Individual Differences in Context:
A Neurolinguistic Investigation of Working Memory and L2 Development

BY

MANDY FARETTA-STUTENBERG
B.A., University of Wisconsin-Madison, 2007
M.A., University of Illinois at Chicago, 2009

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Hispanic Studies
in the Graduate College of the
University of Illinois at Chicago, 2014

Chicago, Illinois

Defense Committee:

Kara Morgan-Short, Chair and Advisor
Luis López
Susanne Rott, Germanic Studies
Jessica Williams, Linguistics
Darren Tanner, University of Illinois - Urbana-Champaign

This thesis is dedicated to Dr. Thomas Mount, whose generous support of my academic pursuits allowed me to discover my passion for learning and follow it from Oconomowoc to Madison, Madrid, and Chicago. Your confidence in me has made all the difference.

A special kind of beauty exists which is born in language, of language, and for language.

- Gaston Bachelard

ACKNOWLEDGMENTS

The following people have supported, inspired, and encouraged me as I worked toward completing this dissertation and degree. I am delighted to have this opportunity to acknowledge their various contributions now, and to express my gratitude and appreciation.

To the participants of this study - thank you for your willingness to share your time with me, and to put on an electrode cap (twice!). This project is made possible by your participation.

To my incredible advisor Kara Morgan-Short - you have been a phenomenal mentor and role model since my first semester as an MA student. Thank you for your encouragement at every stage of this project and of my graduate career. You are a truly gracious and inspiring mentor – thank you for everything.

To the members of my dissertation committee, all of whom have been generous with their time and expertise. Luis López, my first professor and DGS at UIC, thank you for a warm welcome to the program and for always providing a challenging question to tackle. Susanne Rott, thank you for your encouraging words and thoughtful criticism. Jessica Williams, thank you for your early comments that greatly improved my methodology, and for your unwavering support and encouragement; you have been a great inspiration to me and your comments on my manuscript have made revisions much more enjoyable! Darren Tanner, thank you for sharing your passion and for helping me to continue to appreciate the magic of event-related potentials.

To members of the Cognition of Second Language Acquisition Laboratory, past and present – thank you for your perspective, and for offering critical feedback and encouragement. Your interest and ideas have inspired me and have improved the quality of my work. To the truly fantastic, dedicated undergraduate research assistants with whom I have had the privilege of working on this (and other) projects: Erin Kohnke - for your help with the seemingly unending task of data processing, as well as with participant testing; Erin O'Connor, Luz Orozco, Nabiha Quadri for your help with data collection, transcribing, double-, triple- and quadruple-scoring…Your assistance in this project has been indispensable!

Deserving of special mention are my friends, colleagues, and fellow COGSLA lab members Bernard Issa and Janire Zalbidea. Bernie - you've shared your time, your Excel tricks, your humor, your best recipes, and, most importantly, your friendship with me. I immensely enjoy our conversations (from SLA theories and research ideas to Chicago restaurants and new music) and time together (even hours and hours in the lab running *yet another* set of analyses). Thanks for your constant willingness to talk through ideas, double-check the numbers, or grab a cup of coffee; everything is better with a good friend. Janire - your motivation, critical thinking, and consistent interest and support have meant the world to me. Thank you for always making time to discuss this and other projects, and for your enthusiasm and friendship. Bonus points for Skyping in to my defense – you're amazing. I couldn't ask for a better support team, and I look forward to many more collaborations!

ACKNOWLEDGMENTS (continued)

To my UIC & graduate school family of colleagues, particularly Jason Sarkozi & Bicho Azevedo – for getting me through my first year in Chicago with a smile; Xabier Granja & Junice Acosta – por su amistad y apoyo desde el primer día; Bryan Koronkiewicz – for invigorating playlists & interdisciplinary chats. Mil gracias, 'pañeros.

To the School of Literature, Cultural Studies and Linguistics and Department of Hispanic and Italian Studies staff, Rocío García, Kaitlin Roeldofs, Montanee Wongchinsri, María Gonzalez, whose help allowed this project to run smoothly.

To my generous funding sources - UIC Provost's Award for Graduate Research, Language Learning Dissertation Grant, and Audrey Lumsden-Kouvel Dissertation Fellowship.

To Ann Kuta & Peter Felitti, my basement presentation audience, for providing playdates for Lia and bowl after bowl of delicious soup. Thank you for all of the time you spent proof-reading and copy editing, but even more, thank you for the enthusiasm you have consistently shown for my work.

To Michelle Johnston, for her love, encouragement, and well-timed delivery of chocolates.

To my dear friend Joanna Parrin, for her positivity, Panera dates, and steadfast support.

To Linda Newman, the first person who uttered "PhD" to me when I was considering the possibility of a Master's. You told me I could do it, and have been unfailingly kind and encouraging since we first met, gently challenging me to see things from a different perspective.

To the members of my academic family tree, especially Michael Ullman & Cristina Sanz, who have made me feel valued and have paved an extraordinary path to follow.

To Dr. Thomas Mount & the Oconomowoc Scholarship Fund Future Leader Award, for getting me started in academia and believing in me from day one.

Al Profesor Eugenio Bustos Gisbert, Universidad Complutense de Madrid, quien me puso en este camino: gracias por la inspiración y motivación.

To Kevin & Lia, for their love, support, and comedic relief: you make home my favorite place to be.

With much appreciation -

MFS

TABLE OF CONTENTS

CHAPTER PAGE

TABLE OF CONTENTS (continued)

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

LIST OF FIGURES (continued)

FIGURE PAGE

LIST OF FIGURES (continued)

FIGURE PAGE

LIST OF ABBREVIATIONS

d′	d-prime
EEG	electroencephalogram
ERPs	event-related potentials
fem	feminine
LCP	Language Contact Profile
L1	first language
L2	second language
masc	masculine
ms	milliseconds
RDI	Response Dominance Index
RMI	Response Magnitude Index
SAE	study abroad experience
SOA	stimulus onset asynchrony, measure of time between onsets of two stimuli in a sequence
SLA	second language acquisition as a field of research
WM	working memory
μV	microvolts

SUMMARY

This thesis examines the interplay between external and internal factors in second language acquisition by analyzing the role of individual differences in experiential (language contact) and cognitive (working memory) factors in linguistic development that takes place in traditional, at home classroom settings and during study abroad. The study aims to provide a multi-dimensional perspective on these relationships by assessing both behavioral and neural evidence.

Previous research has revealed that second language acquisition among adults is a complex and difficult process that results in a great deal of individual variability. Research that examines linguistic development in natural learning settings, such as university-level classrooms at home and during study abroad, has indicated that a single context may not effectively lead to linguistic development for all learners (e.g., Ellis & Larsen-Freeman, 2006) and researchers have called for a theory of second language acquisition that examines relationships between external and internal factors (e.g., Collentine & Freed, 2004; Sanz, 2005).

The present study examines changes in linguistic proficiency and processing among intermediate-level learners of Spanish as a second language who are studying the language in either an at home or study abroad context. The study employs a three-session, longitudinal design in which cognitive abilities are assessed prior to the semester of study, linguistic abilities (behavioral and processing) are assessed prior to and immediately following the semester of study, and language contact is reported each week during the semester of study. This design allows for (1) a measurement of changes in second language abilities that are isolated to the semester of study and (2) analysis of changes in proficiency and processing with regard to individual differences in cognitive abilities and language contact. Behavioral tasks include a grammaticality judgment task and communicative production task designed to assess learner knowledge and use of grammatical gender agreement on articles and adjectives. Additionally, participants completed a standard measure of overall proficiency, which allows for a more global examination of changes in linguistic abilities and also facilitates comparisons between the present learner group and previously studied learners. Event-related potentials are used to measure second language processing of grammatical gender agreement.

The principle findings of the study were (1) learners in both the At Home and Study Abroad groups evidenced behavioral gains and processing changes from pre- to post-semester, (2) language contact and working memory accounted for changes in second language behavior and processing, and (3) individual differences in language contact and working memory impacted linguistic development differently for At Home and Study Abroad learners.

Results of this study contribute data to context-based and neurocognitive approaches to second language acquisition research. They also provide preliminary answers to Collentine and Freed's (2004) call for theories of language acquisition and processing to take into consideration cognitive abilities and context of learning. Future research that utilizes a multidimensional approach informed by the fields of second language acquisition and cognitive neuroscience is likely to provide further insights into the relationships between external and internal factors in linguistic development and have significant implications for identifying the predictors of successful second language acquisition among adult learners.

1 INTRODUCTION

Adult second language (L2) acquisition occurs in highly variable contexts, and with varying degrees of success. Many adults acquire their L2 in classroom settings in their native language (L1) environment. Increasingly, however, adults are choosing to supplement traditional classroom study with immersion or study abroad experiences. The relationship between such experiences and L2 development has been studied extensively using behavioral assessments, but with an emphasis on fluency and oral abilities. Furthermore, the impact of natural context of learning on the neural mechanisms used to process L2 grammatical structures has been largely ignored in the literature. The role of individual differences in explaining variability in learning outcomes – as assessed by both behavioral and processing measures – constitutes another, largely open, question.

1.1 Summary of the Problem

Despite a large number of studies in the past five decades, research in the field of second language acquisition (SLA) has yet to suggest a particular methodology or context that most effectively facilitates acquisition for all learners (e.g., Collentine & Freed, 2004; Norris & Ortega, 2001). Indeed, recent approaches to the study of SLA suggest that a single methodology or context may *not* effectively lead to L2 development for all learners (e.g., Ellis, 1998; Ellis & Larsen-Freeman, 2006). In other words, the story of adult L2 acquisition is complex and there is not one ideal solution for all learners. Current calls in the literature stress the importance of taking an approach that examines relationships among a variety of factors, including learning context and individual differences, in order to advance our understanding of L2 development (Collentine & Freed, 2004; Ellis, 1998; Ellis & Larsen-Freeman, 2006; Lafford, 2007; Long, 1997; Robinson, 2001; Sanz, 2005). One promising way to address the dynamic relationships that may be relevant to adult L2 acquisition is to examine the interplay between external and internal factors (Sanz, 2005).

1.1.1 External Factors

The importance of *external factors* (that is, factors that exist outside of an individual, such as teaching methodology, instructional intervention, or training condition) in the L2 acquisition process is widely accepted (e.g., Norris & Ortega, 2001; Spada & Tomita, 2010). One external factor that is receiving increasing attention is context of learning, which often refers to whether L2 study takes place in “at home” or “study abroad” settings (Collentine & Freed, 2004). Researchers have addressed the effects of study abroad on various aspects of L2 development (e.g., Brecht, Davidson, & Ginsberg, 1995; Collentine, 2004; DeKeyser, 1991, 2010; Freed, 1995a; Freed, Segalowitz, & Dewey, 2004; Harley & Hart, 2002; Hernández, 2010; Lafford, 1995; Lapkin, Hart, & Swain, 1995; Segalowitz & Freed, 2004), but this body of research has produced conflicting results, such that the nature of linguistic gains in this context remains unclear (Rees & Klapper, 2008). Research that considers the role of context of learning in language development is fundamental to understanding adult L2 acquisition; in order to explain why a particular setting holds advantages for some learners but not others, however, studies that examine *interactions* between context of learning and learner-internal individual differences are essential (Segalowitz & Freed, 2004).

1.1.2 Internal Factors

The importance of *internal factors* (that is, factors that vary by learner, such as individual differences in cognitive abilities, learning style, or motivation) in the L2 acquisition process is also widely accepted; individual differences in various domains are posited to account for a large amount of variability in success among L2 learners. Indeed, a number of internal factors have been examined in relation to L2 success including for example age, experience and proficiency with second languages, motivation, anxiety, intelligence, aptitude, attention control, and learning ability within different memory systems (e.g., Bialystok & Frohlich, 1978; Carpenter, 2008; Carroll, 1958, 1962, 1981; DeKeyser, 2000, 2010; Dörnyei, 2005; Ehrman & Oxford, 1995; Harley & Hart, 1997, 2002; Horwitz, Horwitz, & Cope, 1986; Mackey, Adams, Stafford, & Winke, 2010; Miyake & Friedman, 1998; Robinson, 2002, 2003; Sanz, 2000; Segalowitz & Frenkiel-Fishman, 2005; Skehan, 1991). The internal factors investigated in this study are individual differences in (1) amount of L2 contact[1] and (2) working memory (WM) ability.

Language Contact

Increased contact with the L2 is often credited as holding the "key to success" in linguistic development during study abroad (Freed, 1995b). Studies assessing L2 development in a study abroad context have quantified L2 contact hours in order to test this claim (e.g., Freed, 1995b; Freed, Segalowitz, & Dewey, 2004; Freed, Dewey, Segalowitz, & Halter, 2004; Isabelli-Garcia, 2010; Segalowitz & Freed, 2004). Despite the use of a standardized questionnaire to assess L2 contact among learners (Language Contact Profile; Freed, Dewey et al., 2004), results from studies that directly examine the relationship between L2 contact and linguistic development are inconsistent (e.g., Freed, Segalowitz, & Dewey, 2004; Isabelli-Garcia, 2010).

[1] For the purposes of this report, language contact is discussed as an "internal factor" that varies by learner.

There is little research examining the role of L2 contact in linguistic development in at home contexts.

Working Memory

Another internal factor posited to play a role in adult L2 acquisition is a learner's ability to process and store information, that is, a learner's working memory (WM; e.g., McDonald, 2006; Williams, 2012). Several empirical studies provide evidence of a positive relationship between WM and L2 development and performance (Mackey et al., 2010; Sagarra & Herschensohn, 2010; Sunderman & Kroll, 2006; Tokowicz, Michael, & Kroll, 2004); although contrasting results have also been found (Juffs, 2004). A number of laboratory-based studies have addressed the relationship between WM and L2 development and performance; however, there has been a paucity of such research in more naturalistic learning contexts.

1.1.3 Brain-Based Measures

The methods used to assess L2 abilities a crucial component to answering questions of variability in SLA. A number of studies have explored the role of external and internal factors in L2 development using behavioral or performance-based measures (e.g., Bialystok & Frohlich, 1978; Brecht & Robinson, 1995; Carpenter, 2008; Ellis & Larsen-Freeman, 2009; Freed, 1995b; Freed, Segalowitz, & Dewey, 2004; Harley & Hart, 1997, 2002; Hernández, 2010; Lafford, 1995; Lapkin, Hart, & Swain, 1995; Larsen-Freeman, 2006; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; Robinson, 1997, 2001, 2002, 2005). The inclusion of brain-based processing measures, such as event-related potentials (ERPs) to explore these relationships has been far more limited (e.g., *External factors*: Morgan-Short, Sanz, Steinhauer,

& Ullman, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; *Internal factors:* Bond, Fiorentino, Gabriele, & Alemán Bañón, 2011; Tanner, Inoue, & Osterhout, 2014; *External and Internal factors*: Carpenter, 2008). ERPs are sensitive to individual differences between learners (e.g., Bond et al., 2011; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013) and have been shown to reveal effects not detected by behavioral measures alone (e.g., Gabriele, Fiorentino, & Alemán Bañón, 2013; McLaughlin, Osterhout, & Kim, 2004; Morgan-Short, et al., 2010; Morgan-Short, Finger, Grey, & Ullman, 2012; Morgan-Short, Steinhauer, et al., 2012; Tokowicz & MacWhinney, 2005). As such, the inclusion of this brain-based processing measure in addition to behavioral assessments is expected to provide rich data that can speak to the complex relationships between external and internal factors and L2 development.

1.2 Open Question and Present Study

The question of how external and internal factors influence adult L2 acquisition emerges independently from SLA research that adopts context-based and cognitive approaches. Within the context-based literature, researchers have called for an examination of "the dynamics of learner-context interactions" (Segalowitz & Freed, 2004); within the cognitive SLA literature, researchers have called for careful assessment of interactions between WM abilities and type of exposure to the L2 (Williams, 2012). The nature of learner-context interactions and the specific questions of whether L2 contact or WM serve as predictors of successful L2 development in different settings have scarcely been addressed.

The present study addresses these open questions by experimentally examining how individual differences in L2 contact hours and WM abilities relate to linguistic development that takes place in two commonly occurring L2 learning contexts: at home and study abroad.

Specifically, the study employs a short-term longitudinal design in order to evaluate changes in L2 performance and processing among learners in these contexts in order to provide an ecologically valid account of language learning that contributes to debates regarding the successful predictors of successful L2 acquisition and can also speak to practical, educational questions related to interactions between learner and context.

2 REVIEW OF LITERATURE

The present study synthesizes questions of interest to three primary areas of research: (1) external factors relevant to L2 acquisition, specifically naturalistic contexts of learning, (2) internal factors relevant to L2 acquisition, specifically L2 contact hours and WM abilities, and (3) online assessments of L2 processing, specifically electrophysiological measures (event-related potentials; ERPs). This chapter provides a review of literature in each of these areas, beginning with research that has examined the role of context of learning in L2 development (Section 2.1). Section 2.2 provides an overview of studies that have examined the role of the internal factors L2 contact and WM in L2 acquisition. This section also includes a review of literature that has addressed interactions between external and internal factors in L2 acquisition. An overview of ERPs in language research as well as a review of empirical studies of L2 processing and of the use of ERPs to investigate individual variability are provided in Section 2.3. Within each section, particular focus is given to studies related to morphosyntactic knowledge and processing, as the target structure of the present study – grammatical gender agreement – is a morphosyntactic structure (see Section 3.3 for details). This chapter concludes with a description of the motivation for the present study and its research questions (Section 2.4).

2.1 External Factors

The relationship between external factors and L2 development has been examined extensively in laboratory-based settings, with numerous studies addressing the relative effectiveness of various types of instruction and interventions on the acquisition of a wide variety of linguistic skills and structures (e.g., Doughty & Williams, 1998; Lightbrown & Spada, 1990; Morgan-Short et al., 2010; Morgan-Short, Finger et al., 2012; Morgan-Short, Steinhauer et al., 2012; Sanz & Morgan-Short, 2004). Meta-analyses of this set of literature have shown that learning can occur under implicit conditions, but that more explicit types of instruction lead to greater linguistic gains, particularly for some language skills and linguistic structures (e.g., Norris & Ortega, 2000, 2001; Spada & Tomita, 2010). However, this conclusion may be qualified by general limitations in the literature that may bias results towards explicit training conditions (Norris & Ortega, 2000, 2001). In fact, brain-based research has shown more native-like processing[2] as a result of implicit, but not explicit training (Morgan-Short et al., 2010, Morgan-Short, Finger et al., 2012; Morgan-Short, Steinhauer et al., 2012). In addition, these conclusions run contrary to popular opinion that immersion settings, which arguably constitute a more implicit learning context, are most beneficial to language acquisition (e.g., Carroll, 1967; Freed, 1995a, 1995b). Importantly, because explicit and implicit training conditions, and the various learning conditions examined in laboratory-based research, do not necessarily map directly onto real-world learning settings, it is largely unclear whether these conclusions extend to more naturalistic contexts, leaving the field with an important open question: how does context of learning influence L2 development?

Many adults begin the task of learning an L2 in a classroom context. A significant number of learners choose to complement traditional classroom experiences with immersion, or study abroad experiences in the L2 environment. According to the Institute of International Education's *Open Doors 2013 Report on International Educational Exchange*, more than 283,000 U.S. students studied abroad for academic credit in the 2012-2013 academic year. A second Institute of International Education report states that across the globe, more than 3.3 million students are studying in a country beyond their own (Macready & Tucker, 2011). These

[2] See Section 2.3 for more discussion of language processing signatures.

numbers reveal a growing trend for students to complement classroom L2 experiences with study abroad experiences, a trend that may be linked, in part, to a widely held assumption that immersion experience results in language acquisition for all learners (Freed, 1998). The external factor context of learning, then, is important on both a theoretical level to researchers interested in providing an explanatory account of SLA, and also to educators and administrators charged with developing programs and curricula that facilitate language acquisition (Collentine & Freed, 2004).

Empirical research has begun to address the call for a “broader, context-sensitive” theory of SLA (Long, 1997). Context-based work has examined the linguistic impact of study abroad experiences, in which learners (1) attend formal classes that require them to use the L2, and (2) are immersed in an L2 speech community, which presumably offers consistent opportunities for use of the target language in more natural, communicative settings, with native speakers of the L2 (Freed, 1995).

Empirical studies have addressed the role of context of learning in the development of a variety of linguistic skills, including oral and grammatical abilities (see Table I for selected review). In general, study results indicate that study abroad learners improve in measures of fluency and oral skills (Freed, 1995; Lafford, 1995, 2004; Lord, 2006; Magnan & Back, 2007; Segalowitz & Freed, 2004; Simões, 1996), and that these gains may be related to phonological working memory in both at home and study abroad contexts (O’Brien, Segalowitz, Collentine & Freed, 2006; O’Brien, Segalowitz, Freed & Collentine, 2007). Conflicting results are particularly prevalent in studies examining grammatical development and accuracy: there is some evidence that study abroad groups show gains in L2 accuracy and grammatical abilities (Isabelli, 2004, *cf*. Isabelli-García, 2010). Comparisons between at home and study abroad groups in terms of grammatical development have shown that gains among at home groups that are equal to or superior than study abroad groups (Collentine, 2004; DeKeyser, 1991; Isabelli-García, 2010), although the opposite has also been found (Howard, 2001, 2005; Isabelli & Nishida, 2005).

Despite the relevance to theoretical claims and educational and administrative decisions, relatively few studies have addressed the linguistic impact of study abroad in a carefully-controlled and in-depth manner (DeKeyser, 2010; Rees & Klapper, 2008). Among the limitations that have been noted by various researchers are (1) lack of control or comparison at home group, making interpretation of the role of study abroad difficult and precluding conclusions regarding the role of context (Cheng & Mojica, Diaz, 2006; Isabelli, 2004), (2) lack of an objective measure of initial proficiency (Collentine, 2004; DeKeyser, 1991; Howard, 2001, 2004; Isabelli & Nishida, 2005), and (3) failure to employ a systematic assessment of L2 contact to determine how study abroad experiences differ from each other and from the at home context (Cheng & Mojica, Diaz, 2006; Collentine, 2004; DeKeyser, 1991; Howard, 2001, 2004; Isabelli, 2004; Isabelli & Nishida, 2005). Moreover, limited previous research has considered interactions between context and learner-internal factors in order to gain a better understanding of the role of context in L2 development (Collentine & Freed, 2004; Segalowitz et al., 2004).

This empirical work reveals that L2 acquisition in a study abroad context, just as with learning that takes place in at home or laboratory settings, is complex. There is significant individual variation in linguistic development during study abroad: some learners show substantial gains during even short study abroad programs, whereas others experience minimal gains after a full semester abroad (Freed, 1998a; Stewart, 2010). It seems that a number of factors may affect which learners experience linguistic development during study abroad and which learners do not; the following sections consider possible learner-internal factors and characteristics that have been posited to play a role in this variability.

