Participants also filled out a questionnaire every other week during the semester in which they estimated the amount of time that they had spent interacting with native speakers of Spanish.

Results show that participants experienced significant gains on most measures of fluency during study abroad. This finding was especially true for participants who began their time abroad with low L2 fluency. Nevertheless, students who began the semester abroad with high L2 fluency still had significantly higher fluency at the end of the semester than students who began with low L2 fluency.

Looking at the relationship between L2 fluency and L2 linguistic knowledge
(vocabulary and grammar scores) and language processing speed (picture-naming and sentence-matching scores), the study found a moderate relationship between pretest measures of L2 fluency and pretest measures of linguistic knowledge and processing speed. However, the results show no relationship between pre-SA linguistic knowledge and processing speed and gains in L2 fluency, and little relationship between gains in linguistic knowledge and processing speed and gains in L2 fluency. The best predictor of gains in L2 fluency was pre-SA L2 fluency. These results suggest that although there is a relationship between L2 linguistic knowledge and L2 fluency, having more advanced L2 linguistic knowledge prior to study abroad does not necessarily give students an advantage in the area of fluency development during study abroad.

Regarding the measurement of fluency, the data show that learners with low and high levels of lexical–grammatical competence significantly differed from one another on all measures of rates of pauses (short and long pauses, filled and unfilled pauses, and mid-clause and end-of-clause pauses) as well as in the percent of pauses occurring in the middle of a clause. However, they did not significantly differ from one another in the percent of filled pauses. The findings suggest that measuring all of these pauses may be useful in examining L2 fluency. However, there is perhaps little or nothing to be gained from counting filled and unfilled pauses separately, as speakers' tendency to use more of one or the other appears to be more closely related to personal speaking style than to L2 ability.
PUBLIC ABSTRACT

Because study abroad programs continue to grow in popularity, it is valuable to examine the outcomes of these programs. This study looks at changes in college students’ speaking fluency in a second language during a semester abroad. Thirty-nine English-speaking students learning Spanish in Argentina participated in the study.

Near the beginning and end of the semester, they were audio recorded while speaking in Spanish about a list of topics. They also completed tests measuring their knowledge of Spanish grammar and vocabulary as well as computerized tasks measuring how quickly they processed language.

Students experienced significant gains in fluency in Spanish during the semester. Those who began with lower levels of fluency tended to have larger gains than those who began with higher levels of fluency. However, many students who began with higher levels of fluency also improved, and they were still significantly more fluent at the end of the semester than students who began with lower levels of fluency. This finding suggests that study abroad provides a good opportunity for students to improve their fluency, but students should also be encouraged to take advantage of other ongoing opportunities to improve their fluency and not view study abroad as the magic key to fluency.

The study found no relationship between students’ knowledge of grammar and vocabulary at the beginning of the semester and their gains in fluency, which suggests that students can benefit from study abroad at both earlier and later points in the process of learning a language.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... x

LIST OF FIGURES ........................................................................................................... xii

CHAPTER 1: INTRODUCTION ......................................................................................... 1

Background ................................................................. 1
What is fluency? ............................................................ 4
Research questions ....................................................... 6

The development of fluency ........................................ 6
The measurement of fluency .......................................... 11
Overview of the dissertation ....................................... 13

CHAPTER 2: PREVIOUS RESEARCH ON FLUENCY ..................................................... 14

Fluency: Measurement and development .................... 14

Broad definitions of fluency ........................................ 14
Perceived, utterance, and cognitive fluency ................. 15
Models of fluent speech production ............................ 24

The relationship between L1 and L2 fluency ............... 26
Knowledge and skills underlying fluency ..................... 28

Fluency development and study abroad ..................... 31

Overall proficiency: Improvement on the ACTFL scale 31
Detailed analyses of changes in fluency ...................... 33
Changes in fluency and underlying skills .................... 38
Areas for further research ......................................... 40

Issues in measuring pauses ....................................... 41
Summary .................................................................. 47

CHAPTER 3: METHODOLOGY ..................................................................................... 48

Research questions .................................................... 48

The development of fluency ....................................... 48
The measurement of fluency ....................................... 50

Participants ............................................................... 51
Background .............................................................. 51
Materials and tasks .................................................... 52

Speaking tasks .......................................................... 53
Grammar test ........................................................... 55
Vocabulary test ........................................................ 55
Processing tasks ......................................................... 56

Background questionnaire ......................................... 58
Language contact questionnaire ................................. 58

Data collection ......................................................... 59
LIST OF TABLES

Table 1: Measures of pausing in previous studies.................................................. 43
Table 2: Speaking tasks .......................................................................................... 53
Table 3: Data transformations................................................................................ 70
Table 4: Fluency measures.................................................................................... 80
Table 5: Mean pretest and posttest scores for all fluency measures......................... 81
Table 6: Mean L1–L2 difference scores.................................................................. 84
Table 7: Minimum and maximum values for L1 and L2 fluency measures .............. 86
Table 8: Correlations between pre-SA L2 fluency and changes in L2 fluency .......... 88
Table 9: Mean changes for low and high pre-SA fluency groups............................ 90
Table 10: Correlations between pretest measures .................................................. 93
Table 11: Correlations between pretest measures of grammar, vocabulary, and processing scores and posttest measures of L2 fluency ............................ 94
Table 12: Regression model predicting pre-SA L2 fluency .................................... 97
Table 13: Correlations between pre-SA grammar, vocabulary, and processing scores and changes in fluency from pre- to post-SA.......................................... 99
Table 14: Correlations between pre-SA grammar, vocabulary, and processing scores and changes in L2 fluency, controlling for pre-SA L2 fluency ................. 100
Table 15: Correlations between changes in grammar, vocabulary, and processing scores and changes in L2 fluency ..................................................... 102
Table 16: Regression model predicting changes in L2 fluency ............................... 104
Table 17: Mean pause scores for low and high groups.......................................... 105
Table 18: Hours per week interacting with native speakers of Spanish..................... 107
Table 19: Other interactions with native speakers of Spanish................................ 108
Table 20: Participants' most meaningful language practice ..................................... 109
Table 21: Comparison across studies of the relationship between fluency and grammar, vocabulary, and processing measures ........................................... 123
Table 22: Mean pre- and post-SA L2 fluency T-scores ........................................ 132
Table 23: Percent filled pauses in Spanish based on rate of filled pauses in English.................................................................................................................................................. 147
LIST OF FIGURES

Figure 1: Marking pauses in PRAAT .............................................................................................................. 62

Figure 2: Changes in pausing from pre- to post-study abroad ................................................................. 82

Figure 3: Changes in articulation rate, repair, and mean length of run from pre- to post-study abroad ............................................................................................................................. 82

Figure 4: L1–L2 difference scores for pausing .......................................................................................... 85

Figure 5: L1–L2 difference scores for articulation rate, repair, and mean length of run .......................... 85

Figure 6: Mean amount of change for high and low pre-SA fluency groups: pausing ................................ 91

Figure 7: Mean amount of change for high and low pre-SA fluency groups: articulation rate, repair, and mean length of run .................................................................................................................. 91

Figure 8: Mean pause scores for low and high groups ............................................................................. 105

Figure 9: Picture matching and picture naming response times, pre- and post-SA .................................. 143
Chapter 1: Introduction

Background

The nature and development of fluency in a second language (L2) has been the focus of a wide range of second language acquisition (SLA) research (De Jong et al., 2013b; Derwing et al., 2009; Freed et al., 2004b; Hilton, 2008, 2009; Kormos & Denes, 2004; Lennon, 1990; O’Brien et al., 2007; Osborne, 2011; Riggenbach, 2000; Schmidt, 1992). One particular strand of this research has looked at the development of fluency in the context of study abroad (Freed et al., 2004b; Isabelli-García, 2003; Lennon, 1990; Segalowitz et al., 2004; Segalowitz & Freed, 2004; Towell, 2002; Trenchs-Parera, 2009; Valls Ferrer, 2008). A common assumption in language learning is that studying abroad is very helpful, if not essential, for developing fluency in an L2 (Freed, 1998). Although research shows that learners often do experience growth in fluency while studying abroad (Freed, 1995; Freed et al., 2004b; Segalowitz & Freed, 2004), this growth is not necessarily as predictable, consistent, or dramatic as popular belief might suggest (Freed, 1995; Freed et al., 2004b; Segalowitz & Freed, 2004).

The continued growth of study abroad programs highlights the need for research examining the outcomes of study abroad. During the 2012-2013 academic year,¹ a record-high 289,408 students from the United States participated in programs abroad for which they received academic credit, a 65% increase from 10 years earlier (Open Doors, 2014). Of course, not all students who go abroad do so for the purpose of improving their language skills. Other commonly advertised

---

¹ The most recent year for which data are available
benefits include gaining cultural understanding, growing in adaptability and independence, developing new relationships, and learning about oneself (“Career benefits of study abroad,” 2014; “Reasons to study abroad,” 2014; “Why study abroad?,” 2014). However, since language acquisition—with a particular emphasis on speaking—continues to be one of the main benefits used to promote study abroad, it is valuable to research the specific changes that take place in participants’ speaking skills during study abroad.

Some studies have examined whether variable outcomes in L2 fluency development are related to differences in students’ experiences while abroad, including the amount of time spent using the L2 in a variety of activities outside of class (Freed et al., 2004b; Perez-Vidal & Juan-Garau, 2011; Segalowitz & Freed, 2004; Valls Ferrer, 2008). Several studies have found such a relationship (Freed et al., 2004b; Perez-Vidal & Juan-Garau, 2011; Valls Ferrer, 2008), suggesting that time on task while abroad explains some of the differences in outcomes. However, there are other factors potentially related to L2 fluency development that have rarely been considered in study abroad research, including differences in first language (L1) fluency (Towell & Dewaele, 2005), knowledge of L2 grammar (Segalowitz et al., 2004), knowledge of L2 vocabulary, processing speed (Segalowitz & Freed, 2004), and pre-study abroad L2 fluency (Towell, 2002).

A separate, more recent area of L2 fluency research has investigated the relationship between L2 fluency and L2 grammar, vocabulary, and processing speed. De Jong et al. (2013a) found that each of six measures of grammar, vocabulary, and processing speed significantly correlated with at least two
measures of fluency, with the strongest correlations for vocabulary knowledge and sentence building, a measure of grammatical processing speed. Hilton (2008) found moderate to strong correlations between each of eight measures of fluency and scores on grammar and vocabulary tests. Although more research is needed to confirm the results, these findings suggest that grammar, vocabulary, and processing speed are all related to fluency. However, to date, this research has focused on the relationship at one point in time rather than look at changes over time.

The present study aims to address these gaps in the research by bringing together longitudinal research on L2 fluency development in study abroad and research on knowledge and skills related to L2 fluency. The primary purpose of the study is to explore the development of L2 fluency during a semester abroad and its relationship to the development of grammar, vocabulary, and processing speed, while also considering the influence of individual participants’ L1 and pre-SA L2 fluency.

A challenging factor in carrying out such a study is that the measurement of fluency is not a straightforward task. Many different measures have been used in SLA research, and there is no clear consensus regarding which measures to use or when and why to use them (see, for example, Freed et al., 2004b; García-Amaya, 2009; Gilabert & Muñoz, 2010; Guará Tavares, 2009; Hilton, 2008; Iwashita et al., 2008; Kormos & Denes, 2004; Llanes & Muñoz, 2009; Mochizuki & Ortega, 2008; Skehan & Foster, 2005; Tajima, 2003; Wigglesworth, 1997). Thus, a secondary
purpose of the present study is to examine issues in the measurement of fluency, specifically focusing on questions related to measuring pauses in L2 speech.

This first chapter briefly surveys definitions of fluency, followed by a discussion of the research questions and an overview of the rest of the dissertation.

What is fluency?

In a study of the development of L2 fluency, it is important to begin by clarifying what is meant by the term fluency. Researchers, teachers, and learners all use the term with a variety of meanings (Freed et al., 2004b; Koponen & Riggenbach, 2000). In a broad sense, fluency is sometimes used as a synonym of overall proficiency in a language (Chambers, 1997; Freed et al., 2004b; Hilton, 2008; Koponen & Riggenbach, 2000). This use of the term is common among learners expressing the desire to become fluent in a language (Hilton, 2008). In research on speaking, however, fluency often has a more narrow meaning, referring to quantifiable aspects of speech such as speech rate, pauses, repair, and mean length of run. Speech rate is generally calculated as the number of words (Freed et al., 2004b; Hilton, 2008), semantic units (words and partial words; Mota, 2003), or syllables (Kormos & Denes, 2004; Mehnert, 1998) spoken per second or per minute. Pauses refer to both silence and non-lexical fillers in speech (Chambers, 1997). Repair includes repetitions, replacements, reformulations, and false starts (Iwashita et al., 2008; see chapter 2 for a more detailed description of these measures). Mean length of run is a measure of the average number of words or syllables between pauses (Chambers, 1997). Fluency measured by these features of speech is frequently called utterance fluency (De Jong et al., 2013a, 2013b; Prefontaine, 2010).
Segalowitz (2000, 2010) has also introduced the concept of cognitive fluency. Cognitive fluency describes an individual's ability to efficiently plan and assemble an utterance, including elements such as its content, vocabulary, and grammatical form. Since it involves mental processes, it cannot be measured directly like utterance fluency (De Jong et al., 2013b).

Perceived fluency refers to a listener's impression of a speaker's fluency (Lennon, 2000). It is a “subjective reaction to speech” (Koponen & Riggenbach, 2000: 12), and therefore, different listeners may have different reactions to the same speech (Freed, 2000). Both linguistic and non-linguistic features may contribute to this impression. Linguistic features include temporal variables such as speech rate and pausing as well as other factors such as grammar, vocabulary, and pronunciation (Freed, 2000; Rossiter, 2009). Non-linguistic features include the perceived quality of interaction between the learner and an interviewer (Freed, 2000).

In the present study, fluency refers to utterance fluency, or temporal aspects of speech. Although this study does not measure perceived fluency, previous research has consistently found that utterance fluency and perceived fluency are related (Bosker et al., 2013; Derwing et al., 2009; Kormos & Denes, 2004; Riggenbach, 1991; Rossiter, 2009). Chapter 2 includes a more detailed discussion of each type of fluency and the relationship between them.

With the definition of fluency used in the study now clarified, the following section presents the research questions to be addressed, related to both the development and measurement of fluency.
Research questions

As previously indicated, the purpose of the study is twofold:

1. To explore the development of L2 fluency during a semester abroad and its relationship to the development of grammar, vocabulary, and processing speed, while also considering the influence of individual participants’ L1 and pre-SA L2 fluency

2. To examine issues in the measurement of fluency, specifically focusing on questions related to measuring pauses in L2 speech

The research questions are divided into two broad categories according to these two goals.

The development of fluency

Within the category of fluency development, there are six main questions, most of which have several sub-questions. An explanation follows each set of questions.

1. Are there changes in participants’ L2 articulation rate, pausing, repair, and mean length of run from pre- to post-study abroad (SA)?
   a. What changes are evident in terms of raw numbers?
   b. What changes are evident in terms of L1–L2 difference scores?

Question 1a looks at the mean scores for L2 articulation rate, pausing, repair, and mean length of run and compares those means for pre- and post-SA data. Question 1b examines changes in L2 fluency as measured against participants’ L1 fluency, using effect sizes to quantify L1–L2 differences. In other words, how much more similar to (or different from) their L1 fluency is participants’ L2 fluency after
studying abroad? As Segalowitz explains, “difference scores indicate something about how much more difficult the L2 is for the speaker, compared to the L1” (2010: 36). This measurement is based on the assumption that for L2 learners at the levels of proficiency of the participants in the study (i.e., learners not yet approaching native-like proficiency), greater fluency in the L2 more closely resembles their L1 fluency.2

2. Is there a relationship between pre-SA L2 fluency and changes in L2 fluency?
   a. Is there a correlation between pre-SA L2 fluency scores and change in L2 fluency scores from pre- to post-SA?
   b. Is there a significant difference between the amount of change in L2 fluency scores for participants with low pre-SA L2 fluency scores (below the median score) and the amount of change in L2 fluency scores for participants with high pre-SA L2 fluency scores (above the median score)?
   c. If there are differences in the amount of change in L2 fluency scores for participants with low and high pre-SA L2 fluency scores, in what areas of fluency (articulation rate, pausing, repair, or length of run) are the differences most evident?

2 This assumption does not take into account the fact that native-like fluency may differ across languages. For example, Campione and Véronis (2002) found that mean silent pause duration among native speakers varies from one language to another. Thus, the assumption that learners’ L2 fluency more closely resembles their L1 fluency as they increase in L2 proficiency may not hold true for learners at very advanced levels of L2 proficiency if their speech becomes fine-tuned to reflect features that characterize the specific L2. It seems unlikely, however, for such fine-tuning to take place among learners at lower levels of proficiency.
Question 2a addresses whether participants who begin at different levels of L2 fluency show similar or different amounts of change in their L2 fluency from pre- to post-SA. These measures were calculated both with and without controlling for L1 fluency. Previous studies have addressed a somewhat related question by examining whether learners who begin at different levels of overall L2 proficiency experience different amounts of growth in overall proficiency from pre- to post-SA, finding that learners who begin at lower levels of proficiency tend to experience greater measurable growth in proficiency (Freed, 1990; Lindseth, 2010; Milleret, 1991). It was expected that findings for growth in fluency might be similar, as one previous study has found (Towell, 2002). Question 2c examines the possible differences in fluency development between participants with low and high pre-SA fluency scores in greater detail, looking at differences in specific sub-constructs of fluency.

3. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and pre- and post-SA L2 fluency scores?

   a. Is there a correlation between participants’ pre-SA grammar scores and their pre-SA fluency scores? Is there a correlation between their pre-SA grammar scores and their post-SA fluency scores? Is one correlation stronger than the other?

   b. Is there a correlation between participants’ pre-SA vocabulary scores and their pre-SA fluency scores? Is there a correlation between their pre-SA vocabulary scores and their post-SA fluency scores? Is one correlation stronger than the other?
c. Is there a correlation between participants’ pre-SA processing scores and their pre-SA fluency scores? Is there a correlation between their pre-SA processing scores and their post-SA fluency scores? Is one correlation stronger than the other?

d. Considering as possible factors pre-SA grammar, vocabulary, picture naming, and sentence matching scores, what regression model best predicts pre-SA L2 fluency?

Questions 3a through 3c examine whether the relationship between present grammar, vocabulary, and processing scores and present fluency scores is stronger or weaker than the relationship between present grammar, vocabulary, and processing scores and future fluency scores. In other words, does the relationship between each of these areas of knowledge or skill and fluency appear to be immediate or delayed? All measures were calculated both with and without controlling for L1 fluency. Question 3d asks which factors best predict pre-SA L2 fluency, allowing for comparisons with previous research looking at these relationships at one point in time.

4. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and changes in fluency?

   a. Is there a correlation between participants’ pre-SA grammar scores and change in fluency scores from pre- to post-SA?

   b. Is there a correlation between participants’ pre-SA vocabulary scores and change in fluency scores from pre- to post-SA?
c. Is there a correlation between participants’ pre-SA processing scores and change in fluency scores from pre- to post-SA?

d. What is the relative strength of each of these correlations?

Questions 4a through 4c examine whether there are certain levels of pre-SA grammar, vocabulary and processing speed that are most strongly associated with changes in fluency during SA. Again, measures were calculated both with and without controlling for L1 fluency. Question 4d asks whether the relationship between one of these areas of knowledge/skill and changes in fluency is stronger than the others.

5. Is there a relationship between changes in grammar, vocabulary, and/or processing scores and changes in L2 fluency?

   a. Is there a correlation between change in grammar scores and change in fluency scores from pre- to post-SA?

   b. Is there a correlation between change in vocabulary scores and change in fluency scores from pre- to post-SA?

   c. Is there a correlation between change in processing scores and change in L2 fluency scores from pre- to post-SA?

   d. What is the relative strength of each of these correlations?

Questions 5a through 5c examine the relationship between changes in grammar, vocabulary, and processing speed and changes in L2 fluency. As with the previous research questions, these measures were calculated both with and without controlling for L1 fluency. Question 5d asks whether the relationship is stronger in one of these areas than in others.
6. Considering as possible factors L1 fluency, pre-SA L2 fluency, changes in grammar from pre- to post-SA, changes in vocabulary from pre- to post-SA, changes in processing speed from pre- to post-SA, and time spent interacting with native speakers of the L2 during the semester abroad, what regression model best predicts changes in L2 fluency during a semester abroad?

Question 6 asks which of a number of variables considered in the study predict changes in L2 fluency from pre- to post-SA, and how strongly each of these variables contributes to predicting changes in L2 fluency.

**The measurement of fluency**

As previously mentioned, there is a lack of consensus regarding the measurement of L2 fluency, as evidenced by the use of many different measures in studies examining similar questions (Foster & Skehan, 1996; Gilabert, 2007; Mehnert, 1998; Michel et al., 2007; Skehan & Foster, 1997; Tavakoli & Foster, 2011; Wigglesworth & Elder, 2010). One particular area where work is needed is in the measurement of pausing. Although the large number of studies including one or more measures of pausing suggests some level of agreement that pausing should be measured, the specific ways in which it has been measured vary considerably. For example, the minimum length of pauses measured ranges from 0.2 seconds (Kormos & Denes, 2004) to 2 seconds (Nakakubo, 2011; Tajima, 2003). Additionally, researchers vary in how they treat filled and unfilled pauses, with several options available: measuring them together, measuring them separately, or measuring only one or the other. Whether or not filled and unfilled pauses differ in what they reveal about fluency is not clear from the few studies that have measured them separately.
(Iwashita et al., 2008; Kormos & Denes, 2004). Finally, very few studies have distinguished between mid-clause and end-of-clause pauses (Skehan & Foster, 2005; Tavakoli & Foster, 2011), but L1 data suggest that pausing at the end of clauses is natural, whereas pausing in the middle of a clause is less common in native speaker speech (Chambers, 1997; Davies, 2003; Skehan, 2009). Exploring patterns of mid-clause and end-of-clause pausing across participants with different levels of lexical–grammatical competence may be valuable in determining what pause location reveals in L2 speech.

In light of these limitations in previous research, the present study addresses the following measurement-related questions:

1. Are there significant differences between the speech of learners at different levels of lexical–grammatical competence in the number of pauses 0.5 seconds and longer? Are there significant differences between these groups in the number of shorter pauses (0.25 to 0.49 seconds)?

2. Do filled and unfilled pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

3. Do mid-clause and end-of-clause pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

---

3 For the purpose of addressing these questions, participants were divided into high and low groups based on their scores on pre-study abroad measures of grammar and vocabulary, and their speech samples also came from pre-study abroad data.

4 The use of 0.5 seconds as a dividing line comes from Goldman-Eisler’s (1961) finding that short pauses of less than 0.5 seconds are a common feature of native speaker speech in contexts including discussions and giving summaries.
Overview of the dissertation

This first chapter has discussed the motivation for the study, definitions of fluency, and the research questions to be addressed. Chapter 2 begins with a more in-depth discussion of fluency, followed by a review of other relevant literature on the relationship between L1 and L2 fluency, knowledge and skills underlying fluency, changes in fluency in the context of study abroad, and issues in the measurement of fluency. Chapter 3 describes the methodology of the study, including information about the participants, materials, tasks, and data collection and analysis. Chapter 4 presents the results, and chapter 5 discusses these findings.
Chapter 2: Previous research on fluency

This chapter begins with a discussion of fluency and how it has been measured, building on the brief discussion of fluency in chapter 1. Next, models of fluent speech production in both L1 and L2 are described. Following the presentation of this background information, several areas of research related to the topic of the present study are examined, including research on the relationship between L1 and L2 fluency, knowledge and skills underlying fluency, and fluency development in the context of study abroad. The final section of the chapter discusses issues in the measurement of pauses in previous research on fluency.

Fluency: Measurement and development

As mentioned in chapter 1, fluency has been defined and measured in a variety of ways. This section first explores broad definitions of fluency and then describes how more specific types of fluency—perceived fluency, utterance fluency, and cognitive fluency—have been measured in previous studies, as well as what research suggests about the relationship between these types of fluency.

Broad definitions of fluency

In its broad definitions, fluency is viewed as similar to overall speaking proficiency. An early and often-cited example of such a view is found in Fillmore’s (1979) essay on fluency. Fillmore describes four ways in which speakers may be fluent. The first is “the ability to fill time with talk,” speaking with few pauses (1979: 93). Fillmore gives the example of disc jockeys and sports announcers who need to speak this way on the job. The second aspect of fluency that Fillmore describes is the ability to speak coherently using “semantically dense” sentences, with few
superfluous words or fillers (1979: 93). The third dimension involves knowing what
to say in a large variety of contexts, not becoming flustered in unfamiliar situations.
Lastly, Fillmore considers the ability to be creative with language—displayed
through jokes, metaphors, and the use of different styles—as a dimension of fluency.
Fillmore concludes that a person with all four of these abilities is “maximally gifted”
in their use of language (1979: 93).

   Whereas Fillmore’s view of fluency focuses on native speakers, Koponen and
Riggenbach (2000) present a broad view of fluency that primarily applies to non-
native speakers. They point out that the terms employed in a variety of languages to
describe overall L2 speaking proficiency convey the idea of smooth or fluid speech.
For example, German, French, and Russian all use terms that mean in a running
manner, and Swedish and Finnish use words that mean flowingly. Thus, it seems that
the ability to speak with smooth and connected speech is closely associated with
common views of what it means to speak a language well.

   Although each of these views of fluency reflects widely used definitions of the
term, in research, fluency often refers to something more specific. The next section
describes ways in which the term has been defined and operationalized in previous
studies.

**Perceived, utterance, and cognitive fluency**

**Definition and measurement**

*Perceived fluency*

   Researchers have approached measuring fluency in several different ways.
Perceived fluency refers to a listener’s impression of a speaker’s fluency (Lennon,
2000) and has been measured by asking teachers (Freed, 2000; Kormos & Denes, 2004; Riggenbach, 1991; Rossiter, 2009) or other native (Bosker et al., 2013; Derwing et al., 2009; Freed, 2000) or nonnative (Rossiter, 2009) speakers to rate speech samples. Since L2 learners presumably want to sound fluent to others, using listeners’ perceptions to rate speakers’ fluency intuitively makes sense. However, this approach creates certain challenges in research. One important issue is whether and how to define fluency for the raters. Some studies have allowed raters to make judgments based on their own definitions (Freed, 2000; Kormos & Denes, 2004), which may lead to raters basing their ratings on different criteria (Rossiter, 2009). Others (Bosker et al., 2013; Derwing et al., 2009; Rossiter, 2009) have provided definitions based on temporal aspects of fluency. For example, Bosker et al. (2013) instructed raters to listen for speech rate, pauses, and self-corrections. This approach, then, seems like an indirect measure of utterance fluency, in which case measuring utterance fluency directly would appear to provide more objective measures.

**Utterance fluency**

Utterance fluency refers to quantifiable aspects of speech such as rate, pauses, repair, and mean length of run. The two main measures of rate include the unpruned and pruned speech rates. Unpruned speech includes everything that a participant produces (although often excluding non-word fillers), while pruned speech eliminates instances of repair from the calculations (Mota, 2003; Yuan & Ellis, 2003). Within these categories, there are minor variations in operationalizing speech rate, such as measuring words (Freed et al., 2004b; Hilton, 2008), semantic
units (words and partial words; Mota, 2003), or syllables (Kormos & Denes, 2004; Mehnert, 1998). Additionally, some or all pauses are sometimes excluded from calculations (Iwashita et al., 2008; Mochizuki & Ortega, 2008; Préfontaine, 2010).

There are a number of variations in how pauses are measured. Some studies use a count of the total number of pauses (Foster & Skehan, 1996, Mehnert, 1998; Rossiter, 2009; Tavakoli & Skehan, 2005), while others include ratios of pauses to units, such as pauses per T-unit5 (Bygate, 2001), or pauses to time, such as pauses per minute (Iwashita et al., 2008; Kormos & Denes, 2004; Llanes & Muñoz, 2009; Skehan & Foster, 2005). Additionally, some researchers calculate pausing time as a percent of total speaking time (Hilton, 2008; Iwashita et al., 2008; Mehnert, 1998; Sangaran, 2005) or calculate the phonation-time ratio, the inverse of this measure (Kormos & Denes, 2004; Préfontaine, 2010). If all participants speak for the same amount of time, a simple count may be adequate, whereas ratios or percentages are important to facilitate meaningful comparisons if speech samples differ in length.

In addition to how pauses are measured, another variation between studies is which pauses are measured. Some researchers count only filled (García-Amaya, 2009) or only unfilled (Guará Tavares, 2012) pauses, whereas others count both but distinguish between them (Kormos & Denes, 2004; Rossiter, 2009), and still others make no distinction between them (Foster & Skehan, 1996; Mehnert, 1998; Tavakoli et al., 2011). It has been noted that not all pauses sound equally dysfluent, with unfilled pauses perhaps creating a greater sense of dysfluency (Chambers, 1997).