Table I

Overview of Select Context-Related SLA Research

Study	Language	Skill/Structure	Groups	Length of Program	Assessment	Initial Proficiency	Results
				Oral Abilities			
Díaz-Campos (2004)	Spanish	Pronunciation	SA (*N*=26) AH (*N*=20)	10 weeks	Reading task	Varied	No effect of context
Freed (1995)	French	Fluency	SA (*N*=15) AH (*N*=15)	1 semester	Pre/Post OPI	Varied	SA more fluent than AH
Freed, Segalowitz, & Dewey (2004)	French	Fluency	SA (*N*=8) AH (*N*=8) IM (*N*=12)	1 semester	Pre/Post oral interviews; Post LCP	2-4 years prior instruction	IM more fluent than SA and AH
Lafford (1995)	Spanish	Communicative Strategies (CS)	SA (*N*=28) AH (*N*=*13*)	1 semester	Post OPI (Mexico, AH); Pre/Mid/Post OPI (Spain)	No prior university-level study	SA groups better at using CS than AH
Lafford (2004)	Spanish	Communicative Strategies (CS)	SA (*N*=26) AH (*N*=20)	1 semester	Pre/Post OPI; Pre/Post LCP	2+ semesters of prior study	Greater reduction in CS use for SA than AH
Lord (2000)	Spanish	Pronunciation	SA (*N*=8)	8 weeks	Production task	Intermediate	Positive effect of previous instruction
Lord (2006)	Spanish	Fluency	SA (*N*=19)	6 weeks	Pre/Post Sentence repetition task	3rd year of study	Improved ability to reproduce more syllables

Note. *Assessments*: Pre denotes assessment administered prior to or at the beginning of semester of study; Post denotes assessment administered at the end of the semester; OPI = Oral Proficiency Interview; LCP = Language Contact Profile, GJT = Grammaticality Judgment Task.
Groups: SA = Study Abroad, AH = At Home, IM = Domestic Immersion.

Study	Language	Skill/Structure	Groups	Length of Program	Assessment	Initial Proficiency	Results
			Oral Abilities, continued				
Magnan & Back (2007)	French	Fluency	SA (*N*=24)	1 semester	Pre/Post OPI, Self-assessment, Post LCP	4-6+ semesters prior instruction	Maintained or improved OPI scores
O'Brien, Segalowitz, Collentine & Freed (2006)	Spanish	Fluency	SA (*N*=25) AH (*N*=18)	1 semester	Pre/Post OPI, PM Test	2 semesters of prior study	No effect of context; role of PM varied by item
O'Brien, Segalowitz, Freed & Collentine (2007)	Spanish	Fluency	SA (*N*=18) AH (*N*=25)	1 semester	Pre/Post OPI, Phonological Memory (PM) Test	2+ semesters of prior study	No effect of context; Higher PM, more fluency gains
Segalowitz & Freed (2004)	French	Fluency	SA (*N*=22) AH (*N*=18)	16 weeks	Pre/Post OPI	3rd semester of study	SA more fluent than AH
Simões (1996)	Spanish	Pronunciation	SA (*N*=5)	5 weeks	Production task	Intermediate-low and advanced	Improved pronunciation skills

Note. *Assessments*: Pre denotes assessment administered prior to or at the beginning of semester of study; Post denotes assessment administered at the end of the semester; OPI = Oral Proficiency Interview; LCP = Language Contact Profile, GJT = Grammaticality Judgment Task; PM = Phonological Memory. *Groups*: SA = Study Abroad, AH = At Home, IM = Domestic Immersion.

Study	Language	Skill/Structure	Groups	Length of Program	Assessment	Initial Proficiency	Results
			Grammatical Abilities				
Cheng & Mojica-Diaz (2006)	Spanish	Subjunctive	SA (*N*=6)	2 months	OPI	Advanced	No improvement in subjunctive use
Collentine (2004)	Spanish	Morphosyntax, Lexical abilities	SA (*N*=12) AH (*N*=12)	16 weeks	Pre/Post Oral interviews	3rd semester of study	AH greater improvement in accuracy; SA greater improvement in narrative abilities
DeKeyser (1991)	Spanish	Grammar, Communicative Strategies	SA (*N*=7) AH (*N*=5)	1 quarter	Pre/Middle/Post Picture description task and Oral Interview	Intermediate	No significant differences between groups
Howard (2001, 2005)	French	Past tense	PreSA (*N*=6) PostSA (*N*=6) AH (*N*=6)	1 year	Sociolinguistic interviews	7-9 years of prior study	SA better use & accuracy for some forms
Isabelli (2004)	Spanish	Null Subject Parameter	SA (*N*=29)	9 months	Pre/Post GJT and Oral Interview	Intermediate	Partial resetting of null subject parameter

Note. *Assessments*: Pre denotes assessment administered prior to or at the beginning of semester of study; Post denotes assessment administered at the end of the semester; OPI = Oral Proficiency Interview; LCP = Language Contact Profile, GJT = Grammaticality Judgment Task.
Groups: SA = Study Abroad, AH = At Home, IM = Domestic Immersion.

Study	Language	Skill/Structure	Groups	Length of Program	Assessment	Initial Proficiency	Results
			Grammatical Abilities				
Isabelli & Nishida (2005)	Spanish	Subjunctive	SA (*N*=29) AH (*N*=32)	9 months	Pre/Middle/Post OPI for SA; Cross-sectional OPI data for AH	3rd year of study	SA groups better oral production of subjunctive than AH groups
Isabelli-Garcia (2010)	Spanish	Grammatical Gender Agreement	SA (*N*=12) AH (*N*=12)	4 months	Pre/Post Stimulated OPI and GJT, Pre/Post LCP for SA, Pre LCP for AH	Intermediate	No significant pre-post GJT gains for either group
Lafford & Ryan (1995)	Spanish	Acquisition of lexical meaning (*por/para*)	SA (*N*=9)	1 semester	Pre/Middle/Post OPI	Novice	Correct use of *por* and *para* varied with proficiency
Ryan & Lafford (1992)	Spanish	Acquisition of Spanish copula (*ser/estar*)	SA (*N*=16)	1 semester	Pre/Middle/Post OPI	Novice	Order of acquisition different in SA (compared to AH: VanPatten, 1987)

Note. *Assessments*: Pre denotes assessment administered prior to or at the beginning of semester of study; Post denotes assessment administered at the end of the semester; OPI = Oral Proficiency Interview; LCP = Language Contact Profile, GJT = Grammaticality Judgment Task.
Groups: SA = Study Abroad, AH = At Home, IM = Domestic Immersion.

2.2 **Internal Factors**

Adult L2 acquisition is highly variable and may be constrained by a wide range of internal factors or "individual differences" in learner-specific characteristics that affect the way an individual learner acquires and processes the L2. Accordingly, a number of internal factors have been examined in relation to L2 success, including age, intelligence, aptitude, experience and proficiency with second languages, attention control, and learning ability within different memory systems (e.g., Bialystock & Frohlich, 1978; Carpenter, 2008; Carroll, 1958, 1962, 1981; Dörnyei, 2005; Miyake & Friedman, 1998; Morgan-Short et al., 2014; Robinson, 2002, 2003; Sanz, 2000; Segalowitz & Frenkiel-Fishman, 2005; Skehan, 1991).

Given the variability in L2 outcomes evidenced during study abroad, there is a growing demand for a deeper understanding of the factors that are relevant to successful L2 development in the study abroad context. Several significant predictors of successful L2 development in the study abroad context have been identified, including experience with multiple L2s (Brecht et al., 1995), participant gender (Brecht et al., 1995), preprogram L2 abilities and knowledge (Davidson, 2010), and cognitive processing abilities (Segalowitz et al., 2004). Two particular factors that have emerged from multiple studies as potentially relevant to gains in the study abroad context are (1) reported L2 contact hours (Freed, Segalowitz & Dewey, 2004; Segalowitz et al., 2004) and (2) WM abilities (LaBrozzi, 2009, 2012; Sunderman & Kroll, 2009; Tokowicz, Michael, & Kroll, 2004). The following sections present SLA research that has considered the relationship between these factors (L2 contact, Section 2.2.1; WM, Section 2.2.2) and linguistic development.

2.2.1 <u>Language Contact</u>

Use of the L2 outside of the classroom in interactions with native speakers is often considered to be the key to success in study abroad experiences (Freed, 1995b). Empirical considerations of whether informal out-of-class L2 contact enhances L2 acquisition, however, are limited, and the results of studies that have assessed L2 contact are ambiguous. The theorized importance of contact hours in the target language in the study abroad context, coupled with mixed results from empirical studies, make plain the importance of assessing L2 contact more rigorously in order to better understand its relationship with linguistic development.

2.2.1.1 <u>Measuring Language Contact</u>

In order to facilitate the collection of language use data from learners, researchers have developed, and made publicly available, a standardized questionnaire with pre- and post-test versions: the Language Contact Profile (LCP; Freed, Dewey, Segalowitz, & Halter, 2004). This questionnaire is designed to probe factors such as living arrangement, time spent speaking the L2 and L1 inside and outside of the classroom, the nature of interlocutors, use of material learned in the classroom in non-classroom settings, and the pragmatic nature of L2 exchanges. A number of studies have included the LCP in their experimental designs in order to assess individual differences between learners in terms of L2 contact hours during study abroad (Freed et al., 2004; Isabelli-Garcia, 2010; Magnan & Back, 2007; Segalowitz et al., 2004).

2.2.1.2 <u>Empirical Studies: Language Contact</u>

Freed, Segalowitz, and Dewey (2004) examined interactions between language contact and context of learning in relation to gains in oral fluency, as measured by oral interviews.

Learners of L2 French in study abroad (*n* = 8), at home (*n* = 8), and intensive domestic immersion (*n* = 12) contexts provided detailed reports of out-of-class L2 contact at the beginning and again at the end of the semester via a modified version of the LCP. Mean number of hours per week of French out-of-class contact (speaking, writing, reading, listening, as well as the sum of these four activities) were calculated for each student and group; between-group comparisons revealed that students in the immersion group reported significantly more L2 contact hours than the study abroad and at home groups.

The relationship between L2 contact hours and oral performance gains was examined using multiple regression analyses that collapsed across the three learner groups. Results indicated that gains in oral fluidity were predicted by out-of-class L2 writing hours. Critically, the group that made the most gains in oral fluidity, the domestic immersion group, reported more than four times the amount of out-of-class writing in the L2 than the study abroad and at home groups. The analyses reveal significant differences between learning contexts in terms of both L2 contact and oral fluency gains. The authors suggest that writing in the L2 (producing output) may have led learners to process more deeply, thus leading to greater gains (Swain 1993, 2000). This analysis, however, cannot distinguish between the role of L2 contact and the role of learning context, given that the significant relationship between L2 writing hours and oral fluidity gains is driven by the domestic immersion group. The authors note the importance of considering individual differences that may allow some students to benefit more from L2 contact activities.

Subsequent research has examined language contact together with other internal variables, such as prior course work (Magnan & Back, 2007), measures of individual variability in confidence using the target language (Isabelli-García, 2010), and cognitive abilities

(Segalowitz et al., 2004) in order to contribute to our understanding of how these internal factors related to linguistic development.

Magnan and Back (2007) investigated the relationships between prior language coursework, L2 contact, and gains in speaking abilities among 20 intermediate-level learners participating in semester-long study abroad programs in France. The researchers examined changes in learner-reported self-confidence in speaking the L2, as well as correlations between levels of improvement on an oral proficiency interview (OPI) administered pre- and post-semester and responses on a modified LCP administered at the end of the study abroad program. Post-program assessments revealed that overall, students reported increased self-confidence in their speaking abilities and that all maintained or improved their level of speaking proficiency.

Correlational analyses of reported language use and improvement in oral proficiency yielded one significant result: reported time speaking French with American classmates correlated negatively with improvement on the OPI, suggesting that speaking the L2 with nonnative peers may impede proficiency development. This unexpected finding may reflect an overall tendency among these students to spend more time with other English speakers; the authors suggest that the critical language contact factor seems to be *with whom* students spend their time speaking the L2. Additional analyses with learner background variables revealed a positive relationship between pre-study abroad coursework and gains on the OPI. Given that all of the learners in this study tested at intermediate levels on the OPI prior to the semester abroad, the authors conclude that coursework may be a more important factor than initial proficiency in preparing students to make gains in the study abroad context.

The relationship between initial proficiency and gains during study abroad is the subject of considerable discussion. There is evidence that learners at higher proficiency levels at the outset of the study abroad program have an advantage, being more likely to engage in use of the L2 while abroad (Brecht et al., 1995), but it is also true that lower-proficiency learners tend to make more obvious gains in language skills in comparison to more advanced learners (Regan, 2003). As in the laboratory-based literature, it is important to consider the methods used to assess L2 abilities and gains. Whereas many studies interested in linguistic development during study abroad utilize standardized measures that divide learners into discrete proficiency categories, or assess L2 abilities in the domain or oral production, recent work has also made use of objective linguistic assessments that yield continuous variables.

Isabelli-García (2010) examined whether L2 contact and "individual variability" (i.e., learner apprehension or confidence about learning the language; Individual Variability Questionnaire, Ely, 1986) affect the acquisition rate of grammatical gender agreement in L2 Spanish. L1 English students studying Spanish in a semester-long study abroad (n = 12) or at home (n = 12) context completed a 65-question grammaticality judgment task that asked learners both (1) to indicate whether each sentence was correct or incorrect and (2) to modify incorrect sentences. Isabelli-García assessed the accuracy of learner corrections to attributive and predicative adjectival agreement for gender-marked and non-gender-marked nouns, disregarding corrections made to any other part of the sentence. Analyses revealed that hours spent reading and writing in the L2 outside of class correlated with improved corrections of predicative adjectival agreement for unmarked nouns only. No significant correlations were found between individual variability and pre- or post-test gender agreement accuracy. The author concludes that individual variability and language contact abroad have minimal influence on acquisition rate for gender agreement.

Within this sample, no significant improvements in accuracy for pre- to post-test were evidenced in either the at home or study abroad groups. The absence of a relationship between language contact and improvement in accuracy may have been influenced by an overall lack of gains in L2 abilities from pre- to post-test, the limited number of items in each of the agreement categories (non-gender-marked nouns were further separated into three categories, yielding 8 to 13 tokens per category), and the relatively small sample size.

In a large study whose full results are described in detail in multiple reports (Collentine, 2004; Lafford, 2004; Segalowitz & Freed, 2004), Segalowitz and colleagues assessed linguistic gains made by students in study abroad and at home contexts, and examined three categories of factors that may predict such gains: (a) reported language use on a post-semester LCP, (b) prior experience learning Spanish and pre-test language knowledge, and (c) cognitive processing abilities, specifically, speed and efficiency of lexical access, and speed and efficiency of attention control.

With regard to reported language contact, a negative relationship between language use and oral fluency was revealed within the study abroad group. This effect, however, was limited to one measure of oral fluency and one reported L2 activity (time spent speaking with home-stay family). On the other hand, within the study abroad group, overall outside of class L2 contact hours were negatively correlated with use of communication strategies, indicating fewer communication breakdowns among study abroad students who reported greater L2 contact time (Lafford, 2004). No relationship between L2 contact and linguistic development was found within the at home group, and no significant relationships between L2 contact hours and overall oral proficiency were found for either group.

With regard to prior knowledge and experience learning Spanish, neither pre-test grammar scores nor years of previous study of Spanish were significantly correlated with gains in L2 abilities in either group, in contrast to Magnan and Back (2007). A positive relationship between pre-test scores and reported out-of-class L2 use was revealed within the study abroad group only (Segalowitz & Freed, 2004), which corroborates findings from a previous, large-scale report (Brecht et al., 1995).

With regard to cognitive processing abilities, results revealed that faster, more efficient L2 lexical access at pre-testing was related to fluency gains, with no significant effect of context of learning. The authors emphasize the importance of considering a learner's "language learning readiness" (Segalowitz et al., 2004) and preparation for the challenges he or she will face in a given learning context. These data reveal a complex interaction between initial cognitive and linguistic abilities, language use, and gains in different linguistic skills that appears to play out differently in at home and study abroad contexts.

Taken together, empirical findings related to the relationship between L2 contact and linguistic development reveal conflicting results. Research has found positive relationships with out-of-class L2 contact and linguistic development (Freed, et al., 2004; Segalowitz et al., 2004), limited to no relationship (Isabelli-García, 2010; Magnan & Back, 2007), and even a negative relationship (e.g., Magnan & Back, 2007; Segalowitz & Freed, 2004). Studies have found an interaction between type of out-of-class contact (e.g., *interactive contact*: time spent speaking with friends and family versus *non-interactive contact:* time spent reading books, watching television) and learner proficiency level (Freed, 1990; Spada, 1986), and have reported that out-of-class L2 contact predicts linguistic gains in one context, but not another (Magnan & Back, 2007). In general, the shortage of easily interpretable, significant effects and the lack of consistency in language use research should be interpreted with care. The following section provides a discussion of concerns and considerations related to the assessment of language contact in SLA research.

2.2.1.3 Language Contact: Questions and Considerations

The relationship between language contact and context of learning is under-investigated in SLA literature, and it is difficult to determine the level of precision in language use reports (Freed, Segalowitz, & Dewey, 2004). In the majority of studies that include an assessment of L2 contact, learners complete the LCP after returning from study abroad programs, and are asked to provide a retrospective report of the average number of days per week and hours per day that they engaged in various activities in the L2 during the preceding months. Such retrospective reports may not provide the most accurate representation of L2 contact over the course of the entire semester, as learners' experiences may vary from week to week, and as it can be challenging for learners to provide accurate estimates of average time spent in different types of interactions weeks or months after those interactions occurred.

Given the complex findings related to language contact, it is also important for research to address the ways in which initial cognitive abilities may predict or interact with language contact in different learning contexts (Segalowitz & Freed, 2004; Segalowitz et al., 2004). The following section describes research related to a cognitive ability that has been posited to play a role in learner behavior, L2 processing, and L2 development: working memory.

2.2.2 **Working Memory**

Working memory (WM) is understood to be a dual-component system that is responsible for both the storage and processing of information in short-term memory (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Just & Carpenter, 1992; Miyake & Shah, 1999). More specifically, WM[3] is posited to comprise a domain-specific storage component, responsible for the temporary storage of information (short-term memory), and a domain-general processing component, responsible for controlling the information stored in short-term memory in order to carry out complex tasks (Baddeley, 2007; Kane, Bleckley, Conway, & Engle, 2001). WM has been posited as a possible factor affecting L2 processing (e.g., Harrington & Sawyer, 1992; Kroll, Michael, Tokowicz, & Dufour, 2002; Miyake & Freidman, 1998; Robinson, 2002; Sunderman & Kroll, 2009; Tokowicz, Michael, & Kroll, 2004), although results regarding the relationship between WM abilities and L2 performance have not always been consistent (e.g., Daneman & Carpenter, 1980; *cf.* with Juffs, 2004).

Higher WM abilities are associated with the efficient allocation of cognitive resources, such as suppressing the L1 in order to store and process information in the L2 (e.g., Just & Carpenter, 1992; Engle, 2002; Meuter, 2005); this ability may be particularly critical when an individual is surrounded by L2 input in an immersion setting and must continuously inhibit the L1 in order to understand and produce the L2. Indeed, researchers have posited that WM abilities may be particularly relevant in study abroad contexts (Lafford, 2006; Sunderman & Kroll, 2009; Tokowicz et al., 2004); however, given that L2 processing has been shown to make greater

---

[3] Note that the present study employs the term "working memory" to denote *dynamic* working memory *abilities,* which includes both memory and attention abilities (Redick et al., 2012). This construct is also labeled "executive attention" in cognitive psychology literature, e.g., Unsworth & Engle, 2007) as opposed to simple working memory *capacity*. Section 2.2.2.1 provides additional discussion of these constructs. Section 3.4.1.1 provides detailed descriptions of the tasks employed to assess individual differences in WM in the present study.

cognitive demands of an individual than L1 processing (Hasegawa, Carpenter, & Just, 2002), it may be the case that individual differences in WM abilities explain variability in L2 development among learners in any context (e.g., Williams, 2012).

In order to meaningfully address the relationship between WM and L2 acquisition, one must first obtain a reliable measure of individual WM abilities. The following section describes the tasks commonly used to measure WM abilities and discusses some methodological considerations related to those tasks.

2.2.2.1 Measuring Working Memory

Complex span tasks are assessments that combine the storage element of a simple span task (remember a series of words, digits, letters in order) with a processing task (reading a sentence). Whereas simple span tasks (e.g., Word Span, Digit Span, Corsi Blocks) measure short-term memory capacity (storage), complex span tasks are posited to provide a measure of both processing and storage, or of dynamic *working* memory abilities (Redick et al., 2012). Importantly, recall on complex span tasks has been shown to predict performance in linguistic measures whereas recall in simple span tasks has not (e.g., Daneman & Carpenter, 1980; but see Juffs, 2004).

There are multiple considerations regarding testing of WM abilities, including contention regarding (a) whether WM is language dependent or independent (see Berquist, 1997; Osaka & Osaka, 1992), (b) if there are multiple pools of resources or a single pool for both linguistic and non-linguistic tests (see Just & Carpenter, 1992; MacDonald & Christiansen, 2002; Shah & Miyake, 1996; Waters & Caplan, 1996), and (c) whether WM scores should be split into high/low designations or considered as continuous measures (Friedman & Miyake, 2005;

Sagarra, 2000). Nevertheless, several established tests of WM abilities have been used. Below are described three widely used tests of WM abilities, two of which are considered linguistic measures and one that is considered a nonlinguistic measure.

Perhaps the most commonly used test of WM abilities in SLA research is the Reading Span Test (Daneman & Carpenter, 1980). In this linguistic test of WM abilities, participants read or hear a set of sentences (processing component) and are asked to remember the last word of each sentence (storage component). Sentences are presented one at a time, and the number of sentences in each set varies from two to six. At the end of each set, participants recall the final words from each of the sentences in the preceding set. In other versions of this task, a response is required during the processing task in order to ensure that participants are engaged in the task; for example, participants may be asked to provide a judgment of the plausibility (Waters & Caplan, 1996) or sensicality (Unsworth, Heitz, Schrock, & Engle, 2005) of each sentence. Reaction times and error rates during the processing component of the test can be measured and used to determine whether there is a relationship between speed and accuracy, a question of interest in some cognitive research. The storage component of this task can also be manipulated, such that participants are asked to remember and later recall a letter presented after each sentence instead of the last word of the sentence (Unsworth et al., 2005).

The Operation Span Test (Turner & Engle, 1989; Michael & Gollan, 2005) is another commonly-used linguistic WM test. The processing component of this task consists of solving math problems, and the storage component involves remembering a word or letter presented after each math problem. Participants are presented with math problems in sets ranging in size from two to six that must be solved as they are presented. At the end of each set, participants must recall all of the words or letters that appeared during that set, in the order in which they were presented.

A non-linguistic test of WM designed to eliminate the possibility of reliance on linguistic processing, is the Symmetry Span Test (Kane et al., 2004; Shah & Miyake, 1996; Unsworth, Heitz, Schrock, & Engle, 2005), which requires the simultaneous processing and maintenance of spatial information. In one version of this task, participants are presented with a series of two to six abstract designs and are asked to determine whether the design is symmetrical about its vertical axis. This symmetry judgment comprises the processing component of the task. For the storage component, after each symmetry judgment, participants are presented with a 4x4 matrix in which one cell is filled in red. Participants need to remember and recall which cells appeared in red, in order of presentation, after each set of symmetry judgments (see Section 3.4.1.2 for more information on each of these tasks).

Although the WM tests previously described are widely considered to be valid measures, and have been shown to be reliable, no single task can provide a perfect measure of the construct that it purportedly represents. In order to obtain a comprehensive measure of WM abilities, the optimal research strategy is to administer multiple WM tasks and then calculate a latent variable, or composite score, based on the average of scores on all tasks (Conway et al., 2005). Variations of the tasks described above, among others, have been used to study individual differences in SLA. The following section provides a discussion of the extant research on WM and L2 development.

2.2.2.2 Working Memory: Empirical Studies

Researchers have examined the relationship between WM and a variety of skills in L2. Empirical studies have provided evidence of a positive relationship between WM abilities and L2 vocabulary acquisition (e.g., Daneman & Green, 1986), word-naming and translation tasks (e.g., Kroll et al., 2002; Tokowicz et al., 2004), lexical retrieval and fluency (e.g., Michael & Gollan, 2005), noticing and benefitting from interactional feedback (e.g., Mackey et al., 2010; Mackey, Philp, Fujii, Egi, & Tatsumi, 2002), reading comprehension (e.g., Harrington & Sawyer, 1992; Leeser, 2007), and processing of morphological cues (e.g., LaBrozzi, 2009; Sagarra, 2007; Sagarra & Herschensohn, 2010) and sentence-level structural ambiguities (Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009, subject-object ambiguities; but *cf*. Juffs, 2004, garden-path sentences). The notion that variability in learner WM abilities impacts L2 learning has also been noted by Williams (2012), who suggests that it is important to consider the role of WM in different learning conditions, as the relationship between WM abilities and linguistic gains may not be the same across language skills or learning contexts. The studies described below reveal the complex nature of the relationship between WM and L2 abilities in regard to morphosyntactic development (Sagarra, 2007) and processing (Sagarra & Herschensohn, 2010).

Sagarra (2007) examined the relationship between WM and development of noun-adjective agreement in L2 Spanish. Participants completed sentences by providing an adjective that matched in gender and number with the noun it modified. Participants in the experimental group received computer-delivered recasts (the correct construction when the learner response was incorrect), whereas participants in the control group received no feedback. WM abilities were assessed using a reading span test, and linguistic gains were measured with immediate and delayed written post-tests and an oral information exchange task that required production of the target form. The results provided evidence that (a) recasts positively affected grammatical development and that (b) WM proved to be a good predictor of learners' ability to benefit from the recasts. The relationship between WM and processing of these morphosyntactic structures was addressed in subsequent research that included learners at different levels of proficiency, and employed online and offline assessments.

Sagarra & Herschensohn (2010) investigated the role of language proficiency and WM in gender and number agreement processing using both an online and offline task. Participants included English-speaking learners of Spanish enrolled in a university Spanish course (beginners: third-semester course; intermediates: seventh- or eighth-semester course), as well as monolingual native speakers. Theses participants completed a self-paced reading task (online) and a grammaticality judgment task (offline) containing sentences with noun–adjective gender/number agreement/disagreement; WM was assessed using a reading span test. The results revealed that all participants were highly accurate in the offline task, with both learner groups more accurate in their judgments of sentences containing number than gender agreement violations. The online task revealed sensitivity to gender and number violations among intermediate-level learners and Spanish monolinguals, but not beginner-level learners. In regard to the role of WM, intermediate-level learners with higher WM were more sensitive to gender (but not number) disagreement than those with lower WM, as indexed by greater accuracy and faster reaction times to adjectives that represented gender-agreement violations within this subgroup. Worth noting, however, is the fact that the mean WM score within the intermediate-level learner group was significantly higher than beginner-level learners and monolingual Spanish-speakers, which may partially explain the lack of effect of WM among the beginning learners. These findings suggest that WM facilitates sensitivity to L2 grammatical gender agreement, but does not necessarily play a role in number agreement, which was less difficult for this group of learners. Evidence for a role for WM abilities in grammaticality judgments, and particularly that this effect may vary by structure, corroborates findings from previous research (McDonald, 2006), indicating that WM play a greater role when processing demands are highest.

Indeed, there is mounting evidence from laboratory-based research for a relationship between WM and L2 learning, and this relationship appears to be instantiated differently depending on the learning target or grammatical structure being assessed (e.g., McDonald, 2006; Sagarra & Herschensohn, 2010; Williams, 2012). Theoretical positions and empirical research suggest that the role of WM may also vary with relation to learning conditions (Williams, 2012), perhaps playing a larger role under more explicit conditions because of the requirement to retain metalinguistic information in memory while simultaneously comprehending and producing language (e.g., Mackey et al., 2002; Roehr, 2008; Sagarra, 2008). In order to better understand of how individual differences in WM affect L2 acquisition, investigations of interactions between the internal variable WM and external variables such as learning condition or context of learning are needed.

Interestingly, the primary evidence of a relationship between WM and learning in the study abroad context comes from studies that have assessed the relationship between internal cognitive abilities and previous study abroad experiences, rather than from experimental analyses of L2 gains that occur during immersion in the L2 setting. To date, there have been three studies that have assessed the relationship between WM abilities and study abroad or immersion experience. These studies have examined the interaction between these internal and external factors with regard to translation skills (Tokowicz et al., 2004), lexical comprehension and production (Sunderman & Kroll, 2009), and online processing of temporal cues (LaBrozzi, 2009). Each of these studies is described in detail below.

Tokowicz, Michael, and Kroll (2004) examined the effects of amount of previous study abroad experience (SAE) and WM abilities on the types of errors (non-response vs. meaning errors) made during single-word translations from the L1 to the L2. Participants included 15 native speakers of Spanish (tested in Mexico) and 22 native speakers of English (tested in the United States), all of whom were relatively proficient in both languages. Participants were grouped based on their reported amount of previous SAE: “less SAE” (8 or fewer months of SAE, n = 20) or “more SAE” group (15 or more months of SAE, n = 17). Based on their scores from an operation span test in their native language, participants were classified as “lower” or “higher” WM, such that four distinct groups were created (more/less SAE, lower/higher WM). In order to control for confounding factors such as higher overall WM scores in one group than another, the researchers excluded 5 participants from the analyses. The final groups were as follows: less SAE/lower WM (n = 6), more SAE/lower WM (n = 6), less SAE/higher WM (n = 12), more SAE/higher WM (n = 8). Participants completed a translation task in which they were presented with words in the L1 (English or Spanish) and asked to provide an appropriate L2 translation. Participants were asked to respond as quickly and accurately as possible and to say “no” when they were unable to provide a translation for an item.

The researchers hypothesized that individuals with more SAE and/or greater WM abilities would have larger vocabularies, and thus better overall translation accuracy, due to greater exposure to L2 words and to a facilitative role for WM in lexical development. Furthermore, the researchers hypothesized that learners with more SAE would be less prone to non-response errors, and more prone to making meaning errors, due to their experiences in communicating without precise translations during SAE. As meaning errors require a stimulus to be maintained in memory while related words are activated, the authors predicted an interaction between SAE and WM, expecting a higher percentage of meaning errors among individuals with more SAE and greater WM abilities than individuals with lower WM abilities and/or less SAE. In order to test these hypotheses a series of hierarchical linear regression analysis were conducted.

Results revealed that (1) neither SAE nor WM, nor an interaction between SAE and WM predicted overall task accuracy, and that (2) only the more SAE/higher WM made the same percentage of meaning as non-response errors; individuals in the other groups made more non-response than meaning errors. The researchers interpret these results as evidence that higher WM learners make use of the ability to hold multiple items in memory and develop a strategy of using approximate translations to communicate in during SAE, hence the higher levels of meaning errors within the higher WM/more SAE group. Despite small group sizes, this study provides compelling evidence for an interaction between the external factor context of learning and the internal factor WM, suggesting that WM plays an important role in lexical development during SAE. This relationship was addressed by subsequent research carried out by Sunderman and Kroll (2009), which examined lexical abilities among learners with and without SAE.

Sunderman and Kroll's (2009) study provides additional evidence of a relationship between WM and lexical comprehension and production during SAE. Given the limited research related to the intersection between internal and external factors in lexical processing, the following four hypotheses (and their predictions) were considered: (1) *internal resource hypothesis*: internal L1 cognitive resources are related to an individual's ability to comprehend and produce the L2; learners with greater WM will be faster and more accurate in their L2 comprehension and production than learners with lower WM; (2) *external cue hypothesis*: context of learning is related to an individual's ability to comprehend and produce the L2; individuals with SAE will be faster and more accurate in their L2 comprehension and production than those without SAE; (3) *interaction hypothesis*: cognitive resources modulate the influence of external factors such that learner outcomes reflect a mix of external and internal factors; as WM increases, an individual will increasingly be able to benefit from SAE; (4) *threshold hypothesis*: the benefits of an immersion setting might not hold for all individuals; only learners with WM abilities above an certain threshold will benefit from SAE, whereas learners whose WM abilities fall below the threshold will not see benefit from SAE.

In order to test these hypotheses, native English-speakers studying Spanish at the university level were tested (14 with SAE, and 34 without SAE). WM ability was measured using a reading span test; linguistic benefits of external (SAE) and internal (WM) factors were indexed by faster reaction times and higher accuracy scores on a translation recognition task (comprehension measure) and a picture naming task (production measure).

Descriptively, participants with SAE were faster and more accurate than learners without SAE on both measures, and the impact of SAE greater for production than comprehension. Multivariate regression analyses of performance on the comprehension measure, which treated WM as a continuous variable, support both the internal resource and external cue hypotheses: SAE and WM resources independently, positively affected the speed and accuracy of lexical comprehension. Analysis of performance on the production measure revealed a positive effect of SAE on reaction time (faster reaction times for those with SAE than without) and a marginally significant interaction between SAE and WM on accuracy, indicating that internal resources may mediate benefits of learning context such that those with higher WM resources benefit more from SAE, providing support for the interaction hypothesis. The threshold hypothesis was supported in terms of production accuracy only. The estimated effect of SAE on production accuracy was plotted for each level of WM abilities (along with 95% confidence levels for the relationship) in order to reveal the point at which interactions between WM and SAE become significant. This analysis revealed that SAE has a significant effect on production accuracy only for participants with sufficiently high WM scores (at least 24 out of 80 words recalled on the reading span test). For learners below this threshold, the effect of SAE on production accuracy was not statistically different from zero.

The findings of Sunderman and Kroll (2009) make important contributions to the exploration of interactions between external and internal factors in L2 development. First, these results indicate a potential threshold level of WM resources required to realize benefits during SAE, at least for production accuracy. The notion of a cognitive threshold had been discussed in previous research (e.g., Segalowitz & Freed, 2004), but a specific threshold had not been identified prior to this study. These results also reveal that the effects of external and internal factors may vary by type of linguistic task, specifically comprehension or production. It is important to note, however, that the treatment of SAE as a discrete variable, as was done in the analysis used to test the threshold hypothesis, is potentially problematic in this study given the variability in duration of SAE. Within the 14 participant "with SAE" group, SAE duration ranged from two to ten months. Considering empirical findings that duration of SAE is related to linguistic development during SAE (e.g., Davidson, 2010), this variability should be considered in analyses regarding the impact of SAE on L2 development. Furthermore, these learners were not tested during or even immediately following their experiences in a study abroad context. As such, it may be relevant to measure and control for other factors that may vary between learners with and without SAE (e.g., motivation, proficiency).

LaBrozzi (2009, 2012) contributed to discussion of the relationship between external factors and internal cognitive factors by investigating how WM, inhibitory control, and SAE help classroom learners rely on morphological (rather than lexical) cues during sentence processing. Adult English-Spanish classroom learners with (n = 36) and without (n = 24) an immersion experience (16 week program in Spain) completed a series of screening tasks (including a standardized proficiency test), as well as a measure of L2 processing, a test of WM, and a test of inhibitory control.

In order to obtain a measure of online processing, participants' eye movements were recorded while they read sentences in Spanish and answered a comprehension question. The experimental sentences each contained a past tense adverb, which served as a lexical temporal cue. Half of the sentences also contained a past tense verb (morphological cue), and half contained a present tense verb (conflicting morphological cue). Reading times for the critical regions (adverbs and verbs) were assessed only for sentences in which the comprehension questions were answered correctly in order to decrease the probability of increased reading time due to lack of comprehension. WM abilities were assessed using a reading span test, and inhibitory control was measured using the Simon task, which required participants to ignore the position of a target stimulus in order to respond only to its color.

Results of the screening tests revealed that the sample pool was homogeneous in terms of L2 proficiency and knowledge of the target verbs and tenses. Analyses of the eye-tracking data indicated that classroom learners without SAE tended to use lexical cues to resolve tense conflict, whereas those with SAE showed a decreased reliance on lexical cues and an increased reliance on morphological cues, the more native-like processing pattern. Reading span test results showed that WM ability modulates the processing of morphological cues in learners with SAE, but not in learners without SAE. This finding is interpreted as suggesting that when in the SA setting, learners with higher WM abilities are able to focus on morphological cues when lexical cues are absent. Inhibitory control was not related to processing of lexical or morphological cues for either group, a result the author suggests may be attributed to the data being collected after SAE learners had returned from their immersion experience. Taken as a whole, the findings from this study suggest that (1) the external factor SAE and (2) the combination of SAE and WM abilities help adult classroom learners to process morphological cues in the input.

Important to note is that the learners in both the SAE and no-SAE groups were at the same proficiency level, as revealed by not only the objective proficiency screening test, but also by accuracy on comprehension questions during the eye-tracking task. This careful assessment of proficiency is particularly important given the lack of a pre-test/post-test design. The experimental design employed by LaBrozzi (2009, 2012) reveals the benefit of including a sensitive online processing measure, such as eye-tracking, in order to capture subtle, yet meaningful, differences between individuals and groups. Findings from these three studies provide empirical evidence of the need for research designs that consider both context of learning and individual differences in cognitive abilities (Collentine &Freed, 2004).

2.2.2.3 Working Memory: Questions and Considerations

A considerable amount of research has investigated the role of WM in L2 acquisition. To date, only three studies have directly addressed the relationship between WM and context of learning (LaBrozzi, 2009; Tokowicz et al., 2004; Sunderman & Kroll, 2009). The results from these studies that have addressed the interaction between WM and SAE have provided substantial evidence of a relationship between WM and the development of various L2 skills, however, there are several methodological issues to consider when interpreting this body of research, and certainly, several open questions remain.

In these studies, as in the vast majority of SLA studies that examine WM abilities, a single, linguistic measure of WM is employed. In order to obtain a more "pure" measure of the cognitive ability, multiple assessments of WM (both linguistic and non-linguistic) should be administered (Conway et al., 2005). An additional consideration is the lack of an objective test of proficiency (Sunderman & Kroll, 2009; Tokowicz et al., 2004), which would allow for a more complete assessment of the relationship between WM and L2 development in different contexts. Critically, none of these studies has employed a pre-test/post-test design in order to assess L2 development that occurs specifically during the SAE. Instead, each of these studies has assessed the relationship between WM and previous SA experience, examining both factors in a post hoc manner in learners who were engaged in at home language study at the time of testing. Thus, these studies are unable to differentiate between learning that occurs during study abroad and learning that takes place during subsequent study.

Previous research suggests a complex relationship between WM and context of learning; in order to facilitate assessment of the predictive effects of WM on gains during study abroad or at home study, a pre-test measure of WM abilities is essential. Furthermore, the experimental design must include assessments of language skills immediately prior to and immediately following the period of study. Future research could also benefit by considering the relationship between WM and other internal factors, such as language contact, that may be relevant to L2 development in different contexts in order to gain a more complete view of how various factors interact and contribute to learning. It is also important that these relationships be examined with regard to the learning and processing of various linguistic and grammatical forms, among multiple and large groups of learners, and in different L1-L2 pairings. Finally, only one study to date has addressed the relationship between WM and L2 development and processing using an online measure (LaBrozzi, 2009, 2012), which can provide important insight into the cognitive processes that may be in use at a given moment in time. Future research that includes an online measure of L2 processing, as well as measures of multiple internal factors, has the potential to enhance our understanding of the role of WM in L2 acquisition.

2.3 Online Processing of Language: Event-Related Potentials

Online measures of language processing can serve as a powerful and informative complement to behavioral data in L2 studies. One online measure used to quantify real-time neurocognitive processes underlying language comprehension is event-related potentials (ERPs), which reflect online electrophysiological brain activity related to cognitive processing. Indeed, in several studies, ERPs have been shown to detect and reveal effects that are not evidenced in behavioral measures alone (e.g., Alemán Bañón, Fiorentino & Gabriele, in press; Gabriele, Fiorentino & Alemán Bañón, 2013; McLaughlin et al., 2004; Morgan-Short et al., 2010; Morgan-Short, Steinhauer, et al., 2012; Tokowicz & MacWhinney, 2005).

The following subsections provide information regarding this brain-based measure. Section 2.3.1 provides the basic principles behind ERPs, followed by a description of the ERP components that are commonly associated with particular types of linguistic stimuli in L1 and L2 processing in Section 2.3.2. Section 2.3.3 provides a description of recent empirical work that makes use of ERPs as a measure of individual differences in language learning. In Section 2.3.4, open questions and considerations for the use of ERPs in SLA research are detailed, with particular emphasis on the use of ERPs to investigate the effects of external and internal factors on L2 development.

2.3.1 Event-Related Potentials: The Basics

ERPs provide a highly sensitive measure online cognitive processing. The human brain continuously produces electrical potentials; these electrical potentials form an electroencephalogram (EEG) signal that can be detected and recorded using electrodes placed near the scalp (Luck, 2005; Morgan-Short & Tanner, 2014). External stimuli can evoke small voltage changes in the naturally-occurring, ongoing EEG; in experimental designs, EEG recordings are time-locked to these stimuli, allowing for analysis of neural responses in relation to specific stimuli or events. EEG that is amplified and averaged across stimuli of the same type, both within and across subjects, yields a waveform with a series of positively and negatively deflected peaks, which reflects the central tendency of brain responses across trials and individuals. From these waveforms, we can identify ERP components based on the polarity (positive or negative), latency (time after the presentation of a stimulus that the component starts or peaks), scalp distribution (location where the signal is detected), and amplitude (maximal height or overall size) of the deflection (Luck, 2005).

In linguistic ERP experiments, the waveform elicited by an ungrammatical word is often compared to the waveform elicited by a grammatical word in a matched sentence or context, and components are described based on the difference between the two waves (violation paradigm; e.g., Morgan-Short et al., 2010). These data can be represented graphically, in which the *x*-axis represents time or latency and the *y*-axis corresponds to both polarity (with negative voltage generally plotted up) as well as amplitude. Graphical representations of ERP data from a violation paradigm study will either plot the grammatical and ungrammatical waves in separate lines on the same graph, or create a single, difference wave by subtracting the grammatical wave from the ungrammatical. Another way to represent the ERP data from either a single condition (e.g., grammatical) or a difference wave is using a voltage map. Voltage maps are schematic representations of the scalp that can display overall voltages across all electrodes on the scalp in a single figure. ERP data displayed in waveform graphs or voltage maps allows for the identification of ERP components. These components are believed to reflect distinct brain responses associated with different types of cognitive processes, although a single ‘surface component’ may reflect one or many underlying ‘latent components’ (Luck, 2005, Morgan-Short & Tanner, 2014). In language research, three ERP components are widely attested: the N400, the P600 and the AN. These components, and the type of linguistic stimuli they are associated with, are described below with regard to both L1 and L2 processing.

2.3.2 Event-Related Potential: Components in Language

ERPs have been used to address a variety of linguistic phenomena. It has been widely observed that lexical-semantic and grammatical (syntactic, morphosyntactic) manipulations in linguistic input elicit qualitatively different brain responses, reflected in distinct ERP components (the N400, and the (AN)/P600, respectively). These components, and the processes they are believed to represent, are described below.

The N400 component is generally believed to be associated with lexical/semantic processing (Friederici, 2002; Kaan, 2007; Kutas & Federmeier, 2011; Lau, Phillips, & Poeppel,

2008) and appears as a centro-parietal, bilateral negativity between 250 and 600 ms post-stimulus (e.g., Kutas & Hillyard, 1980). The waveform and voltage map provided in Figure 1 and Figure 2 illustrate the N400 effect.

The N400 component has been found in L1 and L2 processing in multiple languages (e.g., Kaan, 2007; Kutas & Federmeier, 2011). The amplitude, duration, and distribution of the N400 can be affected by word class and increases in difficulty of semantic integration or lexical access (Kutas & Hillyard, 1983; King & Kutas, 1998), as well as by mode of presentation, priming, and a stimulus word's position in a sentence (Hinojosa et al., 2001; Nobre, Allison, & McCarthy, 1994; Van Petten, 1995).

In L2 research, N400s are reliably elicited in response to semantic anomalies, such as a word that violates the semantic context established by the sentence (e.g., *I drink my coffee with cream and *shoes*), across learners, including novice and low proficiency learners (Bowden, Steinhauer, Sanz, & Ullman, 2013; McLaughlin, et al., 2004; Ojima, Nakata & Kakigi, 2005; Weber-Fox & Neville, 1996). Semantic processing among late L2 learners may elicit an N400 that is similar to native speakers (e.g., Bowden et al., 2013), but L2 N400s can also have a reduced amplitude or delayed onset (Ojima et al., 2005; Weber-Fox & Neville, 1996) and can also be more broadly distributed and longer in duration (Morgan-Short, 2007), particularly among low proficiency learners. Elicitation of an N400 effect is not limited to semantic processing; N400 effects have also been found for syntactic and morphosyntactic processing among learners at both low proficiency (e.g., Morgan-Short et al., 2010; Morgan-Short, Steinhauer et al., 2012; Mueller, Oberecker & Friederici; 2009; Tanner et al., 2013) and higher proficiency (e.g., Guo, Guo, Yan, Jiang, & Peng, 2009; Foucart & Frenck-Mestre, 2012;

Morgan-Short et al., 2010; Weber & Lavric, 2008), and even among native speakers (Tanner & Van Hell, 2014).

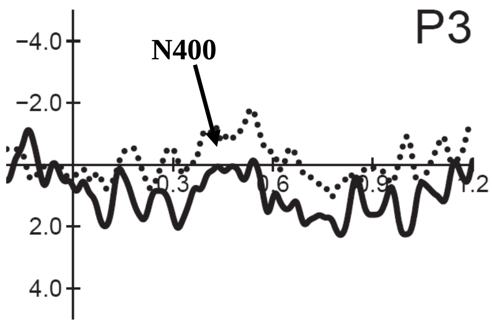


Figure 1. Event-related potential waveform representative of the N400; dotted line represents violation wave and solid line represents correct wave (adapted from Morgan-Short, 2007).