---

5 A T-unit is “one main clause plus whatever subordinate clauses happen to be attached to or embedded within it” (Hunt, 1965: 305).
Others regard them as similar, arguing that both filled and unfilled pauses equally “reflect an inability to handle speech during uninterrupted time” (Skehan & Foster, 2005: 199). Skehan and Foster (2005) do, however, advocate for distinguishing between mid-clause and end-of-clause pauses, given that studies suggest that native speakers tend to pause at the end of clauses. Further discussion of this point occurs in the later section on issues in measuring pauses.

Length of pause is another aspect of measurement that varies between studies. Some researchers have used the average pause length as a measure of breakdown\(^6\) fluency (Hilton, 2008; Kormos & Denes, 2004; Tavakoli & Skehan, 2005; Wendel, 1997). In studies counting the number of pauses, not all indicate the length of pauses included in this total, but those that do vary from 0.2 seconds or longer (Kormos & Denes, 2004) to 2 seconds or longer (Nakakubo, 2011; Tajima, 2003).

In addition to speech rate and pausing, repair is sometimes included as a dimension of utterance fluency. As measured in previous studies, repair includes repetitions (Bygate, 1996; Foster & Skehan, 1996; García-Amaya, 2009; Hilton, 2008; Iwashita et al., 2008; Kormos & Denes, 2004; Rossiter, 2009; Tavakoli & Foster, 2011), replacements (Foster & Skehan, 1996; García-Amaya, 2009; Iwashita et al., 2008; Kormos & Denes, 2004; Rossiter, 2009), reformulations (Foster & Skehan, 1996; Iwashita et al., 2008; Rossiter, 2009; Skehan & Foster, 2005; Tavakoli & Foster, 2011), and false starts (Foster & Skehan, 1996; Iwashita et al., 2008; Kormos & Denes, 2004; Rossiter, 2009; Skehan & Foster, 2005; Tavakoli & Foster, 2011).
2011). Repetitions are exact duplications of the same syllable, word, or phrase (Iwashita et al., 2008). Replacements involve the substitution of one lexical item for another (Inoue, 2010), while reformulations generally refer to grammatical corrections (Iwashita et al., 2008). A false start occurs when a speaker begins to express an idea but abandons it without completing it (Inoue, 2010).

Length of run is the final commonly used measure of utterance fluency. Mean length of run has been operationalized in slightly different ways across studies. The main variation is in how researchers have divided speech samples into runs. As in measures of pausing, the length of pause signaling the end of one run and the beginning of another differs, ranging from 0.25 seconds (Kormos & Denes, 2004; Préfontaine, 2010) to 1 second (Mehnert, 1998). Freed et al. (2004b) calculated mean length of run in two ways, using filled pauses to separate speech into runs in one measure and using false starts and repetitions in the other. Additionally, as an alternative to mean length of run, some researchers have measured only the longest run (Llanes & Muñoz, 2009; Segalowitz & Freed, 2004) or the 15 longest turns (García-Amaya, 2009) of each participant.

**Cognitive fluency**

Although the approach to its measurement has not been consistent, it is evident from the preceding examples that utterance fluency has been measured in many studies. In contrast, very few studies have attempted to measure cognitive fluency. Cognitive fluency describes an individual's ability to efficiently plan and assemble an utterance, including elements such as its content, vocabulary, and grammatical form (Segalowitz, 2010). Since it involves mental processes, it cannot
be measured directly (De Jong et al., 2013b). In one study that did address cognitive fluency, De Jong et al. (2013a) operationalized cognitive fluency in terms of grammar, vocabulary, and processing abilities. They measured knowledge of grammar and vocabulary in untimed tasks and also used several tasks to measure processing speed, including speed of lexical retrieval, speed of sentence completion, and speed of articulation. They determined that such tasks could “tap most aspects of L2 linguistic skills that supposedly underlie L2 cognitive fluency” (De Jong et al., 2013a: 7).

This section has examined the measurement of perceived, utterance, and cognitive fluency; the following section describes studies that have explored the relationship between these types of fluency.

The relationship between perceived, utterance, and cognitive fluency

Several studies have looked at both utterance fluency and perceived fluency, and all have found a significant relationship between the two. One of the first studies to examine this relationship was Riggenbach’s (1991) study of six nonnative speakers (NNS) of English. Teachers judged three of the participants to be fluent and three to be non-fluent. Riggenbach found significant differences between the fluent and non-fluent groups in their speech rate and the number of unfilled pauses 0.5 seconds and longer. There were no significant differences for short pauses (less than 0.5 seconds), filled pauses, total pauses, or repair; however, in some cases, the lack of significance may be due to the small number of participants.

Kormos and Denes (2004), Rossiter (2009), Derwing et al. (2009), and Bosker et al. (2013) have all conducted larger studies on this topic, using
correlations to explore the relationship between quantitative measures of fluency and raters’ judgments. Kormos and Denes’s (2004) study included 16 NNS of English and found significant correlations for speech rate, phonation-time ratio, mean length of run, mean length of pauses, and number of words. Correlations were not significant for the articulation rate, the number of silent pauses per minute, the number of filled pauses per minute, or the number of disfluencies per minute. Twenty-four NNS of English participated in Rossiter’s (2009) study. Ratings by novice ESL teachers, experienced ESL teachers, and NNS of English all correlated significantly with pauses per second and speech rate, the two measures of utterance fluency used in the study. Similarly, in Derwing et al.’s (2009) study of 32 NNS of English, ratings correlated with all three measures of utterance fluency: pausing, the unpruned speech rate, and the pruned speech rate. Finally, Bosker et al. (2013) studied 30 NNS and 8 native speakers (NS) of Dutch and found significant correlations between ratings and speech rate, number of silent pauses, mean length of silent pauses, number of repetitions, and number of corrections. The correlation was not significant for number of filled pauses.

Speech rate stands out as the one measure of utterance fluency that significantly correlated with judges’ ratings in every study. Although the findings of these studies vary in the other measures of utterance fluency that correlated with ratings, the fact that all of them found a significant relationship between utterance fluency and perceived fluency would appear to provide strong evidence that the two are indeed related. However, De Jong et al. (2013a) mention one necessary caution in interpreting the results: the studies in which fluency was defined for raters in
terms of utterance fluency (Bosker et al., 2013; Derwing et al., 2009; Rossiter, 2009) might be viewed as circular. When raters are told to pay attention to the particular aspects of speech that are also measured quantitatively, this approach may increase the likelihood that the two types of measures will correlate significantly.

Nevertheless, given that the studies that did not define fluency for raters also found significant correlations between ratings and quantitative measures, there still seems to be evidence of a relationship between utterance fluency and perceived fluency.

Unlike utterance fluency and perceived fluency, the relationship between utterance fluency and cognitive fluency has rarely been examined, and according to Segalowitz (2010), it has never been studied with an appropriate research design. Segalowitz and Freed (2004) looked at the relationship between fluency and underlying cognitive skills, finding significant correlations between mean length of run without filled pauses and lexical access speed ($r = .375$) and lexical access efficiency ($r = .377$). Lexical access speed was measured by reaction time in a semantic classification task, and lexical access efficiency is a measure of variation in reaction times. Segalowitz (2010), however, does not consider the findings to be reliable evidence of a relationship between utterance fluency and cognitive fluency for two reasons. First, the study included no measures of L1 fluency, with the possible consequence that “individual speech differences unrelated to the L2 [perhaps] have provided an unwanted source of noise that may have masked important L2 fluency phenomena” (Segalowitz, 2010: 35). Second, it used the ACTFL Oral Proficiency Interview (OPI) to collect speech samples, which Segalowitz (2010)
believes introduces too much variability in the samples since the OPI does not include exactly the same speaking tasks for every participant.

One other study has looked at the relationship between cognitive fluency and utterance fluency. De Jong et al. (2013a) defined cognitive fluency as linguistic knowledge and processing skills, and they calculated the correlations between these measures and measures of utterance fluency. Based on their findings, they concluded that there is a relationship between most measures of utterance fluency (number of silent pauses, filled pauses, corrections, repetitions, and mean syllable duration, but not mean silent pause duration) and cognitive fluency. The design of this study avoided one of the problems that Segalowitz (2010) identified by having all of the participants complete the same set of speaking tasks. However, it did not address Segalowitz's concern that failing to take L1 fluency into account may be a serious limitation.

In summary, while initial findings suggest that cognitive fluency and utterance fluency are related, more research is needed prior to drawing firm conclusions about their relationship. A complicating factor is that since cognitive fluency cannot be measured directly, it is difficult to determine exactly what constitutes research on cognitive fluency.

The present study measures fluency as utterance fluency; the specific measures used are described in chapter 3. Given the findings of the research discussed in this section, it seems likely that utterance fluency as measured in the present study is related to perceived fluency and possibly cognitive fluency.
However, no claims about either of these types of fluency are made based on measures of utterance fluency in the study.

The following section surveys previous work on how fluent speech is produced, in both L1 and L2 speech.

Models of fluent speech production

Levelt's (1989) model is perhaps the most influential model of L1 speech production. Given that L2 models have built on Levelt's work (De Bot, 1992; Kormos, 2006), it is fitting to begin with a brief overview of his model.

Levelt (1989) describes speech production in three stages: conceptualizing, formulating, and articulating. Conceptualizing refers to selecting and ordering the information that one wants to convey. This process produces what Levelt calls the *preverbal message*, which contains the ideas but not the structure of the utterance that is produced.

The second stage, formulating, takes the preverbal message and gives it linguistic structure. This process includes both grammatical and phonological encoding. Grammatical encoding involves retrieving lemmas, which are lexical items together with relevant syntactic information for assembling them. When the lemmas are retrieved and assembled, the result is a *surface structure* ready for phonological encoding. Phonological encoding then creates a plan for how the intended utterance will be articulated. In the final stage, articulation, this plan is physically carried out, producing *overt speech*.

In L1 speech, these processes operate with a high degree of automaticity, which allows them to function in parallel. Levelt argues that the ability of these
processes to work in parallel is "a main condition for the generation of uninterrupted fluent speech" (1989: 2). Models of bilingual or L2 speech production may help explain why L2 speech is often less fluent than L1 speech by identifying ways in which the automatic functioning of speech production processes can break down while producing L2 speech.

In adapting Levelt's (1989) model for bilingual speech production, De Bot (1992) describes several aspects of speech affecting bilingual speakers that do not affect monolingual speakers. First, a person who speaks two or more languages must decide in which language to produce an utterance. De Bot believes that the information determining this selection is included in the preverbal message created during the process of conceptualizing. Presumably, this difference does not affect the fluency of speech production.

However, in formulating the utterance, there are differences between monolingual and bilingual speech that may affect the automaticity with which speech is produced. For example, as De Bot (1992) indicates, the preverbal message may contain concepts requiring lexical items that do not exist in the L2, or which the speaker has never learned in the L2, or which the speaker cannot quickly retrieve in the L2. Although lexical retrieval problems do occur in L1 speech (Hartsuiker & Notebaert, 2010), they are much more common among bilingual speakers, particularly when speaking an L2 in which the speaker has not attained a high level of proficiency. Another problem more likely to occur in L2 speech than in L1 speech is that a speaker may be able to retrieve lexical items but not the relevant syntactic information for assembling them. When lexical items are retrieved without the
necessary syntactic information, the formulator encounters difficulty while constructing the surface structure.

In her bilingual adaptation of Levelt’s (1989) L1 model, Kormos (2006) also identifies differences between L1 and L2 speech production at the level of formulating that may influence fluency. She proposes that, whereas L1 syntactic and phonological rules are believed to be automatized and stored together with lexical items as lemmas, a portion of these rules in the L2 are stored separately as declarative knowledge rather than automatized knowledge. Given Levelt’s (1989) emphasis on the importance of automaticity for fluent speech, it is a logical conclusion that having a portion of the required knowledge for producing an utterance stored as declarative knowledge would interfere with the fluency of speech production.

In view of these different sources of L2 dysfluency—limited lexical and grammatical knowledge, slow lexical retrieval, and declarative rather than automatized knowledge of rules—the present study includes both untimed measures of lexical and grammatical knowledge as well as timed measures of speed of lexical retrieval and syntactic processing. The purpose of employing this range of measures is to allow for a closer examination of possible sources of L2 dysfluency for the participants in the study.

**The relationship between L1 and L2 fluency**

In spite of differences between L1 and L2 speech production that may contribute to dysfluency in L2 speech, many of the underlying processes of L1 and L2 speech are assumed to be similar, as reflected in De Bot’s (1992) and Kormos’s
use of Levelt’s (1989) L1 model as the starting point for their bilingual models. This view would predict that although L1 and L2 fluency differ in some respects, there is also a relationship between the two. Studies measuring both L1 and L2 fluency have generally found evidence of such a relationship.

For example, Towell and Dewaele (2005) measured the L1 and L2 fluency of 12 L1 English learners of French before and after they spent six months abroad. Operationalized as speech rate, L1 and L2 fluency significantly correlated both before (r = 0.81) and after (r = 0.73) study abroad.

Derwing et al. (2009) conducted a more extensive study, measuring the L1 and L2 fluency of 16 L1 Mandarin and 16 L1 Russian or Ukrainian learners of English at several points in time: two months into the study—when the participants had been in Canada for six months or less—and again one and two years into the study. For the L1 Russian and Ukrainian learners, temporal measures of fluency (pausing, speech rate, and pruned speech rate) were all significantly correlated (r = 0.548 to 0.722) in the L1 and L2 each time they were measured, except for the pruned speech rate two years into the study. For the L1 Mandarin learners, the L1 and L2 correlations were significant (r = 0.524 to 0.618) two months into the study only; none were significant later on. The authors suggest that among the L1 Mandarin learners, those with the greatest L1 fluency may have had an initial advantage in speaking the L2. However, these individuals appear to have improved little in L2 fluency over time, whereas others made more progress.

De Jong et al. (2013b) also studied learners from two different L1 backgrounds, 29 L1 English and 24 L1 Turkish learners of Dutch. They used seven
temporal measures including aspects of fluency such as mean syllable duration, filled and unfilled pauses, repetitions, and corrections. Each of these measures significantly correlated between the L1 (English or Turkish) and the L2 (Dutch), ranging from $r = 0.37$ for mean syllable duration to $r = 0.76$ for mean silent pause duration between Analysis of Speech Units, with most correlations at $r = .60$ and above. The relationship between L1 and L2 fluency did not differ for the groups with different L1 backgrounds.

These studies demonstrate that there does appear to be a relationship between L1 and L2 fluency. However, most studies of L2 fluency do not take L1 fluency into account. Segalowitz (2010) argues that the lack of L1 fluency data is a significant oversight in studies of L2 fluency, because differences between individuals’ L2 fluency may be partly due to their general speaking style. He suggests collecting L1 fluency data to use as “baseline measures” (2010: 35) when examining L2 fluency, which the present study does.

This section has surveyed the research on the relationship between L1 and L2 fluency; the next section examines research on the relationship between fluency and underlying knowledge and skills.

**Knowledge and skills underlying fluency**

Whereas a number of studies have looked at L1 and L2 fluency, research on the relationship between fluency and underlying knowledge and skills is more limited. As previously indicated in the discussion of models of speech production, these models would predict that L2 dysfluency is related, at least in part, to limitations in areas such as vocabulary and grammatical knowledge as well as a lack
of automatization of L2 knowledge. Two studies (De Jong et al., 2013a; Hilton, 2008) examining this topic have found a relationship between grammar and vocabulary and fluency, and one of these studies (De Jong et al., 2013a) also found a relationship between processing skills and fluency.

In Hilton’s (2008) study, 47 participants from various L1 backgrounds completed monologic speaking tasks in English, their L2. Their knowledge of English grammar and vocabulary was also tested using DIALANG, a diagnostic test designed to help language learners estimate their level of proficiency and identify their strengths and weaknesses in a variety of subskills (Alderson, 2005). Hilton looked at the correlations between grammar and vocabulary scores and seven measures of fluency, including speech rate, pausing, and repair. She found significant correlations, ranging from $r = .390$ to $r = .733$, between every measure of fluency and both grammar and vocabulary scores, although every correlation was stronger for grammar than for vocabulary. For both grammar and vocabulary, the strongest correlation was with mean length of run. However, Hilton’s results should be interpreted with caution given the use of DIALANG to measure grammar and vocabulary knowledge. DIALANG was created to provide feedback to individual learners about their language skills; it was not intended as a research tool or certification exam. Also, the project to develop DIALANG was suspended in 2006 due to a lack of funding, and the version that has been available to the public since then is a prototype (“Information about DIALANG,” 2013).

De Jong et al.’s (2013a) participants included 179 learners of Dutch from a variety of L1 backgrounds. They completed eight monologic speaking tasks as well
as six measures of grammar, vocabulary, and processing speed. Their speech samples were analyzed using six measures of utterance fluency, including number of silent pauses, mean silent pause duration, filled pauses, corrections, repetitions, and mean syllable duration. The researchers found significant correlations between every measure of grammar, vocabulary and processing and at least two measures of utterance fluency. The measures most strongly correlated with fluency were vocabulary knowledge and sentence building speed; both correlated with every measure of utterance fluency except mean silent pause duration, with correlations as high as $r = .66$ between sentence building speed and mean syllable duration.

Although the correlations were lower ($r = .16$ to .32), speed of lexical retrieval also significantly correlated with every measure of utterance fluency.

The findings of these two studies suggest that grammar, vocabulary, and processing speed are all related to fluency. However, given the limited number of studies and the fact that Hilton's (2008) study is subject to criticism since DIALANG is not intended as a research instrument, more research is needed to better understand the relationship. The present study contributes to this line of research by selecting participants learning a different L2, Spanish. Additionally, unlike the studies reviewed here, the present study looks at the relationship over time (pre- and post-SA) rather than at only one point in time.

---

7 Other measures of pausing, including silent pauses per 100 words and filled pauses per 100 words, significantly correlated with vocabulary knowledge and sentence building speed.
While rarely examining the relationship to underlying skills, a number of studies have looked at fluency development over time in the context of study abroad. The following section surveys this research.

**Fluency development and study abroad**

Previous research provides fairly consistent evidence of growth in general speaking proficiency (Brecht et al., 1995; Hernández, 2010; Lindseth, 2010; Magnan & Back, 2007) as well as fluency (D'Amico, 2010; Lennon, 1990; Llanes & Muñoz, 2009; O’Brien et al., 2007; Towell, 2002; Valls Ferrer, 2008) through study abroad. This section describes the research that has been done and identifies areas where more research is needed, such as studies exploring whether growth in fluency during SA differs for participants who begin with different levels of fluency as well as studies looking at changes in grammar, vocabulary, and processing speed along with changes in fluency.

**Overall proficiency: Improvement on the ACTFL scale**

Several studies have examined growth in oral proficiency during study abroad using the American Council on the Teaching of Foreign Languages (ACTFL) scale (Swender et al., 2012). The ACTFL scale provides detailed descriptions of proficiency at five levels, ranging from Novice to Distinguished. The three lower levels—Novice, Intermediate, and Advanced—are each broken down into the sublevels Low, Mid, and High. Although studies measuring growth on this scale do not directly provide evidence of growth in fluency as defined in the present study, they are still somewhat relevant, as fluency is one aspect of proficiency included in the level descriptors. For example, speech at the Advanced Low level may be
“marked by an irregular flow, and containing noticeable self-correction” (Swender et al., 2012: 6). Intermediate Low speech is “characterized by frequent pauses, ineffective reformulations and self-corrections” (Swender et al., 2012: 8).

Studies using the ACTFL scale have found that most students who study abroad for at least a semester experience measurable growth on this scale. In Brecht et al.’s (1995) analysis of 646 learners of Russian who studied abroad for one semester, 452 (70%) students improved by at least one sublevel from pre- to post-SA, with 138 (21%) improving by two or more sublevels. Studies in other languages have included far fewer participants but have found similar results. Magnan and Back (2007) tracked the progress of 20 learners of French during a semester in France, finding that 12 students (60%) had measurable gains. Of these 12 students, half improved by one sublevel, while the other half improved by two sublevels. In Lindseth's (2010) study of learners of German, 31 of 38 students (81.6%) who spent a semester in Germany improved by at least one sublevel, including seven (18.4%) who improved by two sublevels. Finally, in Hernández’s (2010) study, 16 of 20 learners of Spanish (80%) who spent a semester in Mexico improved by at least one sublevel, with five (25%) improving by two sublevels.

The findings of these studies demonstrate that most students who study abroad for a semester improve their speaking abilities. Given that fluency is a factor in the level descriptors for the ACTFL scale, it seems likely that growth in fluency was part of the overall growth in proficiency for at least some participants in these studies. However, the data presented do not allow for an examination of this
assumption. More specific measures of fluency are needed in order to make such a claim, and the studies presented in the following section included these measures.

**Detailed analyses of changes in fluency**

A number of studies have examined changes in fluency from pre- to post-SA using multiple measures of fluency, allowing for a detailed analysis. Most have found statistically significant changes (D’Amico, 2010; Lennon, 1990; Llanes & Muñoz, 2009; O’Brien et al., 2007; Towell, 2002; Valls Ferrer, 2008), although there is variation between studies in the particular measures of fluency for which the changes were significant. This section presents the results of these studies, suggests possible reasons for the variation in results, and also considers their limitations.

Several studies have examined changes in the fluency of participants who studied abroad for at least one semester. Lennon (1990) conducted one of the first of such studies, a small-scale study involving four advanced learners of English from Germany. All participants were English majors who had previously studied English for approximately 10 years. They completed a picture-elicited story narration task before and after a six-month stay in England. Lennon found a significant increase in their pruned speech rate and a significant decrease in filled pauses and the percent of T-units followed by a filled or unfilled pause. Changes were not significant for unpruned speech rate, repetitions, self-corrections, mean length of run, percent of total pause time at T-unit boundaries, or mean pause time at T-unit boundaries. However, the lack of significant differences in some of these categories may have been due to the small number of participants.
Towell (2002) conducted a similar study of 12 intermediate learners of French from the United Kingdom who spent six months in France. The participants completed two monologic speaking tasks, telling a personal story and giving a continuation of a fictitious story, before and after studying abroad. Results indicated a significant increase in speech rate, mean length of run, and phonation-time ratio. There was not a significant change in average length of pause. Towell also found that participants who exhibited lower fluency at the beginning of the study generally had the largest gains in fluency.

O’Brien et al. (2007) studied a larger group of participants, 25 learners of Spanish from the United States who had previously completed at least two semesters of Spanish courses. Excerpts from OPIs before and after participants’ 13 weeks in Spain provided data for the study. There was a significant increase in speech rate, mean length of run without filled pauses, and mean length of run without any pauses (filled or unfilled). Change in mean length of run without unfilled pauses was not significant.

Valls Ferrer’s (2008) study is the largest of its kind to date. Her study included 30 advanced learners of English from Spain, all of whom were students in a Translation and Interpreting program. Before and after studying in the United Kingdom for three months, they completed a paired interview in which the students asked each other questions. There were significant changes from pre- to post-SA for all variables measured, including speech rate, mean length of run, phonation-time
ratio, articulation rate, dysfluency ratio,\textsuperscript{8} pause frequency, pause duration ratio,\textsuperscript{9} and internal pause duration ratio.\textsuperscript{10}

Each of the four studies discussed above found significant changes in participants' fluency. However, one other study that examined fluency before and after study abroad found no significant differences. Freed et al. (2004b) studied eight learners of French from the United States who had previously studied French for two to four years. Before and after spending 12 weeks in France, they completed an interview similar to the OPI. Measures of fluency included speech rate and several measures of speech runs.\textsuperscript{11} Explaining the lack of significant differences, the researchers suggest that the lack of growth may be partly due to the fact that participants reported speaking more English than French during their time abroad.

In addition to these studies of students who spent a semester or more abroad, two other similar studies examined changes in fluency among students who participated in shorter study abroad programs. D'Amico's (2010) participants were nine intermediate learners of Spanish from the United States who studied in Spain for six weeks. Data were collected through informal interviews that included questions primarily about participants' families, likes and dislikes, past experiences,\textsuperscript{8,9,10}

\textsuperscript{8} The mean number of dysfluencies per minute
\textsuperscript{9} The mean length of all pauses over 0.4 seconds
\textsuperscript{10} The author does not define internal pauses.

\textsuperscript{11} These measures included hesitation-free runs (the average number of words spoken without a silent pause 400 ms or longer), filler-free speech runs (the average number of words spoken without a French or English filler such as “uh”), fluent runs (the number of words in the longest segment of speech with neither a pause nor a filler), repetition-free runs (the average number of words spoken without a false start or repeat), and grammatical-repair-free speech run (the number of words in the longest segment of speech without self-correction).
and future plans. From pre- to post-SA, there was a significant increase in their speech rate, mean length of run, and repair frequency. There were no significant changes in filled and unfilled pauses or dysfluency clusters (defined as two or more dysfluencies grouped together).

Llanes and Muñoz (2009) also studied participants in short-term SA programs. Their study included 24 learners of English who spent three to four weeks abroad; all had studied English for at least five years and about half had previously been abroad for the purpose of improving their English. They completed a picture-elicited story narration task before and after their time abroad. Results indicated a significant increase in speech rate and longest fluent run, with no significant changes in filled and unfilled pauses.

Together, the results of these studies show that students generally do experience growth in fluency during study abroad; six of the seven studies found significant changes for at least some measures of fluency. These studies provide fairly consistent evidence of increases in speech rate and mean length of run. However, results are less conclusive for other areas, such as measures of pausing.

There are a number of possible reasons for the variation in findings. First, the studies measured pausing in different ways, including the frequency of filled pauses, the frequency of unfilled pauses, the frequency of all pauses, and the average length of pause. Additionally, the studies included participants at a variety of proficiency levels. Although it is difficult to quantify proficiency using the background information provided in the studies, it seems likely that students who had studied the L2 for as little two semesters before going abroad (O'Brien et al., 2007) began
their study abroad experience with substantially more limited abilities than students who had previously studied the L2 for 10 years (Lennon, 1990). The studies also included a range of length of study abroad programs, including programs as short as three weeks (Llanes & Muñoz, 2009) and as long as six months (Lennon, 1990; Towell, 2002). Another potential explanation lies in the use of different types of tasks to collect data, including monologues (Lennon, 1990; Llanes & Muñoz, 2009; Towell, 2002), student-to-student interviews (Valls Ferrer, 2008), and student-to-researcher or teacher interviews (D’Amico, 2010; Freed et al., 2004b; O’Brien et al., 2007). Finally, as Freed et al. (2004b) suggest, differences in students’ experiences abroad may lead to different outcomes. Given these variations in study design and findings, more research is needed to better understand the specific ways in which fluency changes during study abroad.

There are several areas in which future research could contribute to a better understanding of this topic. First, as previously noted, there is a lack of clarity regarding changes in pausing. Future research that includes a wider range of measures of pausing, rather than select only one or two (which differ between studies), may be helpful. The present study measures both filled and unfilled pauses and further divides pauses into categories of short (less than 0.5 seconds) versus long (0.5 seconds or more) pauses as well as pauses appearing in the middle of a clause versus pauses between clauses. Another area where research is quite limited is in comparisons of students who begin their time abroad with different levels of L2 ability. Only one of the studies described in this section, Towell (2002), examined
this question. Although there seems to be a range of L2 ability levels in the studies cited here, with some focusing on advanced learners (Lennon, 1990; Valls Ferrer, 2008) and others focusing on learners with less previous L2 experience (D’Amico, 2010; Freed et al., 2004b; O’Brien, 2007; Towell, 2002), there are too many other differences between the studies to compare them with one another for the purpose of analyzing whether or not learners at different levels differed in their fluency gains. In light of this gap in the research, the present study examines whether there are differences in the fluency gains of participants who begin their semester abroad with higher and lower levels of fluency.

In addition to including more measures of pausing and comparing the fluency gains of students who begin with different levels of fluency, the present study also looks at the development of grammar, vocabulary, and processing abilities during study abroad and their relationship to fluency. The following section discusses previous research in this area.

Changes in fluency and underlying skills

Research on the development of both oral skills and grammatical knowledge, vocabulary knowledge, or processing abilities during study abroad is quite limited. However, there are a couple of previous studies related to this topic. Segalowitz et al. (2004) studied 46 learners of Spanish from the United States, 26 of whom studied abroad, while the other 20 were an at-home (AH) group that took Spanish classes at a university in the United States. All participants completed a series of

---

12 One other study, Perez-Vidal and Juan-Garau (2011), touches on this topic, but their study included only one measure of fluency, which does not allow for detailed comparisons of students at different levels.
tasks at the beginning and end of the semester, including an OPI, a grammar test, a test of lexical access (how quickly they judged words as referring to living or nonliving objects), and a test of attention control measuring their ability to shift their attention in a word-selection task in which the category of words changed on every trial. The researchers found that students with measurable gains on the OPI from the beginning to end of the semester began the semester with better lexical access abilities than those without measurable OPI gains; this finding held true for both the SA and AH groups. In contrast, there was no significant correlation between pretest grammar scores and OPI gains for either group. Additionally, gains in attention control did not correlate with measures of fluency.