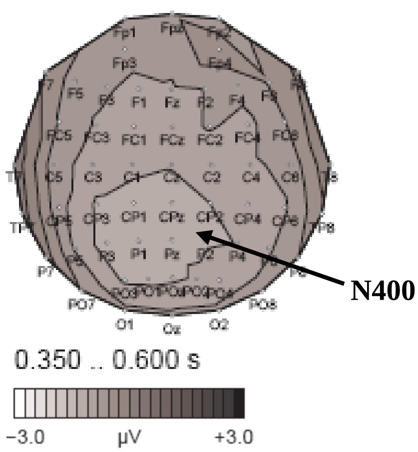


Figure 2. Event-related potential voltage map representative of the N400, violation minus correct condition (adapted from Morgan-Short, 2007).

The ERP components associated with grammatical processing, including syntactic and morphosyntactic structures, are the AN (Neville, Nicol, Barss, Forster, & Garrett, 1991) and the

P600 (Osterhout & Holcomb, 1992). The AN, if present, occurs between 150-500 ms and has a left-lateralized or bilateral anterior distribution (e.g., Coulson, King, & Kutas, 1998; Osterhout & Mobley, 1995). The waveform and voltage map provided in Figure 3 and

Figure ***4*** illustrate the AN effect.

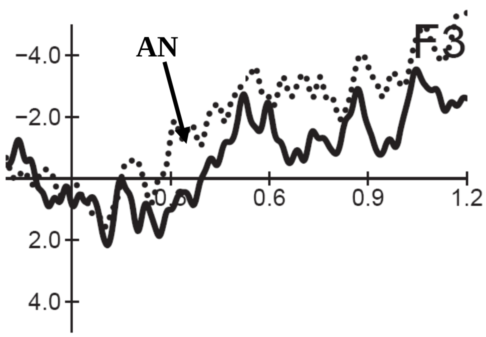


Figure 3. Event-related potential waveform representative of the AN; dotted line represents violation wave and solid line represents correct wave (adapted from Morgan-Short, 2007).

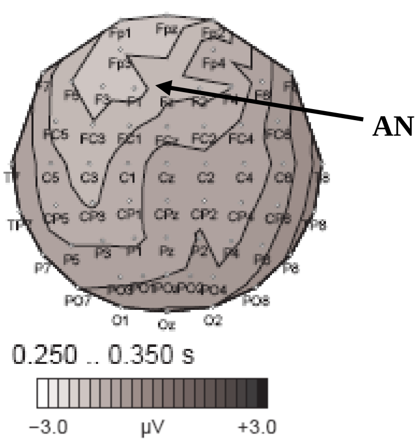


Figure 4. Event-related potential voltage map representative of the AN, violation minus correct condition (adapted from Morgan-Short, 2007).

In L1 processing, there is evidence that the distribution of the AN can be affected by individual differences, including working memory (Fiebach et al., 2002) and native speaker proficiency (Pakulak & Neville, 2010). This component has been evidenced in both syntactic, word-order violations (e.g., *I *coffee my drink in the morning*) and morphosyntactic, agreement violations (e.g., *I *drinks my coffee in the morning*), including subject-verb (e.g., Molinaro, Vespignani, & Job, 2008), determiner-noun, and noun-adjective violations (e.g., Barber & Carreiras, 2005). The AN is posited to reflect the early, automatic detection of violations of morphosyntactic information (Friederici, 2002; Molinaro et al., 2011; Steinhauer & Connolly, 2008); however, important to note is the absence of the AN effect in many studies of grammatical processing in native speakers (e.g., Frenck-Mestre et al., 2008, 2009; Hagoort et al., 1993; Kim & Osterhout, 2005; Sabourin & Stowe, 2008), which draws into question the interpretation of the AN component as reflective of “native-like” processing. Furthermore, recent empirical work indicates that AN effects can be a variant of the N400 (Tanner & Van Hell, 2014).

In studies of L2 processing, the presence of the AN during grammatical processing has occasionally been evidenced among learners at high proficiency levels (Morgan-Short, Steinhauer et al., 2012; Ojima et al., 2005; Rossi, Gugler, Friederici, & Hahne, 2006), but high-proficiency learners may also show no early effect (e.g., Frenck-Mestre, Foucart, Carrasco & Herschensohn, 2009) or an N400-like effect (e.g., Hahne, Mueller, & Clahsen, 2006). Studies of noun-determiner or noun-adjective agreement violations usually fail to find ANs (Gillon Dowens, Barber, Guo, Guo, & Carreiras, 2011; Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005); however, highly proficient learners in Gillon Dowens et al. (2010) did show ANs in response to nominal agreement violations. These learners have been living in an L2 setting for 12 to 33 years, suggesting that immersion experience may influence the emergence of the AN effect. Additional support for this idea comes from research with artificial language that has shown emergence of an AN at high proficiency for learners trained in immersion-like (uninstructed) conditions, but not classroom-like (instructed) conditions (Morgan-Short, Steinhauer et al., 2012).

The second (and often only) component elicited by syntactic and morphosyntactic processing difficulties is the P600, a positivity that occurs bilaterally in centroparietal regions between 500 and 1000 ms (Friederici & Mecklinger, 1996; Osterhout & Holcomb, 1992; Steinhauer & Connolly, 2008). The waveform and voltage map provided in Figure 5 and Figure 6 illustrate this effect. There are different interpretations of the significance of the P600; this component correlates with processing of syntactic, rule-based, combinatorial constraints (e.g., Coulson, King, & Kutas, 1998; Morgan-Short & Tanner, 2014). The P600 is a robust effect that is reliably elicited in L1 processing of various types of syntactic violations and morphosyntactic violations (e.g., Frenck-Mestre, Foucart, Carrasco, & Herschensohn, 2009; Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008; Osterhout & Holcomb, 1992) as well as correct sentences that are syntactically complex or otherwise difficult to parse (e.g., Kaan, Harris, Gibson, & Holcomb, 2000; Osterhout & Holcomb, 1992). The amplitude of P600s can be modulated by sentence complexity (e.g., Kaan et al., 2000) as well as by experimental tasks (Osterhout, Holcomb & Swinney, 1994; Steinhauer & Connolly, 2008).

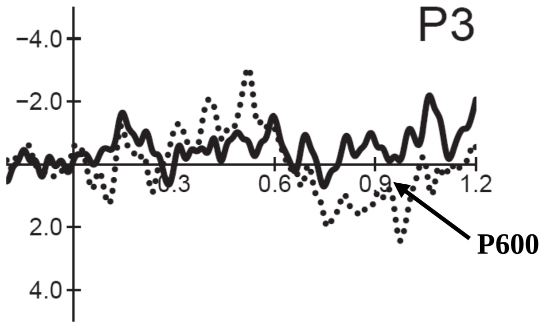


Figure 5. Event-related potential waveform representative of the P600; dotted line represents violation wave and solid line represents correct wave (adapted from Morgan-Short, 2007).

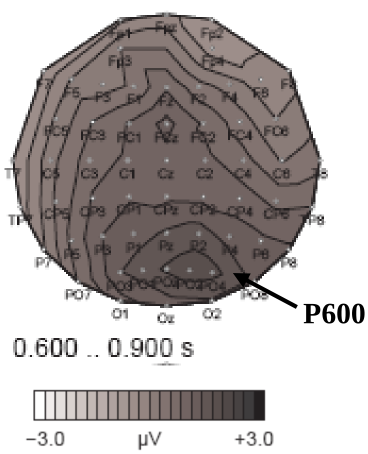


Figure 6. Event-related potential voltage map representative of the P600, violation minus correct condition (adapted from Morgan-Short, 2007).

Studies of syntactic and morphosyntactic processing in L2 reveal more variability in responses, with learners showing P600s, N400s, and sometimes no ERP effects. A number of factors may impact learners' neural responses to grammatical structures, as indicated by the discussion that follows. For syntactic violations, P600s have been found among learners at high levels of proficiency (Bowden et al., 2013; Rossi et al., 2006; Morgan-Short, Steinhauer et al., 2012), and in intermediate level learners (Hahne & Friederici, 2001). P600 amplitude may be reduced in low proficiency learners (Bowden et al., 2013; Mueller, 2009). Late L2 learners with low proficiency may show a delayed P600 (Hahne, 2001; Weber-Fox & Neville, 1996), an N400 (Morgan-Short, Steinhauer et al., 2012), or no ERP response to syntactic violations (Morgan-Short, Steinhauer et al., 2012; Mueller, Hahne, Fujii, & Friederici, 2005; Weber-Fox & Neville, 1996).

Particularly relevant for comparison to the present study are results of investigations of morphosyntactic processing, and, more specifically, agreement on adjectives and articles. Among studies that have investigated the relationship between learner L1 and L2 processing, some have found P600s for morphosyntactic violations when for learners whose L1 shares a similar feature, but not for learners whose L1 does not instantiate agreement features (Ojima, et al., 2005; Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008). P600s, however, have also been found when structures are unique to the L2 (Frenck-Mestre et al., 2009; Gillon Dowens et al., 2010; Gillon Dowens et al., 2011; Tokowicz & MacWhinney, 2005), although sometimes only within a subgroup of learners who demonstrate behavioral sensitivity to the violation (Foucart & Frenck-Mestre, 2011). Finally, evidence suggests that the elicitation of the P600 in L2 learners may depend on the structure being phonologically realized (Frenck-Mestre et al., 2008).

Morphosyntactic violations can also elicit N400 effects in L2 processing. Foucart and Frenck-Mestre (2012) found that advanced learners of L2 French had different neural responses to adjective agreement violations in different positions: no ERP effects for predicative adjectives, an N400 effect for pre-posed adjectives-noun violations (where the violation was realized on the noun), and a P600 effect for post-posed adjectives. The N400 effect for morphosyntactic violations has also been evidenced in intermediate Chinese-English learners (Guo et al., 2009) and L1 German learners of a miniature version of Italian (Mueller et al., 2009).

In a review of several L2 studies, Steinhauer, White, and Drury (2009) show that, cross-sectionally, (a) low-proficiency learners show no ERP effects or display an N400 to grammatical violations, (b) intermediate-proficiency learners display a P600 to the same violations, and (c) high-proficiency learners display a native-like P600 or possibly the biphasic AN/P600 pattern. Results from longitudinal studies seem to support the idea that a neural responses to grammatical violations are related to proficiency, and that the N400 reflects an earlier stage of learning, with learners showing no response/N400s at lower levels of proficiency and exposure, and P600s at higher levels of proficiency and exposure (McLaughlin et al., 2010; Morgan-Short et al., 2010; Morgan-Short, Steinhauer et al., 2012). McLaughlin and colleagues (2010) posit that once an aspect of the L2, such as agreement features, has been "grammaticalized" as rule-based knowledge in a learner's real-time processing system, violations of that element should elicit a P600 effect.

Indeed, several studies have found that learners with lower levels of proficiency or grammatical sensitivity to a particular structure show an N400 effect to violations of that structure, whereas learners with greater grammatical sensitivity show a P600 effect (McLaughlin et al., 2010; Tanner, Osterhout & Herschensohn, 2009; Tanner et al., 2013). More recent analyses, however, suggest that N400/P600 variability exists in both L1 and L2 processing (Tanner, Inoue & Osterhout, 2014; Tanner & Van Hell, 2014), and that this individual variability can be hidden by analyses that average across participants. These findings underscore the importance of examining individual ERP patterns in addition to group-level patterns.

Collectively, results from L2 ERP research reveal that processing of grammatical structures poses a significant challenge for L2 learners. As previously discussed, there are a number of learner-internal factors that may help determine where on the N400/P600 continuum learners will fall when processing their L2. The following section describes studies that have treated individual ERP responses as a measure of variability between learners.

2.3.3 Event-Related Potentials as a Measure of Individual Differences

Inter-learner variability is widely evidenced in performance on behavioral measures of L2 abilities; it should not be surprising, then, that this variability extends to L2 processing patterns. Historically, ERP research has treated variability within a group of learners as “noise” that is reduced through averaging processes. Recent analyses, however, propose that this cross-subject variability in ERP data be treated, instead, as a source of evidence that can inform theories of SLA (McLaughlin et al., 2010; Tanner et al., 2013). Researchers have shown quantitative variability in ERP responses in both L1 and L2 research.

The amplitude and onset of ERP effects in L1 processing can be sensitive to individual differences in WM abilities (King & Kutas, 1995; Vos, Gunter, Kolk & Mulder, 2001), as well as to L1 proficiency (Pakulak & Neville, 2010). Qualitative differences in ERP patterns have also been seen among monolinguals under certain conditions, such that some participants show an N400 and others a P600 (Nieuwland & Van Berkum, 2008; Osterhout, 1997; Osterhout, McLaughlin, Kim, Greewald & Inoue, 2004; Tanner & Van Hell, 2014). This pattern among monolingual participants in Nakano, Saron and Swaab’s (2010) study was found to be related to individual differences in WM: individuals with lower WM abilities showed N400 effects to verb-argument animacy violations, whereas those with higher WM abilities showed P600 effects.

As previously discussed, individual differences in processing have been addressed in ERP research using cross-sectional comparisons between different groups of L2 learners that represent particular internal factors of interest, such as L1-L2 similarity (Sabourin & Stowe, 2008; Sabourin, Stowe, & de Haan, 2006) or L2 proficiency (Bowden et al., 2013; Ojima et al., 2005; Rossi et al., 2006). Relatively new to SLA research is the use of regression analysis to examine individual differences in processing in relation to internal factors, a practice commonly used in behavioral research. A limited number of studies have used such techniques to examine semantic processing in L2 and modulations of N400 in L2 (Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville & Ullman, 2012; Ojima, Matsuba-Kurita, Nakamura, Hoshino & Hagiwara, 2011), as well as to examine the effects of individual differences in relation to morphosyntactic processing patterns in an L2 (Bond, Gabriele, Fiorentino & Alemán Bañón; 2011; McLaughlin et al., 2010; Tanner et al., 2013; Tanner et al., 2014). A detailed discussion of the recent studies of morphosyntactic processing signatures as a measure of individual differences, along with the insights and conclusions drawn from this body of research, is provided below.

In an ERP investigation of morphosyntactic processing in L2 Spanish, Bond, Gabriele, Fiorentino and Alemán Bañón (2011) addressed the impact of individual differences in verbal and nonverbal aptitude and proficiency on grammatical structures that represented different levels of (dis)similarity from the L1. The target structures included three types of agreement: (a) subject-verb agreement, (b) noun-adjective number agreement and (c) noun-adjective gender agreement. Participants were 24 L1-English speakers enrolled in fourth-semester Spanish classes at an American university, all of whom scored in the "low proficiency" range on an objective measure of Spanish proficiency, as well as a group of 8 native Spanish speaker controls. In order to assess individual differences in aptitude, learners completed a standard test of verbal aptitude (Short Version of the Modern Language Aptitude Test, MLAT; Carroll & Sapon, 1959) as well more domain-general test of nonverbal intelligence and reasoning skills (The Raven's Advanced Progressive Matrices). In a second experimental session, participants read Spanish sentences presented one word at a time on a computer screen and completed a grammaticality judgment task while ERP data were collected.

Behavioral results showed that both the native speakers and learners were sensitive to all three types of agreement violations, but that the learners' acceptance rate for ungrammatical noun-adjective gender agreement sentences was significantly higher than that of the native speakers. In terms of ERP results, grand mean averages for each condition were analyzed, revealing a P600 for both the native speaker and L2 groups for each of the three violation types. These effects were qualitatively similar for the subject-verb or noun-adjective number agreement conditions; for the noun-adjective gender agreement condition, however, the distribution of the P600 effect differed by group. For native speakers, the effect was right-lateralized, and for L2 learners, it was left-lateralized.

The researchers then examined relationships (1) between aptitude (verbal and nonverbal) and proficiency, and (2) between aptitude and P600 effect size. Verbal aptitude was indexed by scores on each of the three MLAT subtests, as well as a participant's overall score, and nonverbal aptitude was quantified as the participant's score on the Raven's test. Proficiency was operationalized as sensitivity to agreement violations in the L2 stimuli, as reflected by *d'* scores for each agreement type; P600 effect size was calculated based on mean amplitude differences in the 450-950 ms time window in posterior regions for each violation type.

Results of correlational analyses revealed a significant correlation between verbal aptitude and *d'* scores for subject-verb agreement, as well as a marginally significant correlation between verbal aptitude and *d'* scores for noun-adjective number agreement. The relationship between verbal aptitude and noun-adjective number agreement was also reflected in a significant correlation with P600 effect size for that violation type.

This study contributes to our understanding of L2 morphosyntactic processing, and provides favorable evidence regarding the use of ERPs to investigate the role of individual differences in aptitude in L2 processing. The researchers call for further research into the effects of individual differences on brain responses in L2 learners, suggesting that such research may ultimately inform pedagogical approaches to SLA.

Tanner, McLaughlin, Herschensohn and Osterhout (2013) further promote the consideration and analysis of cross-subject variability in ERP data and advance the statistical procedures that can be used to analyze the relationship between individual differences and neural responses. The researchers performed an investigation of morphosyntactic processing within L1 English learners of L2 German enrolled in first-year (n = 20) and third-year (n = 13) university-level language courses, as well as a group of native German speakers (n = 13). The researchers were interested in whether amount of instruction in German and sensitivity to subject-verb agreement errors would be related to ERP effects. Participants read German sentences presented one word at a time on a computer screen and completed a grammaticality judgment task while ERP data were collected.

Behavioral results revealed no behavioral differences between learner groups: all learners performed quite well on the grammaticality judgment task, although accuracy for the native speakers was significantly higher than for both learner groups. In order to analyze the ERP data, the authors first performed a grand mean analysis for each of the three groups. Both the native speakers and third-year learners showed a P600 effect for subject-verb errors. Within the first-year learner group, a broadly distributed N400 as well as a small P600 were found, although the latter effect did not reach full significance. Visual inspection of ERP waveforms for individuals in the first-year learner group, however, did not seem to support the biphasic response revealed by the grand mean analysis. Although some learners did show the biphasic response, there was a continuous distribution between N400-dominant and P600-dominant responses, with most learners showing *either* an N400-like negativity *or* a P600-like positivity in response to the ungrammatical verbs.

This individual variation was investigated by calculating the magnitude of the N400 and P600 effects for each individual by subtracting the mean amplitude of the ungrammatical condition waveform from the grammatical condition waveform (order of subtraction reversed for P600 magnitude) in the time window of interest (300-500 ms for the N400, 500-800 ms for the P600) and averaging across midline electrodes. Consistent with previous research (McLaughlin et al., 2010), the magnitudes of the two effects were significantly negatively correlated, such that for each participant, as one response increased, the other decreased. First-year learners were then divided into two groups based on whether their response was N400-dominant (n = 9) or P600-dominant (n = 11); grand mean waveforms for each of these groups revealed a statistically significant N400 or P600, respectively, demonstrating that the biphasic response seen in the overall first-year learner grand mean waveform was actually the result of averaging over individuals with qualitatively different brain responses to the morphosyntactic violation. The next set of analyses addressed the factors that may have predicted the type and magnitude of brain response individuals had to the disagreeing verbs.

Across L2 learners, the magnitude of the P600 effect (calculated over electrode Pz, a posterior, midline electrode) correlated with *d'* scores from the acceptability judgment task, revealing that P600 effect magnitudes increased with learner ability to detect agreement violations. A weak, but statistically significant, correlation was also found between N400 effect magnitude and performance, such that enhanced negativities were related to poorer performance on the judgment task. No significant relationships were found between hours of instruction in

German and either effect magnitude, suggesting that individual brain responses were related to grammatical learning rather than classroom exposure to the L2.

The results of this cross-sectional study provide additional evidence of variability between L2 learners who have had the same amount of language exposure and instruction, and illustrate the informative capabilities of ERPs as a tool in investigating this inter-learner variability. The authors suggest that these data provide evidence that L2 learners rely on qualitatively different processing mechanisms at different stages of development, and that language proficiency may be one factor in determining the neural substrates of syntactic processing.

Tanner, Inoue and Osterhout (2014) considered the impact of individual differences in L2 related background variables, such as age of exposure to the L2, age of arrival to the L2 environment, length of residence in the L2 setting, and motivation to sound like a native speaker in determining neural response to L2 morphosyntactic violations. Data from 20 L1 Spanish speakers who had moved to an English-speaking country at age 15 or later, and had at least five years of L2 English immersion were analyzed. In addition to the experimental acceptability judgment task, participants completed a language background and contact questionnaire, motivation questionnaire, and standardized English proficiency test.

Group-level analyses revealed a biphasic N400-P600 response that was not reflected in the majority of individual waveforms. Following Tanner et al. (2014), N400 and P600 effect magnitudes were calculated. The N400 and P600 effect magnitudes were negatively correlated, revealing a continuous distribution of neural responses. In addition to effect magnitude, Tanner and colleagues introduced two new measures of individual neural response: Response Dominance Index (RDI; relative response type – N400 or P600) and Response Magnitude Index

(RMI; overall size of neural response over the N400 and P600 time windows). These measures allow for investigation into the relationship between individual difference variables and type of neural response (RDI) and overall level of sensitivity to violations (RMI).

Regression analyses indicated that L2 proficiency predicted RMI, but did not predict N400 or P600 effect magnitude, suggesting that higher proficiency was related to greater neural sensitivity to agreement violations, but not to a particular type of response. Brain response dominance (RDI) was predicted by age of arrival to the L2 environment and motivation to sound like a native speaker. Specifically, learners who reported earlier immersion in an English-speaking environment and higher motivation to sound like native speakers of English showed stronger P600-dominant neural responses, whereas those with later onset of immersion and lower motivation were more likely to show N400-dominant responses.

Taken together, the results from these studies indicate that ERPs provide evidence of systematic, meaningful variation within groups of L2 learners and are sensitive to individual differences among L2 learners, including verbal aptitude (Bond et al., 2011), hours of instruction (McLaughlin, Osterhout & Kim, 2004), grammatical sensitivity (McLaughlin et al., 2010; Tanner et al., 2013), age of arrival to an L2 environment (Tanner et al., 2014), and motivation to sound like a native speaker (Tanner et al., 2014). Collectively, these findings illustrate the potential for ERPs to be used as an index of cross-subject variability, and inspire future investigations to measure this variability and interpret it with respect of individual differences. The assessment of additional internal and external variables could complement the findings of this study, and provide valuable insights into the factors that affect L2 development and processing.

2.3.4 Event-Related Potentials: Questions and Considerations

Although natural language ERP studies have not yet systematically examined the relationship between different exposure conditions and L2 processing, some studies have limited their L2 participant pool to those who have had exposure to the L2 in natural, immersion settings (Gillon Dowens, et al., 2010; Sabourin & Stowe, 2008). In addition, ERPs have been collected from L2 learners immersed in an L2 context at the time of testing (study abroad or learners residing in a target-language country; e.g., Foucart & Frenck-Mestre, 2011, 2012; Frenck-Mestre, et al., 2009; Hahne, 2001; Hahne, & Friederici, 2001; Hahne, Mueller, & Clahsen, 2006; Sabourin et al., 2006; Sanders & Neville, 2003; Tanner et al., 2014; Weber & Lavric, 2008; Weber-Fox & Neville, 1996) as well as from those in an L1 context (at home; e.g., Gillon Dowens et al., 2011; Tanner et al, 2013; Tokowicz & MacWhinney, 2005). Furthermore, there is evidence that type of exposure can impact ERP responses elicited during morphosyntactic processing (Morgan-Short et al., 2010; Morgan-Short, Steinhauer et al., 2012). Some insight into how context may influence L2 processing can be gained by comparing results from independent studies conducted in different contexts, but a systematic examination of processing changes that occur over the course of a semester or year of study in at home and study abroad contexts is needed in order to provide a faithful assessment of the relationship between learning context and L2 processing changes.

Recent research suggests that a more fine-grained approach to the analysis of ERP data can reveal brain responses for individuals that differ from grand-mean waveforms (e.g., McLaughlin et al., 2010; Tanner et al., 2013; Tanner et al., 2014; Tanner & van Hell, 2014). The use of ERPs as a measure of individual differences in a study that addresses internal factors beyond L2 proficiency, such as WM and language contact, and external factors, such as context of learning, could provide important insights into the factors that influence the neural representation of L2.

2.4 Motivation and Research Questions for the Study

Despite suggestions that a comprehensive theory of SLA must consider both external and internal factors that impact acquisition (Collentine & Freed, 2004) and empirical findings that indicate complex relationships between external and internal factors in L2 development (e.g., Segalowitz et al., 2004; Sunderman & Kroll, 2009), relatively little research has investigated the specific effects of learner-internal factors on L2 development in naturalistic contexts of learning.

Independent analyses of the relationships between factors such as language contact, WM, and context of learning and L2 development are relevant and interesting, but it appears that a more informative approach is to consider how these factors interact to influence L2 acquisition. Consideration of multiple factors together in SLA research reflects the reality that L2 acquisition occurs in complex, diverse settings, with great variability in learners themselves, in their L2 experiences, and in opportunities for L2 contact.

Although it is essential that particular factors be examined in controlled settings, it is also critical that research in more naturalistic settings be conducted in order to maximize the ecological validity of findings in the field. Currently lacking from the field of SLA is research in real-world settings that takes into account multiple individual difference variables and goes beyond behavioral measures of L2 development by examining changes in L2 processing. Accordingly, the current study considers the role learning context (at home and study abroad) and individual differences in language contact and WM in explaining linguistic change (in both behavior and processing), in order to provide an account that examines language learning as a complex interaction among external and internal variables.

Informed by the fields of SLA and cognitive neuroscience, the current study attempts to address open questions that emerge from both fields. The study also considers limitations inherent in each field, employing a research design developed to minimize such limitations. This study investigates whether the internal factors language contact and WM contribute to linguistic development in L2 Spanish among learners in at home and study abroad contexts. The current study contributes to existing literature and addresses open questions in the fields of SLA and cognitive neuroscience in several ways. First, the study employs a longitudinal design, allowing for an experimental assessment of changes in L2 abilities that occur during a semester of study in at home or study abroad contexts. This design also allows for within-subjects analyses, which have been rare in both SLA and cognitive neuroscience. Second, the study examines L2 development among learners engaged in study in natural, authentic settings, maximizing the ecological validity of the design and relevance of outcomes to applied questions. Finally, assessment of L2 abilities is enhanced by the use of both behavioral and neurocognitive measurements within subjects. Given this design, research outcomes are expected to elucidate the possible effects of individual differences in L2 contact and WM in distinct, natural contexts of learning on behavioral and neurocognitive aspects of L2 abilities.

2.4.1 <u>Research Questions</u>

Given the study's goal of investigating whether individual differences in language contact and working memory contribute to L2 development among learners in at home and study abroad settings, the following specific research questions are addressed:

RQ1a: Does language contact account for behavioral development over the course of one semester for learners in an at home or study abroad context?

RQ1b: Does working memory ability account for behavioral development over the course of one semester for learners in an at home or study abroad context?

RQ2a: Does language contact account for changes in online processing of morphosyntactic structures over the course of one semester for learners in an at home or study abroad context?

RQ2b: Does working memory ability account for changes in online processing of morphosyntactic structures over the course of one semester for learners in an at home or study abroad context?

2.4.2 <u>Hypotheses</u>

With regard to Research Question 1a, the existing work on the relationship between L2 contact and L2 grammatical development has primarily focused on learners in a study abroad context (Freed et al., 2004; Isabelli-Garcia, 2010; Magnan & Back, 2007; Segalowitz et al., 2004). Although the findings regarding the specific relationship between language contact and linguistic gains are somewhat conflicting, multiple studies have found a positive correlation between language use and behavioral development (e.g., Freed et al., 2004; Isabelli-Garcia, 2010; Segalowitz et al., 2004). These results, coupled with the popular claim that interactions with native speakers are the key to success in a study abroad context (Freed, 1995b), lead to the prediction that overall language contact will account for a behavioral development for learners in the study abroad context. Context-based research to date has not found a relationship between outside-of-class language contact and linguistic development in the at home context; the results of this study will provide new insights into this relationship. The relationship between L2 contact and behavioral gains may also vary by type of behavioral measure, such that L2 contact accounts for more change in production abilities than grammatical sensitivity or overall proficiency (e.g., Magnan & Back, 2007).

H1a: Language contact will account for behavioral gains for learners in both at home and study abroad contexts.

With regard to Research Question 1b, a number of previous studies provide evidence of a positive relationship between working memory and grammatical development among learners in an at home context (Sagarra, 2007; Sagarra & Herschensohn, 2010), and studies have also shown that working memory facilitates the use of redundant morphological cues for learners with SAE, but not for learners without (LaBrozzi, 2009). Based on previous research, the following hypotheses is made:

H1b: Working memory ability will account for behavioral gains for learners in both at home and study abroad contexts. Working memory will account for more variation in behavioral change among learners in a study abroad context.

With regard to Research Question 2a, this is a paucity of longitudinal research that examine changes in ERP processing signatures upon which to base predictions. Indeed, there have been no experimental investigations of the relationship between context of learning and changes in neural processing or of the relationship between L2 contact and processing changes using ERPs. Within-subject design ERP studies, however, have shown that with increased L2 exposure, learner neural responses to morphosyntactic violations can change from no effect or an N400-effect to a P600 effect (McLaughlin et al., 2010; Morgan-Short et al., 2010). In line with this work, the following prediction is made:

H2a: Language contact will account for a positive change in type of neural response (more positive RDI) for learners in both at home and study abroad contexts.

With regard to Research Question 2b, empirical studies reveal a relationship between working memory and neural signatures, particularly in L1, where the amplitude, latency, and distribution of ERP components can be sensitive to individual differences in working memory abilities (Fiebach et al., 2002; King & Kutas, 1995; Vos, Gunter, Kolk & Mulder, 2001). Indeed, results from Nakano, Saron and Swaab's (2010) study of L1 processing indicate that lower working memory ability speakers may show an N400 effect to violations that elicit a P600 effect from higher working memory ability speakers. Within L2 research, a positive relationship between working memory and online processing of morphological cues has been found using eye-tracking methodology, and indeed, this relationship was instantiated for learners with SAE but not for learners without (LaBrozzi, 2009).

H2b: Working memory ability will account for positive change in size (increase in RMI) and type (more positive RDI) of neural response for learners in both at home and study abroad contexts. Working memory ability will account for more variation in processing changes among learners in a study abroad context.

Regardless of the direction of the findings, the results of the present study will contribute to open questions that emerge from various areas of adult L2 acquisition research. The study design provides an experimental assessment of L2 development in an at home and study abroad context, addressing a call in the literature to assess interactions between cognitive abilities and type of exposure to the L2 (Segalowitz & Freed, 2004; Williams, 2012). Specifically, results will contribute to the literature that has found evidence of a relationship between working memory abilities and language experience in natural settings (LaBrozzi, 2009; Sunderman & Kroll, 2009; Tokowicz et al., 2004), as well as to debates regarding the importance of language contact to L2 development (e.g., Freed, Segalowitz, & Dewey, 2004; Magnan & Back, 2007).

Results from this study will also contribute to questions of neural processing of features unique to the L2 (e.g., Bond et al., 2011; Frenck-Mestre et al., 2009; Gillon Dowens et al., 2010; Gillon Dowens et al., 2011; Morgan-Short et al., 2010; Tokowicz & MacWhinney, 2005) and to questions about the role of internal and external factors in SLA (e.g., Ellis & Larsen-Freeman, 2009; Norris & Ortega, 2000, 2001; Sanz, 2005; Spada & Tomita, 2011). This research has implications for theoretical models of adult L2 acquisition that posit a role for cognitive abilities in L2 development and processing (Clahsen & Felser, 2006; Hopp, 2007), as well as for practical, applied questions related to interactions between learner and context.

3 METHODS

This chapter provides a detailed description of the experimental design of the study. Section 3.1 provides a description of the overall research design. Subsequent sections describe each element of the design in detail in the following manner: the participants are described in Section 3.2 and the target structure (grammatical gender agreement) in Section 3.3. Section 3.4 provides detailed information about the materials and procedures for each experimental session and task, and Section 3.5 explains the statistical analyses that were performed to address each of the research questions.

3.1 Overall Research Design

This study examines the relationship between L2 contact, working memory, and linguistic development in "at home" and "study abroad" language-learning contexts. The study employs short-term longitudinal design, including two pre-semester[4] experimental sessions (comprising cognitive assessments and baseline language assessments), weekly L2 contact surveys completed online throughout the semester of study, and one post-semester experimental session (follow-up language assessments). This design allows for the measurement of changes in L2 Spanish abilities and online processing over the course of 12-18 weeks of university-level study, and for an experimental consideration of L2 behavioral and processing changes with regard to both pre-semester individual differences in working memory (WM) abilities and L2 contact during the semester. The experimental design is represented graphically in Figure 7.

[4] Note that for the At Home group, *pre-semester* sessions also took place during the first two weeks of the semester.

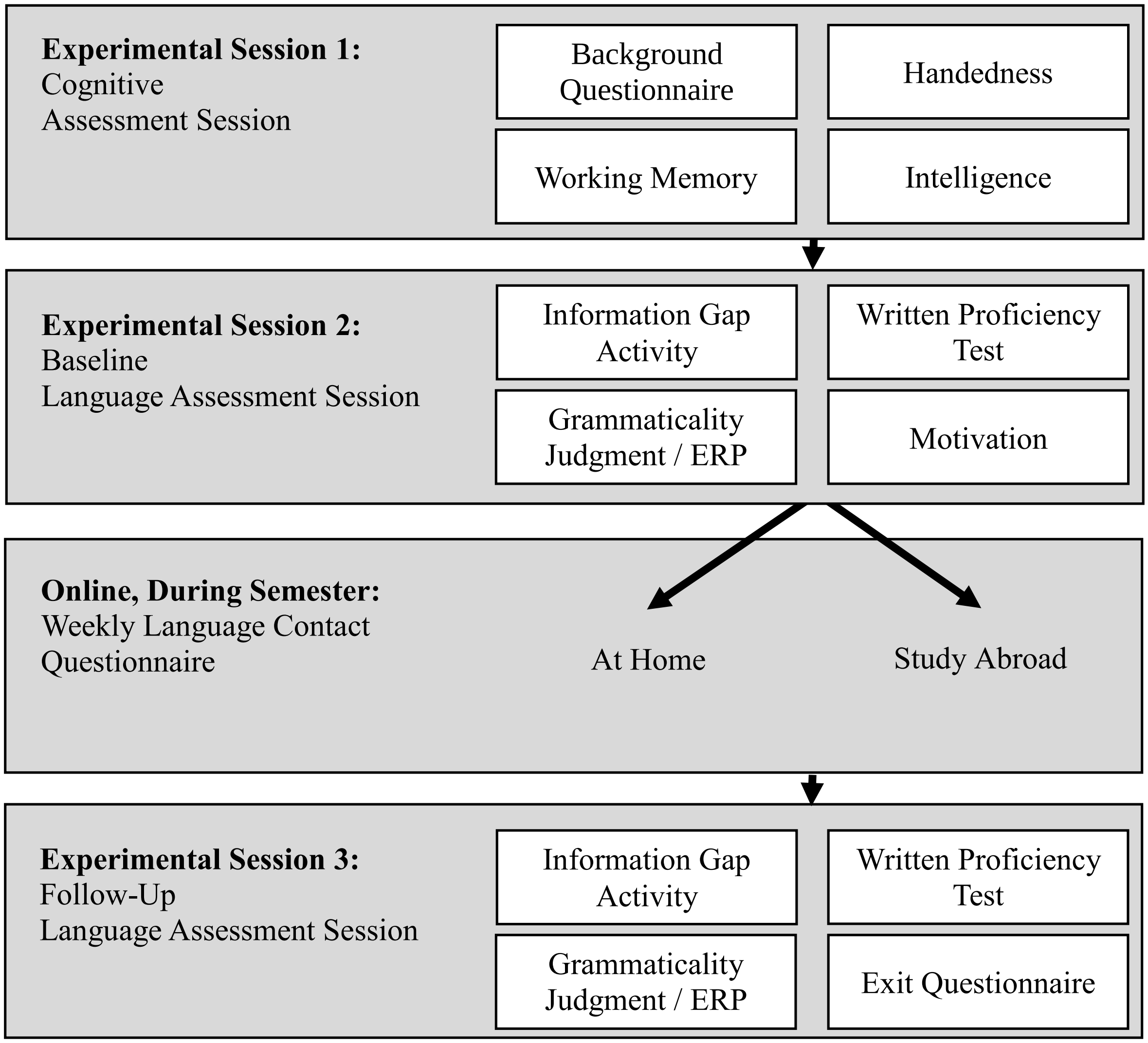


Figure 7. Overview of study design.

3.2 Participants

Participants included 38 native speakers of English studying Spanish as an L2 at the university level. During the semester of study, participants were enrolled in at least one university-level Spanish course at a fifth semester level of above (i.e., intermediate/advanced Spanish courses, including advanced grammar review, introductory linguistics, and literary analysis) at either their home university (“At Home”) or in study abroad programs in Spanish-speaking countries (“Study Abroad”). Participants were screened for language history and experience using a modified version of the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007; the language background questionnaire used in this study is provided in Appendix A) to ensure that they had only classroom experience with Spanish (that is, no significant immersion experience prior to beginning the study). The language requirement was established to ensure that learners in the current study had sufficient L2 knowledge to allow them to make gains during study abroad (DeKeyser, 2010; Golonka, 2006), but who have not yet mastered the target structure: grammatical gender agreement (e.g., Montrul, 2004; Montrul et al., 2008; target structure described in detail in Section 3.3).

All participants were right-handed, as assessed by an abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, and reported no history of drug or alcohol dependence or of psychiatric, neurological, or learning disorders. A total of 14 participants were excluded from analysis for the following reasons: Nine participants failed to complete all experimental sessions (2 Study Abroad), one was excluded after the final session after reporting having spent six weeks in a Spanish-speaking country prior to beginning the study (At Home), and six had excessive artifacts in EEG data (3 Study Abroad; criteria reported in detail in Section 3.5). All participants provided written informed consent to participate at each session, and received monetary compensation for their time. Background information for the participants included in analyses (N = 22; 14 At Home, two male; 8 Study Abroad, all female) is provided in Table II.

Table II

Participant Background Information by Group

Variable	At Home	Study Abroad
Age[a]	20.36 (1.65) 19 – 25	20.88 (1.81) 20 – 25
Years Formal Education[b]	15.36 (1.03) 14 – 17.5	15.44 (.417) 15 – 16
Number of L1s	1.36 (.497) 1 – 2[c]	1.13 (.354) 1 – 2[d]
Number of L2s	1.71(.914) 1 – 4[e]	1.50 (.535) 1 – 2[f]
Age of Exposure – Spanish	13.07 (3.75) 6[g] – 21	11.50 (2.67) 6[g] – 14
Years Formal Instruction – Spanish	6.50 (2.67) 2 – 12	7.63 (2.43) 3.5 – 11
Age of Exposure – First L2	11.43 (2.56) 6 – 15	11.50 (2.67) 6 – 14[h]
Years Formal Instruction – All L2s[i]	7.55 (2.28) 5 – 12	8.00 (2.55) 4 – 11.5
IQ[j]	104.07 (13.42) 86 – 135	112.00 (10.99) 101 – 132
Motivation – Spanish[k]	6.43 (.697) 4.67 – 7.00	6.58 (.812) 4.67 – 7.00

Note. SD provided in parentheses; Range provided below *M* and *SD.*
[a] Age at Session 1.
[b] Years of Formal Education includes kindergarten through completed college semesters.
[c] Five participants in the AH group reported a second native language: Chinese, Gujarati, Igbo, Tagalog, and Urdu.
[d] One participant in the SA group was a native speaker of both English and Polish; all others native speakers of English only.
[e] Additional L2s included ASL (1), French (2), Hindi (1), Italian (1), Japanese (1), Latin (1), Portuguese (1), and Punjabi (1).
[f] Additional L2s included ASL (2) and Korean (1).
[g] One participant in each group reported minimal exposure to Spanish at age 6 in the form of weekly Spanish story hour or similar.
[h] Spanish was the first L2 for all participants in the SA group.
[i] Sum of reported formal instruction in all L2s.
[j] Composite IQ score, calculated during Session 1.
[k] Motivation on a range of 0 – 7, composite of responses to three questions to assess overall motivation to learn Spanish, questionnaire completed at the end of Session 2.

3.3 Target Structure

The target structure assessed in this study is grammatical gender agreement, a morphosyntactic form and feature that is present in the L2 (Spanish) but absent from the dominant L1[5] (English). The Spanish language has a two-gender system (i.e., masculine and feminine) in which determiners and adjectives must agree with the noun they modify. Spanish has (1) distinct canonical endings that mark a noun or adjective as either masculine (*-o*) or feminine (*-a*) and (2) masculine and feminine forms of the determiner (singular forms *el* and *la*, respectively), although there are also irregular forms (e.g., *el lápiz* 'the pencil$_{masc}$', *la canción* 'the song$_{fem}$', *el planet**a*** 'the planet $_{masc}$', *la man**o*** 'the hand$_{fem}$'). Despite its status as an arguably simple linguistic form, grammatical gender agreement is widely known to be difficult for late L2 learners to master (e.g., Montrul, 2004; Montrul, Foote, & Perpiñán, 2008; White, 2003). Indeed, errors in inflectional morphology among L2 learners, particularly for grammatical gender

---

[5] One participant in the Study Abroad group, and two participants in the At Home group reported a second native language (heritage language) that encodes grammatical gender – Polish (Study Abroad) and Gujarati and Urdu (At Home). All three participants had spent their entire lives in the United States, and reported that English was their dominant language in terms of proficiency and amount of use. Given the substantial differences between the target language (Spanish) and the native languages reported (Polish, Gujarati and Urdu), and the lack of evidence for an advantage for heritage speakers in terms of grammatical gender agreement judgments, even in the heritage language (Ranjan, 2013), I decided not to exclude these participants based on their additional L1s. Additional details on grammatical gender in each of these languages are provided below.

Polish has a three-gender system for nouns (masculine, feminine, neuter) and attributive adjectives agree in gender with the noun they modify. Attributive adjectives in Polish usually precede the noun, and Polish does not have articles (Sadowska, 2012). This participant considered English to be her strongest language, but rated her ability to speak and understand Polish as fully-native (5/5), and her ability to read and write in Polish at intermediate (3/5)

Gujarati has a three-gender system for nouns and variable adjectives, which precede nouns, agree in gender with the noun they modify. Gujarati does not have articles, and employs a non-Roman script (Ogilvie, 2009). This participant reported being able to speak and understand Gujarati (5/5), but not read or write (1/5).

Urdu has a two-gender system for nouns, and there is concord with adjectives, although not all adjectives inflect with respect to gender (invariable), and articles do not encode grammatical gender. Urdu employs a non-Roman script. A recent study revealed that heritage speakers of Urdu did not have an advantage over L2 learners on a grammaticality judgment task that probed for implicit knowledge of grammatical gender (Ranjan, 2013). This participant reported being able to speak (4/5) and understand (5/5) Urdu, but not read or write (1/5).

agreement, are widely evidenced in the literature and often persist even among learners who have reached advanced levels of proficiency (e.g., Dewaele & Veronique, 2001; Franceschina, 2005).

Of particular interest is the fact that while many learners seem to fossilize with regard to these errors, other learners overcome such errors (Montrul, 2004). Grammatical gender agreement, therefore, is a target structure that has the potential to allow for the observation of development among learners who have substantial L2 knowledge and experience. Furthermore, it is a structure (1) whose acquisition may be impacted by individual differences between learners and (2) that could be expected to reveal within-learner disparities in terms of production, interpretation, and online processing. The acquisition of L2 Spanish grammatical gender agreement in different learning contexts has been addressed by only one previous study (Isabelli-Garcia, 2010); as such, the role of context of learning in morphosyntactic development is not well understood.

Grammatical gender agreement processing among different groups, in both L1 and L2, is fairly well documented in ERP literature, allowing for comparisons between the present group of learners and previously studied groups, as well as for an extension of this body of literature. Previous studies have found that a variety of factors may affect the processing of grammatical gender agreement among L2 learners: L1-L2 similarity (Bond et al., 2011; Foucart & Frenck-Mestre, 2011, 2012; Gillon Dowens, Guo et al., 2011; Gillon Dowens, Vergara et al., 2010; Tokowicz & MacWhinney, 2005), L2 proficiency (Gabriele et al., 2013; Morgan-Short et al., 2010), amount of exposure to the L2 (Gillon Dowens, et al., 2010), and structural and linear distance between agreeing elements (Alemán Bañón et al., in press; Foucart & Frenck-Mestre, 2011, 2012; Gabriele et al., 2013). Indeed, this body of literature has revealed inter-learner variability in processing of agreement violations (e.g., Bond et al., 2011; McLaughlin et al., 2010).

The present study adds to previous work by addressing the role of (a) context of learning and (b) individual differences in L2 contact and WM in the acquisition and processing of grammatical gender agreement among intermediate learners of L2 Spanish within a longitudinal design. Furthermore, although it is not one of the primary aims of the study, behavioral and ERP data attained can contribute to debates surrounding competing accounts of the role of the L1 in L2 acquisition; specifically, the Interpretability Hypothesis (Tsimpli & Mastropavlou, 2007; Tsimpli & Dimitrakopoulou, 2007), which posits that learners cannot acquire uninterpretable features in the L2 that are not instantiated in the L1 – such as grammatical gender agreement on articles and adjectives in English L1 learners of Spanish – and the Full Transfer/Full Access hypothesis (Schwartz & Sprouse, 1996), which posits that the L1 provides the initial state for L2 acquisition, but that with sufficient exposure and proficiency, learners should be able to acquire features unique to the L2.

3.4 Materials and Procedure

In the following subsections, the procedures followed and materials used for each experimental session and task are described. Section 3.4.1 presents the Cognitive Assessment Session procedure and tasks; Section 3.4.2 describes the procedure and tasks used during the Baseline and Follow-Up Language Assessment Sessions, and Section 3.4.3 provides details regarding the Weekly Language Contact Questionnaire. The specific data used in analyses for each task are described within the relevant subsection.