Towell and Dewaele (2005) focused on the relationship between grammatical knowledge and fluency. In their study, 12 students from the United Kingdom who spent six months in France completed grammaticality judgment tests (GJT) and a speaking task in which they recounted the storyline of a short film. Measuring fluency as speech rate, the researchers did not find a significant relationship between pre-SA GJT scores and fluency scores either pre- or post-SA.

A consideration of these studies reveals that there is considerable room for further research in this area. Neither study included an untimed measure of vocabulary knowledge. Only one study, Segalowitz et al. (2004), included measures of processing abilities. Both studies looked at grammar, but they treated it quite differently. Segalowitz et al. (2004) used a multiple-choice grammar test, whereas Towell and Dewaele (2005) used GJTs. Additionally, they looked at its relationship to fluency differently. Segalowitz et al. (2004) examined whether there was a
correlation between pre-SA scores and OPI gains from pre- to post-SA; Towell and Dewaele (2005), in contrast, examined whether there was a correlation between pre-SA scores and fluency measures either pre- or post-SA (not changes from pre- to post-SA). For these reasons, there is a clear need for more research on this topic.

**Areas for further research**

To summarize, this review of the literature has identified several areas for further research related to the development of fluency during study abroad. First, most previous studies have used few measures of pausing, and the use of different measures in different studies makes it difficult to compare their results. Therefore, further research is needed to better understand how this aspect of fluency develops during study abroad. There is also little previous research examining whether there are differences in changes in fluency—either in amount of change or in the aspects of fluency in which change is evident—for learners who differ in their level of fluency prior to studying abroad. Additionally, there is a need for more research examining whether and how grammatical knowledge, vocabulary knowledge, and processing skills change along with fluency during a semester abroad.

A final limitation of most previous research on L2 fluency and study abroad is that there has been little consideration of participants’ L1 fluency, despite findings that L1 and L2 fluency are related (Derwing et al., 2009; De Jong et al., 2013b; Towell & Dewaele, 2005). As Segalowitz points out, “a great deal of fluency-related phenomena (hesitations, speech rate) may be characteristic of the way individuals speak in general and not just characteristic of their L2 speech” (2010: 35), and it is
therefore important to take into account L1 fluency in studies of L2 fluency. To address this gap in the research, the present study includes measures of L1 fluency.

In addition to examining fluency development during study abroad, the present study also addresses questions related to the measurement of fluency, specifically in the area of pausing. The following section presents and analyzes the measurement of pauses in previous studies.

**Issues in measuring pauses**

Pausing is a sub-construct of fluency measured in many studies on a variety of topics. As noted above, measures of pauses have been used in studies of the relationship between L1 and L2 fluency (De Jong et al., 2013b; Derwing et al., 2009), the relationship between utterance fluency and perceived fluency (Bosker et al., 2013; Derwing et al., 2009; Kormos & Denes, 2004; Riggenbach, 1991; Rossiter, 2009) and fluency development during study abroad (D'Amico, 2010; Lennon, 1990; Llanes & Muñoz, 2009; Towell, 2002; Valls Ferrer, 2008). In addition to these topics, measures of pausing have also been used in studies of the effects of planning (Foster & Skehan, 1996; Guará Tavares, 2009; Mehnert, 1998; Nakakubo, 2011; Sangarun, 2005; Skehan & Foster, 1997; Tajima, 2003; Tavakoli & Skehan, 2005; Wendel, 1997; Wigglesworth & Elder, 2010), task type (Foster & Skehan, 1996; Gilabert, 2007; Mehnert, 1998; Michel et al., 2007; Skehan & Foster, 1997; Tavakoli & Foster, 2011; Tavakoli & Skehan, 2005), repetition of tasks (Bygate, 2001), and working memory (Mota, 2003) on L2 speech performance. It is evident, then, that measuring pauses in a meaningful and appropriate way is important for a variety of SLA research.
One of the primary issues in measuring pauses, as well as other sub-constructs of fluency, is that the lack of consistency across studies makes it difficult to compare results. As noted in the earlier section on defining and measuring fluency, there is considerable variation in which pauses are measured as well as how they are measured. Regarding which pauses are measured, some researchers count only filled (García-Amaya, 2009) or only unfilled (Guará Tavares, 2012) pauses, whereas others count both but distinguish between them (Kormos & Denes, 2004; Rossiter, 2009), and still others make no distinction between them (Foster & Skehan, 1996; Mehnert, 1998; Tavakoli et al., 2011). As an example of differences in how pauses are measured, some studies count the total number of pauses (Foster & Skehan, 1996; Mehnert, 1998; Rossiter, 2009; Tavakoli & Skehan, 2005), whereas others include ratios of pauses to units, such as pauses per T-unit (Bygate, 2001), or pauses to time, such as pauses per minute (Iwashita et al., 2008; Kormos & Denes, 2004; Llanes & Muñoz, 2009; Skehan & Foster, 2005). Table 1 summarizes these measures of pausing that have been used in previous studies.

Considering this lack of consistency across studies, there are several aspects of pausing that deserve further attention in research. First, there has been substantial variation in the length of pauses included in calculations, which displays a lack of agreement regarding which pauses indicate dysfluency. On the surface it would appear that two studies in which the researchers counted the number of unfilled pauses per minute used the same measure, but when one researcher counted all pauses of 0.2 seconds or more in length (Kormos & Denes, 2004) and the other counted only pauses of 2 seconds or more in length (Tajima, 2003),
Table 1: Measures of pausing in previous studies

<table>
<thead>
<tr>
<th>Total number of pauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 0.4 seconds or longer (Rossiter, 2009; Tavakoli &amp; Skehan, 2005)</td>
</tr>
<tr>
<td>• 1 second or longer (Foster &amp; Skehan, 1996; Mehnert, 1998; Skehan &amp; Foster, 1997; Tavakoli et al., 2011)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pause-to-time ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Filled pauses per second or per minute (Bosker et al., 2013; D'Amico, 2010; Iwashita et al., 2008; Kormos &amp; Denes, 2004; Llanes &amp; Muñoz, 2009; Skehan &amp; Foster, 2005)</td>
</tr>
<tr>
<td>• Unfilled pauses per second or per minute</td>
</tr>
<tr>
<td>o 0.2 seconds or longer (Kormos &amp; Denes, 2004)</td>
</tr>
<tr>
<td>o 0.25 seconds or longer (Bosker et al., 2013)</td>
</tr>
<tr>
<td>o 0.4 seconds or longer (D'Amico, 2010; Llanes &amp; Muñoz, 2009)</td>
</tr>
<tr>
<td>o 0.5 seconds or longer (Mota, 2003)</td>
</tr>
<tr>
<td>o 1 second or longer (Iwashita et al., 2008)</td>
</tr>
<tr>
<td>• Hesitations per minute: unfilled pauses less than 0.5 seconds + filled pauses (Mota, 2003)</td>
</tr>
<tr>
<td>• Both filled and unfilled pauses per minute</td>
</tr>
<tr>
<td>o 0.4 seconds or longer (Derwing et al., 2009; Valls Ferrer, 2008)</td>
</tr>
<tr>
<td>o 2 seconds or longer (Nakakubo, 2011; Tajima, 2003)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pause-to-unit ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Filled pauses per turn (García-Amaya, 2009)</td>
</tr>
<tr>
<td>• Filled pauses per 100 words (De Jong et al., 2013a; Michel et al., 2007)</td>
</tr>
<tr>
<td>• Unfilled pauses per 100 words (De Jong et al., 2013a)</td>
</tr>
<tr>
<td>• Pauses per 1,000 words (Hilton, 2008)</td>
</tr>
<tr>
<td>• Filled pauses per T-unit (Lennon, 1990)</td>
</tr>
<tr>
<td>• Unfilled pauses per T-unit (Bygate, 2001)</td>
</tr>
<tr>
<td>• Unfilled pauses per c-unit13 (Guará Tavares, 2009)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total pausing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Measured as percent of total speaking time</td>
</tr>
<tr>
<td>o Unfilled pauses (Iwashita et al., 2008)</td>
</tr>
<tr>
<td>o Both filled and unfilled pauses (Lennon, 1990; Mehnert, 1998; Sangarun, 2005)</td>
</tr>
<tr>
<td>o Unspecified (Hilton, 2008)</td>
</tr>
<tr>
<td>• Measured in number of seconds or unspecified (Foster &amp; Skehan, 1996; Tavakoli &amp; Skehan, 2005)</td>
</tr>
<tr>
<td>• Phonation-time ratio: percent of time speaking rather than pausing (Kormos &amp; Denes, 2004; Préfontaine, 2010)</td>
</tr>
</tbody>
</table>

13 A c-unit is a “grammatical independent clause with any of its modifiers” (Loban, 1963: 7).
Table 1: Continued

<table>
<thead>
<tr>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mid-clause vs. end-of-clause pauses (Skehan &amp; Foster, 2005; Tavakoli &amp; Foster, 2011)</td>
</tr>
<tr>
<td>• Mean length of pauses</td>
</tr>
<tr>
<td>o All pauses or all unfilled pauses (Bosker et al., 2013; De Jong et al., 2013a; Hilton, 2008; Kormos &amp; Denes, 2004; Tavakoli &amp; Skehan, 2005; Towell, 2002; Valls Ferrer, 2008; Wendel, 1997)</td>
</tr>
<tr>
<td>o Internal pauses (Valls Ferrer, 2008)</td>
</tr>
<tr>
<td>o Pauses at T-unit boundaries (Lennon, 1990)</td>
</tr>
<tr>
<td>o Pauses within AS-units14 (De Jong et al., 2013b)</td>
</tr>
<tr>
<td>o Pauses between AS-units (De Jong et al., 2013b)</td>
</tr>
<tr>
<td>• Percent of T-units followed by a filled or unfilled pause (Lennon, 1990)</td>
</tr>
<tr>
<td>• Percent of total pause time at T-unit boundaries (Lennon, 1990)</td>
</tr>
<tr>
<td>• Unspecified measures of filled and unfilled pauses (Wigglesworth &amp; Elder, 2010)</td>
</tr>
</tbody>
</table>

the measures are actually quite different. Considering the speech of native speakers may provide guidance in deciding which pauses to measure. Goldman-Eisler (1961) found that short pauses of less than 0.5 seconds are a common feature of native speaker speech in contexts including discussions and giving summaries.

   In L2 research, Riggenbach (1991) compared pausing in the speech of low- and high-proficiency learners of English, finding a difference between groups in the frequency of pauses lasting 0.5 seconds or more, but no differences in the frequency of shorter pauses. Given the small number of participants in her study (3 per group) and the fact that the low- and high-proficiency groups were formed based on subjective impressions of fluency, Riggenbach’s study should be replicated with larger numbers to inform both future research and the interpretation of findings related to pausing in previous studies. The present study addresses the question of

14 An AS-unit (Analysis of Speech Unit) is "a 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: 365).
whether there are significant differences between the speech of learners with different levels of lexical–grammatical competence\textsuperscript{15} in the number of pauses 0.5 seconds and longer and whether there are there significant differences between these groups in the number of shorter pauses (0.25 to 0.49 seconds).

Another related question is whether or not to measure filled and unfilled pauses separately. As noted above, researchers vary in how they treat filled and unfilled pauses, with several options available: measuring them together, measuring them separately, or measuring only one or the other. Whether or not filled and unfilled pauses differ in what they reveal about fluency is not clear from the few studies that have measured them separately. In Iwashita et al.’s (2008) study, advanced learners had significantly less unfilled pauses than learners at lower levels, whereas advanced learners had slightly more filled pauses than learners at lower levels, although this difference was not significant. Cenoz (2000) also found that advanced learners used more filled pauses, while learners at lower proficiency levels used more unfilled pauses; it appears that the statistical significance of this difference was not tested. In Kormos and Denes’s (2004) data, in contrast, low-intermediate and advanced learners had nearly identical numbers of unfilled pauses, but the low-intermediate group had almost twice as many filled pauses (this difference, however, was not statistically significant). Given these mixed results, future studies should continue to investigate whether filled and unfilled pauses

\textsuperscript{15} Since the present study does not include an overall measure of proficiency, lexical–grammatical competence (based on scores on grammar and vocabulary tests administered as part of the study) is used instead of proficiency to divide participants into high and low groups.
pattern differently across proficiency levels. The present study addresses this question.\textsuperscript{16}

A final issue related to pauses regards their location. Very few studies have distinguished between mid-clause and end-of-clause pauses (Skehan & Foster, 2005; Tavakoli & Foster, 2011), but L1 data suggest that pausing at the end of clauses is natural, whereas pausing in the middle of a clause is less common in native speaker speech (Chambers, 1997; Davies, 2003; Skehan, 2009). Although distinguishing between the two types of pauses does not appear to have impacted the conclusions drawn in either Skehan and Foster's (2005) or Tavakoli and Foster's (2011) studies, neither of these studies included learners with different levels of L2 ability. Exploring patterns of mid-clause and end-of-clause pausing among learners at different levels may be valuable in determining what pause location reveals in L2 speech, which could then guide decisions about if and when it is important to measure them separately. The present study measures mid-clause and end-of-clause pauses separately for the purpose of comparing them across levels of lexical–grammatical competence.

Addressing these measurement-related questions may contribute to a better understanding of measuring pauses that may inform measurement decisions not only in research on study abroad, but also in other areas of fluency research.

\textsuperscript{16} Again, it should be noted that dividing participants into high and low groups in the present study is based on lexical–grammatical competence rather than overall proficiency. However, classifications of students as “intermediate” or “advanced” in previous research have not always been based on proficiency testing, either.
Summary

This chapter has described previous research on fluency and its development during study abroad. It has also identified areas for further research on fluency development and its relationship to underlying knowledge and skills, as well as the measurement of fluency. The following chapter explains the design of the present study.
Chapter 3: Methodology

This chapter explains the design of the study. It begins by restating the research questions described in Chapter 1, followed by information about the participants, materials, tasks, and data collection and analysis.

Research questions

There are two sets of research questions, one examining the development of fluency and the other focusing on the measurement of fluency.

The development of fluency

1. Are there changes in participants’ L2 articulation rate, pausing, repair, and mean length of run from pre- to post-study abroad (SA)?
   a. What changes are evident in terms of raw numbers?
   b. What changes are evident in terms of L1–L2 difference scores?
2. Is there a relationship between pre-SA L2 fluency and changes in L2 fluency?
   a. Is there a correlation between pre-SA L2 fluency scores and change in L2 fluency scores from pre- to post-SA?
   b. Is there a significant difference between the amount of change in L2 fluency scores for participants with low pre-SA L2 fluency scores (below the median score) and the amount of change in L2 fluency scores for participants with high pre-SA L2 fluency scores (above the median score)?
   c. If there are differences in the amount of change in L2 fluency scores for participants with low and high pre-SA L2 fluency scores, in what
areas of fluency (articulation rate, pausing, repair, or length of run) are the differences most evident?

3. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and pre- and post-SA L2 fluency scores?
   a. Is there a correlation between participants’ pre-SA grammar scores and their pre-SA fluency scores? Is there a correlation between their pre-SA grammar scores and their post-SA fluency scores? Is one correlation stronger than the other?
   b. Is there a correlation between participants’ pre-SA vocabulary scores and their pre-SA fluency scores? Is there a correlation between their pre-SA vocabulary scores and their post-SA fluency scores? Is one correlation stronger than the other?
   c. Is there a correlation between participants’ pre-SA processing scores and their pre-SA fluency scores? Is there a correlation between their pre-SA processing scores and their post-SA fluency scores? Is one correlation stronger than the other?
   d. Considering as possible factors pre-SA grammar, vocabulary, picture naming, and sentence matching scores, what regression model best predicts pre-SA L2 fluency?

4. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and changes in fluency?
   a. Is there a correlation between participants’ pre-SA grammar scores and change in fluency scores from pre- to post-SA?
b. Is there a correlation between participants’ pre-SA vocabulary scores and change in fluency scores from pre- to post-SA?

c. Is there a correlation between participants’ pre-SA processing scores and change in fluency scores from pre- to post-SA?

d. What is the relative strength of each of these correlations?

5. Is there a relationship between changes in grammar, vocabulary, and/or processing scores and changes in L2 fluency?

   a. Is there a correlation between change in grammar scores and change in fluency scores from pre- to post-SA?

   b. Is there a correlation between change in vocabulary scores and change in fluency scores from pre- to post-SA?

   c. Is there a correlation between change in processing scores and change in L2 fluency scores from pre- to post-SA?

   d. What is the relative strength of each of these correlations?

6. Considering as possible factors L1 fluency, pre-SA L2 fluency, changes in grammar from pre- to post-SA, changes in vocabulary from pre- to post-SA, changes in processing speed from pre- to post-SA, and time spent interacting with native speakers of the L2 during the semester abroad, what regression model best predicts changes in L2 fluency during a semester abroad?

**The measurement of fluency**

1. Are there significant differences between the speech of learners at different levels of lexical–grammatical competence in the number of pauses 0.5
seconds and longer? Are there significant differences between these groups in the number of shorter pauses (0.25 to 0.49 seconds)?

2. Do filled and unfilled pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

3. Do mid-clause and end-of-clause pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

Participants

Background

Study abroad participants

Forty native speakers of English from the United States, Canada, and Australia studying abroad in Buenos Aires, Argentina, completed the pretest. Of these 40 participants, 39 also completed the posttest. Data from two additional students who completed the pretest were not used because their background questionnaires revealed that one was a heritage speaker of Spanish and one was a native speaker of French.

All study abroad participants were university students between the ages of 19 and 29 (mean age at pretest = 20.6). They had a wide range of previous experiences with Spanish. The number of years of formal study of Spanish ranged from two to 12 (mean = 6.0). Most (31 of 40) reported experience with Spanish outside of the classroom in the context of work, volunteer experience, extracurricular activities, friendships, or previous travel abroad. Of these participants, eight had previously spent three months or more in a Spanish-speaking country. Additionally, they had varying exposure to other languages. Approximately
half (19) had previously studied one or more additional languages for at least one semester, most commonly French, German, or Portuguese.

At the time of the study, these students were studying at private universities in Buenos Aires, Argentina; students at three different universities participated. At two of the universities, all participants were enrolled in classes with local students. At the third university, a few of the participants were enrolled only in classes for international students.

**Non-study abroad participants**

A second group of participants included eight students studying Spanish at a large public university in the United States. The data collected from this group of participants was used only to address the research questions related to the measurement of fluency. The students ranged in age from 20 to 46 (mean = 25.1). They had previously studied Spanish for one to 12 years (mean = 5.9). Four of the eight students had previous experience in a Spanish-speaking country, but only one had been abroad for more than a month. They reported little experience using Spanish outside of the classroom in the United States; three mentioned having occasional conversations in Spanish. They also had little exposure to other languages. One student reported studying French for two years, whereas the rest had no experience with languages other than English and Spanish.

**Materials and tasks**

To measure fluency and factors that may be related to fluency, a number of tasks were used to collect data, including speaking, grammar, vocabulary, and processing tasks; a questionnaire about participants' interactions with native
speakers of Spanish during their semester abroad; and a questionnaire to gather background information about participants. This section describes these tasks and the materials that each required.

**Speaking tasks**

To collect speech samples that allowed for the measurement of utterance fluency in the participants' L2 and L1, each participant completed three speaking tasks in Spanish and three similar tasks in English. Table 2 describes these tasks.

All of these tasks were monologues. Although monologues have the disadvantage of being somewhat artificial, an important advantage of monologues is that they avoid the variability that interacting with another person introduces into the task (Segalowitz, 2010). The purpose of including several tasks was to elicit different speech functions, providing a broader sample of participants’ speech.

Table 2: Speaking tasks

<table>
<thead>
<tr>
<th>Task type</th>
<th>Prompt for L2 (Spanish) task</th>
<th>Prompt for L1 (English) task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Describe personal activities</td>
<td>Describe what you do on a typical weekday during the school year.</td>
<td>Describe what you do on a typical weekend during the school year.</td>
</tr>
<tr>
<td>2. Explain advantages and disadvantages</td>
<td>Explain the advantages and disadvantages of working part time while in college.</td>
<td>Explain the advantages and disadvantages of going to college immediately after graduating from high school.</td>
</tr>
<tr>
<td>3. Narrate in past time</td>
<td>Look at the series of pictures below. Using the pictures, tell the story as a sequence of events that occurred in the past.</td>
<td>Look at the series of pictures below. Using the pictures, tell the story as a sequence of events that occurred in the past.</td>
</tr>
</tbody>
</table>

17 See Appendix A for the pictures used in the narration tasks.
For each task, the instructions were presented in SuperLab. Following the presentation of the instructions, participants were given 30 seconds to prepare their answer. Then, they had a maximum of two minutes to respond to the prompt. Previous research suggests that the amount of preparation time included in speaking tasks may influence certain aspects of fluency, with fewer pauses (Skehan & Foster, 1997), less total pausing time (Foster & Skehan, 1996; Mehnert, 1998), increased speech rate (Gilabert, 2007; Guará Tavares, 2009; Mehnert, 1998; Ortega, 1999; Tajima, 2003), and longer mean length of run (Mehnert, 1998) when participants are given more time to prepare. Most of these studies examined the effects of 10 minutes of planning time compared to no planning time; however, Mehnert (1998) found an increase in speech rate and a decrease in total pausing time with as little as one minute of planning time. Following De Jong et al.’s (2013a, 2013b) design, 30 seconds was chosen as the amount of preparation time in the present study to allow participants enough time to briefly consider the content that they wished to include (for the first two tasks) or to familiarize themselves with the pictures (for the narration tasks), but not enough time to carefully plan or rehearse their responses.

As Table 2 shows, the L1 and L2 tasks were similar, but not identical. The use of tasks that were not identical was intended to limit the effects of practice. However, the tasks were intentionally similar to facilitate comparisons of L1 and L2 fluency. To verify the similarity of the pairs of tasks, they were piloted with three native speakers of English and four native speakers of Spanish. The native English speakers completed all of the tasks in English, and the native speakers of Spanish
completed all of the tasks in Spanish. Several measures of fluency were calculated, including speech rate, filled pauses (0.5 seconds and longer) per minute, unfilled pauses (0.5 seconds and longer) per minute, and instances of repair per 100 syllables. No significant differences were found between the two tasks for each task type.

**Grammar test**

To measure grammatical knowledge, participants completed a 30-item test (see Appendix B). Each item consisted of a sentence containing an error, such as a verb in the wrong tense or mood, an adjective that does not agree with the noun that it describes, or an incorrect pronoun. The participants’ task was to find the error and write the corrected form of the word containing the error. The test was administered in Microsoft Word.

**Vocabulary test**

Participants completed a vocabulary test adapted from the Diploma de Español como Lengua Extranjera (DELE). Portions of this test, which includes a 30-item multiple-choice vocabulary section and a 20-item multiple-choice cloze passage measuring both grammar and vocabulary, have previously been used in SLA research to assess the proficiency level of nonnative speakers of Spanish (Duffield & White, 1999; Montrul & Slabakova, 2003; Slabakova et al., 2011, 2012; White et al., 2004). As the purpose of this task in the present study was to measure vocabulary knowledge, only the 30-item vocabulary section was used. The test was administered in Microsoft Word.
Processing tasks

To measure processing speed, participants completed two tasks, a picture-naming task and a sentence-picture verification task. In the picture-naming task, participants were presented a series of 50 pictures of common objects and instructed to name each one in Spanish as quickly as possible. The pictures were presented in SuperLab, and participants’ responses were recorded in Audacity. Pictures for the task were selected from the Snodgrass and Vanderwart (1980) picture set, which has been used in a variety of L1 (Barry et al., 2001; Bonin et al., 2002; Cycowicz et al., 1997; Hartsuiker & Notebaert, 2010; Mitchell & Brown, 1988) and L2 (Bloem et al., 2004; De Bot & Stoessel, 2000; De Jong et al., 2012, 2013a; Finkbeiner & Nicol, 2003; Sunderman & Kroll, 2009) research. The pictures included in the study were among those found to have at least 90% naming agreement among native speakers of Spanish (Cuetos et al., 1999; Sanfeliu & Fernandez, 1996). The task also included an eight-item warm-up set to familiarize participants with the procedure.

The second processing task was a sentence–picture verification task. In such tasks, a picture and a sentence are presented to participants. They must indicate, as quickly as possible, whether or not the sentence accurately describes the picture. Sentence–picture verification has primarily been used in L1 research, examining topics such as cortical activation during sentence processing (Mack et al., 2013; Neubauer et al., 1995) and the effect of aging on sentence comprehension (López-Higes Sánchez et al., 2008). It has also been used in L2 reading research to measure the effects of word spacing on reading speed (Bassetti, 2009).
In sentence–picture verification tasks, judging whether or not the sentence matches the picture often depends on processing vocabulary. For example, participants may see the sentence “These dots are black” and a picture of dots that are black or of another color (Liu, 1980). Given that the purpose of this task in the present study was to measure speed of morphosyntactic processing, the task was somewhat different from sentence-picture verification tasks used in previous studies. In the present study, judging whether or not a picture and a sentence matched required processing of features such as pronouns and verb endings. For example, participants heard the sentence *Mira la televisión* ("He/she watches television"). A mismatched picture showed two people watching television. A matching picture would show one person watching television. See Appendix C for more examples of sentences and pictures used in this task.

The sentences were created using vocabulary often taught in first-year Spanish courses. Additionally, word frequency was checked using NIM.\(^\text{18}\) All forms included in the sentences have a frequency of at least 3.55 occurrences per million words. The lowest-frequency forms are conjugated verbs, which have a considerably higher frequency if other forms of the verb are counted. For example, *acuesta* (he/she puts to bed) is the least frequent form; counting all forms of the verb *acostar* increases the frequency to 24.9 occurrences per million words.

Previous research using sentence–picture verification has presented the sentences in either written (Bassetti, 2009; Liu, 1980; Neubauer et al., 1995) or

\(^{18}\) NIM is a web application that searches corpora in several languages. The Spanish corpus is the LEXESP database, which contains 5.63 million words.
auditory (Engelen et al., 2011; Hirschfeld et al., 2012; Mack et al., 2013) form. In the present study, the sentences were presented in auditory form since the study did not intend to measure reading speed. Participants first heard a sentence, and then the picture appeared. Reaction time was measured from the point at which the picture appeared. The task included 34 sentence–picture pairs in the main set and six pairs in the warm-up set.

**Background questionnaire**

The background questionnaire (see Appendix D), administered in Microsoft Word, consisted of a short series of questions about the participants and their previous experiences with Spanish and other languages. This information was collected for the purpose of reporting basic data about the groups of participants.

**Language contact questionnaire**

A language contact questionnaire (see Appendix E) was used to gather information about the amount of time that participants spent interacting with native speakers of Spanish outside of class during their semester abroad. It included questions about participants’ interactions with their host families, with friends and acquaintances, in service encounters, and in other contexts. It was administered every other week during the semester.

Previous research has also looked at motivation as a factor in study abroad outcomes (Anderson, 2012; Isabelli-García, 2006; Martinsen, 2010, 2011; Yager, 1998). It was decided not to include a separate measure of motivation in the present study given that motivation questionnaires often ask participants how likely they are to use the L2 in certain situations or how much effort they make to learn the L2
(Anderson, 2012; Martinsen, 2010, 2011; Yashima et al., 2004). Rather than have participants describe their general level of effort, it seemed more meaningful to ask them to record, on a regular basis, how much time they had actually spent using the L2 recently. The language contact questionnaire fulfilled this purpose.

**Data collection**

As previously noted, participants included students studying abroad in Buenos Aires, Argentina, as well as students studying Spanish in the United States. This section describes the data collection procedures for each of these groups.

**Study abroad participants**

**Pre-study abroad**

The first data collection took place near the beginning of the participants’ semester abroad. At that time, they completed the background questionnaire followed by the speaking, grammar, vocabulary, and processing tasks. Participants met with the researcher in individual sessions lasting approximately 75 minutes.

**During study abroad**

During the semester, participants were asked to fill out the language contact questionnaire every other week. Participants received and returned the questionnaire by e-mail. In several previous studies (Martinsen, 2010; Perez-Vidal & Juan-Garau, 2011; Segalowitz et al., 2004), similar data has been gathered from students only at the end of their time abroad. Although any self-reported data is still open to the charge that it cannot be verified, it seemed likely that participants would be able to give more accurate estimates of their activities on a biweekly basis as
opposed to estimating their activities for the whole semester at the end of their time abroad.