3.4.1 Cognitive Assessment Session

During the Cognitive Assessment Session (first experimental session), participants completed a language background and experience questionnaire (modified version of the Language Experience and Proficiency Questionnaire; Marian, Blumenfeld & Kaushanskaya, 2007; see Appendix A), handedness questionnaire (modified version of the Edinburgh Handedness Inventory, Oldfield, 1971), three assessments of working memory abilities (described in detail in Section 3.4.1.1), and an intelligence (IQ) assessment (described in Section 3.4.1.2). All participants completed the language background and handedness questionnaires at the beginning of the session, then completed the working memory and IQ assessments along with four additional cognitive assessment tasks (not reported here). The order of the cognitive tests was counterbalanced across participants. The Cognitive Assessment Session was completed in approximately three hours.

3.4.1.1 Working Memory

In order to obtain a measurement of WM ability among the participants, automated versions (experimentally shown to correlate strongly with traditional versions, Unsworth et al., 2005) of three widely-used WM measures were employed during the Cognitive Assessment Session: Operation Span, Reading Span, and Symmetry Span tasks. These tasks are described in detail in the following sections, and represented in Figure 8. Multiple sample trials for each task are provided in Appendix B.

Automated Operation Span Task

The Automated Operation Span Task (Ospan; Unsworth, Heitz, Schrock, & Engle, 2005) is a linguistic task that requires participants to solve math operations (processing) while trying to remember letters that are presented after each math problem (storage). Throughout the task, participants provide all responses using the computer mouse. First, participants complete a three-part practice session, beginning with the storage element of the task: a letter span task. Participants are presented with a series of letters on the computer screen for 800 ms each followed by a 4 x 3 matrix that includes all 12 possible letters (F, H, J, K, L, N, P, Q, R, S, T, and Y). Participants are asked to click the box next to each of the letters that was presented in the order of presentation. Participants then receive feedback on the number of letters correctly recalled from the current set. In the second part of the practice session, participants practice the processing portion of the task by completing 15 math operations. A math operation is presented (e.g., *(1*2) + 1 = ?*) along with the following on-screen instructions: *Click the mouse to continue.* On the next screen, a possible answer for the operation is presented, (e.g., *3*) and participants are required to click on either "true" or "false" depending on whether the value presented matches their answer to the math operation. After each operation, participants receive feedback on their accuracy. Prior to beginning this practice set, participants are instructed to solve the problems as quickly as possible and to click the mouse to advance to the next screen.

This portion of the practice session serves two purposes: (a) to familiarize participants with the math portion of the task and (b) to determine the mean time required for a participant to solve the math operations. The time limit for the math portion of the experimental session for each participant is that individual's mean solution time during math practice plus 2.5 *SD*. The final practice task combines the processing and storage components of the task: participants see and solve a math operation, then indicate whether their answer matches the number provided on the next screen; following this response, a screen with a single letter appears, followed by another math operation, solution, and letter. After a set of math operations and letters has been presented, the participant is asked to recall all of the letters that appeared during that set, in the order of presentation. To prevent participants from rehearsing the letters instead of processing the math operations, if the participant exceeds his or her personal mean solution time (determined during practice), the program automatically advances to the letter presentation and counts that trial as a speed error. Participants complete three practice sets with two trials each before moving on to the experimental sets.

The experimental task consists of a total of 75 letters and 75 math problems, presented exactly as in the final stage of practice. These trials are divided into sets ranging from 3 to 7 trials each, with three sets of each size presented at random for each participant, and approximately half of the solutions provided for the math operations are correct. After letter recall, the participant's overall percentage of correctly solved math operations is presented in the upper right corner of the screen in red font. In the center of the screen, the number of correctly recalled letters and the number of math errors for that set are reported (e.g., *You recalled 3 letters correctly out of 5; You made 3 math errors for this set of trials*).

Five scores are reported at the end of the task: Overall Ospan score (sum of perfect trials, math accuracy and letter recall), total number correct (total number of letters recalled in correct position), speed errors (number of trials that advanced automatically because participant exceeded her personal time limit), accuracy errors (incorrect responses to math operations), and math errors (total number of processing task errors – speed and accuracy combined). The maximum score for each metric (sum of perfect trials, total number of letters recalled in correct position, and total speed/accuracy/math errors) is 75. The task takes around 25 minutes to complete.

Automated Reading Span Task

The Automated Reading Span Task (RSpan; Daneman & Carpenter, 1980; Unsworth et al., 2005) is a linguistic task that requires participants to read sentences and judge whether or not they make sense (processing) while trying to remember letters that are presented after each sentence judgment (storage). Throughout the task, participants provide all responses using the computer mouse. As in the Ospan, the participants complete a three-part practice session, beginning with a letter span task (identical to the OSpan). In the second part of the practice session, participants practice the processing portion of the task by reading and judging 15 sentences. A sentence is presented (e.g., *The prosecutor's dish was lost because it was not based on fact.*) along with the following on-screen instructions: *When you have read the sentence, click the mouse to continue.* When the participant clicks to continue, the prompt *This sentence makes sense* is presented, and the participant is required to click on either "true" or "false" to indicate his or her judgment. After each sentence judgment, participants receive feedback on their accuracy. Prior to beginning this practice set, participants are instructed to work as quickly and accurately as possible.

This portion of the practice session serves two purposes: (a) to familiarize the participants with the sentence judgment portion task and (b) to determine the mean time required for a participant to read and judge the sentences. The time limit for the sentence portion of the experimental session for each participant is that individual's mean solution time during practice plus 2.5 *SD*. The final practice task combines the processing and storage components of the task:

participants read a sentence and determine whether it makes sense, then indicate their judgment on the next screen; following this response, a screen with a single letter appears, followed by another sentence, judgment, and letter. After a set of sentences and letters has been presented, the participant is asked to recall all of the letters that appeared during that set, in the order of presentation. To prevent participants from rehearsing the letters instead of processing the sentences, if the participant exceeds his or her personal mean reading time (determined during practice), the program automatically advances to the letter presentation and counts that trial as a speed error. Participants complete three practice sets with two trials each before moving on to the experimental sets.

The experimental task consists of a total of 75 sentences and letters, presented exactly as in the final stage of practice. These trials are divided into sets ranging from 3 to 7 trials each, with three sets of each size presented at random for each participant. Half of the sentences are nonsensical, created by changing one word (e.g., “dish” from “case”) from an otherwise normal sentence (the unaltered sentences used to create the nonsensical sentences are never presented during the task), and each sentence contains between 10 and 15 words. After letter recall, the participant’s overall percentage of correctly judged sentences is presented in the upper right corner of the screen in red font. In the center of the screen, the number of correctly recalled letters and the number of sentence errors for that set are reported (e.g., *You recalled 5 letters correctly out of 5; You made 0 sentence errors for this set of trials*).

Five scores are reported at the end of the task: Overall RSpan score (sum of perfect trials, sentence judgment accuracy and letter recall), total number correct (total number of letters recalled in correct position), speed errors (number of trials that advanced automatically because participant exceeded her personal time limit), accuracy errors (incorrect responses to sentence judgments), and sentence errors (total number of processing task errors – speed and accuracy combined). The maximum score for each metric (sum of perfect trials, total number of letters recalled in correct position, and total speed/accuracy/sentence errors) is 75. The task takes around 25 minutes to complete.

Automated Symmetry Span Task

The Automated Symmetry Span Task (SymSpan; Kane et al., 2004; Unsworth et al., 2005) is a non-linguistic task that requires participants to perform a symmetry-judgment task (processing) while trying to remember the location of red squares presented within a matrix after each image judgment (storage). Throughout the task, participants provide all responses using the computer mouse. First, participants complete a three-part practice session, beginning with the storage element of the task. Participants are presented with a 4x4 matrix with one of the cells filled in red for 650 ms. After a set of 2 to 4 red squares has been presented, participants are prompted to recall the location of each red square, in order of presentation, by clicking on each location within a blank matrix. After recall, participants receive feedback on the number of squares correctly recalled from the current set. A total of 17 squares are presented for recall during this practice task. In the second part of the practice session, participants practice the processing portion of the task by completing 10 symmetry judgments. Participants are presented with an 88-square matrix with some squares filled in black to form an image (sample images provided in Appendix B); participants decide whether the design is symmetrical about its vertical axis, then click to continue. On the following screen, the prompt *This image is symmetrical* is presented, and the participant is required to click on either “true” or “false” to indicate their judgment. After each symmetry judgment, participants receive feedback on their accuracy. Prior to beginning this practice set, participants are instructed to work as quickly and accurately as possible.

This portion of the practice session serves two purposes: (a) to familiarize the participants with the symmetry judgment portion task and (b) to determine the mean time required for a participant to view and judge the images. The time limit for the symmetry portion of the experimental session for each participant is that individual's mean solution time during practice plus 2.5 *SD*. The final practice task combines the processing and storage components of the task: participants view an image and determine whether it is symmetrical, then indicate their judgment on the next screen; following this response, a screen with a single red square in a matrix appears, followed by another image, symmetry judgment, and red square. After a set of images and red squares has been presented, the participant is asked to recall the locations of all of the red squares that appeared during that set, in the order of presentation. To prevent participants from rehearsing the locations of the red squares instead of processing the images for symmetry judgments, if the participant exceeds her personal mean solution time (determined during practice), the program automatically advances to the red square presentation and counts that trial as a speed error. Participants complete three practice sets with two trials each before moving on to the experimental sets.

The experimental task consists of a total of 42 symmetry judgments and 42 red squares, presented exactly as in the final stage of practice. These trials are divided into sets ranging from 2 to 5 trials each, with three sets of each size presented at random for each participant, and the pattern is symmetrical approximately half of the time. After recall, the participant's overall percentage of correct symmetry judgments is presented in the upper right corner of the screen in red font. In the center of the screen, the number of correctly recalled red square locations and the number of symmetry judgment errors for that set are reported (e.g., *You recalled 2 squares correctly out of 3; You made 0 symmetry errors for this set of trials*).

Five scores are reported at the end of the task: Overall SymSpan score (sum of perfect trials, symmetry judgment accuracy and square recall), total number correct (total number of squares recalled in correct position), speed errors (number of trials that advanced automatically because participant exceeded her personal time limit), accuracy errors (incorrect responses to symmetry judgments), and symmetry errors (total number of processing task errors – speed and accuracy combined). The maximum score for each metric (sum of perfect trials, total number of letters recalled in correct position, and total speed/accuracy/symmetry errors) is 42. The task takes around 25 minutes to complete.

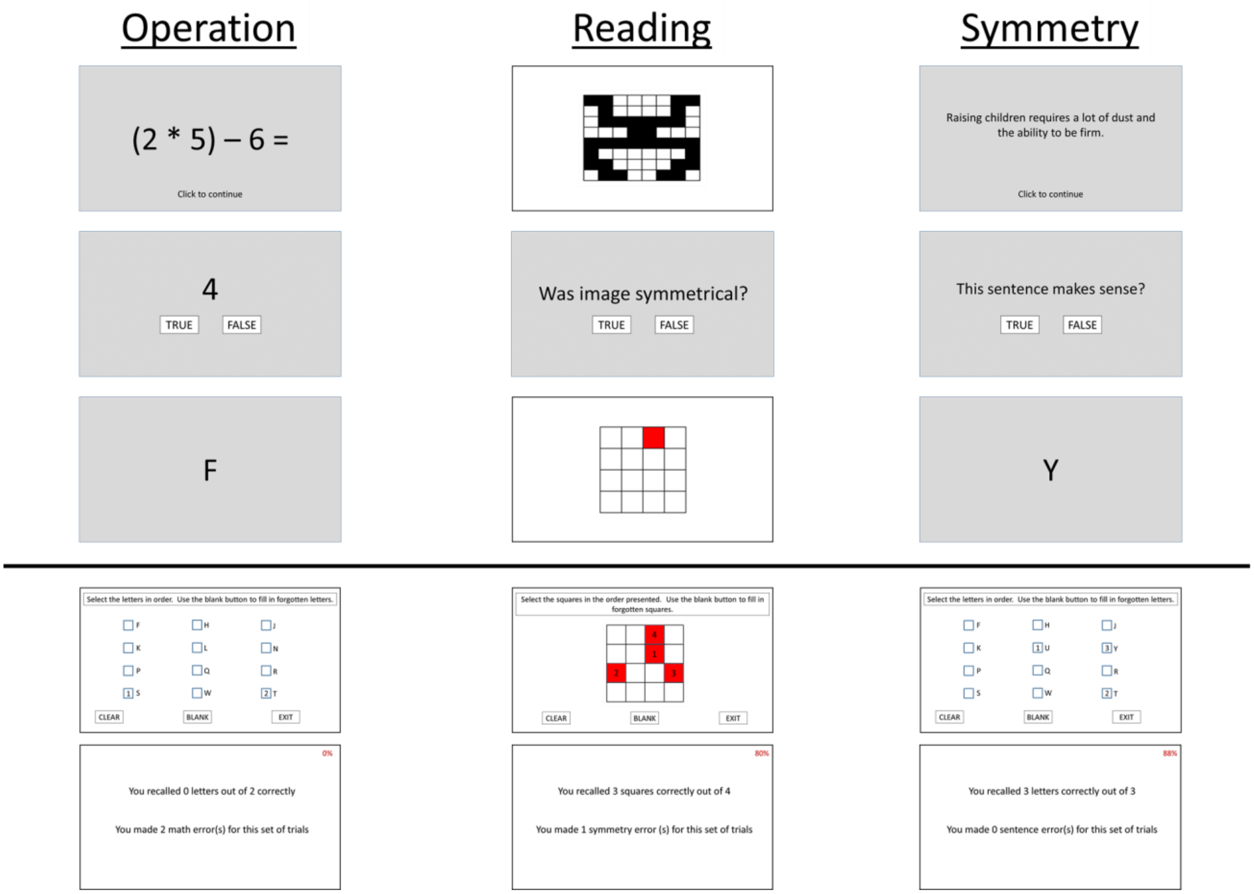


Figure 8. Overview of experimental trials for working memory tasks. For each task, the first three slides show one processing-storage sequence and the last two slides show the recall and feedback screens at the end of each trial.

These three tasks represent "complex span tasks" that engage both processing and storage elements of WM (Unsworth et al., 2005). Performance on these three tasks is posited to reflect individual differences in executive attention (Engle & Kane, 2004; Kane et al., 2007; Redick et al., 2012), which includes both memory and attention abilities (Unsworth & Engle, 2007). These tasks measure learner ability to maintain information related to a goal (temporarily) in primary memory (e.g., solve a math operation) *and* to store and retrieve information from secondary memory (e.g., continually memorize and recall an ordered sequence of letters) under conditions that require the learners to switch quickly between primary and secondary memory during each trial (Redick et al., 2012). The complex span tasks, unlike typical "attention control" tasks, include a clear memory component (the recall task). In order to ensure assessment of WM ability (or executive attention) rather than attentional control alone, 'absolute' scoring (Overall Score), which takes into account sets with correct responses on both processing and recall portions, was used in analyses. Specifically, the overall OSpan, RSpan, and SymSpan scores were converted to z-scores (given the disparity between maximum score between the OSpan/RSpan – 75, and SymSpan – 42); these z-scores[6] were then averaged to calculate a composite WM ability score for each participant.

3.4.1.2 <u>Intelligence</u>

Researchers have identified overall intelligence (IQ) as a factor that may impact L2 acquisition (e.g., Robinson, 2012; Skehan, 2008). Participant IQ was assessed during the Cognitive Assessment Session using the Kaufman Brief Intelligence Test (Kbit-2; Kaufman &

---

[6] Z-score calculations were performed within each group (i.e., At Home and Study Abroad), such that the average and standard deviation used in the z-score calculation [(Participant Score – Average Score) / Standard Deviation] reflects the values for each group, rather than the entire participant pool. Furthermore, only participants included in analyses (reported in Section 3.2) are included in these calculations.

Kaufman, 2004), which provides a measure of crystallized (verbal) knowledge as well as fluid (nonverbal) abilities. In this task, crystallized knowledge is measured through verbal knowledge items and word riddles; fluid IQ or nonverbal abilities are measured through a series of matrices problems that require the participant to solve visual analogies. The test requires 20 to 30 minutes to administer, and results in standardized verbal, nonverbal, and composite IQ scores that can be taken into account during analysis in order to more accurately assess the contributions of the predictor variables on linguistic development. Composite IQ scores for each group are reported in Section 3.2 (Table II).

3.4.2 Language Assessment Sessions

During the Baseline and Follow-Up Language Assessment Sessions (second and third experimental sessions, respectively), participants first completed a Spanish sentence repetition task (not reported here) to help them "warm up" to speaking Spanish. Next, participants completed a communicative task (Information Gap Activity, Section 3.4.2.1), followed by a written proficiency test (Section 3.4.2.2). After completing these tasks, participants completed a grammaticality judgment task (GJT, Section 3.4.2.3) while EEG data were collected (Section 3.4.2.4). At the end of the Baseline session, participants completed a motivation questionnaire (Section 3.4.2.5), and at the end of the Follow-Up session, an exit questionnaire was administered (Section 3.4.2.6). Each of the Language Assessment Sessions was completed in approximately three hours.

3.4.2.1 **Information Gap Activity**

In order to measure accurate use of grammatical gender agreement in spoken production, participants completed a communicative, information gap task (Info Gap) designed to elicit the production of gender agreement on articles and adjectives (adapted from Leeman, 2003). Participants were given a picture of a kitchen containing items of varying sizes (small, medium, large) and colors (red, yellow, black, pink, purple). The researcher, in turn, had a matching, but blank, kitchen, as well as nine options for each of the items in the participant's kitchen (each item in three colors and three sizes), placed in loose piles on a shared desk. Participants were instructed to describe their kitchen aloud to the researcher, using complete sentences in Spanish, so that the researcher could make her kitchen match the participant's kitchen. When necessary, the researcher prompted the participant to provide more information about a particular item without providing any clues regarding the grammatical gender of the item, e.g., *¿Cuál gato?* 'Which cat?' No feedback was given at any point during the task.

In order to control for vocabulary knowledge, participants were provided with a vocabulary sheet containing the names and pictures of all 16 possible target nouns and a list and translation of colors, size words, copular verbs, prepositions (above, below, etc.), and the names of non-target items in the kitchen (counter, table, etc.). In order to avoid gender agreement priming, all nouns on the vocabulary sheet (target items and general kitchen items) were presented without articles. Target nouns were presented in Spanish only, under a picture of the item. Non-target kitchen items as well as adjectives were presented in English with the Spanish equivalent (e.g., *counter – encimera*). All adjectives appeared in masculine, singular form on the vocabulary sheet. A sample participant and researcher kitchen and the vocabulary sheet used during this task are provided in Appendix C.

Four versions of the participant kitchen were used, with a different version used during Baseline and Follow-Up Language Assessments for each participant. Each kitchen contained 12 experimental items drawn from a bank of 16 canonical Spanish nouns (8 masculine ending in *-o*, 8 feminine ending in *-a*). The critical adjectives comprised five canonically agreeing colors and three sizes, two of which have canonical endings and a third that is invariable (*grande* 'large').

The entire activity was recorded using a digital voice recorder. All of the recordings were transcribed independently by two raters; any discrepancies were resolved by a third rater to create a final transcription for each participant. These transcriptions were separated into noun phrases containing target nouns (any of the 16 items included in the kitchens) and scored, independently, by two raters. Again, a third rater resolved any discrepancies. In both transcription and scoring, inter-rater reliability was 100% between two raters (that is to say, when raters one and two did not correspond, there were no instances where the third rater transcription and score did not match one of the first two raters). Scoring consisted of coding each participant utterance containing a target noun for (1) use and (2) accuracy of articles and gender-marked size and color adjectives,[7] such that each participant's proportion of correct usage could be calculated. Separate scores were calculated for article-noun and adjective-noun (collapsing across size and color) agreement. A sample transcription and simplified scoring rubric is represented in Table III.

[7] Following Leeman (2003), invariable adjectives such as *grande* 'large' or *rosa* 'pink' were not analyzed; false starts were also excluded (e.g., for a participant utterance such as "*El cebolla – no, la cebolla…*" only the second article would be included in the participant's overall accuracy score.) Accuracy is defined as article/adjective agreement with the target noun, regardless of agreement between article and adjective within a noun phrase.

Table III

Information Gap Activity – Sample Transcription and Scoring

Participant Production (Noun Phrase)	Noun Gender	Article Use	Article Correct	Color Adjective Use	Color Adjective Correct	Size Adjective Use	Size Adjective Correct
El **zapato** negro es a la izquierda de…	m	1	1	1	1	0	
la **piña** rosado, pequeña …	f	1	1	1	0	1	1
cebolla es a la derecha …	f	0		0		0	
el **gato** mediana y amarillo está…	m	1	1	1	1	1	0

Note. Target nouns presented in **bold**; articles and adjectives for scoring are underlined.

3.4.2.2 **Overall Proficiency Test**

In order to provide a measure of overall proficiency, participants completed a modified version of the written *Diploma de Español como Lengua Extranjera* test (DELE; 'Diploma of Spanish as a Foreign Language', Spanish Embassy, Washington D.C.), which examines general knowledge of the L2. Specifically, two 50-point versions of the test were used, each comprising a cloze passage with four options per answer (20 points) and a multiple choice vocabulary and grammar portion (30 points). The DELE is the official accreditation degree of fluency in the Spanish language, issued and recognized by the Ministry of Education, Culture and Sport of Spain. Similar 50-point versions of the DELE have been used in a number of L2 acquisition studies to determine proficiency level for L2 learners (e.g., Bond et al., 2011; Montrul, 2005;

LaBrozzi, 2012; Alemán Bañón et al., 2013), facilitating comparison between the participants in the present study and previous studies.

The order of administration of the two versions of the test was counterbalanced across participants. Each DELE test was independently scored by two raters, with any discrepancies resolved by a third rater. Participant scores reflect number of items correct (maximum score of 50).

3.4.2.3 Grammaticality Judgment Task

During the grammaticality judgment task (GJT), participants read sentences and indicated, using a computer mouse, whether each sentence was "good" or "bad" in Spanish. EEG data were recorded while participants completed this task (discussed in Section 3.4.2.4).

The GJT includes five violation types – two experimental conditions (article-noun and noun-adjective grammatical gender agreement) and three distractor conditions (subject-verb agreement, semantic, and phrase structure). Each condition comprises 60 violation and 60 matched, correct control sentences, yielding a total of 600 stimuli sentences. Two stimuli lists were created using a Latin square design such that (1) only one version (violation or correct) of each sentence was included in each list, and (2) participants read and judged 300 sentences during each session (half correct). The order of administration of the two lists (Baseline or Follow-Up Language Assessment Session) was counterbalanced across participants. Stimuli sentences in each list are presented in a pseudo-randomized order such that no more than three correct or violation sentences appear together, nor more than three stimuli from the same condition. Stimuli were divided into five blocks, containing 60 sentences each (half correct, balanced across all conditions).

In order to ensure participant familiarity with the vocabulary, all words used in the stimuli sentences appear in at least one of two introductory Spanish language textbooks (*Dicho y hecho,* 8th ed., 2008; *Sol y viento,* 2nd ed., 2009). These words represent vocabulary that is typically covered during the first three semesters of university-level Spanish study.

3.4.2.3.1 Experimental Stimuli

The experimental stimuli conditions are designed to assess participant sensitivity to grammatical gender agreement violations on articles and adjectives, respectively. In these two conditions, half of the target nouns (that is, nouns that trigger agreement on critical articles and adjectives in the stimuli sentences) are masculine and half are feminine. Following previous studies of grammatical gender agreement, all target nouns are inanimate and have canonical endings (e.g., Alemán Bañon et al., in press; Bond et al., 2011; Gabriele et al., 2013; Gillon Dowen et al., 2010; Gillon Dowen et al., 2011). All sentences are written in the third-person form and include names that do not end in canonical masculine (-*o*) or canonical feminine (-*a*) endings (e.g., Felipe, Juan, Pilar, Inés). The semantic gender of the names is balanced across conditions and across grammatical gender of the critical noun.

In the Adjective condition, in order to avoid that potential confound of receiving additional information related to grammatical gender prior to a post-nominal adjective, the non-gender-matched, third-person possessive determiner *su* 'his/her' (invariable) is used in place of the gender-matched article. In both the Article and Adjective conditions, no additional adjectives that encode grammatical gender are included in the sentences, such that the only grammatical gender agreement cues present in a stimuli sentence are (1) the target noun and (2) the critical article or adjective. These controls are designed to eliminate the possible influence of non-experimental grammatical gender agreement cues on GJT responses. Additionally, as these stimuli were designed for use during EEG data collection, critical nouns and adjectives are always preceded and followed by at least three words to avoid initial word effects and wrap-up effects (Hagoort, 2003; Osterhout, Holcomb, & Swinney, 1994; Osterhout & Nicol, 1999; see Section 3.4.2.4 for additional information).

In the Article condition, agreement violations were created by replacing the singular, definite article (*el* or *la*) with the opposite gender article (e.g., *la*$_{fem}$ *calculadora*$_{fem}$ versus **el*$_{masc}$ *calculadora*$_{fem}$ 'the calculator'). Sentences in this condition are between 7 and 15 words in length.

In the Adjective condition, agreement violations were created by changing the gender of an attributive adjective, which appears directly following the noun (e.g., *calculadora*$_{fem}$ *rota*$_{fem}$ versus *calculadora*$_{fem}$ **roto*$_{masc}$ 'calculator broken'). The same 30 target nouns were included in the Article and Adjective conditions; each noun is used twice in each condition (see Table IV for example stimuli, and Appendix D for a full list of stimuli). The noun-adjective condition includes a total of 15 critical adjectives, each used with two masculine and two feminine nouns. Sentences in this condition are between 8 and 15 words in length.

Table IV

GJT Experimental Stimuli

Condition	**Experimental Stimuli**
Article	*Según Montse *el / la* ***<u>falda</u>*** *que lleva cuando hace frío es de lana.* 'According to Montse *the$_{masc}$ / the$_{fem}$ skirt$_{fem}$ that (she) wears when it is cold is made of wool.' *Maribel piensa que *el / la* ***<u>falda</u>*** *que lleva cuando sale le queda bien.* 'Maribel thinks that *the$_{masc}$ / the$_{fem}$ skirt$_{fem}$ that (she) wears when (she) goes out fits her well.'
Adjective	*Luz lleva su <u>falda</u> **___sencillo/sencilla___ *cuando sale con sus amigos.* 'Luz wears her *simple$_{masc}$ / simple$_{fem}$ skirt$_{fem}$ when (she) goes out with her friends.' *Beatriz se pone su <u>falda</u> **___largo/larga___ *cuando hace frío.* 'Beatriz puts on her *long$_{masc}$ / long$_{fem}$ skirt$_{fem}$ when it is cold.'

Note. **Bold** typeface marks the critical word, that is, the word where violation becomes evident in each sentence. ERPs are time-locked to the onset of the critical word. The word that constitutes the violation is indicated with an asterisk* (note that for Article condition, the article is the word that constitutes the violation, but the violation is not apparent until the target noun). For demonstration purposes, the target noun is underlined.

3.4.2.3.2 <u>Distractor Stimuli</u>

The distractor stimuli serve to discourage participants from detecting the linguistic form under investigation. As with the experimental stimuli, these sentences are all in the present tense. Distractor stimuli conditions are described below and examples are presented in Table V.

Semantic. Violation sentences in the semantic condition are created by replacing a noun in a correct sentence with a semantically odd noun. Each semantically odd noun is matched in gender and number with the corresponding noun in the correct, control sentence to avoid agreement violations with articles and/or adjectives. To control for frequency, each noun used in a correct sentence is also used in a violation sentence. Sentences in this condition are between 7 and 12 words in length.

Phrase structure. Violations in the phrase structure condition are created from correct sentences containing a noun and an infinitive; in the violation, the noun and the infinitive have switched places. Each sentence contains one or two words between the switched elements. Sentences in this condition are between 7 and 13 words in length.

Subject-verb agreement. Violations in the subject-verb condition are created by replacing a third-person singular verb with a third-person plural verb, and vice versa. All sentences are in the present tense, and all critical verbs are –*ar* verbs with regular conjugation (ending in –*a* for the third-person singular form and –*an* for the third-person plural form). Each of the 30 target verbs appears with both a singular and plural subject. Sentences in this condition are between 6 and 11 words in length.

Table V

GJT Distractor Stimuli

Condition	**Distractor Stimuli**	
Subject-Verb	*La secretaria* ****contestan / contesta*** *el teléfono en la oficina.* 'The secretary *answer / answers the telephone in the office.	
	Ellos ****contesta / contestan*** *las preguntas del policía.* 'They *answers / answer the questions of the police officer.'	
Phrase Structure	*Tú sueles* ***beber*** *mucho té durante el invierno.* 'You usually drink a lot of tea during the winter.'	*Tú sueles ****té*** *mucho beber durante el invierno.* 'You usually *tea a lot of drink during the winter.'
	Tenemos muchos tipos de ***té*** *para beber en la oficina.* 'We have many types of tea to drink in the office.'	*Tenemos muchos tipos de* ****beber*** *para té en la oficina.* 'We have many types of *to drink for tea in the office.'
Semantic	*Los gatos comen ****espejo / atún*** *por la tarde.* 'The cats eat *mirror / tuna in the afternoon.'	
	La niña se mira en el ****atún / espejo*** *cuando se maquilla.* 'The girls looks at herself in the *tuna / mirror when she puts on makeup.'	

Note. **Bold** typeface marks the critical word, that is, the word where violation becomes evident in each sentence. For all stimuli, ERPs are time-locked to the onset of the critical word. The word that constitutes the violation is indicated with an asterisk.* For demonstration purposes, both the infinitive and noun are underlined in the Phrase Structure stimuli examples.

3.4.2.4 EEG Procedure

EEG data were collected while participants completed the GJT. Sentences were presented visually through EPrime (Psychology Software Tools, Inc.) on a CRT monitor approximately 120 cm from the participant. The presentation began with a fixation cross at the center of the screen for 1000 ms, followed automatically by presentation of the stimulus sentence one word at a time (Rapid Visual Serial Presentation). Each word appeared in the center of the screen for 400 ms (SOA of 800 ms). The first word of each sentence begins with a capital letter, and the last word of each sentence ends with a period. After the last word of the sentence, the screen was blank for 1000 ms, after which a question mark appeared to prompt participants to provide their judgment. The question mark remained on the screen until the participant responded via mouse click (up to 5000 ms). Following participant response, a blink prompt[8] (####) was presented in the center of the screen until the participant clicked the mouse to indicate that he/she was ready for the next sentence. After the blink screen, the cycle repeated, beginning with presentation of the fixation cross preceding the next stimulus sentence.

Participants were instructed to read each sentence for acceptability and indicate whether it was ‘good’ or ‘bad’ in Spanish using a left and right mouse click, respectively (following Alemán Bañón et al., in press; Morgan-Short et al., 2010; Tokowicz & MacWhinney, 2005). Participants completed practice sentences both in English and Spanish prior to seeing experimental stimuli. After the brief set of English practice trials (three sentences), participants were given feedback from the researcher on their responses; this served to ensure that participants understood the task. After the Spanish practice block (25 sentences), participants

[8] Participants were instructed to minimize facial movements during presentation of the sentences. The self-timed blink break between each trial served to provide participants with sufficient time to blink or move, reducing participant fatigue and improving data quality.

were given biofeedback to ensure quality EEG data (e.g., *Your blinking looks good; try to limit facial movements to the blink screens, etc*.). The 300 stimuli were presented over five blocks (approximately 10 to 12 minutes each); participants were given a short break after each block.

3.4.2.4.1 EEG Data Acquisition

Scalp EEG was continuously recorded in DC mode at a sampling rate of 512 Hz using ASA-lab (ANT) 4.7.9 software. Participants were fitted with a Waveguard Cap (ANT) comprising 32 Ag/AgCl electrodes placed according to the extended 10-20 system, as illustrated in Figure 9 (Section 3.5). The impedance for each electrode was reduced to below 5 kΩ, and impedances were monitored after each block to ensure that they were held below this threshold. Scalp electrodes were referenced online to the average of all electrodes. The signal was amplified by an AMP-TRF40AB Refa-8 amplifier with a gain of 22-bit. The amplifier automatically filters EEG data with a digital FIR filter with a cutoff frequency of .27 times the sampling rate (512), resulting in a cutoff frequency of 138.24 Hz. The vertical electrooculogram (VEOG) was recorded from electrodes above and below the right eye, and the horizontal electrooculogram (HEOG) was recorded from electrodes on the left and right temples.

3.4.2.4.2 EEG Data Processing

After recording, data processing and analyses were completed in MatLab (version R2009a) using EEGlab (version 9.0.7.6b) and ERPlab (version 2.0.0.0) plug-ins. First, 1400 ms epochs were extracted from the continuous EEG (200 ms pre-critical word to 1200 ms post-critical words). All data were re-referenced offline to the mean of the left and right mastoids, then filtered using an IIR Butterworth filter with a high pass of .10 Hz and a low pass of 20.0 Hz.

In order to detect eye blinks and other artifacts, stepwise artifact rejections were performed on both EEG and EOG channels using a 40μV threshold, a 10 ms step, and a 400 ms moving window. In order to reject epochs containing drift, an additional stepwise artifact rejection was performed on EEG channels using a 40μV threshold over the entire 1400 ms time window (one 1400 ms step). Participants with greater than 25% of trials rejected overall, or greater than 25% rejection in either of the experimental conditions (article-noun and noun-adjective, violation or correct) during either Baseline or Follow-Up assessments were excluded from analyses. After these participants were excluded, artifacts led to rejections of less than 4% of experimental trials in either group, at either testing session. Rejection rates for each group by session and condition are provided in Table VI. Following previous research, all remaining trials, regardless of behavioral responses, were included in the main analyses (e.g., Bowden et al., 2013; Gabriele et al., 2013; Morgan-Short et al., 2010; Tanner et al., 2013).

Table VI

EEG Artifact Rejection Rates by Group

	At Home		Study Abroad	
Condition	Baseline	Follow-Up	Baseline	Follow-Up
Article-Noun, violation	5.24%	3.33%	2.09%	4.20%
Article-Noun, correct	3.10%	2.86%	2.93%	5.02%
Noun-Adjective, violation	4.75%	3.81%	2.93%	2.93%
Noun-Adjective, correct	2.61%	3.57%	2.51%	2.94%
Total Experimental Trials	3.92%	3.39%	2.62%	3.77%

Note. Values represent percentage of rejected trials in each condition.

3.4.2.5 <u>Motivation</u>

Motivation has been posited to play a role in development in both At Home (Dörnyei, 2005) and Study Abroad contexts (e.g., DeKeyser, 2010; Hernandez, 2010). Participants completed a modified version of the International Attitude/Motivation Test (Gardner, 2004; see Appendix E for full questionnaire), which includes 12 statements rated on a scale of 1 (weak) to 7 (strong). For the purposes of this analysis, participant responses to three items probing for motivation to learn Spanish (*My motivation to learn Spanish is…, My Motivation to learn Spanish to communicate with Spanish-speaking people is…,* and *My motivation to learn Spanish for practical purposes (e.g., to get a good job) is…*) were averaged in order to provide a measure of initial learner motivation (values provided in Table II in Section 3.2). Participants completed this questionnaire at the end of the Baseline Language Assessment Session. The questionnaire took around five minutes to complete.

3.4.2.6 <u>Exit Interview</u>

At the end of the Follow-Up Language Assessment Session, participants completed an exit interview during which the researcher clarified any questions that came up during examination of weekly language use surveys (described in the following section) or the initial language background questionnaire. This interview was also used to confirm the total number of Spanish courses (or courses taught in Spanish) each participant completed during the semester. The interview took between 5 and 15 minutes to complete.

3.4.3 <u>Language Contact Questionnaire</u>

Over the course of the semester, participants were sent a weekly language contact questionnaire via email to probe time spent engaged in use of the L2 for various activities or interactions. Two versions of the survey – one for At Home participants and another for Study Abroad participants – were created by adapting the Language Contact Profile (Freed, Dewey, Segalowitz, & Halter, 2004; see Appendix F for versions used in this study). For all items, participants were asked to select (a) the number of days during the preceding week, and (b) the typical number of hours per day they engaged in each type of activity. Responses to all surveys were combined to calculate the average number of hours per week each participant reported speaking, listening, reading, and writing in Spanish. A participant's average weekly L2 contact hours were calculated by summing reported weekly averages for the four language skills.[9] A link to the survey was emailed every Monday, and the surveys were completed at the participant's convenience. Each survey required 5 to 10 minutes to complete.

3.5 <u>Analysis</u>

The procedure and materials employed in this study yielded a rich set of cognitive, behavioral, and electrophysiological data. In this section, I describe the data analysis procedures employed for behavioral (Section 3.5.1) and ERP (Section 3.5.2) measures. All statistical analyses were conducted using SPSS, version 21 (IBM Corporation).

[9] Two participants, both in the Study Abroad group, failed to complete the weekly online language contact questionnaires. Thus, six participants comprise the Study Abroad group for analyses including L2 Contact.

3.5.1 <u>Behavioral Analysis</u>

As described in the previous sections, participants completed three behavioral measures of L2 abilities (a GJT, an Info Gap Activity, and an overall proficiency test) at the Baseline and Follow-Up Language Assessment Sessions. Data analysis for each of these tasks is described below.

GJT. Participant responses to article-noun and noun-adjective condition trials during the GJT were recorded automatically by EPrime (Psychology Software Tools, Inc.). In addition to calculating accuracy on these trials, participant responses were transformed to *d'* (d-prime) scores, which provide an unbiased measure of a participant's ability to discriminate between violation and correct items (e.g., Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999; Wickens, 2002). *D*' scores for Article and Adjective violations at Baseline and Follow-Up were calculated using the following formula provided in (1). The values are the index of grammatical sensitivity (judgment) used in statistical analyses.

(1) $d' = z(\textit{hit rate}) - z(\textit{false alarm rate})$

Info gap activity. Participant accuracy in production of gender-matched articles and adjectives during the information gap activity were calculated by dividing total number of correct articles or adjectives by total number of attempted articles or adjectives. The proportion correct for each structure (article and adjective) was calculated for Baseline and Follow-Up; these values are the index of production accuracy used in statistical analyses.

Overall proficiency test. Participant accuracy on a modified version of the DELE test was calculated (total number correct out of a possible 50 points). Total score at Baseline and Follow-

Up were recorded; these values provide the index of overall proficiency used in statistical analyses.

An individual's change in performance on each task was calculated by subtracting Baseline scores from Follow-Up scores (e.g., Article d'_{Change} = Article $d'_{\text{Follow-Up}}$ – Article d'_{Baseline}). Change values for each metric served as the dependent variables in regression analyses used to evaluate RQ1a (L2 Contact as a predictor of behavioral change) and RQ1b (WM as a predictor of behavioral change).

3.5.2 ERP Analysis

ERPs time-locked to the onset of critical words were averaged off-line for each participant, at each electrode site, using a 200 ms pre-stimulus baseline. ERP components of interest were quantified by computer as mean voltage amplitudes within a time window of activity. Two time windows of interest, corresponding roughly to the N400 and P600 effects, respectively, were selected based on previous research: 300-500 ms and 600-900 ms. Procedures followed for group-level (grand average) and individual-level (participant average) ERP analyses are described below.

Group-level analysis

Individual ERPs were averaged across participants and entered into grand ERP averages for each group (At Home, Study Abroad), session (Baseline, Follow-Up), condition (Article violation and correct, Adjective violation and correct), and time window (300-500 ms, 600-900 ms). This procedure yielded eight grand averages per group.

For grand average analyses, ANOVAs were calculated within each time window with Violation (violation, correct) as a within-subjects factor. Data for the midline (FPz, Fz, Cz, Pz, Oz), medial (left hemisphere: F3, FC1, C3, CP1, P3; right hemisphere: F4, FC2, C4, CP2, P4), and lateral (left hemisphere: F7, FC5, T7, CP5, P7; right hemisphere: F8, FC6, T8, CP6, P8) electrode sites were analyzed separately in order to determine scalp distribution of effects. ANOVAs on the midline included Electrode as an additional within-subjects factor (5 levels).

When a significant Violation by Electrode interaction was found ($p < .05$), follow-ups were conducted to compare Violation and Correct at each electrode. ANOVAs on medial and lateral electrodes included Hemisphere (2 levels) and Electrode (5 levels) as additional within-subjects factors. When a significant Violation by Electrode interaction was found ($p < .05$), the effect was examined within each hemisphere; where significant Violation by Electrode effects were found within a hemisphere, follow-ups were conducted to compare violation and correct at each electrode within that hemisphere. The Greenhouse-Geisser correction for inhomogeneity of variance was applied to all repeated measures with greater than one degree of freedom in the numerator (any interaction with Electrode). In such cases, the corrected *p*-value is reported. Electrode regions for group-level ERP analyses are indicated in Figure 9.

Individual-level analysis

Individual participant processing signatures were quantified using four metrics: N400 effect magnitude, P600 effect magnitude, overall Response Magnitude Index (RMI), and overall Response Dominance Index (RDI; following Tanner et al., 2014).

Effect magnitudes provide a measurement of the size of an individual's response to violation versus correct stimuli, that is, the "size" of an individual's N400 or P600 effect. Following previous research, N400 and P600 effect magnitudes were calculated using participants' mean amplitudes in a centro-parietal region of interest (ROI; C3, Cz, C4, P3, Pz, P4), where these two effects are typically largest (Tanner et al., 2014). N400 effect magnitude was calculated as mean activity in the correct minus violation condition between 300 ms and 500 ms. P600 effect magnitude was calculated as mean activity in the violation minus correct condition between 600 ms and 900 ms, as shown in equations (2) and (3).

(2) $\textit{N400effect magnitude} = \mathrm{N400}_{\mathrm{correct}} - \mathrm{N400}_{\mathrm{violation}}$

(3) $\textit{P600 effect magnitude} = \mathrm{P600}_{\mathrm{violation}} - \mathrm{P600}_{\mathrm{correct}}$

RMI provides a measure of the overall sensitivity each individual shows to violations within the N400 and P600 time windows. Greater RMI values indicate larger neural responses to violations across both time windows, regardless of the type of response. RMI was computed using equation (4), which calculates each individual's Euclidian distance from zero. RDI provides an index of an individual's relative response dominance (N400 or P600) by fitting that participant's least squares distance from the equal effect sizes line (e.g., dashed line in Figure 22) using perpendicular offsets. A participant with relatively equal-sized N400 and P600 effects would have an RDI value near zero. More negative or positive RDI values reflect relatively larger negativities or positivities across the N400 and P600 time windows, respectively. RDI was calculated using the equation (5), below.

(4) $RMI = \sqrt{N400\ effect\ magnitude^2 + P600\ effect\ magnitude^2}$

(5) $RDI = \dfrac{P600\ effect\ magnitude - N400\ effect\ magnitude}{\sqrt{2}}$

N400 and P600 effect magnitudes, RMI, and RDI were calculated for each participant, for both conditions (Article and Adjective) and both sessions (Baseline and Follow-Up). Additionally, an individual's change in each processing metric was calculated by subtracting the value for the metric at Baseline from the value for that metric at Follow-Up (e.g., Article RMI_{change} = [Article $RMI_{Follow\text{-}Up}$ – Article $RMI_{Baseline}$]). These change values served as the dependent variables in regression analyses used to evaluate RQ2a (L2 contact as a predictor of processing change) and RQ2b (WM as a predictor of processing change). The centro-parietal region of interest used for individual-level analyses is indicated in Figure 10.

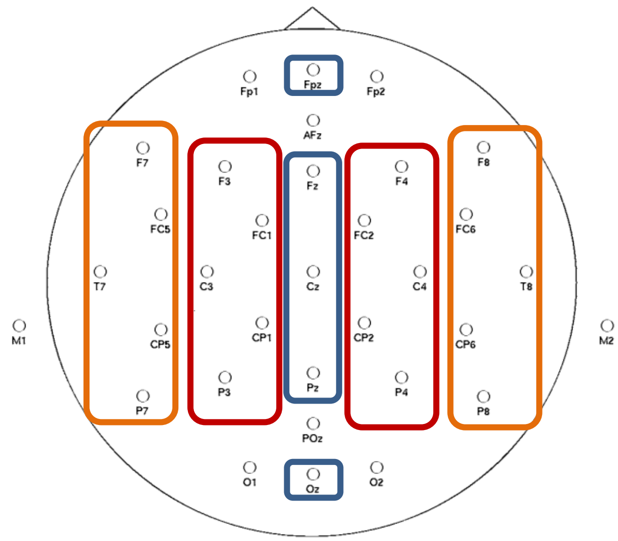


Figure 9. Electrode layout and electrodes included in group-level ERP analyses. Lateral electrodes are enclosed in orange, medial electrodes in red, and midline electrodes in blue.

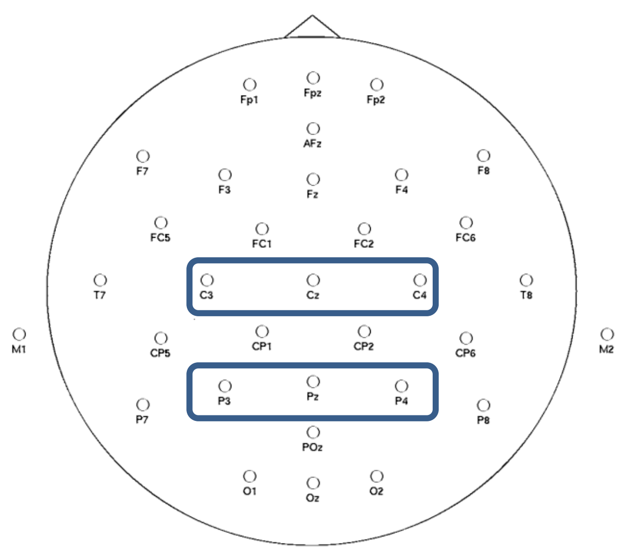


Figure 10. Electrode layout and electrodes included in individual-level ERP analyses. Electrodes in blue represent the centro-parietal region of interest.