Post-study abroad

The final data collection took place at the end of the participants’ semester abroad. They completed the same grammar, vocabulary, processing, and L2 speaking tasks as at the beginning of the semester. They did not repeat the L1 speaking tasks. Participants met with the researcher in individual sessions lasting approximately 60 minutes.

Non-study abroad participants

Non-study abroad participants met once with the researcher in individual sessions lasting approximately 50 minutes. They completed the background questionnaire followed by the speaking, grammar, and vocabulary tasks.

Data analysis

Initial analyses

Once data were collected, they were analyzed as described in the following sections.

Speaking tasks

Thirty-second excerpts were taken from all speaking tasks in both English and Spanish. Typically, each excerpt was taken from second 5 to second 35 of each task. Beginning at second 5 eliminated fillers (umm, so, okay, well, etc.) that were sometimes present prior to the start of content that addressed the prompt. If a participant spoke for at least 31 but less than 35 seconds in response to a given prompt, the excerpt was taken from second 1 to second 31. In cases of short
responses such as these, fillers were rarely present at the beginning of the sample. In a few cases (8 of 240 pretest responses for study abroad participants, 2 of their 117 posttest responses, and 1 of 48 responses for non-study abroad participants), participants spoke for less than 30 seconds in response to a prompt. In these cases, additional time was taken from the preceding or following task so that the excerpts for each participant totaled exactly 90 seconds in each language.

Additionally, following Riggenbach (1991, 2000), unfilled pauses were limited to three seconds.\(^1\)\(^9\) Prior to taking the 30-second excerpts, unfilled pauses longer than three seconds were reduced to three seconds. There were a total of 15 such pauses in the 240 pretest responses for study abroad participants, 7 in their 117 posttest responses, and 5 in the 48 responses for non-study abroad participants.

Once the excerpts were taken, pauses ≥0.25 seconds were marked in each excerpt using Praat. As a first step, a script (Lennes, 2002b) was used to mark silent pauses. These silent pauses were then verified and adjusted manually as needed, and all filled pauses were marked manually. Each pause was labeled as unfilled (u) or filled (f), long (l; ≥0.5 seconds) or short (s; ≥0.25 seconds but <0.5 seconds), and mid-clause (m) or end-of-clause (e). After this step was completed, these notations were copied into a second tier and modified to combine filled and unfilled pauses. A filled and an unfilled pause occurring back to back were counted separately for the purpose of calculating rates of filled and unfilled pauses, but they were counted as a single pause for the purpose of calculating rates of long and short pauses and mid-

\(^{19}\) All filled pauses in the data were less than three seconds in duration.
clause and end-of-clause pauses. Figure 1 displays an excerpt of Participant 3’s pre-SA recording in Spanish and shows how the pauses were marked.

Figure 1: Marking pauses in PRAAT

Pauses were considered end-of-clause pauses not only if they occurred between clauses as formally defined, but also if they occurred at a natural break, where a comma would be found in written text. For example, in the sentence *Usually the rest of the night [pause] after eating dinner with friends [pause] I’ll do homework*, both pauses were marked as end-of-clause pauses even though *usually the rest of the night* is not a clause. Additionally, when a conjunction occurred between clauses, pauses on either side of the conjunction were marked as end-of-clause pauses, as in the sentence *I pack my bag after getting dressed and putting on my makeup [pause] and [pause] then I go to a class*. See Appendix F for more examples of pauses marked
as mid-clause and end-of-clause. After the pauses were marked, another script (Lennes, 2002a) was used to produce a list of all pauses and their durations.

The researcher also listened to each excerpt and counted the number of syllables and instances of repair. After these steps were completed, the following measures of fluency were calculated:

1. Articulation rate: the number of syllables per second, with pauses ≥0.25 seconds removed from the calculation. Using the articulation rate, rather than the overall speech rate with pauses included, separates the speed of delivery from pausing (De Jong et al., 2013a).

2. Repair per 100 syllables, including:
   a. Repetitions: the number of exact duplications of the same syllable, word, or phrase (Iwashita et al., 2008).
   b. Replacements: the number of times one lexical item is substituted for another (Inoue, 2010).
   c. Reformulations: the number of grammatical self-corrections (Iwashita et al., 2008).
   d. False starts: the number of times a speaker abandons an idea without completing it (Inoue, 2010).

3. Pauses
   a. Percent pausing time: the total percent of speaking time taken up by filled or unfilled pauses lasting 0.25 seconds or more.
   b. Rate of all pauses: the number of pauses lasting 0.25 seconds or more, per 100 syllables.
c. Rate of mid-clause pauses: the number of pauses lasting 0.25 seconds or more occurring in the middle of a clause, per 100 syllables.

d. Rate of end-of-clause pauses: the number of pauses lasting 0.25 seconds or more occurring at the end of a clause, per 100 syllables.

e. Rate of long pauses: the number of pauses lasting 0.5 seconds or more, per 100 syllables.

f. Rate of short pauses: the number of pauses lasting between 0.25 seconds and 0.49 seconds, per 100 syllables.

g. Rate of filled pauses: the number of filled pauses lasting 0.25 seconds or more, per 100 syllables.

h. Rate of unfilled pauses: the number of unfilled (silent) pauses lasting 0.25 seconds or more, per 100 syllables.

4. Mean length of run: mean number of syllables spoken between pauses (filled or unfilled) 0.5 seconds or longer.

Grammar test

Each correct answer on the grammar test received one point. The highest possible score was 30 points. Since this test had not previously been used for research purposes, an item analysis was conducted to assess its reliability. The test was found to have an acceptable level of reliability (α = .853), and deleting one or more questions would not substantially affect its reliability. Therefore, all questions were scored.

---

20 After data collection had begun, it was discovered that one item (Question 2) had two errors due to a typographical error. If participants successfully corrected either error, they were given credit for this item.
Vocabulary test

On the vocabulary test, each correct answer received one point. The highest possible score was 30 points.

Processing tasks

Each item in both the picture-naming and sentence-picture matching tasks was analyzed using the pretest data. In the picture-naming task, 7 of 50 items had an accuracy rate of 75% or less. These items were removed for all participants, and the mean accuracy rate for the remaining 43 items was 94.80% on the pretest. In the sentence-matching task, 6 of 34 items had an accuracy rate of 70% or less. These items were removed for all participants, and the mean accuracy rate for the remaining 28 items was 86.24% on the pretest.

For the picture-naming task, certain data points were also eliminated for individual participants. Following Sunderman and Kroll (2009) and De Jong et al. (2012, 2013a, 2013b), outliers (reaction times greater than two standard deviations above or below each participant’s mean) were excluded. As Sunderman and Kroll explain, very short reaction times suggest the influence of “anticipatory processes” (2009: 86), whereas very long times may reflect lapses of attention. This process led to the removal of a total of 74 data points (4.30% of the data) on the pretest and 74 data points (4.41% of the data) on the posttest. Additionally, 0.99% of the pretest data and 1.49% of the posttest data had to be eliminated due to false activation of the voice key or failure of the voice key to detect a response.21

---

21 These percentages include only data that otherwise would have been usable; the percentages would be slightly higher if they included data that was removed for other reasons, such as incorrect responses.
After these adjustments were made, the mean reaction time in milliseconds was calculated for each participant’s correct responses. All participants approached or exceeded 75% accuracy on the pretest,\(^{22}\) so data was used from all participants with the exception of one, whose repeated pressing of the space bar interfered with accurate measurement of her reaction times. On the posttest, repeated pressing of the space bar resulted in unusable data from two participants.

For the sentence-picture matching task, participants’ reaction times were log transformed prior to calculating a mean score for correct responses for each participant. This transformation was performed because the data were positively skewed, and log transformations have been used to correct this issue in previous language-related research using reaction time measures (Bott & Noveck, 2004; Edwards & Lahey, 1996; Feldman, O’Connor, & Moscoso del Prado Martín, 2009; Gordon & Caramazza, 1982). On the pretest, three participants responded with less than 75% accuracy,\(^{23}\) and their data for this task were not used. On the posttest, there were again three participants whose accuracy was less than 75%,\(^{24}\) so their data were not used.

**Language contact questionnaire**

Study abroad participants were asked to fill out the language contact questionnaire five times over the course of the semester, once every other week. The mean number of hours that each participant reported spending on each activity

---

\(^{22}\) The lowest score was 74.42%.

\(^{23}\) Two had scores of 71.43% and one had a score of 67.86%.

\(^{24}\) Again, two had scores of 71.43% and one had a score of 67.86%.
were calculated. Additionally, the mean number of total hours in contact with Spanish each week were calculated for each participant. Averages were calculated for all participants who completed the questionnaire at least three times; data were treated as missing for participants who completed it only once or twice.

To calculate the mean number of hours that each participant spent interacting with native speakers of Spanish each week, it was necessary to assign estimated values to certain activities for which participants did not record the duration. Routine transactions in public places were counted as one minute each, and an additional minute was added for each transaction including a complicating factor. Brief or superficial exchanges with host family members or roommates were counted as 30 seconds each. Short conversations with host family member or roommates as well as instances of engaging in friendly chitchat with strangers were counted as five minutes each. For longer interactions, participants were asked to give an estimate of how long each interaction lasted, and the responses that they recorded were used in the calculations.

The language contact questionnaire included two open-ended questions. To analyze this data, the researcher coded participants’ responses. The first item prompting an open-ended response said, “If you had other interactions with native speakers of Spanish not included in the categories above, please describe them here.” To code these responses, the researcher read through the data and made a list of 10 codes that seemed to capture the interactions that participants described. These codes included travel/tourism, class/tutoring, resolving a problem, non-routine transaction, texting or online communication, time with extended family or
friends of host, academic communication outside of class, leading/teaching a class, work/internship, and sports. The researcher then read through the data again and coded it according to these categories. Two codes, resolving a problem and teaching/leading a class, had only one instance each. The instance of resolving a problem was recoded as a non-routine transaction, and the instance of teaching/leading a class was recoded as work/internship. This process resulted in eight remaining coding categories, all with at least three instances each. See Appendix G for examples.

The second open-ended question asked, “Out of the interactions that you’ve recorded in questions 1–9, which do you think provided the most meaningful language practice? Why?” After reading through the data, the researcher decided to code each response according to the size of the group, the type of relationship, and the reason given to explain why the interaction provided meaningful practice. First, the responses were coded according to the size of the group: one-on-one, larger group, or unspecified. Next, the responses were coded according to the type of relationship: host family, peers/friends, strangers, intercambios, service encounter/work setting, romantic relationship, or unspecified. The researcher then created an initial list of codes for why the interaction provided meaningful practice: learning vocabulary/phrases, error correction, forced to speak Spanish, topic (interesting and/or challenging), comfortable setting, extended time, and listening practice. The data were coded according to these categories, but some of the data

---

25 Intercambio refers to a meeting with a native speaker of Spanish specifically for the purpose of language practice.
did not fit any of these categories. The following categories were then added: more focused interaction, variety, pronunciation/intonation, quick pace, and practicing routinized language. The researcher coded the data again with these categories included. See Appendix G for examples.

All data

Prior to carrying out other statistical analyses, all data were tested for normality using the Shapiro-Wilk test. For data found not to be normally distributed, transformations were applied, as shown in Table 3. Transformations were applied in the order shown in the table (i.e. a log transformation was performed first; if a log transformation did not result in normally distributed data, a square root transformation was used instead).
Table 3: Data transformations

<table>
<thead>
<tr>
<th>Normally distributed; no transformation needed</th>
<th>Log transformation</th>
<th>Square root transformation</th>
<th>Square root transformation + outliers replaced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Grammar</td>
<td>• Picture naming</td>
<td>• Filled pauses</td>
<td></td>
</tr>
<tr>
<td>• Vocabulary</td>
<td>• Repair (English and Spanish)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sentence matching</td>
<td>• Mean length of run (English)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Syllables per second</td>
<td>• Mid-clause pauses (English and Spanish)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(English and Spanish)</td>
<td>• Unfilled pauses (English and Spanish)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Percent pausing time</td>
<td>• Long pauses (Spanish)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(English and Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mean length of run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• All pauses (English and Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• End-of-clause pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(English and Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Long pauses (English)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Short pauses (English and Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Fluency T scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(English and Spanish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Grammar</td>
<td>• Repair</td>
<td>• Filled pauses</td>
<td></td>
</tr>
<tr>
<td>• Vocabulary</td>
<td>• Mean length of run</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Picture naming</td>
<td>• Mid-clause pauses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sentence matching</td>
<td>• Unfilled pauses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Syllables per second</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Percent pausing time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• All pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• End-of-clause pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Long pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Short pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Fluency T scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Repair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mean length of run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mid-clause pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Unfilled pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Picture naming</td>
<td>• Vocabulary</td>
<td>• Grammar</td>
<td></td>
</tr>
<tr>
<td>• Sentence matching</td>
<td>• All pauses</td>
<td>• Filled pauses</td>
<td></td>
</tr>
<tr>
<td>• Syllables per second</td>
<td>• Unfilled pauses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Percent pausing time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mid-clause pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• End-of-clause pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Long pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Short pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Repair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Fluency T scores</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Addressing the research questions

Following these initial analyses, the data were used to address each research question as described in this section. Each question is listed, followed by an explanation of how the data were used to address it.

Questions about the development of fluency

1. Are there changes in participants’ L2 articulation rate, pausing, repair, and mean length of run from pre- to post-study abroad (SA)?
   a. What changes are evident in terms of raw numbers?
   b. What changes are evident in terms of L1–L2 difference scores?

   To address Question 1a, all of the measures of fluency described in the previous section on the initial analysis of the speaking tasks were compared from pre- to post-SA for each participant’s L2 speech samples. To address Question 1b, first, the differences between each participant’s L1 and L2 fluency pre-SA as well as post-SA were quantified using effect sizes. Then, the effect sizes from pre- to post-SA were compared.

2. Is there a relationship between pre-SA L2 fluency and changes in L2 fluency?
   a. Is there a correlation between pre-SA L2 fluency scores and change in L2 fluency scores from pre- to post-SA?
b. Is there a significant difference between the amount of change in L2 fluency scores for participants with low pre-SA L2 fluency scores (below the median score) and the amount of change in L2 fluency scores for participants with high pre-SA L2 fluency scores (above the median score)?

c. If there are differences in the amount of change in L2 fluency scores for participants with low and high pre-SA L2 fluency scores, in what areas of fluency (articulation rate, pausing, repair, or length of run) are the differences most evident?

To address question 2a, correlations were calculated between pre-SA scores and amount of change from pre- to post-SA for each measure of fluency. All correlations were calculated as Pearson correlations. As previously indicated, these correlations were calculated both with and without controlling for L1 fluency. For calculations controlling for L1 fluency, partial correlations were used. All correlations were calculated in SPSS, which also tested the statistical significance of each correlation. To address question 2b, a composite pre-SA fluency score was calculated for each participant by first calculating pre-SA T-scores for articulation rate, mean length of run, percent of time spent pausing, rate of all pauses, and rate of mid-clause pauses.26 Then, the mean T-score for these measures was calculated for

26 Since including all eight measures of pausing in the composite score seemed redundant, percent of time spent pausing and rate of all pauses were chosen as the most general measures of pausing. Rate of mid-clause pauses was also included, as it was possible for participants to have a significant shift in the location of their pauses from pre- to post-SA without this change necessarily being reflected in either of the more general measures of pausing.
each participant. Using these T-scores, participants were divided into high and low pre-SA fluency groups. To create a composite measure of the amount of change in fluency from pre- to post-SA for each participant, T-scores were also calculated for the amount of change in articulation rate, mean length of run, percent of time spent pausing, rate of all pauses, and rate of mid-clause pauses. Then, the mean T-score for these measures was calculated for each participant. To test for a significant difference in the overall amount of change in fluency between participants with low pre-SA fluency scores and high pre-SA fluency scores, an independent samples t-test was used. To address question 2c, independent samples t-tests were used to check for significant differences between the low and high pre-SA fluency groups in the amount of change in fluency for each measure of fluency separately.

3. Is there a relationship between pre-SA grammar, vocabulary, and processing scores and pre- and post-SA L2 fluency scores?
   
a. Is there a correlation between participants’ pre-SA grammar scores and their pre-SA fluency scores? Is there a correlation between their pre-SA grammar scores and their post-SA fluency scores? Is one correlation stronger than the other?

b. Is there a correlation between participants’ pre-SA vocabulary scores and their pre-SA fluency scores? Is there a correlation between their pre-SA vocabulary scores and their post-SA fluency scores? Is one correlation stronger than the other?

c. Is there a correlation between participants’ pre-SA processing scores and their pre-SA fluency scores? Is there a correlation between their
pre-SA processing scores and their post-SA fluency scores? Is one correlation stronger than the other?

d. Considering as possible factors pre-SA grammar, vocabulary, picture naming, and sentence matching scores, what regression model best predicts pre-SA L2 fluency?

Correlations with pre- and post-SA fluency scores were calculated for each measure of fluency (articulation rate, rate of repair, mean length of run, and various measures of pausing) separately. To address Question 3a, grammar scores were taken from the grammar test administered at the beginning of the semester. For Question 3b, vocabulary scores came from the pre-SA DELE vocabulary test. To address Question 3c, correlations between processing and fluency were calculated separately for the picture-naming task and the sentence-picture verification task.

4. Is there a relationship between pre-SA grammar, vocabulary, and processing scores and changes in fluency?

a. Is there a correlation between participants’ pre-SA grammar scores and change in fluency scores from pre- to post-SA?

b. Is there a correlation between participants’ pre-SA vocabulary scores and change in fluency scores from pre- to post-SA?

c. Is there a correlation between participants’ pre-SA processing scores and change in fluency scores from pre- to post-SA?

d. What is the relative strength of each of these correlations?
The pre-SA grammar, vocabulary, and processing scores were the same as those used to address Question 3. As in the previous questions, partial correlations were used for calculations controlling for L1 fluency.

5. Is there a relationship between changes in grammar, vocabulary, and processing scores and changes in L2 fluency?
   a. Is there a correlation between change in grammar scores and change in fluency scores from pre- to post-SA?
   b. Is there a correlation between change in vocabulary scores and change in fluency scores from pre- to post-SA?
   c. Is there a correlation between change in processing scores and change in L2 fluency scores from pre- to post-SA?
   d. What is the relative strength of each of these correlations?

The calculations for Questions 5a through 5c were the same as the calculations for Questions 4a through 4c, except that the grammar, vocabulary, and processing scores used reflected changes from pre- to post-SA. These scores were calculated by subtracting each participant’s pre-SA score for each task from his or her post-SA score on the same task.

6. Considering as possible factors L1 fluency, pre-SA L2 fluency, changes in grammar from pre- to post-SA, changes in vocabulary from pre- to post-SA, changes in processing speed from pre- to post-SA, and time spent interacting with native speakers of the L2 during the semester abroad, what regression model best predicts changes in L2 fluency during a semester abroad?
To address this question, composite scores were used for L1 fluency, pre-SA L2 fluency, and changes in L2 fluency. For pre-SA L2 fluency and changes in L2 fluency, the same composite scores that were calculated to address question 2 were also used for question 6. To create a composite L1 fluency score, the same process was followed, first calculating L1 T-scores for articulation rate, mean length of run, percent of time spent pausing, rate of all pauses, and rate of mid-clause pauses. Then, the mean T-score for these measures was calculated for each participant.

Scores for changes in grammar, vocabulary, and processing speed were the same as the scores used to address question 5. The average amount of time per week spent interacting with native speakers of the L2 was calculated using information that participants provided through the Language Contact Questionnaire. After all of these scores were calculated, regression equations were calculated in SPSS using backward elimination. This method was chosen to allow for comparisons between a model that included all of the factors and models that eliminated factors that did not strongly predict changes in fluency.

The use of correlations to address several of the research questions may raise questions about whether or not this type of analysis is the most effective way to examine the topic. However, previous studies on related topics (Bosker et al., 2013; De Jong et al. 2012, 2013a, 2013b; Derwing et al., 2009; Hilton, 2008; Kormos & Denes, 2004; Rossiter, 2009) have taken the same approach. Also, it should be noted that the purpose of using correlations is to look at which factors are related to one another, and, in the case of changes from pre- to post-SA, to look at which
factors change together. It is acknowledged that correlations do not prove a causal relationship between two factors.

The measurement of fluency

1. Are there significant differences between the speech of learners at different levels of lexical–grammatical competence in the number of pauses 0.5 seconds and longer? Are there significant differences between these groups in the number of shorter pauses (0.25 to 0.49 seconds)?

2. Do filled and unfilled pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

3. Do mid-clause and end-of-clause pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

To address these questions, data from both study abroad and non-study abroad participants were used. Participants were divided into high and low groups based on their scores on the grammar and vocabulary measures, using the pre-SA data for study abroad participants. A composite score from these two tests was calculated for each participant. Those whose score fell below the median (median = 38.5) were considered to have low lexical–grammatical competence, and those whose score fell above the median were considered to have high lexical–grammatical competence.

To address Question 1, several measures of pausing were combined to create two categories, brief pauses (between 0.25 and 0.49 seconds) and longer pauses (0.5 seconds or more), measured as the number of pauses 100 syllables. Each of these categories contained both filled and unfilled pauses and mid-clause and end-
of-clause pauses. T-tests were used to see if there were significant differences
between the high and low groups in the number of brief pauses per 100 syllables in
their speech and the number of longer pauses per 100 syllables in their speech.

To address Question 2, measures were again combined to create two
categories, filled and unfilled pauses. Each of these categories contained both brief
and longer pauses and mid-clause and end-of-clause pauses. T-tests were used to
see if there were significant differences between the high and low groups in the
number of filled pauses per 100 syllables in their speech and the number of unfilled
pauses per 100 syllables in their speech.

Measures were also combined to address Question 3. For this question, the
categories of mid-clause pauses and end-of-clause pauses contained both filled and
unfilled pauses and brief and longer pauses. T-tests were again used to see if there
were significant differences between the high and low groups in the number of mid-
clause pauses per 100 syllables in their speech and the number of end-of-clause
pauses per 100 syllables in their speech.

**Summary**

This chapter has described the design of the study, including the participants,
materials, tasks, data collection, and methods of data analysis. The following chapter
presents the results of the study.
Chapter 4: Results

This chapter presents the results of the study, beginning with the research questions on the development of fluency. Next, results related to the measurement of fluency are presented. Finally, this chapter also presents findings from participants’ responses to the language contact questionnaire.

The development of fluency

Question 1: Changes in fluency during study abroad

The first research question related to the development of fluency asks:

1. Are there changes in participants’ L2 articulation rate, pausing, repair, and mean length of run from pre- to post-study abroad (SA)?
   a. What changes are evident in terms of raw numbers?
   b. What changes are evident in terms of L1–L2 difference scores?

Table 4 reviews the fluency measures that were used in the study. Table 5 summarizes study abroad participants’ mean pretest and posttest scores on each fluency measure. Paired t-tests were used to check for significant differences between pretest and posttest scores; Table 5 also shows the p-value for each measure. Figures 2 and 3 display these results visually.
Table 4: Fluency measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulation rate (syllables per second)</td>
<td>The number of syllables per second, with pauses ≥0.25 seconds removed from the calculation</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>The total percent of speaking time taken up by filled or unfilled pauses lasting 0.25 seconds or more</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>The number of pauses lasting 0.25 seconds or more, per 100 syllables</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>The number of pauses lasting 0.25 seconds or more occurring in the middle of a clause, per 100 syllables</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>The number of pauses lasting 0.25 seconds or more occurring at the end of a clause, per 100 syllables</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>The number of pauses lasting 0.5 seconds or more, per 100 syllables</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>The number of pauses lasting between 0.25 seconds and 0.49 seconds, per 100 syllables</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>The number of unfilled (silent) pauses lasting 0.25 seconds or more, per 100 syllables</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>The number of pauses consisting of non-lexical fillers lasting 0.25 seconds or more, per 100 syllables</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>The number of repetitions (exact duplications of the same syllable, word, or phrase), replacements (lexical substitutions), reformulations (grammatical self-corrections), and false starts (abandoning an idea without completing it) per 100 syllables</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>The mean number of syllables spoken between pauses (filled or unfilled) 0.5 seconds or longer</td>
</tr>
</tbody>
</table>
Table 5: Mean pretest and posttest scores for all fluency measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean pretest score (SD)</th>
<th>Mean posttest score (SD)</th>
<th>Significance (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables per second</td>
<td>3.90 (0.68)</td>
<td>4.08 (0.58)</td>
<td>.002**</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>39.59 (11.16)</td>
<td>35.47 (10.01)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate27 of all pauses</td>
<td>16.84 (5.85)</td>
<td>14.83 (4.64)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>8.94 (4.61)</td>
<td>7.03 (3.55)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>7.90 (2.08)</td>
<td>7.80 (2.11)</td>
<td>.780</td>
</tr>
<tr>
<td>Rate of long pauses (≥0.5 seconds)</td>
<td>12.98 (5.78)</td>
<td>10.85 (4.38)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of short pauses (between 0.25 and 0.5 seconds)</td>
<td>3.87 (1.58)</td>
<td>3.99 (1.63)</td>
<td>.691</td>
</tr>
<tr>
<td>Rate of unfilled pauses28</td>
<td>18.78 (7.75)</td>
<td>15.79 (5.78)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>6.56 (4.31)</td>
<td>4.33 (3.60)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>3.48 (2.34)</td>
<td>3.43 (2.75)</td>
<td>.863</td>
</tr>
<tr>
<td>Mean length of run between pauses ≥0.5 seconds</td>
<td>8.81 (3.74)</td>
<td>10.55 (5.02)</td>
<td>.005**</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

27 Rates are per 100 syllables

28 The rate of unfilled pauses and rate of filled pauses add up to more than the rate of all pauses because of the way the pauses were counted. For the rate of all pauses, mid-clause pauses, end-of-clause pauses, long pauses, and short pauses, instances of a filled pause immediately followed by an unfilled pause (or vice versa) were counted as a single pause. For the rate of unfilled pauses and filled pauses, filled and unfilled pauses were all counted as separate pauses.
Figure 2: Changes in pausing from pre- to post-study abroad

Figure 3: Changes in articulation rate, repair, and mean length of run from pre- to post-study abroad
As Table 5 shows, there were significant changes in participants’ fluency from the pretest to the posttest on a number of measures. First, their mean articulation rate increased from 3.90 to 4.08 syllables per second. There were significant decreases in pausing: percent of pausing time decreased from 39.59% to 35.47%, the rate of all pauses decreased from 16.84 to 14.83 pauses per 100 syllables, the rate of mid-clause pauses decreased from 8.94 to 7.03 pauses per 100 syllables, the rate of long pauses decreased from 12.98 to 10.85 pauses per 100 syllables, the rate of unfilled pauses decreased from 18.78 to 15.79 pauses per 100 syllables, and the rate of filled pauses decreased from 6.56 to 4.33 pauses per 100 syllables. However, the rate of end-of-clause pauses and the rate of short pauses did not significantly change. Looking at rate of repair, there was no significant change. Finally, participants’ mean length of run increased from 8.81 to 10.55 syllables.

Question 1b examines changes in L2 fluency as measured against participants’ L1 fluency, using effect sizes to quantify L1–L2 differences. In other words, how much more similar to (or different from) their L1 fluency is participants’ L2 fluency after studying abroad? First, t-tests were used to evaluate whether or not the difference between participants’ L1 and L2 fluency scores was significant for each measure. With the exception of end-of-clause pauses (for which the difference between the L1 and the L2 was not significant either pre- or post-SA), the differences between the L1 and the L2 were significant at $p < .05$ both before and after study abroad. Effect sizes were then calculated to show the magnitude of the differences. Table 6 summarizes these results, and Figures 4 and 5 present the
results visually. For reference, Table 7 shows the minimum and maximum scores in both the L1 and the L2.

Table 6: Mean L1–L2 difference scores

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD) L1 score</th>
<th>Mean (SD) L2 pretest score</th>
<th>Mean (SD) L1–L2 difference score (pre-SA)</th>
<th>Mean (SD) L2 posttest score</th>
<th>Mean (SD) L1–L2 difference score (post-SA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables per second</td>
<td>4.32 (0.55)</td>
<td>3.90 (0.68)</td>
<td>0.73 (0.79)</td>
<td>4.08 (0.58)</td>
<td>0.31 (0.58)</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>29.09 (8.48)</td>
<td>39.59 (11.16)</td>
<td>1.09 (0.84)</td>
<td>35.47 (10.01)</td>
<td>0.66 (0.98)</td>
</tr>
<tr>
<td>Rate^{29} of all pauses</td>
<td>11.06 (2.92)</td>
<td>16.84 (5.85)</td>
<td>1.36 (1.12)</td>
<td>14.83 (4.64)</td>
<td>0.88 (0.89)</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>3.57 (1.98)</td>
<td>8.94 (4.61)</td>
<td>1.66 (1.16)</td>
<td>7.03 (3.55)</td>
<td>1.07 (0.82)</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>7.53 (1.52)</td>
<td>7.90 (2.08)</td>
<td>0.20 (1.25)</td>
<td>7.80 (2.11)</td>
<td>0.15 (1.26)</td>
</tr>
<tr>
<td>Rate of long pauses (≥0.5 seconds)</td>
<td>7.82 (2.82)</td>
<td>12.98 (5.78)</td>
<td>1.25 (1.10)</td>
<td>10.85 (4.38)</td>
<td>0.73 (0.94)</td>
</tr>
<tr>
<td>Rate of short pauses (between 0.25 and 0.5 seconds)</td>
<td>3.24 (1.18)</td>
<td>3.87 (1.58)</td>
<td>0.44 (1.12)</td>
<td>3.99 (1.63)</td>
<td>0.53 (1.18)</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>11.73 (3.91)</td>
<td>18.78 (7.75)</td>
<td>1.27 (1.08)</td>
<td>15.79 (5.78)</td>
<td>0.73 (0.90)</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>3.15 (2.19)</td>
<td>6.56 (4.31)</td>
<td>1.07 (1.11)</td>
<td>4.33 (3.60)</td>
<td>0.37 (0.92)</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>1.09 (1.14)</td>
<td>3.48 (2.34)</td>
<td>1.21 (0.88)</td>
<td>3.43 (2.75)</td>
<td>1.18 (1.08)</td>
</tr>
<tr>
<td>Mean length of run between pauses ≥0.5 seconds</td>
<td>13.86 (5.95)</td>
<td>8.81 (3.74)</td>
<td>0.96 (0.95)</td>
<td>10.55 (5.02)</td>
<td>0.63 (0.86)</td>
</tr>
</tbody>
</table>

^{29} Rates are per 100 syllables
Figure 4: L1–L2 difference scores for pausing

Figure 5: L1–L2 difference scores for articulation rate, repair, and mean length of run
### Table 7: Minimum and maximum values for L1 and L2 fluency measures

<table>
<thead>
<tr>
<th></th>
<th>Minimum L1 score</th>
<th>Maximum L1 score</th>
<th>Minimum pre-SA L2 score</th>
<th>Maximum pre-SA L2 score</th>
<th>Minimum post-SA L2 score</th>
<th>Maximum post-SA L2 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables per second</td>
<td>3.30</td>
<td>5.24</td>
<td>2.35</td>
<td>5.28</td>
<td>3.04</td>
<td>5.26</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>10.86</td>
<td>53.02</td>
<td>17.69</td>
<td>58.87</td>
<td>17.70</td>
<td>55.57</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>5.78</td>
<td>18.52</td>
<td>7.47</td>
<td>30.40</td>
<td>7.54</td>
<td>25.56</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>0.58</td>
<td>9.88</td>
<td>2.14</td>
<td>18.85</td>
<td>2.48</td>
<td>17.27</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>5.00</td>
<td>10.33</td>
<td>3.74</td>
<td>12.80</td>
<td>4.06</td>
<td>12.72</td>
</tr>
<tr>
<td>Rate of long pauses (≥0.5 seconds)</td>
<td>2.31</td>
<td>15.43</td>
<td>4.98</td>
<td>27.20</td>
<td>3.03</td>
<td>21.80</td>
</tr>
<tr>
<td>Rate of short pauses (between 0.25 and 0.5 seconds)</td>
<td>1.14</td>
<td>6.26</td>
<td>1.32</td>
<td>6.94</td>
<td>1.17</td>
<td>8.08</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>5.78</td>
<td>22.84</td>
<td>7.47</td>
<td>34.78</td>
<td>8.06</td>
<td>32.33</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>0.00</td>
<td>12.25</td>
<td>0.78</td>
<td>16.46</td>
<td>0.00</td>
<td>15.04</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>0.00</td>
<td>5.36</td>
<td>0.00</td>
<td>12.39</td>
<td>0.00</td>
<td>12.92</td>
</tr>
<tr>
<td>Mean length of run between pauses ≥0.5 seconds</td>
<td>6.23</td>
<td>38.44</td>
<td>3.57</td>
<td>18.73</td>
<td>4.43</td>
<td>29.70</td>
</tr>
</tbody>
</table>

Looking at measures of pausing, there were large differences\(^{30}\) between participants’ L1 and L2 fluency before studying abroad on a number of measures: percent pausing time (1.09), rate of all pauses (1.36), rate of mid-clause pauses (1.66), rate of long pauses (1.25), rate of unfilled pauses (1.27), and rate of filled pauses.

---

\(^{30}\) Using Cohen’s *d*, effect sizes between 0.20 and 0.49 are considered small; between 0.50 and 0.79, they are considered moderate; and effect sizes equal to or greater than 0.80 are considered large (Kirk, 2006).
pauses (1.07). After studying abroad, two of these differences were still large: rate of all pauses (0.88) and rate of mid-clause pauses (1.07). Percent pausing time (0.66), rate of long pauses (0.73), and rate of unfilled pauses (0.73) decreased to a moderate difference; and rate of filled pauses decreased to a small difference (0.37). For two measures of pausing, end-of-clause pauses (0.20) and rate of short pauses (0.44), the differences were already small before studying abroad.

Looking at other measures of fluency, there was a moderate difference (0.73) between participants’ L1 and L2 articulation rate before studying abroad, which decreased to a small difference (0.31) after studying abroad. The difference between their L1 and L2 rate of repair was large (1.21) before studying abroad and virtually unchanged (1.18) after studying abroad. The difference between their L1 and L2 mean length of run decreased from a large difference (0.96) before studying abroad to a moderate difference (0.63) after studying abroad.

**Question 2: Pre-SA fluency and changes in fluency**

The second research question related to the development of fluency asks:

2. Is there a relationship between pre-SA L2 fluency and changes in L2 fluency?
   a. Is there a correlation between pre-SA L2 fluency scores and change in L2 fluency scores from pre- to post-SA?
   b. Is there a significant difference between the amount of change in L2 fluency scores for participants with low pre-SA L2 fluency scores (below the median score) and the amount of change in L2 fluency scores for participants with high pre-SA L2 fluency scores (above the median score)?
c. If there are differences in the amount of change in L2 fluency scores for participants with low and high pre-SA L2 fluency scores, in what areas of fluency (articulation rate, pausing, repair, or length of run) are the differences most evident?

Looking at each measure of fluency separately, there does appear to be a relationship between pre-SA L2 fluency and change in L2 fluency. As Table 8 shows, there are significant negative correlations between pretest scores and amount of change for most measures.

Table 8: Correlations between pre-SA L2 fluency and changes in L2 fluency

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation between pretest score and amount of change</th>
<th>With L1 control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllables per second</td>
<td>-0.527**</td>
<td>-0.453**</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>-0.458**</td>
<td>-0.236</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>-0.582**</td>
<td>-0.550**</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>-0.607**</td>
<td>-0.628**</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>-0.516**</td>
<td>-0.530**</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>-0.612**</td>
<td>-0.468**</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>-0.593**</td>
<td>-0.628**</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>-0.619**</td>
<td>-0.484**</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>-0.483**</td>
<td>-0.500**</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>-0.250</td>
<td>-0.327*</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>-0.211</td>
<td>-0.324*</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Dividing participants into low and high pre-SA fluency groups, there is a significant difference in the overall amount of change in fluency from pre- to post-SA, as Table 9 shows. To calculate overall measures of pre-SA fluency and amount of change in fluency, several fluency measures were combined through the use of T-scores, as explained in chapter 3. The mean T-score for amount of change was 52.82.
for the low pre-SA fluency group and 47.47 for the high pre-SA fluency group. Table 9 also compares these two groups on each of the individual measures of fluency and shows the statistical significance of independent t-tests comparing the amount of change for the two groups. Figures 6 and 7 visually display the amount of change for low and high pre-SA fluency groups.

As Table 9 shows, the low pre-SA fluency group improved significantly more than the high pre-SA fluency group on most measures of pausing: percent pausing time, rate of all pauses, rate of mid-clause pauses, rate of unfilled pauses, rate of filled pauses, and rate of long pauses. However, the differences between the low and high groups are not significant for rate of end-of-clause pauses or rate of short pauses. Additionally, the differences are not significant for other aspects of fluency (articulation rate, rate of repair, and mean length of run).
Table 9: Mean changes for low and high pre-SA fluency groups

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD) amount of change for high pre-SA fluency group</th>
<th>Mean (SD) amount of change for low pre-SA fluency group</th>
<th>Significance (2-tail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite fluency score</td>
<td>47.47 (6.01)</td>
<td>52.82 (6.22)</td>
<td>.012*</td>
</tr>
<tr>
<td>Syllables per second</td>
<td>0.14 (0.29)</td>
<td>0.24 (0.39)</td>
<td>.380</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>4.27, 4.41</td>
<td>3.51, 3.74</td>
<td></td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>2.01 (5.12)</td>
<td>6.49 (7.06)</td>
<td>.032*</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>30.52, 28.52</td>
<td>48.64, 42.15</td>
<td></td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>0.64 (2.15)</td>
<td>3.56 (3.37)</td>
<td>.003**</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>11.93, 11.29</td>
<td>21.75, 18.19</td>
<td></td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>0.74 (1.96)</td>
<td>3.08 (3.10)</td>
<td>.009**</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>5.24, 4.50</td>
<td>12.73, 9.64</td>
<td></td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>-0.10 (1.53)</td>
<td>0.46 (2.71)</td>
<td>.441</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>6.68, 6.78</td>
<td>9.01, 8.55</td>
<td></td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>0.68 (2.04)</td>
<td>5.57 (4.16)</td>
<td>.000**</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>12.25, 11.57</td>
<td>25.41, 19.84</td>
<td></td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>1.13 (2.09)</td>
<td>3.31 (3.77)</td>
<td>.036*</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>4.21, 3.08</td>
<td>9.03, 5.72</td>
<td></td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>0.61 (2.08)</td>
<td>3.84 (3.64)</td>
<td>.002**</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>8.21, 7.60</td>
<td>17.76, 13.93</td>
<td></td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>0.03 (1.74)</td>
<td>-0.30 (2.24)</td>
<td>.615</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>3.71, 3.69</td>
<td>3.99, 4.29</td>
<td></td>
</tr>
<tr>
<td>Rate of repair</td>
<td>0.52 (1.47)</td>
<td>-0.29 (2.20)</td>
<td>.195</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>2.87, 2.35</td>
<td>4.23, 4.51</td>
<td></td>
</tr>
<tr>
<td>Mean length of run</td>
<td>1.96 (4.95)</td>
<td>1.65 (1.51)</td>
<td>.796</td>
</tr>
<tr>
<td>Pre and post scores</td>
<td>11.94, 13.90</td>
<td>5.74, 7.39</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

31 The pre and post scores for each group on each measure are on the second line of each row.

32 The composite fluency scores are expressed in T-scores since it was necessary to use standardized scores to pool measures; the other scores are expressed in raw numbers. Also, all changes reflecting increased fluency are expressed as positive numbers, but for percent pausing time, all rates of pauses, and rate of repair, these numbers actually represent a decrease in the percent/rate.
Figure 6: Mean amount of change for high and low pre-SA fluency groups: pausing

Figure 7: Mean amount of change for high and low pre-SA fluency groups: articulation rate, repair, and mean length of run
Question 3: Fluency and grammar, vocabulary, and processing speed

The third research question related to the development of fluency asks:

3. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and pre- and post-SA L2 fluency scores?

   a. Is there a correlation between participants’ pre-SA grammar scores and their pre-SA fluency scores? Is there a correlation between their pre-SA grammar scores and their post-SA fluency scores? Is one correlation stronger than the other?

   b. Is there a correlation between participants’ pre-SA vocabulary scores and their pre-SA fluency scores? Is there a correlation between their pre-SA vocabulary scores and their post-SA fluency scores? Is one correlation stronger than the other?

   c. Is there a correlation between participants’ pre-SA processing scores and their pre-SA fluency scores? Is there a correlation between their pre-SA processing scores and their post-SA fluency scores? Is one correlation stronger than the other?

   d. Considering as possible factors pre-SA grammar, vocabulary, picture naming, and sentence matching scores, what regression model best predicts pre-SA L2 fluency?

Table 10 shows the correlations between pretest measures of grammar, vocabulary, and processing scores and pretest measures of fluency.
Table 10: Correlations between pretest measures

<table>
<thead>
<tr>
<th></th>
<th>Grammar L2 only</th>
<th>Grammar With L1 control</th>
<th>Vocabulary L2 only</th>
<th>Vocabulary With L1 control</th>
<th>Sentence Matching L2 only</th>
<th>Sentence Matching With L1 control</th>
<th>Picture Naming L2 only</th>
<th>Picture Naming With L1 control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulation rate</td>
<td>.543**</td>
<td>.545**</td>
<td>.486**</td>
<td>.381*</td>
<td>.528**</td>
<td>.553**</td>
<td>.551**</td>
<td>.522**</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>.496**</td>
<td>.481**</td>
<td>.417**</td>
<td>.618**</td>
<td>.486**</td>
<td>.451**</td>
<td>.496**</td>
<td>.641**</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>.572**</td>
<td>.519**</td>
<td>.502**</td>
<td>.550**</td>
<td>.660**</td>
<td>.646**</td>
<td>.530**</td>
<td>.563**</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>.629**</td>
<td>.554**</td>
<td>.553**</td>
<td>.513**</td>
<td>.686**</td>
<td>.661**</td>
<td>.513**</td>
<td>.421**</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>.337*</td>
<td>.297</td>
<td>.275</td>
<td>.299</td>
<td>.312</td>
<td>.285</td>
<td>.413**</td>
<td>.438**</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>.633**</td>
<td>.621**</td>
<td>.537**</td>
<td>.670**</td>
<td>.617**</td>
<td>.644**</td>
<td>.566**</td>
<td>.658**</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>.398*</td>
<td>.527**</td>
<td>.354**</td>
<td>.541**</td>
<td>.403*</td>
<td>.499**</td>
<td>.417**</td>
<td>.555**</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>.550**</td>
<td>.540**</td>
<td>.455**</td>
<td>.576**</td>
<td>.555**</td>
<td>.541**</td>
<td>.513**</td>
<td>.604**</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>.342*</td>
<td>.307</td>
<td>.321*</td>
<td>.268</td>
<td>.276</td>
<td>.254</td>
<td>.238</td>
<td>.201</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>.084</td>
<td>.004</td>
<td>.126</td>
<td>.101</td>
<td>.188</td>
<td>.158</td>
<td>.157</td>
<td>.148</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>.559**</td>
<td>.538**</td>
<td>.461**</td>
<td>.557**</td>
<td>.552**</td>
<td>.564**</td>
<td>.512**</td>
<td>.570**</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

---

All correlations representing a positive relationship between higher grammar, vocabulary, and processing scores and greater fluency are expressed as positive correlations, even though some of these correlations are technically negative. For example, participants with higher grammar scores generally had a lower rate of pausing. However, for the sake of consistency, it seemed better to present such correlations as positive since lower rates of pausing indicate higher fluency.
Table 11 shows the correlations between pretest measures of grammar, vocabulary, and processing scores and posttest measures of fluency.

Table 11: Correlations between pretest measures of grammar, vocabulary, and processing scores and posttest measures of L2 fluency

<table>
<thead>
<tr>
<th></th>
<th>Grammar</th>
<th>Vocabulary</th>
<th>Sentence Matching</th>
<th>Picture Naming</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 only</td>
<td>With L1 control</td>
<td>L2 only</td>
<td>With L1 control</td>
</tr>
<tr>
<td>Articulation rate</td>
<td>.406*</td>
<td>.371*</td>
<td>.339*</td>
<td>.198</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>.326*</td>
<td>.276</td>
<td>.343*</td>
<td>.432**</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>.378*</td>
<td>.305</td>
<td>.406*</td>
<td>.450**</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>.483**</td>
<td>.389*</td>
<td>.388*</td>
<td>.318</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>.126</td>
<td>.101</td>
<td>.257</td>
<td>.290</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>.429**</td>
<td>.372*</td>
<td>.415**</td>
<td>.474**</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>.150</td>
<td>.266</td>
<td>.270</td>
<td>.491**</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>.339*</td>
<td>.292</td>
<td>.415*</td>
<td>.480**</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>.166</td>
<td>.124</td>
<td>.048</td>
<td>.019</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>.147</td>
<td>.218</td>
<td>.186</td>
<td>.107</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>.366*</td>
<td>.324*</td>
<td>.369*</td>
<td>.437**</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

34 Again, all correlations representing a positive relationship between higher grammar, vocabulary, and processing scores and greater fluency are expressed as positive correlations.
As Tables 10 and 11 show, there is a relationship between pre-SA grammar scores and pre-SA fluency scores in all areas except for rate of repair; however, the correlations with the rate of end-of-clause pauses and the rate of short pauses are significant only if L1 fluency is not taken into account. There is also a relationship between pre-SA grammar scores and most measures of post-SA fluency. However, the correlations with several of these measures, including the percent of time spent pausing, the rate of all pauses, and the rate of long pauses are significant only if L1 fluency is not taken into account. Additionally, the correlations between pre-SA grammar scores and post-SA scores of rate of repair, rate of end-of-clause pauses, rate of filled pauses, and rate of short pauses are not significant at all. Overall, the correlations are stronger between pre-SA grammar scores and pre-SA fluency scores than between pre-SA grammar scores and post-SA fluency scores.

Similarly, there is a relationship between pre-SA vocabulary scores and pre-SA fluency scores in all areas except for rate of repair and rate of end-of-clause pauses; the correlation with short pauses is only significant if L1 fluency is not taken into account. There is also a relationship between pre-SA vocabulary scores and most measures of post-SA fluency. However, the correlations between pre-SA vocabulary scores and post-SA scores of rate of repair, rate of short pauses, and rate of end-of-clause pauses are not significant. Also, the correlation between pre-SA vocabulary and post-SA rate of filled pauses is only significant if L1 fluency is taken into account, and the correlations between pre-SA vocabulary and post-SA articulation rate and between pre-SA vocabulary and post-SA rate of mid-clause pauses are only significant if L1 fluency is not taken into account. Overall, the
correlations are stronger between pre-SA vocabulary scores and pre-SA fluency scores than between pre-SA vocabulary scores and post-SA fluency scores.

The pattern is similar for processing measures as well. For sentence matching, there is a relationship between pre-SA sentence matching scores and pre-SA fluency scores in all areas except for rate of repair, rate of end-of-clause pauses, and rate of short pauses. There is also a relationship between pre-SA sentence matching scores and most measures of post-SA fluency, with the exception of the rate of repair, the rate of end-of-clause pauses, the rate of short pauses, and the rate of filled pauses. Additionally, the correlation between pre-SA sentence matching and post-SA percent pausing time is only significant if L1 fluency is not taken into account. Most correlations are stronger between pre-SA sentence matching scores and pre-SA fluency scores than between pre-SA sentence matching scores and post-SA fluency scores.

For picture naming, there is a relationship between pre-SA picture naming scores and pre-SA fluency scores in all areas except for rate of repair and rate of short pauses. There is also a relationship between pre-SA picture naming scores and all measures of post-SA fluency except for rate of repair, rate of short pauses, and rate of end-of-clause pauses. Also, the correlation between pre-SA picture naming and post-SA rate of filled pauses is only significant if L1 fluency is taken into account. The correlations are stronger between pre-SA picture naming scores and pre-SA fluency scores than between pre-SA picture naming scores and post-SA fluency scores.
To address the question regarding which areas of linguistic knowledge (pre-SA grammar and vocabulary scores) and processing speed (pre-SA sentence matching and picture naming scores) best predict pre-SA L2 fluency measures, backward elimination regression was used to compare the model including all four factors with models including fewer factors. The processing measures (sentence matching and picture naming) are the best predictors of pre-SA fluency; the model including only these two factors predicts fluency virtually as well as the model that also includes grammar and vocabulary ($R^2 = .522$ for the full model; $R^2 = .507$ for the model including only sentence matching and picture naming). As Table 12 shows, of the processing measures, sentence matching is the strongest single predictor of fluency.

Table 12: Regression model predicting pre-SA L2 fluency

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture naming</td>
<td>-.306</td>
<td>-2.218</td>
<td>.034*</td>
</tr>
<tr>
<td>Sentence matching</td>
<td>-.523</td>
<td>-3.785</td>
<td>.001*</td>
</tr>
</tbody>
</table>

*p < .05

**Question 4: Grammar, vocabulary, and processing and changes in fluency**

The fourth research question related to the development of fluency asks:

4. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and changes in fluency?

---

35 The results were also checked using forward and stepwise regression, and all three methods led to the same results.
a. Is there a correlation between participants’ pre-SA grammar scores and change in fluency scores from pre- to post-SA?

b. Is there a correlation between participants’ pre-SA vocabulary scores and change in fluency scores from pre- to post-SA?

c. Is there a correlation between participants’ pre-SA processing scores and change in fluency scores from pre- to post-SA?

d. What is the relative strength of each of these correlations?

As Table 13 shows, there are negative correlations between pre-SA grammar, vocabulary, and processing scores and several measures of changes in fluency. Pre-SA grammar scores are negatively correlated with changes in the rate of all pauses, the rate of mid-clause pauses, the rate of unfilled pauses, and the rate of long pauses. Pre-SA vocabulary scores are negatively correlated with changes in articulation rate, the rate of mid-clause pauses, and the rate of unfilled pauses. Pre-SA sentence matching scores are negatively correlated with changes in the rate of all pauses, the rate of mid-clause pauses, the rate of unfilled pauses, and the rate of long pauses. Pre-SA picture naming scores are negatively correlated with changes in articulation rate and the rate of unfilled pauses.
Table 13: Correlations between pre-SA grammar, vocabulary, and processing scores and changes in fluency from pre- to post-SA

<table>
<thead>
<tr>
<th></th>
<th>Grammar L2 only</th>
<th>With L1 control</th>
<th>Vocabulary L2 only</th>
<th>With L1 control</th>
<th>Sentence Matching L2 only</th>
<th>With L1 control</th>
<th>Picture Naming L2 only</th>
<th>With L1 control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulation rate</td>
<td>-.311</td>
<td>-.265</td>
<td>-.345*</td>
<td>-.277</td>
<td>-.144</td>
<td>-.097</td>
<td>-.364*</td>
<td>-.296</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>-.273</td>
<td>-.218</td>
<td>-.134</td>
<td>-.182</td>
<td>-.289</td>
<td>-.207</td>
<td>-.233</td>
<td>-.253</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>-.336*</td>
<td>-.383*</td>
<td>-.342*</td>
<td>-.336*</td>
<td>-.488**</td>
<td>-.450**</td>
<td>-.322*</td>
<td>-.306</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>-.377*</td>
<td>-.343*</td>
<td>-.339*</td>
<td>-.310</td>
<td>-.542**</td>
<td>-.508**</td>
<td>-.246</td>
<td>-.209</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>-.157</td>
<td>-.158</td>
<td>.016</td>
<td>.016</td>
<td>-.142</td>
<td>-.137</td>
<td>-.106</td>
<td>-.106</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>-.430**</td>
<td>-.376*</td>
<td>-.351*</td>
<td>-.380*</td>
<td>-.525**</td>
<td>-.479**</td>
<td>-.361*</td>
<td>-.359*</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>-.298</td>
<td>-.315</td>
<td>-.044</td>
<td>-.068</td>
<td>-.217</td>
<td>-.220</td>
<td>-.102</td>
<td>-.112</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>-.340*</td>
<td>-.294</td>
<td>-.146</td>
<td>-.170</td>
<td>-.466**</td>
<td>-.414*</td>
<td>-.211</td>
<td>-.206</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>-.128</td>
<td>-.125</td>
<td>-.211</td>
<td>-.210</td>
<td>-.112</td>
<td>-.111</td>
<td>-.109</td>
<td>-.106</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>.132</td>
<td>.139</td>
<td>.061</td>
<td>.068</td>
<td>.215</td>
<td>.212</td>
<td>.082</td>
<td>.093</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>-.144</td>
<td>-.163</td>
<td>-.093</td>
<td>-.094</td>
<td>-.045</td>
<td>-.125</td>
<td>-.105</td>
<td>-.056</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Given that pre-SA grammar, vocabulary, and processing scores are positively correlated with pre-SA L2 fluency scores (Research Question 3), and pre-SA L2 fluency scores are negatively correlated with changes in L2 fluency (Research Question 2), the negative correlation between pre-SA grammar, vocabulary, and...
processing scores and changes in L2 fluency is logical. Repeating the same analysis while controlling for pre-SA L2 fluency scores, none of the correlations are significant, as Table 14 shows.

Table 14: Correlations between pre-SA grammar, vocabulary, and processing scores and changes in L2 fluency, controlling for pre-SA L2 fluency

<table>
<thead>
<tr>
<th></th>
<th>Grammar</th>
<th>Vocabulary</th>
<th>Sentence Matching</th>
<th>Picture Naming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulation rate</td>
<td>-.064</td>
<td>-.136</td>
<td>.124</td>
<td>-.100</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>-.084</td>
<td>.051</td>
<td>-.064</td>
<td>-.020</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>-.157</td>
<td>-.090</td>
<td>-.140</td>
<td>-.030</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>-.041</td>
<td>-.031</td>
<td>-.044</td>
<td>.087</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>-.005</td>
<td>.171</td>
<td>.005</td>
<td>.132</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>-.092</td>
<td>-.044</td>
<td>-.210</td>
<td>-.045</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>-.165</td>
<td>.130</td>
<td>-.031</td>
<td>.110</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>-.062</td>
<td>.169</td>
<td>-.098</td>
<td>.139</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>.088</td>
<td>-.036</td>
<td>.067</td>
<td>.030</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>.145</td>
<td>.075</td>
<td>.270</td>
<td>.102</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>-.067</td>
<td>-.020</td>
<td>.012</td>
<td>.035</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Question 5: Changes in grammar, vocabulary, and processing and changes in fluency

The fifth research question related to the development of fluency asks:

5. Is there a relationship between changes in grammar, vocabulary, and/or processing scores and changes in L2 fluency?
   a. Is there a correlation between change in grammar scores and change in fluency scores from pre- to post-SA?
b. Is there a correlation between change in vocabulary scores and change in fluency scores from pre- to post-SA?

c. Is there a correlation between change in processing scores and change in L2 fluency scores from pre- to post-SA?

d. What is the relative strength of each of these correlations?

As Table 15 shows, there are few significant correlations between changes in fluency and changes in other measures. A decrease in the rate of mid-clause pauses is significantly correlated with an increase in vocabulary scores, both with \( r = .413 \) and without \( r = .418 \) controlling for the rate of mid-clause pauses in the L1. A decrease in the rate of all pauses and a decrease in the rate of unfilled pauses are also significantly correlated \( (r = .326 \text{ and } r = .328, \text{ respectively}) \) with an increase in vocabulary scores, but only with a control for L1 pauses. An increase in articulation rate is significantly correlated \( (r = .373) \) with a decrease in picture naming times, but only without controlling for L1 articulation rate.
Table 15: Correlations between changes in grammar, vocabulary, and processing scores and changes in L2 fluency\textsuperscript{36}

<table>
<thead>
<tr>
<th></th>
<th>Grammar</th>
<th>Vocabulary</th>
<th>Sentence Matching</th>
<th>Picture Naming</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 only</td>
<td>With L1 control</td>
<td>L2 only</td>
<td>With L1 control</td>
</tr>
<tr>
<td>Articulation rate</td>
<td>.276</td>
<td>.249</td>
<td>- .010</td>
<td>- .038</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>.087</td>
<td>.121</td>
<td>- .054</td>
<td>.036</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>.054</td>
<td>.056</td>
<td>.260</td>
<td>.326*</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>-.013</td>
<td>-.042</td>
<td>.413*</td>
<td>.418**</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>.134</td>
<td>.135</td>
<td>-.204</td>
<td>-.219</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>.095</td>
<td>.095</td>
<td>.191</td>
<td>.328*</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>.188</td>
<td>.185</td>
<td>-.272</td>
<td>-.299</td>
</tr>
<tr>
<td>Rate of long pauses</td>
<td>.064</td>
<td>.089</td>
<td>.077</td>
<td>.197</td>
</tr>
<tr>
<td>Rate of short pauses</td>
<td>.022</td>
<td>.018</td>
<td>.218</td>
<td>.217</td>
</tr>
<tr>
<td>Rate of repair</td>
<td>.051</td>
<td>.044</td>
<td>.048</td>
<td>.062</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>-.034</td>
<td>-.032</td>
<td>.095</td>
<td>.078</td>
</tr>
</tbody>
</table>

\*p < .05; \*\*p < .01

\textsuperscript{36} All correlations representing a positive relationship between increases in grammar, vocabulary, and processing scores and increases in fluency are expressed as positive correlations.
Question 6: Predicting changes in fluency

The sixth research question related to the development of fluency asks:

6. Considering as possible factors L1 fluency, pre-SA L2 fluency, changes in grammar from pre- to post-SA, changes in vocabulary from pre- to post-SA, changes in processing speed from pre- to post-SA, and time spent interacting with native speakers of the L2 during the semester abroad, what regression model best predicts changes in L2 fluency during a semester abroad?

Since there were no significant correlations between changes in L2 fluency and changes in grammar or changes in sentence matching speed (Research Question 5), these factors were not entered into the regression model. Additionally, because eight of the 39 participants did not consistently complete the language contact questionnaire (and therefore their data on this measure could not be used), it seemed best to include the factor time spent interacting with native speakers of the L2 only if it were likely to be a significant predictor of changes in fluency. The correlation between this factor and the T-score for change in L2 fluency is -.051 (p = .785) without controlling for pre-SA L2 fluency and .041 (p = .828) controlling for pre-SA L2 fluency. Therefore, it was not included, and only four factors were entered into the model: L1 fluency, pre-SA L2 fluency, change in vocabulary score, and change in picture naming speed. Backward elimination was used to compare the model including all four factors with models including fewer factors.37 Participants’ L2 fluency prior to study abroad is the strongest predictor of change in L2 fluency,

37 The results were also checked using forward and stepwise regression, and all three methods led to the same results.
and the model including only this factor predicts change in fluency statistically as well as any of the models including more factors (R² = .274 for the full model; R² = .240 for the model including only pre-SA L2 fluency). In the full model, the next strongest predictor is L1 fluency (β = .234; p = .279); change in vocabulary score (β = .047; p = .790) and change in picture naming speed (β = -.076; p = .680) contribute little to the model.

Table 16: Regression model predicting changes in L2 fluency

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-SA L2 fluency</td>
<td>-.490</td>
<td>-3.274</td>
<td>.002*</td>
</tr>
</tbody>
</table>

*p < .05

This section has presented the results related to the development of fluency. The following section presents the results related to the measurement of fluency.

The measurement of fluency

The first research question related to the measurement of fluency asks:

1. Are there significant differences between the speech of learners at different levels of lexical–grammatical competence in the number of pauses 0.5 seconds and longer? Are there significant differences between these groups in the number of shorter pauses (0.25 to 0.49 seconds)?

As Table 17 shows, the low group had a significantly higher rate of long pauses (17.58 per 100 syllables) than the high group (10.72 per 100 syllables) as well as a significantly higher rate of short pauses (5.28 per 100 syllables) than the high group (3.48 per 100 syllables).
Table 17: Mean pause scores for low and high groups

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD) for low group; N=24</th>
<th>Mean (SD) for high group; N=24</th>
<th>Significance (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent pausing time</td>
<td>46.93 (10.16)</td>
<td>34.11 (11.30)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>22.86 (8.85)</td>
<td>14.21 (6.18)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of long pauses (≥0.5 seconds)</td>
<td>17.58 (7.63)</td>
<td>10.72 (6.30)</td>
<td>.001**</td>
</tr>
<tr>
<td>Rate of short pauses (between 0.25 and 0.5 seconds)</td>
<td>5.28 (2.54)</td>
<td>3.48 (1.66)</td>
<td>.006**</td>
</tr>
<tr>
<td>Rate of mid-clause pauses</td>
<td>13.91 (7.80)</td>
<td>6.89 (4.64)</td>
<td>.001**</td>
</tr>
<tr>
<td>Rate of end-of-clause pauses</td>
<td>8.95 (1.88)</td>
<td>7.30 (2.22)</td>
<td>.008**</td>
</tr>
<tr>
<td>Percent mid-clause pauses</td>
<td>58.10 (10.83)</td>
<td>44.93 (13.83)</td>
<td>.001**</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>26.52 (10.89)</td>
<td>15.01 (7.19)</td>
<td>.000**</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>10.15 (8.88)</td>
<td>5.12 (3.85)</td>
<td>.016*</td>
</tr>
<tr>
<td>Percent filled pauses</td>
<td>24.72 (9.90)</td>
<td>22.92 (10.68)</td>
<td>.548</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

Figure 8: Mean pause scores for low and high groups
The second research question related to the measurement of fluency asks:

2. Do filled and unfilled pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

As Table 17 shows, the low group had a significantly higher rate of filled pauses (10.15 per 100 syllables) than the high group (5.12 per 100 syllables) as well as a significantly higher rate of unfilled pauses (26.52 per 100 syllables) than the high group (15.01 per 100 syllables). The two groups did not differ significantly in the percent of pauses that were filled (low group: 24.72%; high group: 22.92%; p=.548).

3. Do mid-clause and end-of-clause pauses pattern differently for participants at higher and lower levels of lexical–grammatical competence?

As Table 17 shows, the low group had a significantly higher rate of mid-clause pauses (13.91 per 100 syllables) than the high group (6.89 per 100 syllables) as well as a significantly higher rate of end-of-clause pauses (8.95 per 100 syllables) than the high group (7.30 per 100 syllables). The percent of pauses occurring in the middle of a clause was also significantly higher for the low group (58.10%) than for the high group (44.93%).

This section has presented the results for the questions related to the measurement of fluency. The following section presents findings from the language contact questionnaire.

**Language contact questionnaire data**

Of the 39 study abroad participants who completed the posttest, 31 also completed the LCQ at least 3 times. Table 18 shows the minimum, maximum and
mean number of occurrences or hours spent engaged in each type of interaction, per week,\textsuperscript{38} for these participants.

Table 18: Hours per week interacting with native speakers of Spanish

<table>
<thead>
<tr>
<th>Type of interaction</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine transactions, such as service encounters (number of occurrences)</td>
<td>4.50</td>
<td>34.00</td>
<td>13.71</td>
</tr>
<tr>
<td>Routine transactions with a complicating factor (number of occurrences)</td>
<td>0.15</td>
<td>14.15</td>
<td>2.42</td>
</tr>
<tr>
<td>Brief, formulaic exchanges with host family/roommates (number of occurrences)\textsuperscript{39}</td>
<td>0.00</td>
<td>30.10</td>
<td>12.67</td>
</tr>
<tr>
<td>Short conversations lasting 10 minutes or less with host family/roommates (number of occurrences)</td>
<td>0.00</td>
<td>15.00</td>
<td>6.28</td>
</tr>
<tr>
<td>Longer conversations with host family/roommates (in hours)</td>
<td>0.00</td>
<td>6.40</td>
<td>1.66</td>
</tr>
<tr>
<td>Conversation with <em>intercambios</em> (in hours)</td>
<td>0.00</td>
<td>3.00</td>
<td>0.31</td>
</tr>
<tr>
<td>Friendly chitchat with strangers (number of occurrences)</td>
<td>0.90</td>
<td>18.20</td>
<td>4.51</td>
</tr>
<tr>
<td>Hanging out in a group setting (in hours)</td>
<td>0.00</td>
<td>10.80</td>
<td>4.12</td>
</tr>
<tr>
<td>Hanging out one on one (in hours)</td>
<td>0.00</td>
<td>16.40</td>
<td>2.60</td>
</tr>
<tr>
<td>Other interactions (in hours)</td>
<td>0.00</td>
<td>2.50</td>
<td>0.31</td>
</tr>
<tr>
<td>Total time per week (in hours)</td>
<td>1.15</td>
<td>25.71</td>
<td>10.26</td>
</tr>
</tbody>
</table>

As Table 18 shows, participants reported spending little time engaged in “other” interactions not included in one of the previous categories. However, some participants did report other types of interactions, which fell into eight categories, as summarized in Table 19. See Appendix G for examples of interactions in each category.

\textsuperscript{38} Participants filled out the questionnaire every other week and gave estimates of their interactions for the previous two-week period; these numbers have been divided in half to give an estimate of their interactions per week.

\textsuperscript{39} Some participants did not live with native speakers of Spanish and therefore had no interactions in the categories related to host families/roommates.
Table 19: Other interactions with native speakers of Spanish

<table>
<thead>
<tr>
<th>Type of interaction</th>
<th>Number of times reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel/tourism</td>
<td>4</td>
</tr>
<tr>
<td>Class/tutoring</td>
<td>6</td>
</tr>
<tr>
<td>Non-routine transaction</td>
<td>3</td>
</tr>
<tr>
<td>Texting or online communication</td>
<td>5</td>
</tr>
<tr>
<td>Time with extended family or friends of host</td>
<td>4</td>
</tr>
<tr>
<td>Academic communication outside of class</td>
<td>4</td>
</tr>
<tr>
<td>Work/internship</td>
<td>5</td>
</tr>
<tr>
<td>Sports</td>
<td>3</td>
</tr>
</tbody>
</table>

The questionnaire also asked participants to explain which interaction provided the most meaningful language practice, in their opinion, and why. Table 20 summarizes participants’ responses to this question. See Appendix G for specific examples of participants’ responses.

---

40 Although attending class was coded because several participants mentioned it, time spent in class was not included in the calculation of mean number of hours per week interacting with native speakers. All participants were enrolled in courses, and the questionnaire specifically asked about interactions outside of class.
Table 20: Participants’ most meaningful language practice

<table>
<thead>
<tr>
<th></th>
<th>Number of times reported</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group size</strong></td>
<td></td>
</tr>
<tr>
<td>One-on-one</td>
<td>72</td>
</tr>
<tr>
<td>Larger group</td>
<td>65</td>
</tr>
<tr>
<td>Unspecified</td>
<td>28</td>
</tr>
<tr>
<td><strong>Type of relationship</strong></td>
<td></td>
</tr>
<tr>
<td>Host family</td>
<td>41</td>
</tr>
<tr>
<td>Peers/friends</td>
<td>45</td>
</tr>
<tr>
<td>Strangers</td>
<td>6</td>
</tr>
<tr>
<td>Intercambios</td>
<td>7</td>
</tr>
<tr>
<td>Service encounter/work setting</td>
<td>8</td>
</tr>
<tr>
<td>Romantic relationship</td>
<td>5</td>
</tr>
<tr>
<td>Unspecified</td>
<td>53</td>
</tr>
<tr>
<td><strong>Why the interaction provided meaningful language practice</strong></td>
<td></td>
</tr>
<tr>
<td>Learning vocabulary/phrases</td>
<td>30</td>
</tr>
<tr>
<td>Error correction</td>
<td>22</td>
</tr>
<tr>
<td>Forced to speak Spanish</td>
<td>13</td>
</tr>
<tr>
<td>Topic (interesting and/or challenging)</td>
<td>17</td>
</tr>
<tr>
<td>Comfortable setting</td>
<td>20</td>
</tr>
<tr>
<td>Extended time</td>
<td>10</td>
</tr>
<tr>
<td>Listening practice</td>
<td>19</td>
</tr>
<tr>
<td>More focused interaction</td>
<td>4</td>
</tr>
<tr>
<td>Variety</td>
<td>8</td>
</tr>
<tr>
<td>Pronunciation/intonation</td>
<td>3</td>
</tr>
<tr>
<td>Quick pace</td>
<td>5</td>
</tr>
<tr>
<td>Practicing routinized language</td>
<td>2</td>
</tr>
</tbody>
</table>

Since participants’ answers to this last question were coded in three ways (group size, type of relationship, and why the interaction provided meaningful language practice), examining the patterns of codes that appeared together led to the following observations about the settings in which meaningful language practice took place:

1. Participants felt that they learned vocabulary in both group (12 mentions) and one-on-one (9 mentions) settings. They most frequently reported

---

41 The numbers in this section may appear not to add up correctly; the reason is that the group size, the type of relationship, or both were sometimes unclear from participants’ responses and were coded as “unspecified.”
learning vocabulary from peers (12), but also from their host families (7), intercambios (3), and occasionally in other types of relationships.

2. According to participants’ responses, one-on-one interactions seem to have provided more opportunities for error correction (18) than group settings (4). Participants most often mentioned being corrected by their host families (11), but also by peers or friends (4).

3. Participants felt that they were forced (in a positive sense) to speak Spanish in both group (9) and one-on-one (8) settings among peers (5) as well as with their host families (3) and in romantic relationships (2).

4. Participants mentioned that they were interested in and/or were challenged by the topic of conversation more often in one-on-one settings (12) than in group settings (6). Such conversations took place most frequently with their host families (11), but also with peers (5) and in work settings (2).

5. Participants more frequently reported feeling comfortable in group settings (12) than in one-on-one interactions (4). These interactions took place both with their host families (10) and with peers (10).

6. The chance to learn by listening was more frequently mentioned in connection with group interactions (17) than with one-on-one interactions (6), primarily in conversations with peers (10).

Summary

This chapter has presented the results of the study. To summarize, the results show that participants experienced significant gains on most measures of fluency during study abroad. Participants who began their time abroad with low L2
fluency experienced greater gains on several measures of pausing than did participants who began their time abroad with high L2 fluency. Although there are moderate correlations between pretest measures of L2 fluency and pretest measures of linguistic knowledge (vocabulary and grammar) as well as processing speed, there is no relationship between pre-SA linguistic knowledge and gains in L2 fluency, and little relationship between gains in linguistic knowledge and gains in L2 fluency. The best predictor of gains in L2 fluency is pre-SA L2 fluency.

Regarding the questions related to the measurement of fluency, the data show that learners with low and high levels of lexical–grammatical competence significantly differed from one another on all of the measures of rates of pauses as well as in the percent of pauses occurring in the middle of a clause. However, they did not significantly differ from one another in the percent of filled pauses.

The following chapter reviews these results and discusses their implications.

---

42 Controlling for pre-SA L2 fluency; without pre-SA L2 fluency as a control variable, there is a negative relationship between pre-SA linguistic knowledge and gains in L2 fluency.
Chapter 5: Discussion

This chapter reviews the results presented in the previous chapter and discusses them in light of the findings of previous research. The discussion begins with the research questions related to fluency development, followed by the questions related to the measurement of fluency. The chapter concludes with a presentation of the study's limitations and ideas for future research on fluency and study abroad.

Fluency development and its relationship to grammar, vocabulary, and processing speed

Question 1: Changes in fluency during study abroad

1. Are there changes in participants’ L2 articulation rate, pausing, repair, and mean length of run from pre- to post-study abroad (SA)?
   a. What changes are evident in terms of raw numbers?
   b. What changes are evident in terms of L1–L2 difference scores?

Participants’ L2 articulation rate and mean length of run significantly increased from pre- to post-SA, and their pausing significantly decreased according to several measures. There were no significant changes in their rate of repair.

The findings regarding articulation rate and mean length of run are consistent with the results of previous research. The previous studies that have included articulation rate as a measure of fluency (Llanes & Muñoz, 2009; Valls Ferrer, 2008) both found an increase from pre- to post-SA. Other studies (D'Amico, 2010; Freed et al., 2004b; Lennon, 1990; O'Brien et al., 2007; Towell, 2002) have used speech rate, which measures the rate of speech with pauses included, rather than articulation rate, which measures the rate of speech with pauses excluded, so comparison with the results of these studies is less precise. However, with the
exception of Freed et al. (2004), who found no significant changes in any aspect of fluency, the results of the studies measuring speech rate also agree, having found a significant increase from pre- to post-SA. What the present study—along with the previous two studies that measured articulation rate rather than speech rate—shows is that participants actually spoke more quickly (in terms of the rate at which they articulated syllables) after studying abroad. With measures of speech rate, the source of the significant change may have been that participants articulated syllables more quickly post-SA, or it may have been due to a decrease in the amount of time that they spent pausing.

The present study’s finding that participants’ mean length of run significantly increased from pre- to post-SA is also in line with the findings of most previous research. D’Amico (2010), Towell (2002), and Valls Ferrer (2008) all found a significant increase in mean length of run. O’Brien et al. (2007) found a significant increase in two of their three measures of mean length of run. Llanes and Muñoz (2009) measured participants’ longest run rather than mean length of run; they also found a significant increase from pre- to post-SA. In contrast to these findings, Freed et al. (2004b) measured mean length of run in several ways and found no significant changes, but, as noted above, their participants exhibited no significant changes in any aspect of fluency from pre- to post-SA. Lennon (1990) also did not find a significant change in mean length of run, but the lack of significance may be due to the small number of participants in his study; three of his four participants did

43 Changes were significant for mean length of run without filled pauses and mean length of run without any pauses; the change was not significant for mean length of run without unfilled pauses.
demonstrate an increase in mean length of run. Overall, then, there is general agreement from previous research that mean length of run increases from pre- to post-SA, and the present study offers further evidence in support of this finding.

As noted in chapter 2, findings are less clear regarding changes in pausing as compared to mean length of run and articulation rate or speech rate. The two studies that looked at short-term study abroad programs (six weeks in D’Amico, 2010; three to four weeks in Llanes & Muñoz, 2009) did not find any significant changes in pausing. Of the five studies that examined longer stays abroad lasting three to six months, two did not measure pausing (Freed et al., 2004b; O’Brien et al., 2007). Results for pausing in the present study, then, can be compared only with the findings of Lennon (1990), Towell (2002), and Valls Ferrer (2008).

Towell (2002) and Valls Ferrer (2008) both measured the phonation-time ratio, which is essentially the same measure as percent pausing time in the present study; the phonation-time ratio expresses the percent of time spent speaking, and percent pausing time expresses the percent of time spent pausing. The present study’s finding that percent pausing time significantly decreased from pre- to post-SA is consistent with Towell’s (2002) and Valls Ferrer’s (2008) findings that the phonation-time ratio significantly increased from pre- to post-SA.

The other overall measure of pausing in the present study, the rate of all pauses, cannot be directly compared with previous research, as it has not been used before in the context of study abroad research. Valls Ferrer’s (2008) measure of pause frequency is perhaps the most similar measure, and the present study’s finding that the rate of all pauses significantly decreased from pre- to post-SA is in
line with Valls Ferrer’s finding that pause frequency significantly decreased.

However, the present study measured the rate of pauses per 100 syllables, whereas Valls Ferrer measured the rate of pauses per minute, so these measures are not identical.

The more specific measures of pausing used in this study—looking at the rates of mid-clause and end-of-clause pauses, filled and unfilled pauses, and short and long pauses—provide a more detailed examination of changes in pausing from pre- to post-SA than most previous studies have given. The study found that the rate of mid-clause pauses significantly decreased from pre- to post-SA, but the rate of end-of-clause pauses did not decrease. This finding suggests that the reduction in pausing occurred for pauses that are less natural, given that pausing at the end of clauses is more common in L1 speech in general (Chambers, 1997; Davies, 2003; Skehan, 2009), and participants in the present study paused more at the end of clauses (7.53 pauses per 100 syllables) than in the middle of clause (3.57 pauses per 100 syllables) in their L1. A notable difference between participants’ L1 and L2 fluency prior to study abroad was that in their L1, they paused mostly between clauses (suggesting that they were often able to formulate and articulate full clauses without stopping mid-clause), whereas in their L2, they paused more frequently in the middle of clauses than at the end of clauses (suggesting that they had difficulty formulating and articulating full clauses without stopping mid-clause). Post-SA, their pattern of mid-clause and end-of-clause pauses in the L2 was more similar to their L1 pattern than it was prior to study abroad, suggesting an improved ability to formulate and articulate full clauses in the L2 without stopping mid-clause.
Measuring filled and unfilled pauses separately showed a significant decrease in both types of pauses from pre- to post-SA. The only previous studies that measured changes in filled and unfilled pauses separately focused on short-term study abroad programs and did not find any changes in pausing (D’Amico, 2010; Llanes & Muñoz, 2009). As is discussed further in the section on measuring fluency, it appears that an L2 learner’s ratio of filled pauses to unfilled pauses may be more a matter of personal speaking style transferred from the L1 than an indicator of L2 ability. If this hypothesis is true, we would expect to see either a decrease in both filled and unfilled pauses, or little change in either (rather than a substantial change in one and little or no change in the other), as the data from the present study as well as previous studies show.

Measuring short and long pauses separately showed that participants’ rate of long pauses significantly decreased from pre- to post-SA, whereas their rate of short pauses did not change. This finding suggested the possibility that participants’ mean length of pause might have decreased, and although mean length of pause was not a measure originally included in the study, it seemed worthwhile to look at it because it allows for additional comparison with previous research. A paired t-test shows that participants’ mean length of pause did decrease significantly, from 1.09 seconds pre-SA to 0.96 seconds post-SA (p = .004). This finding agrees with Valls Ferrer (2008) but is not consistent with the results of two other studies, Lennon (1990) and Towell (2002), which did not find significant changes. However, the difference between the significant changes found in the present study and in Valls Ferrer (2008) as compared to the lack of significant changes found in Lennon (1990) and
Towell (2002) may be more a matter of sample size than actual differences in results. The present study included 39 participants who completed the posttest, and Valls Ferrer’s (2008) study had 30. In Lennon’s (1990) study, the mean pause length at T-unit boundaries was 1.73 seconds pre-SA and 1.23 seconds post-SA, but with only four participants, this change was not statistically significant. Similarly, in Towell’s (2002) study, the mean pause length was 1.03 seconds pre-SA and 0.93 seconds post-SA, but with 12 participants, this change was not statistically significant. Although it is not possible to draw a firm conclusion without additional research, the findings of the present study add to the evidence suggesting that mean length of pause may decrease during study abroad.

As with pausing, previous study abroad research has not led to consistent findings regarding changes in frequency of repair. Repair includes measures such as the frequency of repetitions of words or syllables, the frequency of grammatical and lexical self-corrections, and the frequency of false starts (abandoning an idea without completing it); these measures are thought to indicate breakdowns in fluency. Lennon (1990) measured the number of repetitions per T-unit, the number of self-corrections per T-unit, and the percent of repeated and self-corrected words; he found no significant changes for any of these measures. Valls Ferrer (2008) and D’Amico (2010) both found significant changes, but in opposite directions. Valls Ferrer (2008) measured the number of repairs per minute and found a significant decrease from pre- to post-SA. D’Amico (2010) also measured the number of repairs per minute and found a significant increase from pre- to post-SA. The present study

44 The only pauses for which Lennon (1990) measured mean length
found no significant changes, but looking at the low pre-SA fluency and high pre-SA fluency groups separately, there is a trend toward increasing repair for the low group (4.23 repairs per 100 syllables pre-SA; 4.51 post-SA) and decreasing repair (2.87 repairs per 100 syllables pre-SA; 2.35 post-SA) for the high group.

Valls Ferrer (2008), who found a decrease in repair, had participants who were advanced L2 learners in a translation and interpreting program, whereas D’Amico (2010), who found an increase in repair, had participants were enrolled in an intermediate-level course in the L2. Looking at the data from these two studies together with the present study suggests, as an initial hypothesis, that frequency of repair may develop in a non-linear manner and changes may therefore differ for participants at different stages of their L2 learning. Testing this hypothesis would require further research with L2 learners at different levels.

The present study also examined the changes in each of these aspects of fluency (articulation rate, mean length of run, pausing, and repair) in terms of L1–L2 difference scores. Looking at the data in this manner does not change the results in any way, but it gives a better sense of participants’ pre-SA L2 fluency as well as the size of the change from pre- to post-SA by providing a meaningful point of comparison, participants’ L1 fluency. In this way, the L1 data serve as baseline measures, as Segalowitz (2010) recommends. For example, for the rates of mid-clause and end-of-clause pauses, looking at the L1 data shows that participants’ rate of end-of clauses pauses was already similar in the L1 and L2 before studying abroad (7.53 pauses per 100 syllables in the L1 and 7.90 in the L2; difference score = 0.20), whereas their rate of mid-clause pauses was much greater in the L2 than the
L1 (3.57 pauses per 100 syllables in the L1 and 8.94 in the L2; difference score = 1.66). Although their rate of mid-clause pauses in the L2 significantly decreased to 7.03 pauses per 100 syllables on the posttest, comparison with the L1 data shows that mid-clause pauses are an aspect of L2 fluency that still differed substantially from the L1 even after study abroad (difference score = 1.07). In contrast, looking at another dimension of fluency, there was only a small difference between participants’ L1 and L2 articulation rate post-SA (4.32 syllables per second in the L1 and 4.08 in the L2; difference score = 0.31). The ability to make such comparisons is useful for interpreting the L2 measures.

**Question 2: Pre-SA fluency and changes in fluency**

2. Is there a relationship between pre-SA L2 fluency and changes in L2 fluency?
   
a. Is there a correlation between pre-SA L2 fluency scores and change in L2 fluency scores from pre- to post-SA?

b. Is there a significant difference between the amount of change in L2 fluency scores for participants with low pre-SA L2 fluency scores (below the median score) and the amount of change in L2 fluency scores for participants with high pre-SA L2 fluency scores (above the median score)?

c. If there are differences in the amount of change in L2 fluency scores for participants with low and high pre-SA L2 fluency scores, in what areas of fluency (articulation rate, pausing, repair, or length of run) are the differences most evident?

There are negative correlations between pre-SA L2 fluency scores and
changes in L2 fluency from pre- to post-SA, indicating that participants who began the semester with lower L2 fluency generally experienced greater gains in fluency than participants who began the semester with higher L2 fluency. These results agree with Towell’s (2002) finding that participants who exhibited lower fluency at the beginning of the study generally had the largest gains in fluency. These findings may be at least partly due to ceiling effects for high-fluency participants; beyond a certain point, it would not be natural to continue to increase one’s rate of speech or decrease one’s pausing.

The overall amount of change in L2 fluency from pre- to post-SA was significantly greater for participants who began the semester with low L2 fluency than for participants who began the semester with high L2 fluency. Looking at each dimension of fluency separately, differences in amount of change were not significant for articulation rate, mean length of run, or repair rate. For repair rate, the change from pre- to post-SA was not significant for the group of participants as a whole. For articulation rate and mean length of run, the change from pre- to post-SA was significant for the group of participants as a whole. Particularly in the case of mean length of run, a measure on which the high pre-SA fluency group actually increased slightly more than the low pre-SA fluency group (an increase of 1.96 syllables for the high group and 1.65 syllables for the low group)—though the difference in amount of change is not statistically significant—this finding suggests that it may be a dimension of fluency in which growth continues even among those who have already reached relatively high levels of fluency.

The differences in the amount of change for low and high pre-SA fluency
groups were significant for several measures of pausing, with the low pre-SA fluency group showing greater gains than the high pre-SA fluency group. These measures include percent pausing time and the rates of all pauses, mid-clause pauses, long pauses, unfilled pauses, and filled pauses—all of the measures of pausing on which the participants as a whole significantly improved from pre- to post-SA. The differences between groups in the amount of change were not significant for the rate of end-of-clause pauses or short pauses, but there were no significant changes for the participants as a whole on these measures. Together, these findings clearly show that participants who began the semester with low L2 fluency reduced their pausing significantly more than participants who began the semester with high L2 fluency. However, even with the low group reducing their pausing more than the high group, there were still significant differences between the two groups at the end of the semester on every measure of pausing except for short pauses.

**Question 3: Fluency and grammar, vocabulary, and processing speed**

3. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and pre- and post-SA L2 fluency scores?

   a. Is there a correlation between participants’ pre-SA grammar scores and their pre-SA fluency scores? Is there a correlation between their pre-SA grammar scores and their post-SA fluency scores? Is one correlation stronger than the other?

   b. Is there a correlation between participants’ pre-SA vocabulary scores and their pre-SA fluency scores? Is there a correlation between their
pre-SA vocabulary scores and their post-SA fluency scores? Is one correlation stronger than the other?

c. Is there a correlation between participants’ pre-SA processing scores and their pre-SA fluency scores? Is there a correlation between their pre-SA processing scores and their post-SA fluency scores? Is one correlation stronger than the other?

d. Considering as possible factors pre-SA grammar, vocabulary, picture naming, and sentence matching scores, what regression model best predicts pre-SA L2 fluency?

There were significant correlations between participants’ pre-SA grammar, vocabulary, and processing scores and many measures of fluency, both pre- and post-SA. As noted in chapter 2, a small amount of previous research has found that knowledge of L2 grammar and vocabulary (De Jong et al., 2013a; Hilton, 2008) and language processing speed (De Jong et al., 2013a) correlate with L2 fluency. The results of the present study generally support these findings. Table 21 compares findings across the three studies for areas in which similar measures were used. The correlations presented for the previous study include only the pretest data since the other two studies used data collected at a single point in time.
Table 21: Comparison across studies of the relationship between fluency and grammar, vocabulary, and processing measures

<table>
<thead>
<tr>
<th>Rate of speech47</th>
<th>Hilton (2008)</th>
<th>.679*</th>
<th>.581*</th>
<th>N/A</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Jong et al. (2013a)</td>
<td>.470*</td>
<td>.580*</td>
<td>.660*</td>
<td>.320*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L2 only</td>
<td>.543*</td>
<td>.486*</td>
<td>.528*</td>
<td>.551*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L1 control</td>
<td>.545*</td>
<td>381*</td>
<td>.553*</td>
<td>.522*</td>
<td>N/A</td>
</tr>
<tr>
<td>Percent pausing time</td>
<td>Hilton (2008)</td>
<td>.593*</td>
<td>.551*</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>De Jong et al. (2013a)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L2 only</td>
<td>.496*</td>
<td>.417*</td>
<td>.486*</td>
<td>.496*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L1 control</td>
<td>.481*</td>
<td>618*</td>
<td>.451*</td>
<td>.641*</td>
<td>N/A</td>
</tr>
<tr>
<td>Rate of all pauses</td>
<td>Hilton (2008)</td>
<td>.728*</td>
<td>.661*</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>De Jong et al. (2013a)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L2 only</td>
<td>.572*</td>
<td>.502*</td>
<td>.660*</td>
<td>.530*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L1 control</td>
<td>.519*</td>
<td>.550*</td>
<td>.646*</td>
<td>.563*</td>
<td>N/A</td>
</tr>
<tr>
<td>Rate of unfilled pauses</td>
<td>Hilton (2008)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>De Jong et al. (2013a)</td>
<td>.300*</td>
<td>.390*</td>
<td>.380*</td>
<td>.200*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L2 only</td>
<td>.633*</td>
<td>.537*</td>
<td>.617*</td>
<td>.566*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L1 control</td>
<td>.621*</td>
<td>.670*</td>
<td>.644*</td>
<td>.658*</td>
<td>N/A</td>
</tr>
<tr>
<td>Rate of filled pauses</td>
<td>Hilton (2008)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>De Jong et al. (2013a)</td>
<td>.200*</td>
<td>.330*</td>
<td>.400*</td>
<td>.320*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L2 only</td>
<td>.398*</td>
<td>.354*</td>
<td>.403*</td>
<td>.417*</td>
<td>N/A</td>
</tr>
<tr>
<td>Present study, L1 control</td>
<td>.527*</td>
<td>.541*</td>
<td>.499*</td>
<td>.555*</td>
<td>N/A</td>
</tr>
</tbody>
</table>

45 For the sake of clarity, all correlations representing a positive relationship between greater fluency and higher scores on the other measures are expressed as positive correlations.

46 To measure grammatical processing speed, De Jong et al. (2013a) used a productive measure of sentence building speed, whereas the present study used a receptive measure of how quickly participants could identify whether or not a sentence and a picture matched.

47 Words per minute in Hilton (2008), mean syllable duration in De Jong et al. (2013a), and syllables per second (with pauses removed) in the present study.
<table>
<thead>
<tr>
<th>Rate of repair</th>
<th>Hilton (2008)</th>
<th>De Jong et al. (2013a)</th>
<th>Present study, L2 only</th>
<th>Present study, L1 control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.566*</td>
<td>.516*</td>
<td>.566*</td>
</tr>
<tr>
<td></td>
<td>Hilton (2008)</td>
<td>.330*</td>
<td>.430*</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>De Jong et al. (2013a)</td>
<td>.30*</td>
<td>.240*</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Present study, L2 only</td>
<td>.058</td>
<td>.299</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.06</td>
<td>.065</td>
<td>.300</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.084</td>
<td>.126</td>
<td>.188</td>
</tr>
<tr>
<td></td>
<td>Present study, L1 control</td>
<td>-.026</td>
<td>.230</td>
<td>-.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.168</td>
<td>.203</td>
<td>.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.04</td>
<td>.101</td>
<td>.158</td>
</tr>
<tr>
<td>Mean length of run</td>
<td>Hilton (2008)</td>
<td>.733*</td>
<td>.668*</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>De Jong et al. (2013a)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Present study, L2 only</td>
<td>.559*</td>
<td>.461*</td>
<td>.552*</td>
</tr>
<tr>
<td></td>
<td>Present study, L1 control</td>
<td>.538*</td>
<td>.557*</td>
<td>.564*</td>
</tr>
</tbody>
</table>

*p < .05*

As Table 21 shows, the results are somewhat similar, though certainly not identical, across the three studies. For rate of speech, although there is some variation in the strength of the correlations (which could be partly due to the differences in the way in which the rate of speech was measured), there is agreement across the three studies that there is a relationship with grammar and vocabulary. The results are also consistent between De Jong et al. (2013a) and the

---

48 Hilton’s (2008) category, called “retracing,” included the rate of pauses, repetitions, reformulations, and false starts; De Jong et al. (2013a) measured corrections and repetitions separately; and the present study included repetitions, replacements, reformulations, and false starts. For the sake of better comparison with De Jong et al. (2013a), this table also shows correlations with repair and with corrections (replacements + reformulations) separately.

49 The top row shows the correlation with corrections; the bottom row shows the correlation with repetitions.

50 The top row shows the correlation with corrections (replacements + reformulations), and the second row shows the correlation with repetitions. The bottom row shows the overall correlation with all categories of repair.

51 No distinction is made between p < .05 and p < .01 in this table because De Jong et al. (2013a) did not make this distinction in reporting their results.
present study in showing a relationship with processing measures; Hilton (2008) did not measure processing speed. For percent pausing time, the rate of all pauses, and mean length of run, there is agreement between Hilton (2008) and the present study that there is a relationship with grammar and vocabulary; De Jong et al. (2013a) did not include these measures of fluency.

Looking at the rate of unfilled pauses and the rate of filled pauses, the results of both De Jong et al. (2013a) and the present study show a relationship with grammar, vocabulary, and processing measures. Especially for unfilled pauses, the correlations tend to be stronger in the present study, and the difference is statistically significant for the correlations with grammar and picture naming. However, it is difficult to hypothesize about the reason for the differences, given that a variety of factors, such as differences in the tasks and differences in participants’ backgrounds, could have played a role.

For repair, Hilton’s (2008) correlations appear to be stronger than the other studies; however, she included pauses along with repetitions, replacements, and false starts, so it is not very accurate to compare her results with the two other studies. Comparing De Jong et al. (2013a) with the present study, De Jong et al. found significant correlations between repair and grammar, vocabulary, and processing measures, whereas the present study did not. It should be noted that lower correlations were statistically significant in De Jong et al.’s study than in the present study given their much larger number of participants (179 compared to 39); in some cases, the correlations are fairly similar between the two studies. Nevertheless, given that De Jong et al. found significant correlations between certain
measures for which the present study found essentially no relationship, including grammar and corrections as well as grammatical processing speed and corrections, it seems that further research is needed to clarify this question.

In addition to looking at the relationship between grammar, vocabulary, and processing scores and fluency at one point in time (pre-SA), the present study looked at the relationship between pre-SA grammar, vocabulary, and processing scores and post-SA fluency. Although the relationship was still significant for a number of measures of fluency, the correlations were generally lower than they were for pre-SA grammar, vocabulary, and processing scores and pre-SA fluency. This finding suggests that the relationship between each of these areas of knowledge or skill and fluency is immediate, not delayed. In other words, there is a stronger relationship between current linguistic knowledge and processing speed and current fluency than there is between current linguistic knowledge and processing speed and the future trajectory of learners’ fluency. However, it is also possible that the pre-SA correlations were stronger in part because participants’ L2 fluency was more diverse at the beginning of the semester than at the end; for most measures of L2 fluency, the standard deviations were larger pre-SA than post-SA. The fact that participants’ L2 fluency was more diverse at the beginning of the semester than at the end indicates that participants who began the semester with low L2 fluency made some progress toward catching up to their peers who began the semester with higher L2 fluency; nevertheless, as discussed in relation to Research Question 6, there were still significant differences at the end of the
semester between those who began with low fluency and those who began with high fluency.

**Question 4: Grammar, vocabulary, and processing and changes in fluency**

4. Is there a relationship between pre-SA grammar, vocabulary, and/or processing scores and changes in fluency?
   a. Is there a correlation between participants’ pre-SA grammar scores and change in fluency scores from pre- to post-SA?
   b. Is there a correlation between participants’ pre-SA vocabulary scores and change in fluency scores from pre- to post-SA?
   c. Is there a correlation between participants’ pre-SA processing scores and change in fluency scores from pre- to post-SA?
   d. What is the relative strength of each of these correlations?

The results show negative correlations between pre-SA grammar, vocabulary, and processing scores and several measures of changes in L2 fluency. Since pre-SA grammar, vocabulary, and processing scores are positively correlated with pre-SA L2 fluency scores (Research Question 3), and pre-SA L2 fluency scores are negatively correlated with changes in L2 fluency (Research Question 2), this finding is not surprising.

When the same analysis is conducted while controlling for pre-SA L2 fluency, there are no significant correlations. One might expect that with a control for pre-SA L2 fluency, the correlations would be positive—that is to say, with pre-SA L2 fluency scores held equal, participants who began the semester with more grammar and vocabulary knowledge and the ability to process the L2 more quickly would tend to
have greater gains in L2 fluency. The fact that the results do not support this hypothesis suggests that although there is a relationship between participants’ grammar, vocabulary, and processing scores at a given point in time and their L2 fluency at the same point in time (Research Question 3), there is not an association between grammar, vocabulary, or processing scores and future gains in L2 fluency. In practical terms, this finding suggests that L2 learners with higher levels of grammar and vocabulary knowledge prior to study abroad do not necessarily have an advantage over other learners in the area of fluency gains during study abroad. This finding is consistent with Segalowitz et al.’s (2004) finding that there was no significant correlation between pretest grammar knowledge and language gains over the course of a semester among either at home or study abroad students. However, other previous research on changes in overall speaking proficiency during study abroad has found that pre-SA grammar and vocabulary knowledge are positively correlated with gains in proficiency (Brecht et al., 1995; Golonka, 2006). Proficiency and fluency are different constructs, and it may be that pre-SA grammar and vocabulary knowledge are more closely related to the development of proficiency than fluency since reaching higher levels of proficiency (at least as it is measured by the ACTFL scale) requires a certain degree of accuracy. Nevertheless, additional research may be helpful in clarifying the relationship between pre-SA grammar and vocabulary knowledge and changes in both fluency and proficiency during study abroad. The section on L2 fluency and models of speech production

52 Given that Segalowitz et al. (2004) measured speaking fluency as well as overall speaking proficiency, it is assumed that “language gains” refer to both fluency and proficiency, but the article does not explicitly define “language gains.”
later in this chapter provides further discussion of the distinction between fluency and proficiency.

It is also important to note that all of the participants in the present study had previously studied Spanish for at least two years and had developed some knowledge of grammar and vocabulary. None of the participants were complete beginners, and it is possible that the results would have differed if the study had included learners who began the semester with no knowledge of Spanish. It appears, though, that some participants’ knowledge was quite limited, as demonstrated, for example, by their inability to correct basic verb conjugation and adjective agreement errors on the pre-SA grammar test. It is evident, then, that the lack of relationship between pre-SA grammar and vocabulary knowledge and gains in fluency in the present study cannot be attributed to all participants having highly developed knowledge of the L2 prior to study abroad.

**Question 5: Changes in grammar, vocabulary, and processing and changes in fluency**

5. Is there a relationship between changes in grammar, vocabulary, and/or processing scores and changes in L2 fluency?
   
a. Is there a correlation between change in grammar scores and change in fluency scores from pre- to post-SA?
   
b. Is there a correlation between change in vocabulary scores and change in fluency scores from pre- to post-SA?
   
c. Is there a correlation between change in processing scores and change in L2 fluency scores from pre- to post-SA?
   
d. What is the relative strength of each of these correlations?
Only four significant correlations between changes in fluency and changes in other measures were found: decreases in the rates of all pauses, mid-clause pauses, and unfilled pauses were significantly correlated with an increase in vocabulary scores; and an increase in articulation rate was significantly correlated with a decrease in picture naming times. However, even these correlations were fairly low (the strongest was \( r = .418 \), for mid-clause pauses and vocabulary).

In light of the many significant correlations between fluency and other measures at one point in time that the present study as well as previous studies (De Jong et al., 2013a; Hilton, 2008) have found, it is somewhat surprising that there were not more significant correlations between changes in fluency and changes in other measures. There are a couple of possible explanations for these results.

First, in the case of changes in grammar and vocabulary, the lack of significant correlations with changes in fluency may be at least partly a measurement issue. Since the maximum score on each of these measures was 30 points, the range of change was fairly limited: -4 to +9 on the grammar measure and -1 to +9 on the vocabulary measure. This scale may have been too small to clearly show a relationship between changes on these measures and changes in fluency.

Second, the duration of the study (three months) may have been too short a period of time to observe a relationship between changes in grammar, vocabulary, and processing measures and changes in fluency. Perhaps over a longer period of time—presumably with the opportunity for larger changes to take place in grammar and vocabulary knowledge and processing speed as well as in fluency—a relationship between these changes would become evident. Future research that
focuses on students who study abroad for more than one semester, or research that examines fluency development over a longer period of time in a different context, could test this hypothesis.

**Question 6: Predicting changes in fluency**

6. Considering as possible factors L1 fluency, pre-SA L2 fluency, changes in grammar from pre- to post-SA, changes in vocabulary from pre- to post-SA, changes in processing speed from pre- to post-SA, and time spent interacting with native speakers of the L2 during the semester abroad, what regression model best predicts changes in L2 fluency during a semester abroad?

The best predictor of change in L2 fluency is pre-SA L2 fluency; the higher the L2 fluency before study abroad, the less change during study abroad.

Nonetheless, pre-SA L2 fluency predicts only 24% of the variation in the amount of change in L2 fluency during study abroad, suggesting that there are other factors for which this model does not account. The next best predictor is L1 fluency; however, including it in the model does not change the model’s predictive power a statistically significant amount.

These findings suggest, then, that the L2 learners who are likely to experience the greatest gains in fluency during study abroad are those who have the lowest L2 fluency prior to study abroad, which is likely due in part to ceiling effects for learners who already have relatively high L2 fluency prior to study abroad.

However, it is important to point out that in spite of the greater gains in fluency among the low pre-SA fluency group as compared to the high pre-SA fluency group in the present study, the low pre-SA group still had significantly lower fluency than
the high pre-SA group at the end of the semester, as Table 22 shows. Taken together, these two findings—that the low pre-SA group had greater gains, but the high pre-SA group still significantly outperformed the low pre-SA group at the end of the semester—should help provide hopeful yet realistic expectations for L2 learners at a variety of stages in the L2 learning process who desire to improve their fluency through study abroad. For learners with relatively low L2 fluency going abroad for a semester, it is likely that they will experience noticeable gains in fluency, yet a semester abroad will not magically make them as fluent as learners who have already attained a significantly higher level of fluency. For learners with relatively high L2 fluency, it is quite possible that they will still experience gains in fluency during a semester abroad, but they should not expect these gains to be as dramatic as for learners who begin their time abroad with lower fluency. Additionally, these findings indicate that students should be encouraged to take advantage of opportunities to develop fluency both at home and abroad, and not assume that study abroad alone is the key to fluency development.

Table 22: Mean pre- and post-SA L2 fluency T-scores

<table>
<thead>
<tr>
<th></th>
<th>Low pre-SA fluency group</th>
<th>High pre-SA fluency group</th>
<th>Significance (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean pre-SA fluency T-score</td>
<td>42.29</td>
<td>57.69</td>
<td>.000**</td>
</tr>
<tr>
<td>Mean post-SA fluency T-score</td>
<td>43.32</td>
<td>56.82</td>
<td>.000**</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01
Regarding the observation that other factors (changes in grammar, vocabulary, and processing speed and time spent interacting with native speakers of the L2) did not predict changes in fluency, the lack of a significant relationship between changes in grammar, vocabulary, and processing speed and changes in fluency has been discussed in relation to Research Question 5. As such, only the last factor, time spent interacting with native speakers of the L2 during study abroad, will be discussed in this section.

Most (Freed et al., 2004b; Perez-Vidal & Juan-Garau, 2011; Valls Ferrer, 2008) but not all (Segalowitz et al., 2004) previous research that has looked at gains in fluency and time spent using the L2 outside of class has found a positive relationship between the two. One possible reason for the difference in findings between the present study and most previous research is that previous research has asked participants about all use of the L2 outside of class, whereas the present study focused only on oral interactions with native speakers of the L2. Interestingly, Freed et al. (2004b), whose participants included at-home and intensive immersion students in addition to study abroad students, found that gains in fluency were more closely related to the amount of time spent reading and writing in the L2 outside of class than to the amount of time spent speaking and listening to the L2 outside of class. Of course, the results of a single study do not prove that reading and writing contribute more to the development of oral fluency than do speaking and listening (and Freed et al. do not make such a claim). However, in light of Freed et al.’s findings, it is possible that participants in the present study who did not report spending large periods of time interacting with native speakers of the L2...
nevertheless used the L2 in other ways that contributed to the development of their fluency.

Another potential explanation for the lack of a significant relationship between changes in fluency and time interacting with native speakers of the L2 is that it is difficult to meaningfully quantify interactions that may have been qualitatively different. For example, a number of participants reported spending several hours per week hanging out with native speakers of Spanish in a group setting. Without detailed descriptions of these interactions from each participant, it is hard to know what such interactions involved. Some participants may have spent much of this time actively engaged in conversation in the L2, and others may not. Therefore, the numbers alone might obscure important differences between participants’ experiences.

Finally, it is likely that participants varied in the accuracy of their estimates of time spent interacting with native speakers of the L2. As indicated in chapter 3, it was assumed that asking participants to fill out the questionnaire multiple times during the semester would provide better estimates than asking them to reflect on their whole time abroad only at the end of the semester, as some previous studies have done. Nevertheless, obtaining precise data about how learners spend their time during study abroad would require tracking them much more closely throughout their stay abroad, which is difficult to do.

**L2 fluency and models of speech production**

The previous sections have discussed the results for each research question related to the development of fluency separately. This section considers the
implications of the findings more generally. Returning to the models of L1 and L2 speech production discussed in chapter 2 provides a useful framework in which to consider what the findings of the present study suggest about L2 fluency. As noted in chapter 2, Levelt’s (1989) L1 model is the most influential model of speech production, and L2 models (De Bot, 1992; Kormos, 2006) are adaptations of this model. Levelt describes speech production in three stages: conceptualizing, formulating, and articulating. Breakdowns in fluency may occur at any of these stages.

**Conceptualizing**

Conceptualizing refers to selecting and ordering the information that one wants to convey (Levelt, 1989). In participants’ L1 speech samples, it appears that the longest pauses may have been related to difficulties at this stage. Sometimes, participants made a couple of comments in response to the prompt and then seemed to be unsure what else to say about the topic. For example, about 13 seconds into his description of his routine on a typical weekday, Participant 34 said:

> After class, (0.65) I come home, (0.30) I eat lunch, (0.93), study, (0.41) and (2.64) usually spend some time with my friends.

It is quite likely that some pauses occurred for the same reason in participants’ L2 speech samples. Since conceptualizing involves forming the ideas but not the structure of what one wants to say, the frequency of difficulties at the conceptualizing stage is presumably similar for a given speaker in both the L1 and the L2, provided that the speaker is speaking about similar topics in the L1 and the

---

53 The length of pauses (in seconds) is indicated in parentheses.
L2. In light of the fact that nearly all participants exhibited substantially lower fluency in their L2 than in their L1, it seems, then, that they must have had more breakdowns in fluency at the formulating and/or articulating stages in the L2 than in the L1.

**Formulating**

The formulating stage involves giving linguistic structure to the message conceived in the previous stage. As De Bot (1992) discusses in relation to L2 speech production, difficulties may occur at this stage due to an inability to retrieve needed lexical items or because lexical items are retrieved without the syntactic information required to assemble them. In either case, it may be that the speaker does not have the necessary knowledge (of lexical items or syntax), or it may be that he or she does have the knowledge but cannot retrieve it quickly enough to produce fluent speech.

The present study suggests that all of these factors may contribute to L2 dysfluency at the formulating stage. The study found a relationship between untimed measures of grammar and vocabulary and fluency as well as a relationship between timed measures of lexical and syntactic processing and fluency; there are moderate correlations between scores on the grammar, vocabulary, sentence matching, and picture naming tasks and most measures of fluency (Research Question 3). However, it is primarily the timed measures of processing speed that

---

54 As noted in chapter 3, the similarity of the L1 and L2 tasks in the present study was verified by having native speakers of English complete all of the tasks in English and native speakers of Spanish complete all of the tasks in Spanish, and then calculating fluency measures for these speech samples.
best predict fluency (Research Question 3d). This finding suggests that L2 fluency is more closely connected to how quickly a speaker can use the L2 knowledge that he or she has than to how much L2 knowledge he or she has. Of course, this assertion assumes that the speaker does have some level of L2 knowledge; otherwise, it would be impossible to produce speech in the L2. However, it may be that once one has a certain basic knowledge of the language, fluency is more closely related to the automatization of what one knows than to acquiring new L2 knowledge.

Here, it again seems relevant to note the distinction between fluency and proficiency. Since gains in proficiency require the ability to talk about a wider range of topics, with reasonably accurate syntax and increasingly precise vocabulary (Swender et al., 2012), gains in proficiency almost certainly do require gains in L2 grammar and vocabulary knowledge. Since fluency (at least as defined in the present study) measures only temporal aspects of speech, gains in fluency do not inherently require gains in grammar and vocabulary knowledge.

However, the fact that gains in grammar and vocabulary knowledge are not inherently necessary for gains in fluency does not mean that they are entirely unrelated to gains in fluency. Although the present study does not directly test this idea, a preliminary hypothesis to propose is that the relationship between gains in grammar and vocabulary knowledge (as well as gains in utilizing this knowledge more quickly) and gains in fluency may depend on individual differences in learners’ awareness of and concern for accuracy. A speaker with low attention to accuracy might become quite fluent (though not highly proficient) despite relatively limited L2 grammar and vocabulary knowledge. In cases where he or she lacks the
necessary lexical or syntactic knowledge to express an idea, or perhaps has the knowledge but cannot access it quickly, he or she produces lexically or grammatically inaccurate speech with little or no hesitation. For example, during the pretest task in which participants were asked to narrate a story in the past based on a set of pictures (see Appendix A), Participant 27 said:

\[Y \text{ quiere comer (0.32) el pan. (0.39) } Y \text{ por eso, (0.62) cuando ellos fui a la a su escuela, (0.51) y la madre di dijo adios, (0.77) y ellos tienen el pan...}\]

‘And he\(^{55}\) wants [present tense] to eat (0.32) the bread. (0.39) And therefore, (0.52) when they went [first person singular verb form] to the to their school, (0.51) and the mother sa- said goodbye, (0.77) and they have [present tense] the bread...’

During the pretest task in which participants were asked to discuss the advantages and disadvantages of working part time while in college, Participant 1 said:

\[Trabajando \text{ es mejor para personas que (1.13) que son muy responsabilidad (0.74) y (0.63) necesitan danero para comprar...}\]

‘Working [wrong verb form] is better for [missing article] people who (1.13) who are very responsibility [noun instead of adjective] (0.74) and (0.63) need money [mispronounced] to buy...’

For both of these participants, their fluency scores were well above what would be expected based on their grammar and vocabulary scores.\(^{56}\)

A speaker with higher awareness of or concern for accuracy, in contrast, may abandon an idea or attempt to formulate it differently if he or she has inadequate grammar or vocabulary knowledge for the idea that he or she is trying to express.

\(^{55}\) He refers to a dog; \(they\) refers to two children.

\(^{56}\) Participant 27’s grammar and vocabulary scores were right at the mean (39 out of 60), but her fluency T-score was 57.09 (0.709 SD above the mean). Participant 1 had the lowest grammar and vocabulary scores (22 out of 60) of all participants, but her fluency T-score (47.78) was only slightly below the mean.
If he or she has the necessary linguistic knowledge but cannot produce it quickly, he or she may take the time to search for the item. For example, narrating a story in the past based on a set of pictures, Participant 15 said:

\[
Y \text{ ayudaron a su madre a preparar la (2.50) la (3.36)^57 cómo se llama (3.60) sábana.}
\]

‘And they helped their mother to prepare the (2.50) the (3.36) what’s it called (3.60) sheet.’

In the same task, Participant 7 said:

\[
Un \text{ día, los los chicos (0.99) quieren (1.25) quisier- (0.82) los chicos (3.11) quisieron (1.55) para almorzar (1.23) en el parque.}
\]

‘One day, the the children (0.99) want to (1.25) want- (0.82) the children (3.11) wanted to (1.55) in order to [extra word] eat lunch (1.23) in the park.’

For individuals displaying a high concern for accuracy, reaching high levels of fluency would seem to require greater lexical and syntactic knowledge, and greater ability to utilize that knowledge quickly, since they are less inclined to sacrifice accuracy for fluency.

At this point, a couple of caveats are in order. First, this discussion is certainly not intended to imply that speakers can be neatly divided in categories of either caring greatly about accuracy or not caring about it at all; it is probably more helpful to think of attention to accuracy as existing on a continuum. Second, anecdotal evidence from the data is only a starting point and is of course not sufficient to confirm or disconfirm the hypothesis. However, previous research does suggest a trade-off effect between fluency and accuracy (Skehan, 2009), and future

\[^57\text{Pauses longer than 3.00 seconds represent a combination of filled and unfilled pauses, as unfilled pauses were limited to 3.00 seconds each, as noted in chapter 3.}\]
research could further explore the issue by operationalizing attention to accuracy and/or using stimulated recalls after a speaking task to gain more insight into why learners hesitate (or do not hesitate) at certain points while speaking. This research may need to take into account factors such as learners’ goals in learning the L2 (i.e. do they just want to develop conversational skills, or do they want to learn to speak the language at a professional level?) and the effects of the setting (for example, an academic setting might prompt some students to strive for greater accuracy than they would in a social setting).

It should also be noted that it seems likely that there is a relationship between speaking tasks and the level of grammar knowledge and, perhaps especially, vocabulary knowledge required for fluent formulating. In the present study, one of the tasks in each language (talking about personal activities) would be classified as an Intermediate-level task on the ACTFL scale. Two of the tasks in each language would be classified as Advanced-level tasks (explaining the advantages and disadvantages of two options, and narrating a story in past time). Although the Advanced-level tasks in the L2 appeared to be challenging for many participants, none of the tasks required specialized or technical vocabulary, and they could all be completed using grammar structures typically introduced—though certainly not mastered—in first-year Spanish courses. If the study had included Superior-level tasks, such as developing an argument in favor of a particular viewpoint on a topic unrelated to students’ personal lives, grammar and/or vocabulary knowledge might have played a stronger role in predicting fluency. On the other hand, it is also possible that at least some students would have simplified the tasks and still
exhibited fluency comparable to that which they demonstrated on the Intermediate- and Advanced-level tasks.

**Articulating**

Articulation is the physical production of the message that has been conceptualized and formulated (Levlt, 1989). The present study did not include a separate measure of articulation as it did for knowledge and skills important at the formulating stage (measures of grammar, vocabulary, and processing speed). However, there is still evidence to suggest that breakdowns at the level of articulation played a role in L2 dysfluency for the participants in the study. First, articulation rate (the number of syllables spoken per second with pauses ≥0.25 seconds removed) in the speaking tasks was slower in the L2 than in the L1 and increased in the L2 from pre- to post-SA. Nevertheless, it is possible that in some cases, participants articulated more slowly in the L2 than in the L1 not because of actual difficulties at the level of articulation, but as a stalling strategy due to difficulties at the formulating stage. For example, participants sometime elongated the final (or only) syllable of conjunctions such as porque (*because*), *y* (*and*), *o* (*or*), and *pero* (*but*). These syllables were not counted differently from other syllables since they are part of words, and it is difficult to determine precisely at what point a syllable is long enough that it functions as a filler. Analyzing the data in this way, though, makes it less clear whether an increase in articulation rate is truly related to improved articulation skills in the L2. Therefore, this measure alone does not fully demonstrate that changes from pre- to post-SA were due to gains at the level of articulation.
However, comparing data from the present study with data from a task that the same participants completed as part of a separate study may provide further evidence of changes at the level of articulation from pre- to post-SA. In the present study, the picture-naming task that participants completed required them to both retrieve and articulate a word. As part of a separate study, they also completed a word-matching task in which they heard a word and then saw a picture, and they were asked to indicate as quickly as possible whether or not the word and the picture matched by pressing one of two buttons on the keyboard. In this task, they did not have to articulate the word. Although the two tasks did not use the same words, both used common, high-frequency words. Looking at the pretest data, there is not a significant relationship between participants’ scores on the two tasks (r = .298; p = .098). For the posttest data, in contrast, there is a significant relationship between their scores on the two tasks (r = .558; p = .001). As Figure 9 shows, on the pretest, some participants had relatively slow response times on the picture-naming task (the task involving articulation) despite having relatively quick response times on the picture-matching task (the task not involving articulation). On the posttest, this pattern was less pronounced. This finding might suggest that improved articulation abilities among some participants resulted in a stronger relationship between the two tasks on the posttest.
Figure 9: Picture matching and picture naming response times, pre- and post-SA
Fluency and study abroad

To summarize, the study found that participants’ L2 speech was less fluent than their L1 speech on similar topics in similar contexts, which suggests that more breakdowns in fluency occurred at the formulating and articulating stages in the L2 than in the L1. The data in the present study also suggest that fluency at the formulating stage—at least for the task types and levels of participants included in the study—is more closely connected to the pace at which learners can make use of their L2 knowledge than to the quantity of their L2 knowledge, although there is also a relationship between fluency and quantity of L2 knowledge.

If automaticity at each stage of speech production is necessary to facilitate fluent speech, a claim that Levelt (1989) supports, study abroad seems to be a potentially ideal (though certainly not the only) context for such automaticity to develop, given the opportunity for immersion in the language and daily practice. Even routine transactions require learners to repeatedly retrieve L2 words and phrases from memory—albeit within a fairly limited range of vocabulary—and articulate them. More in-depth interactions push learners to create with the language, requiring them to both retrieve lexical items and assemble them in a comprehensible way, perhaps speeding up their ability to execute skills such as conjugating verbs in context. Although the results of the present study do not show

---

58 Of course, students make use of this opportunity to differing degrees; for example, as participants reported on the Language Contact Questionnaire, the amount of time that they spent interacting with native speakers of the L2 outside of class ranged from just over one hour per week to more than 25 hours per week.

59 See Appendix G for examples from the Language Contact Questionnaire of how students described the benefits of different types of interactions.
a direct relationship between the amount of time spent interacting with native speakers of the L2 and fluency development, and provide only weak evidence of a relationship between gains in linguistic knowledge and processing speed and fluency development, it may be that there are too many variables involved (both those considered in the study as well as others not considered, such as participants’ personality) to isolate any one of these variables and see a clear relationship to fluency development, especially in a timespan of only three months.

This section has considered the implications of findings related to L2 linguistic knowledge, language processing speed, and fluency development. The follow section discusses the results of findings related to the measurement of fluency.

**The measurement of fluency**

The study compared the pauses of learners with low and high levels of lexical–grammatical competence in order to address questions about which pauses to measure in L2 fluency research. Results indicate that the low and high groups significantly differed from one another in both long and short pauses, filled and unfilled pauses, and mid-clause and end-of-clause pauses.

The finding that the two groups differed from one another even in short pauses (0.25 to 0.49 seconds) suggests that it is reasonable to count pauses of this length—along with longer pauses—in studies of L2 fluency, as there seems to be a connection between their frequency and learners’ level of L2 development. Although pauses less than 0.5 seconds are relatively common even in L1 speech (Goldman-Eisler, 1961), an abundance of such pauses may still indicate dysfluency.
This conclusion contrasts with that of Riggenbach (1991), who found no significant differences between low and high groups in the frequency of pauses less than 0.5 seconds. However, the lack of significant differences in Riggenbach’s study may be due to the low number of participants (3 per group).

The finding that the low and high groups in the present study significantly differed from one another in the rate of filled as well as unfilled pauses suggests that both types of pauses should be considered as measures of fluency. However, the fact that the two groups were very similar to one another in the percent of pauses that were filled (24.72% for the low group; 22.92% for the high group) indicates that perhaps there is little to be gained from counting the two types of pauses separately as opposed to counting them together. Previous studies (Cenoz, 2000; Iwashita et al., 2008; Kormos & Denes, 2004) have found varying trends in the percent of filled pauses for low- and high-proficiency learners, but no study has reported statistically significant differences between groups. The data in the present study suggest that using a higher or lower percent of filled pauses may be more a matter of personal speaking style than fluency. Dividing participants into groups based on their use of filled pauses in English, the group with a high rate of filled pauses in English used a significantly higher percent of filled pauses in Spanish (29.35%) compared with the group that had a low rate of filled pauses in English (18.29%), as Table 23 shows.
Table 23: Percent filled pauses in Spanish based on rate of filled pauses in English

<table>
<thead>
<tr>
<th>Percent of filled pauses in Spanish</th>
<th>Mean (SD) for group with high rate of filled pauses in English; N=24</th>
<th>Mean (SD) for group with low rate of filled pauses in English; N=24</th>
<th>Significance (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.35% (8.10)</td>
<td>18.29% (9.17)</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

The finding that the low and high groups in the present study significantly differed from one another in the rate of mid-clause as well as end-of-clause pauses suggests that pauses in both locations should be considered as measures of fluency. Given that the difference in rate was more pronounced for mid-clause pauses (13.91 versus 6.89) than for end-of-clause pauses (8.95 versus 7.30), and the low group had a significantly higher percent of mid-clause pauses (58.10%) than the high group (44.93%), it is evident that these pauses pattern differently across levels. Therefore, it may be valuable to distinguish between them in counts of pauses, as both the data and intuition suggest that mid-clause pauses indicate greater dysfluency than do end-of-clause pauses.

**Limitations and directions for future research**

As with all research studies, the present study has certain limitations, which are discussed in this section. Considering these limitations also leads to ideas about how to improve and expand on this research in future studies.

One limitation is the number and selection of participants. Although 39

---

60 These groups had nearly equal numbers of participants classified as having low and high levels of lexical-grammatical competence in Spanish (13 low and 11 high for the group with a low rate of filled pauses in English; 11 low and 13 high for the group with a high rate of filled pauses in English).
participants is a larger group than most similar studies have had, repeating the study with even more participants may lead to more robust findings. Having a larger number of participants would also make it possible to divide participants into more than two groups based on their pre-SA L2 fluency, which would allow for a more detailed look at how fluency changes for learners with different levels of pre-SA fluency. Additionally, nothing was known about participants’ level of Spanish until they took the pretest. Given the opportunity, it would be ideal to select participants more carefully to ensure a balanced number of students with beginning, intermediate, and advanced abilities in Spanish at the beginning of the semester.

The attempt to keep the length of the pretest and posttest sessions to 75 minutes or less also introduced other limitations. Keeping the sessions to this length was done to avoid discouraging students from participating\textsuperscript{61} and to make sure that the researcher had time to meet with an adequate number of participants during her time in Buenos Aires. However, longer sessions would have allowed for more extensive grammar and vocabulary testing as well as the inclusion of more speaking tasks, all of which may have been beneficial. For example, as previously mentioned, the grammar and vocabulary measures included only 30 items each, and the range of change in scores from pretest to posttest was only -4 to +9 on the grammar measure and -1 to +9 on the grammar measure. This narrow range of change, created in part by the use of short grammar and vocabulary tests, may have contributed to the lack of significant correlations between changes in grammar and

\textsuperscript{61} One study abroad director advised keeping the sessions to an hour or less based on previous experiences in which students were reluctant to participate in studies requiring longer sessions.
vocabulary scores and changes in fluency scores. Using more extensive grammar and vocabulary tests in a similar study would provide a better opportunity to examine the relationship between changes in grammar and vocabulary knowledge and changes in speaking fluency.

Additionally, the vocabulary measure used in the present study was chosen based on its use in previous research as well as the lack of suitable alternatives. However, it would be ideal to create and pilot a Spanish vocabulary test specifically for the purpose of the study, as De Jong et al. (2013a) did in Dutch. The test used in the present study measured receptive vocabulary knowledge, whereas De Jong et al.’s test measured productive vocabulary knowledge, which is preferable in a study that focuses on speaking.

Another methodological consideration is that an individual’s verbal fluency may vary for non-linguistic reasons such as caffeine consumption (Franks et al., 1975; Wesensten et al., 2005) and sleep, or lack thereof (Curcio et al., 2006; Harrison & Horne, 1997). Although the present study’s findings for the group as a whole offer strong evidence that participants’ L2 fluency did significantly increase from pre- to post-SA, it is possible that some individuals’ gains were under- or overestimated due to such variations, which may have affected the findings regarding the relationship between changes in fluency and changes in other measures. To minimize the possibility of such factors affecting the results, it would be best to test participants on multiple days at the beginning and end of the semester. Of course, requiring participants to be present for testing on more days would increase the difficulty of collecting a complete set of data for all participants.
Alternatively, it might be useful to at least ask participants (on the day of the pretest and the day of the posttest) about recent caffeine consumption and sleep patterns in case such information helps explain any unexpected results.

Another factor that should perhaps be taken into account in future research is the quality of task completion for the speaking tasks. In the researcher’s judgment, participants in the present study generally fulfilled the speaking tasks adequately. Nevertheless, in a study of this type, it would be possible for participants to achieve high scores on quantitative fluency measures without speaking coherently, in which case their fluency may not be particularly meaningful. Again, this concern relates to the distinction between fluency and proficiency. The fact that one participant’s response is more detailed or well developed than another participant’s response may lead to differing conclusions about their proficiency, but it does not directly say anything about which participant is more fluent. However, to give an extreme example, if a participant spoke about a topic completely unrelated to the prompt, one could hardly say that he or she gave a fluent response to the prompt, regardless of what the quantitative measures of fluency indicated. Therefore, it seems reasonable to consider some standard of adequate task completion even when the focus of the study is fluency rather than proficiency.

The previous suggestions relate to improving the methodology in future research; there are also ways in which the topic could be fruitfully expanded in future research. In a future study, it would be valuable to consider including other measures of speaking in addition to measures of fluency. Previous research unrelated to study abroad suggests a trade-off between fluency, accuracy, and
complexity (Skehan, 2009). Although examining changes in fluency as well as accuracy and/or complexity has been done in a small amount of previous study abroad research (Llanes & Muñoz, 2009; Perez-Vidal & Juan-Garau, 2011), this research either has not looked at fluency in depth (Perez-Vidal & Juan-Garau, 2011\textsuperscript{62}) or has looked at only two of these dimensions rather than all three (Llanes & Muñoz, 2009).

From the perspective of examining changes in L2 fluency, it would be valuable to measure changes in accuracy and complexity as well, particularly to see if students who do not experience gains—or actually decrease—in fluency during study abroad experience gains in accuracy, complexity, or both. It may be, for example, that a participant increases in knowledge of grammar but (at least temporarily) experiences a decrease in fluency because he or she is focused on accurately applying the grammar that he or she has learned. In this way, considering accuracy and complexity may also help clarify the relationship between fluency and measures of grammar and vocabulary knowledge and processing speed.

To address the questions about measuring pauses, it would be ideal to divide participants into groups based on the results of proficiency testing, such as Oral Proficiency Interviews (OPIs) or Simulated Oral Proficiency Interviews (SOPIs), rather than based on their scores on written grammar and vocabulary tests. Additionally, with a larger number of participants, it would be valuable to divide participants into more than two groups for comparison.

\textsuperscript{62} This study included only one measure of fluency, words per minute.
Conclusion

This study was designed with two main purposes in mind:

1. To explore the development of L2 fluency during a semester abroad and its relationship to the development of grammar, vocabulary, and processing speed, while also considering the influence of individual participants’ L1 and pre-SA L2 fluency

2. To examine issues in the measurement of fluency, focusing on questions related to measuring pauses in L2 speech

In relation to the first purpose of the study, the results show that participants experienced significant gains on most measures of fluency during study abroad. This finding was especially true for participants who began their time abroad with low L2 fluency; they experienced greater gains on several measures of pausing than did participants who began their time abroad with high L2 fluency. Nevertheless, students who began the semester abroad with high L2 fluency still had significantly higher fluency at the end of the semester than students who began with low L2 fluency.

Looking at the relationship between L2 fluency and L2 linguistic knowledge (vocabulary and grammar scores) and language processing speed (picture-naming and sentence-matching scores), the study found a moderate relationship between pretest measures of L2 fluency and pretest measures of linguistic knowledge and processing speed. However, the results show no relationship between pre-SA
linguistic knowledge and processing speed and gains in L2 fluency, and little relationship between gains in linguistic knowledge and processing speed and gains in L2 fluency. The best predictor of gains in L2 fluency was pre-SA L2 fluency.

These results provide encouraging evidence of students’ gains in L2 fluency during study abroad. They also suggest that although there is a relationship between L2 linguistic knowledge and L2 fluency, having more advanced L2 linguistic knowledge prior to study abroad does not necessarily give students an advantage in the area of fluency development during study abroad. Additionally, the finding that students who began the semester with high L2 fluency still demonstrated significantly greater fluency at the end of the semester than students who began with low L2 fluency—despite the greater gains of the students who began with low L2 fluency—indicates that students should be encouraged to take advantage of opportunities to develop fluency both at home and abroad, and not assume that study abroad alone is the key to fluency development.

Regarding the questions related to the measurement of fluency, the data show that learners with low and high levels of lexical–grammatical competence significantly differed from one another on all of the measures of rates of pauses (short and long pauses, filled and unfilled pauses, and mid-clause and end-of-clause pauses) as well as in the percent of pauses occurring in the middle of a clause. However, they did not significantly differ from one another in the percent of filled pauses. The findings suggest that measuring all of these pauses may be useful in

---

63 Controlling for pre-SA L2 fluency; without pre-SA L2 fluency as a control variable, there is a negative relationship between pre-SA linguistic knowledge and gains in L2 fluency.
examining L2 fluency. However, there is perhaps little or nothing to be gained from counting filled and unfilled pauses separately, as speakers’ tendency to use more of one or the other appears to be more closely related to personal speaking style than to L2 ability.
Appendix A: Pictures for narration tasks

Both sets of pictures are from Heaton (1966).

L1 (English) task
L2 (Spanish) task
Appendix B: Grammar test

Each sentence below contains a grammatical error. Your task is to find the error and write the corrected form in the box below the sentence.

Model: Adriana están feliz.

| Estás |

1. Los niños fue a la escuela juntos.

2. Lamento que José no ha vendido hoy.

3. Se le mandé a las diez de la mañana por correo electrónico.

4. Quiero que tú vaya ahora.

5. Las fresas están muy buenos.


7. Conocía a Juan ayer en la fiesta.

8. Me gusta los libros que traes.

9. Ellos sabía que ibas a venir.
10. Te recomendé que no sales con él.


12. Vicente fue al mercado cada fin de semana cuando era niño.

13. El médico quien me trató ya no trabaja en el mismo hospital.

14. Esta caja es por su colección de libros.

15. Yo supe que a mis amigos les gustaría ir al teatro.

16. Alguien le robaba su bolsa a la señora.

17. ¡No comes este chocolate antes de cenar!


19. El contrato se firmaron los nuevos dueños de la casa.

20. Espero encontrar un profesor que me puede ayudar con mis tareas.
21. Anoche iba a la tienda.

22. Está cerrado la ventana.

23. Me gustaría que vienes conmigo.

24. Estaba buena la fiesta anoche.

25. Este niña tiene que regresar con sus papás.

26. Estaban ya las diez cuando Rosario llegó.

27. Por ser exitoso, necesitas trabajar fuerte.

28. Mis amigos y yo pedíamos la cuenta al mesero.

29. Se pueden entrar por esa puerta.

30. Me duele las piernas muchísimo.
### Appendix C: Samples from sentence-picture verification task

<table>
<thead>
<tr>
<th>Matching pictures and sentences</th>
<th>Picture</th>
<th>Sentence</th>
</tr>
</thead>
</table>
| Silvia is washing the dishes.   | Lava los platos.  
    |         | *She washes the dishes.* |
| Nora, Bea, and Elisa are reading a book. | Leen un libro.  
    |         | *They read a book.* |
| Juana is serving drinks to Ana and Manuel. | Juana les sirve bebidas.  
    |         | *Juana serves them drinks.* |
| Lorena is calling Felipe.       | Lorena lo llama.  
    |         | *Lorena calls him.* |
| Víctor is brushing his teeth.   | Se lava los dientes.  
    |         | *He brushes his teeth.* |

<table>
<thead>
<tr>
<th>Mismatched pictures and sentences</th>
<th>Picture</th>
<th>Sentence</th>
</tr>
</thead>
</table>
| Jessica is giving a gift to Maite. | Jessica les da un regalo.  
    |         | *Jessica gives them a gift.* |
| Samuel is talking to Marisa and Lola. | Samuel te habla.  
    |         | *Samuel is talking to you.* |
| Andrea is playing in the park.    | Jugamos en el parque.  
    |         | *We play in the park.* |
| Nina is hugging Micah.            | Nina los abraza.  
    |         | *Nina hugs them.* |
| Adela is putting Rosa to bed.      | Adela se acuesta.  
    |         | *Adela goes to bed.* |
Appendix D: Background questionnaire

1. Gender ____________________

2. Age _____________

3. Native language ________________________________

4. Experience with Spanish
   
a) How old were you when you began learning Spanish? ______________
   
b) How many years of formal study of Spanish have you completed?
      
      Elementary school  0  1  2  3  4  5  6+
      Middle/high school 0  1  2  3  4  5  6+
      College            0  1  2  3  4  5  6+

   c) Have you previously lived or studied in a Spanish-speaking country? If so, please describe when, where, and for how long.
      
   d) Do you have any other experience studying or using Spanish outside of school? If so, please describe.

5. Have you studied or do you speak any languages other than English and Spanish? If so, please list the language(s) and describe your experience with the language(s).
**Appendix E: Language contact questionnaire**

Please answer these questions, thinking about experiences in the past two weeks in which you interacted with others in Spanish. Enter N/A for any questions that are not applicable to your experience.

<table>
<thead>
<tr>
<th>Question</th>
<th>1a. How many routine transactions in public places (such as ordering food in a restaurant or buying something at a store) did you have?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b. How many of these routine transactions included some type of complicating factor (such as needing to return an item or resolve a problem)?</td>
<td>2. If you are staying with a host family or sharing a room/apartment with a native speaker of Spanish, how many brief or superficial exchanges (greetings, “Please pass the salt,” “I’m leaving for school,” etc.) did you have with them?</td>
</tr>
<tr>
<td>3. If you are staying with a host family or sharing a room/apartment with a native speaker of Spanish, how many times did you have a short conversation (10 minutes or less) with them about everyday topics (such as what you did that day)?</td>
<td>4. If you are staying with a host family or sharing a room/apartment with a native speaker of Spanish, how many times did you have a longer or more in-depth conversation with them?</td>
</tr>
<tr>
<td>4b. About how long did each of these conversations last?</td>
<td>5. Did you participate in an intercambio (“intercambio” refers to meeting with a native speaker of Spanish specifically for the purpose of language practice)?</td>
</tr>
<tr>
<td>5b. If so, how many times, and for how long each time?</td>
<td>6. Not counting interaction with your host family or intercambio, how many times did you engage in friendly chitchat with native speakers of Spanish? (Friendly chitchat refers to an interaction that goes beyond a formulaic exchange of greetings, but is not an in-depth conversation)</td>
</tr>
<tr>
<td>7a. Not counting interaction with your host family or intercambio, how many times did you hang out with one or more native speakers of Spanish in a group setting?</td>
<td>17. Not counting interaction with your host family or intercambio, how many times did you hang out with a native speaker of Spanish one on one?</td>
</tr>
<tr>
<td>7b. About how long each time?</td>
<td>8b. About how long each time?</td>
</tr>
</tbody>
</table>

162
9. If you had other interactions with native speakers of Spanish not included in the categories above, please describe them here.

10. Out of the interactions that you've recorded in questions 1-9, which do you think provided the most meaningful language practice? Why?

11. Please highlight or underline the phrase below that most accurately describes your experience: I have spent _____________ time interacting with others in Spanish in the past two weeks compared with the previous two-week period.

   much less  slightly less  about the same amount of  slightly more  much more

12. Please highlight or underline the phrase below that most accurately describes your experience: I have spent _____________ time interacting with others in Spanish in the past two weeks compared with peers in my study abroad program.

   much less  slightly less  about the same amount of  slightly more  much more
Appendix F: Examples of mid-clause and end-of-clause pauses

Mid-clause pauses, English

And you'll also finish your undergraduate [pause] sooner.

I'll have [pause] one or two classes during the day.

There was a boy biking on [pause] a narrow road.

They might know what they want to [pause] to get their degree in.

You get to get into the workforce a lot younger and maybe amass [pause] savings quickly.

Mid-clause pauses, Spanish


Son diferentes de [pause] su [pause] vida normal. = They are different from [pause] your [pause] normal life.

Cuando llegaron al parque encontraron [pause] el perro. = When they arrived at the park they found [pause] the dog.

End-of-clause pauses, English

Taking time off helps them prepare for that [pause] but [pause] I think the advantage is that...

You’re accustomed to that kind of schedule [pause] having classes and then doing work and not a whole lot else.

If I have time for breakfast [pause] I’ll usually make toast.

And then in the evenings [pause] sometimes I will have orchestra practice.

There’s more continuity between semesters [pause] and sometimes it’s harder to go back to school after having had a longer break.

End-of-clause pauses, Spanish

Es mucho mejor trabajar en vez de no trabajar porque [pause] en la vida siempre va a tener conflictos de tiempo. = It’s much better to work instead of not working because [pause] in life you’re always going to have time conflicts.
Los sábados a veces trabajo otra vez [pause] y si no [pause] hago mi tarea. = On Saturdays sometimes I work again [pause] and if not [pause] I do my homework.

Mientras estaban caminando hacia el parque [pause] el perro comió todo el picnic. = While they were walking toward the park [pause] the dog ate the whole picnic.


La única desventaja de que yo puedo pensar es que [pause] gasta mucho tiempo en su trabajo. = The only disadvantage that I can think of is that [pause] you spend a lot of time at your job.
Appendix G: Examples of language contact questionnaire codes

Question 9: If you had other interactions with native speakers of Spanish not included in the categories above, please describe them here.

<table>
<thead>
<tr>
<th>Code</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel/tourism</td>
<td>I used Spanish for traveling purposes when I went out of town, at the bus station and other stops.</td>
</tr>
<tr>
<td></td>
<td>Last week I traveled to Patagonia. With travel, guides, hostel and finding things within the city I would say the amount of Spanish I used greatly increased for that week.</td>
</tr>
<tr>
<td>Class/tutoring</td>
<td>I am now taking classes with students from Argentina that I attend twice a week for four hours each time.</td>
</tr>
<tr>
<td></td>
<td>I met with my tutors for my direct enrollment classes twice for each class (four classes total, each two hours long).</td>
</tr>
<tr>
<td>Non-routine transaction</td>
<td>Trying to explain my visa to airport security was very difficult, not only because of Spanish but because of bureaucratic difficulties.</td>
</tr>
<tr>
<td></td>
<td>I bought a violin, which required extensive contact with the store owner over several days trying out different ones.</td>
</tr>
<tr>
<td>Texting or online</td>
<td>Chatting on Tinder and Badoo.</td>
</tr>
<tr>
<td>communication</td>
<td>Texting with native speakers.</td>
</tr>
<tr>
<td>Time with extended family</td>
<td>When I visited my host mom’s family for Easter. We talked in Spanish.</td>
</tr>
<tr>
<td>or friends of host</td>
<td>I went to a birthday party with my family, and was able to have conversations with many members of their family during the party. None really in depth, but chatting nonetheless.</td>
</tr>
<tr>
<td>Academic communication</td>
<td>I talk frequently with [my program’s] staff about academic matters.</td>
</tr>
<tr>
<td>outside of class</td>
<td>5-10 minutes conversations before or after each class about differences in requirements, etc. for foreign students and explaining my educational/career plans.</td>
</tr>
<tr>
<td>Work/internship</td>
<td>I had a meeting last week with my boss for an internship.</td>
</tr>
<tr>
<td></td>
<td>During my internship, sometimes I will have conversations about work-related topics in Spanish.</td>
</tr>
<tr>
<td>Sports</td>
<td>Played in a soccer tournament (which didn’t involve much actual conversation, but is for a more extended period of time).</td>
</tr>
<tr>
<td></td>
<td>Playing soccer.</td>
</tr>
</tbody>
</table>
Question 10: Out of the interactions that you've recorded in questions 1-9, which do you think provided the most meaningful language practice? Why?

**Group size**

<table>
<thead>
<tr>
<th>Code</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-on-one</td>
<td>With my host mom because she corrects my Spanish and we only speak in Spanish.</td>
</tr>
<tr>
<td></td>
<td>One on one time with a native Spanish speaker. Unimpeled practice of the language.</td>
</tr>
<tr>
<td>Larger group</td>
<td>The extensive hours in a group setting, because not only was this group particularly keen to correct me, but I would listen to them speak among themselves and got exposure to expressions that a one-on-one interaction wouldn’t necessarily expose me to.</td>
</tr>
<tr>
<td></td>
<td>Dinner parties where there are about 5-10 people all speaking in Spanish, because I can hear different expressions and opinions.</td>
</tr>
<tr>
<td>Unspecified</td>
<td>At bars with locals because we usually have good conversations.</td>
</tr>
<tr>
<td></td>
<td>Use of Spanish during traveling, because the information we discussed was important, relating to where I needed to be or how to organize an aspect of the trip.</td>
</tr>
<tr>
<td>Type of relationship</td>
<td>Code</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>Host family</td>
</tr>
<tr>
<td></td>
<td>Peers/friends</td>
</tr>
<tr>
<td></td>
<td>Strangers</td>
</tr>
<tr>
<td></td>
<td>Intercambios</td>
</tr>
<tr>
<td></td>
<td>Service encounter/ work setting</td>
</tr>
<tr>
<td></td>
<td>Romantic relationship</td>
</tr>
<tr>
<td></td>
<td>Unspecified</td>
</tr>
</tbody>
</table>
## Why the interaction provided meaningful language practice

<table>
<thead>
<tr>
<th>Code</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning vocabulary/ phrases</td>
<td>Hanging out with native speakers in a group setting. I get to learn many words and a lot more slang this way.</td>
</tr>
<tr>
<td></td>
<td>Dinner conversation with my host mom because I learn a lot of new vocabulary and talk about daily activities.</td>
</tr>
<tr>
<td>Error correction</td>
<td>One on one time with native Spanish speaker because we talked a lot and the speaker corrected me very openly in a positive way.</td>
</tr>
<tr>
<td></td>
<td>Conversations with my host mom because she corrects me.</td>
</tr>
<tr>
<td>Forced to speak Spanish</td>
<td>Group or one on one interactions with native Spanish speakers because then I don’t have the option of reverting to English if I don’t feel comfortable in Spanish, so it forces me to practice.</td>
</tr>
<tr>
<td></td>
<td>Probably one of the times I was hanging out with my group of porteño friends because they don’t speak English, which forces me to keep up a lot better than when speaking with staff or others – then I can fall back on certain English words if I’m not sure how to say something.</td>
</tr>
<tr>
<td>Topic (interesting and/or challenging)</td>
<td>A chat I had with my coordinators on our trip was good because we talked about technology and I had to translate pretty complicated ideas.</td>
</tr>
<tr>
<td></td>
<td>The long in-depth conversations with my host family, because I had to be able to talk about abstract concepts in Spanish that I couldn’t really explain very well without knowing the exact Spanish translation of the word.</td>
</tr>
<tr>
<td>Comfortable setting</td>
<td>Conversations in social settings with native speakers, because there isn’t much pressure and it is more comfortable.</td>
</tr>
<tr>
<td></td>
<td>Long conversations with the host family because they occur during dinner and provide a comfortable setting to explain more complex thoughts.</td>
</tr>
<tr>
<td>Extended time</td>
<td>Talking at the dinner table with my host mom because it is a long period of time and we talk about real life topics.</td>
</tr>
<tr>
<td></td>
<td>Any of the activities that required me to speak in Spanish for long periods of time – class, relaxed hanging out, etc. The day I had to use my Spanish all day I also thought in Spanish vs. the days that were just short interactions.</td>
</tr>
</tbody>
</table>
### Why the interaction provided meaningful language practice: Continued

| Listening practice | Group of native Spanish speakers, because I can hear a variety of speaking styles and I can participate in the conversation as well, making it valuable for listening and speaking practice. Spending time with native speakers in a casual group setting because when it is only myself and one other person, I don’t necessarily get the ear-training, whereas in a small group, I still speak a great deal, but I also listen to how native speakers interact, their expressions, etc. |
| More focused interaction | One on one conversations with a native speaker. It was much easier to converse directly than to try to keep up with a group conversation, and more personal. I got more speaking time, and we got to engage in more in-depth conversation. One on one because it allows for each person to focus on what the other is saying with minimal distractions/interruptions. |
| Variety | Speaking Spanish while traveling was the most helpful because it diversified my Spanish use. Hanging out with a group of friends was the most useful to me. We talked about a lot of different things, and they are friendly and easy to talk to. |
| Pronunciation/intonation | Hanging out with native speakers in a group setting, I feel that I pick up on the slang, pronunciation, and intonation better in these situations. Talking one on one with a native speaker of Spanish has also proven helpful because I like to listen to their tonality in order to better form my words. |
| Quick pace | Group setting, because I had to keep up with them at their pace. Group setting practice. More fast-paced language and wider span of vocabulary used. |
| Practicing routinized language | In these past two weeks, I felt that the shorter interactions were more meaningful language practice. Although I only spoke briefly about mundane things, I think that I am beginning to get a hold on how to say routine phrases in non-awkward ways. Ordering food, now I’m gaining more confidence to ask for specific things. |
References