4 RESULTS

This chapter presents the results of behavioral and processing measures of L2 abilities among participants in the At Home and Study Abroad groups. The chapter begins with a description of behavioral performance (Baseline, Follow-Up, and change) in each group, followed by statistical analyses of relationships between behavioral change and individual differences in L2 Contact and WM within each group. Processing data are then presented, beginning with a description and statistical analysis of group-level processing patterns at Baseline and Follow-Up for each of the target structures (Article and Adjective). Next, individual-level processing signatures – specifically, changes in processing signatures from Baseline to Follow-Up testing – are explored, and statistical analyses of relationships between processing changes and individual differences in L2 Contact and WM within each group are provided.

4.1 <u>Behavioral Results</u>

Participants completed three behavioral measures of L2 abilities at Baseline and Follow-Up testing: a GJT, an Info Gap production activity, and an overall proficiency test. Performance on these tasks provide, respectively, a measure of participant's (1) sensitivity to grammatical gender agreement violations on articles and adjectives, (2) production of gender-marked articles and adjectives, and (3) overall proficiency. In addition, subtracting Baseline scores from Follow-Up scores yields a measure of change in each of these behavioral metrics. This section presents the results of behavioral data analyses, presenting descriptive results for behavioral metrics and individual difference factors, followed by results of regression analyses aimed at addressing research questions 1a and 1b.

At Home Group

Descriptively, at Baseline testing, the At Home group performed at a level above chance, but with room for improvement, on the GJT and Info Gap activity. Individual DELE scores fell within the Low and Intermediate ranges (the group average fell in the Low range). Within the At Home group, scores improved from Baseline to Follow-Up testing on the GJT and Info Gap task, but decreased slightly on the DELE (the average change in score was less than one point; the group average remained in Low range, and all individual scores remained in the Low and Intermediate ranges). Paired-samples t-tests indicated that participant production of grammatical gender agreement on adjectives improved significantly from Baseline to Follow-Up testing (Info Gap Accuracy, Adjective: $t(13)= -2.168$, $p = .049$). No other significant behavioral changes from Baseline to Follow-Up testing were observed within this group. Descriptive results for all behavioral metrics for the At Home group are provided in Table VII.

Table VII

At Home Group: Descriptive Results for Behavioral Metrics

Variable	Baseline	Follow–Up	Change
GJT *d'*			
Article	.688 (.570) .192 – 1.81	.821 (.869) -.554 – 2.08	.133 (.619) -.757 – 1.21
Adjective	.355 (.471) -.440 – 1.11	.572 (.838) -.992 – 2.26	.217 (.715) -1.29 – 1.41
GJT Accuracy			
Article	.616 (.093) .533 – .817	.640 (.141) .433 – .850	.024 (.100) -.100 – .178
Adjective	.550 (.066) .450 – .683	.585 (.111) .417 – .817	.035 (.093) -.133 – .200
Info Gap Accuracy			
Article	.840 (.145) .533 – 1.00	.861 (.166) .417 – 1.00	.021 (.158) -.194 – .300
Adjective	.685 (.108) .500 – .852	.751 (.166) .455 – 1.00	.066 (.114)* -.120 – .284
DELE[a]	22.50 (5.10) 16 – 32	21.86 (3.57) 17 – 29	-.64 (4.70) -10 – 7

Note. $N = 14$; *SD* presented in parentheses; Range provided below *M* and *SD*.
[a] Maximum score = 50.
*$p < .05$ (paired-samples t-test, Baseline and Follow-Up scores)

Study Abroad Group

Descriptively, at Baseline testing, the Study Abroad group performed at a level above chance, but with room for improvement, on the GJT and Info Gap activity. Individual DELE scores fell within the Low and Intermediate ranges (the group average fell in the Low range). Within the Study Abroad group, scores improved from Baseline to Follow-Up testing on all behavioral metrics. Individual DELE scores at Follow-Up testing fell within the Low, Intermediate, and Advanced ranges (with the group average falling within the Intermediate range). Paired-samples t-tests revealed that participant sensitivity to grammatical gender violations on articles improved significantly from Baseline to Follow-Up testing (GJT *d'*, Article: $t(7) = -4.497$, $p = .003$; GJT Accuracy, Article: $t(7) = -3.916$, $p = .006$), as did overall proficiency (DELE: $t(7) = -4.331$, $p = .003$). Changes in sensitivity to grammatical gender violations on adjectives and in production of gender-marked articles and adjectives were not significant. Descriptive results for all behavioral metrics for the Study Abroad group are provided in Table VIII.

Table VIII

Study Abroad Group: Descriptive Results for Behavioral Metrics

Variable	Baseline	Follow–Up	Change
GJT *d'*			
Article	.653 (.922) -.501 – 2.12	1.46 (1.20) 0.00 – 2.68	.808 (.508)** .331 – 1.80
Adjective	.669 (1.12) -.440 – 3.12	1.29 (1.02) -.390 – 2.94	.622 (1.09) -.516 – 2.70
GJT Accuracy			
Article	.606 (.155) .417 – .850	.717 (.176) .500 – .900	.111 (.080)** .033 – .283
Adjective	.602 (.156) .450 – 9.33	.693 (.150) .467 – 9.33	.091 (.154) -.050 – .381
Info Gap Accuracy			
Article	.905 (.124) .706 – 1.00	.947 (.081) .800 – 1.00	.042 (.075) -.071 – .167
Adjective	.709 (.195) .354 – 1.00	.825 (.199) .522 – 1.00	.116 (.243) -.350 – .450
DELE[a]	19.63 (4.69) 12 – 26	28.88 (7.85) 19 – 41	9.25 (6.04)** 1 – 16

Note. *N* = 8; *SD* presented in parentheses; Range provided below *M* and *SD*.
[a] Maximum score = 50.
** $p < .01$ (paired-samples t-test, Baseline and Follow-Up scores)

The wide ranges in change scores for each behavioral metric within each group provide evidence of variability in L2 development in both contexts. In order to begin to explore this variability with regard to the individual difference factors (L2 Contact and WM), it is important to first examine variability in reported L2 Contact and WM ability scores.

At Home Group

Within the At Home group, participants reported a wide range of weekly L2 contact hours[10] (*Mean* = 13.68, *SD* = 9.55, Range = 2.46 – 33.65). A wide range of WM[11] abilities were also represented within this group (Composite WM scores: *SD* = .80, Range = -1.56 – 1.30; performance on the OSpan, RSpan, and SymSpan tasks are presented in Table IX).

Table IX

At Home Group: Working Memory Task Scores

	OSpan	RSpan	SymSpan
Mean (*SD*)	48.79 (16.94)	43.14 (16.45)	18.29 (9.00)
Range	18 - 75	18 - 69	3 – 34

Note. Mean scores using Absolute scoring protocol; maximum score on OSpan and RSpan = 75; maximum score on SymSpan = 42.

[10] Average weekly L2 Contact comprises the sum of average of weekly hours reported speaking, listening, reading, and writing in Spanish. This is the metric used in all subsequent analyses.

[11] Mean WM composite score = 0; this does not represent a meaningful value on its own, given its calculation (average of z-scores from each of the three tasks, which are standardized scores that rely on group mean and *SD*). WM Composite score is the metric used in all subsequent analyses; the unit of measurement is standard deviations from the group mean.

Study Abroad Group

Within the Study Abroad group, participants reported a wide range of weekly L2 contact hours (*Mean* = 58.63 weekly, *SD* = 24.14, Range = 30.27 – 82.56, *N* = 6[12]). A wide range of WM abilities were also represented within this group (Composite WM scores: *SD* = .69, Range = -.84 – .95; performance on the OSpan, RSpan, and SymSpan tasks are presented in Table X).

Table X

Study Abroad Group: Working Memory Task Scores

	OSpan	RSpan	SymSpan
Mean (*SD*)	42.25 (11.25)	32.5 (9.10)	18.75 (7.70)
Range	22 - 59	19 - 47	9 to 28

Note. Mean scores using Absolute scoring protocol; maximum score on OSpan and RSpan = 75; maximum score on SymSpan = 42.

Given the variability present in (1) L2 Contact, (2) WM abilities, and (3) behavioral changes from Baseline to Follow-Up testing, regression analyses were conducted in order to determine whether individual differences L2 Contact (to address RQ1a) and WM abilities (to address RQ1b) could account for differences in behavioral change. Specifically, simple regressions were conducted in which each of the metrics of behavioral change (grammatical sensitivity to violations on articles and adjectives: *d'*; accurate production of articles and adjectives: Info Gap accuracy; overall proficiency: DELE score) were regressed onto the predictor variables L2 Contact and WM.

[12] Two participants in the Study Abroad group failed to complete weekly L2 contact surveys.

At Home Group

For the At Home group, behavioral change associated with grammatical gender agreement did not appear to be related to either L2 Contact or WM ability. Indeed, results of regression analyses revealed that neither L2 Contact nor WM were significant predictors of change in sensitivity to grammatical gender violations on articles or adjectives (

Table XI) or of change in production of gender-marked articles or adjectives (Table XII).

In terms of changes in overall proficiency, however, relationships with the individual difference factors were revealed. L2 Contact was found to be a significant, negative predictor of change in DELE score from Baseline to Follow-Up testing in this participant group, indicating that high levels of L2 Contact were associated with smaller (and in some cases, negative) changes in overall proficiency. Specifically, for each additional hour of average weekly L2 Contact reported, the change in a participant's DELE score from Baseline to Follow-Up testing decreased by roughly 1/3 of a point (B = -.275). L2 contact accounted for just over 30% of the variance in DELE score change (R^2 = .312). Regression results are presented in Table XIII, and the relationship between L2 Contact and DELE score change is represented graphically in Figure 11a.

WM was found to be a marginally significant, positive predictor of DELE score change within the At Home group, indicating gains in overall proficiency were associated with higher WM abilities. Specifically, for every additional point in WM composite score (on standard deviation above the group mean), a participant's DELE score change increased by 2.7 points (B = 2.767). WM accounted for roughly 22% of the variance in DELE score change (R^2 = .223). Regression results are presented in Table XIII, and the relationship between WM and DELE score change is represented graphically in Figure 11b.

Table XI

At Home Group: Change in Grammatical Sensitivity, Regressions

	Article Judgment Change			Adjective Judgment Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
				L2 Contact		
Constant	.295	.303		-.130	.335	
L2 Contact	-.012	.018	-.183	.025	.020	.339
R^2		.034			.115	
F		.416			1.561	
				Working Memory		
Constant	.133	.163		.217	.193	
Working Memory	.250	.211	.324	.207	.250	.233
R^2		.105			.054	
F		1.411			.687	

Note. Judgment Change = change in Article/Adjective *d'* scores (GJT) from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.

Table XII

At Home Group: Change in Production Accuracy, Regressions

	Article Production Change			Adjective Production Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	.102	.074		.086	.057	
L2 Contact	-.006	.004	-.360	-.001	.003	-.117
R^2		.130			.014	
F		1.786			.167	
			Working Memory			
Constant	.021	.044		.066	.032	
Working Memory	.021	.057	.109	.007	.041	.048
R^2		.012			.002	
F		.144			.028	

Note. Production Change = change in Article/Adjective production accuracy (Info Gap Activity) from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.

Table XIII

At Home Group: Change in Overall Proficiency, Regressions

Variable	DELE Change		
	B	*SEB*	*β*
		L2 Contact	
Constant	3.115	1.943	
L2 Contact	-.275	.118	-.558
R^2		.312	
F		5.434*	
		Working Memory	
Constant	-.643	1.153	
Working Memory	2.767	1.492	.472
R^2		.223	
F		3.441^	

DELE Change = change in overall written proficiency test score from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.
^ $p < .1$; * $p < .05$

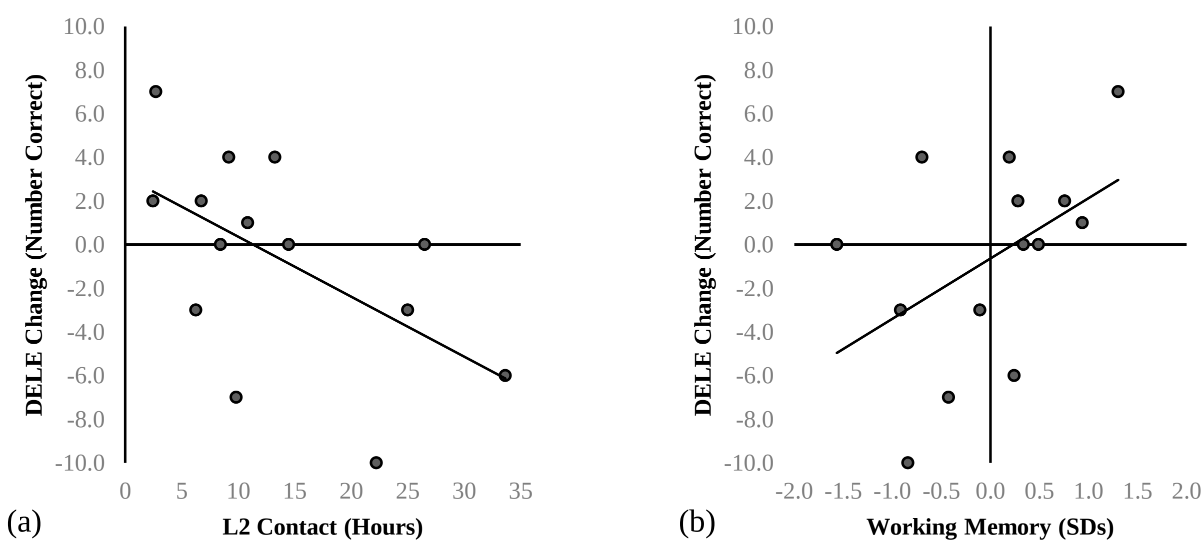


Figure 11. At Home Group: Scatterplots illustrating the relationship between (a) L2 Contact and DELE score change and (b) WM and DELE score change. Each point represents one participant. Solid lines show the best-fit line for the data from the regression analysis.

Study Abroad Group

For the Study Abroad Group, behavioral change appeared to be related to L2 Contact, but not WM ability. Results of regression analyses revealed that L2 Contact was a significant, positive predictor of (1) change in sensitivity to grammatical gender violations on adjectives and (2) change in production of gender-marked adjectives, indicating that learners who reported the highest levels of L2 Contact experienced the greatest amount of linguistic development in terms of grammatical gender agreement on adjectives (judgments and production). Specifically, for every additional hour of average weekly L2 Contact reported, the change in a participant's *d'* score increased by .029 (B = .029) and the change in his or her gender-marked adjective production accuracy increased by nearly 1% (B = .009). L2 Contact accounted for 79.6% of variance in Adjective *d'* change (R^2 = .796) and 68.4% of variance in Adjective production accuracy change (R^2 = .684). Relationships between L2 Contact and Adjective *d'* change and Adjective production accuracy change are represented graphically in Figure 12 and Figure 13, respectively.

Neither L2 Contact nor WM were found to be significant predictors of change in sensitivity to grammatical gender violations on articles, change in production of gender-marked articles, or change in overall proficiency within the Study Abroad group. Regression results for this group are presented in Table XIV (judgment scores), Table XV (production accuracy), and Table XVI (overall proficiency).

Table XIV

Study Abroad Group: Change in Grammatical Sensitivity, Regressions

	Article Judgment Change			Adjective Judgment Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	.495	.465		-1.282	.459	
L2 Contact	.003	.007	.204	.029	.007	.892
R^2		.042			.796	
F		.174			15.600*	
			Working Memory			
Constant	.808	.174		.622	.418	
Working Memory	.325	.270	.441	.045	.647	.028
R^2		.194			.001	
F		1.445			.005	

Note. Judgment Change = change in Article/Adjective *d'* scores (GJT) from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.
* $p < .05$

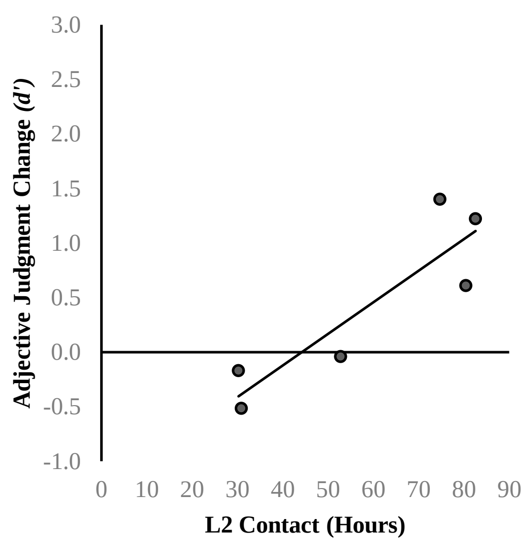


Figure 12. Study Abroad Group: Scatterplot illustrating the relationship between L2 Contact and Adjective Judgment score change. Each point represents one participant. Solid line shows the best-fit line for the data from the regression analysis.

Table XV

Study Abroad Group: Change in Production Accuracy, Regressions

	Article Production Change			Adjective Production Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	.090	.107		-.461	.202	
L2 Contact	-.001	.002	-.164	.009	.003	.827
R^2		.027			.684	
F		.111			8.673*	
			Working Memory			
Constant	.042	.028		.116	.074	
Working Memory	-.029	.043	-.262	-.214	.114	-.609
R^2		.068			.370	
F		.441			3.530	

Note. Production Change = change in Article/Adjective production accuracy (Info Gap Activity) from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.
* $p < .05$

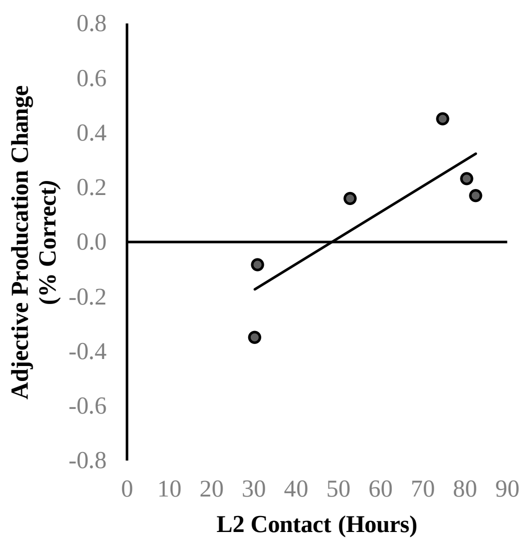


Figure 13. Study Abroad Group: Scatterplot illustrating the relationship between L2 Contact and Adjective Production accuracy change. Each point represents one participant. Solid line shows the best-fit line for the data from the regression analysis.

Table XVI

Study Abroad Group: Change in Overall Proficiency, Regressions

	DELE Change		
Variable	*B*	*SEB*	*β*
		L2 Contact	
Constant	4.126	6.883	
L2 Contact	.100	.110	.415
R^2		.172	
F		.831	
		Working Memory	
Constant	9.250	2.303	
Working Memory	-.517	3.570	-.059
R^2		.003	
F		.021	

DELE Change = change in overall written proficiency test score from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.

To summarize the results from analyses of behavioral change, within the At Home group, significant behavioral change from Baseline to Follow-Up testing was evidenced only for production of gender-marked adjectives. Descriptive results within this group, however, revealed a great deal of variation in behavioral change. Individual differences in L2 Contact and WM predicted changes in overall proficiency, as measured by the DELE: learners who reported higher levels of L2 Contact made fewer gains from Baseline to Follow-Up testing on the DELE, and learners with higher WM abilities made greater gains during the same period. Neither L2 Contact nor WM predicted changes in L2 abilities related to gender agreement.

Within the Study Abroad group, significant behavioral change was evidenced for sensitivity to grammatical gender violations on articles and for overall proficiency. Descriptively, scores on all behavioral metrics increased from Baseline to Follow-Up testing, and there was a great deal of variation in behavioral changes within the group. Individual differences in L2 Contact predicted changes in sensitivity to grammatical gender agreement violations on adjectives, as well as accurate production of gender-marked adjectives. No relationships were found between L2 Contact and change in article judgments, article production, or overall proficiency, and no relationships were found between WM and changes on any behavioral metric.

4.2 <u>Processing Results</u>

This section presents the results of ERP data analyses, presenting group-level results (grand average analysis), followed by individual-level results aimed at addressing research questions 2a and 2b.

4.2.1 <u>Group-Level Processing Results</u>

In order to examine group-level processing effects, repeated-measures ANOVAs were performed for the two time windows of interest (300-500ms and 600-900 ms) for each group (At Home, Study Abroad), session (Baseline, Follow-Up), and condition (Article, Adjective). Grand average ERP analyses are reported by group, below.

At Home Group

Visual inspection of At Home group grand average waveforms at Baseline testing indicated that, in the Article condition, violation waveforms were slightly more negative than correct waveforms in centro-parietal electrodes in the N400 time window (300-500 ms). No clear differences between violation and correct waves were observed in the later time window (600-900 ms) or elsewhere. This small negativity, however, did not reach statistical significance; no significant differences between violation and correct waveforms were found for the Article condition at Baseline. Grand average ERPs for the At Home group's processing of Articles at Baseline are presented in Figure 14.

Visual inspection of grand average waveforms for the Adjective condition at Baseline revealed no clear differences between violation and correct waveforms. Statistical analyses confirmed this observation: no significant effects differences were found between violation and correct waveforms in the Adjective condition. Grand average ERPs for the At Home group's processing of Adjectives at Baseline are presented in Figure 15.

At Follow-Up, visual inspection of At Home group grand average waveforms for the Article condition revealed no clear effects. Statistical analyses supported this observation: no significant differences between violation and correct waveforms in the Article condition were revealed. Grand average ERPs for the At Home group's processing of Articles at Follow-Up are presented in Figure 16.

Visual inspection of grand average waveforms for the Adjective condition at Follow-Up did not reveal any clear differences between violation and correct waveforms. Statistical analyses revealed a significant Violation by Electrode interaction at lateral electrodes in the 300-500 ms time window ($F(4, 100) = 4.044$, $p = .023$); follow-ups on this effect, however, did not reach significance. Grand average ERPs for the At Home group's processing of Adjectives at Follow-Up are presented in Figure 17.

Study Abroad Group

For the Study Abroad group, visual inspection of grand average waveforms at the Article condition at Baseline showed an N400-like negativity (violation waveforms were more negative than correct waveforms in centro-parietal electrodes). This negativity, however, did not reach statistical significance; no significant differences between violation and correct waveforms were found for the Article condition at Baseline. Grand average ERPs for the Study Abroad group's processing of Articles at Baseline are presented in Figure 18.

Visual inspection of grand average waveforms for the Adjective condition at baseline showed a P600-like late positivity (violation waveforms were more positive than correct waveforms in centro-parietal and parietal electrodes). This positivity, however, did not reach statistical significance; no significant differences between violation and correct waveforms were found for the Adjective condition at Baseline. Grand average ERPs for the Study Abroad group's processing of Adjectives at Baseline are presented in Figure 19.

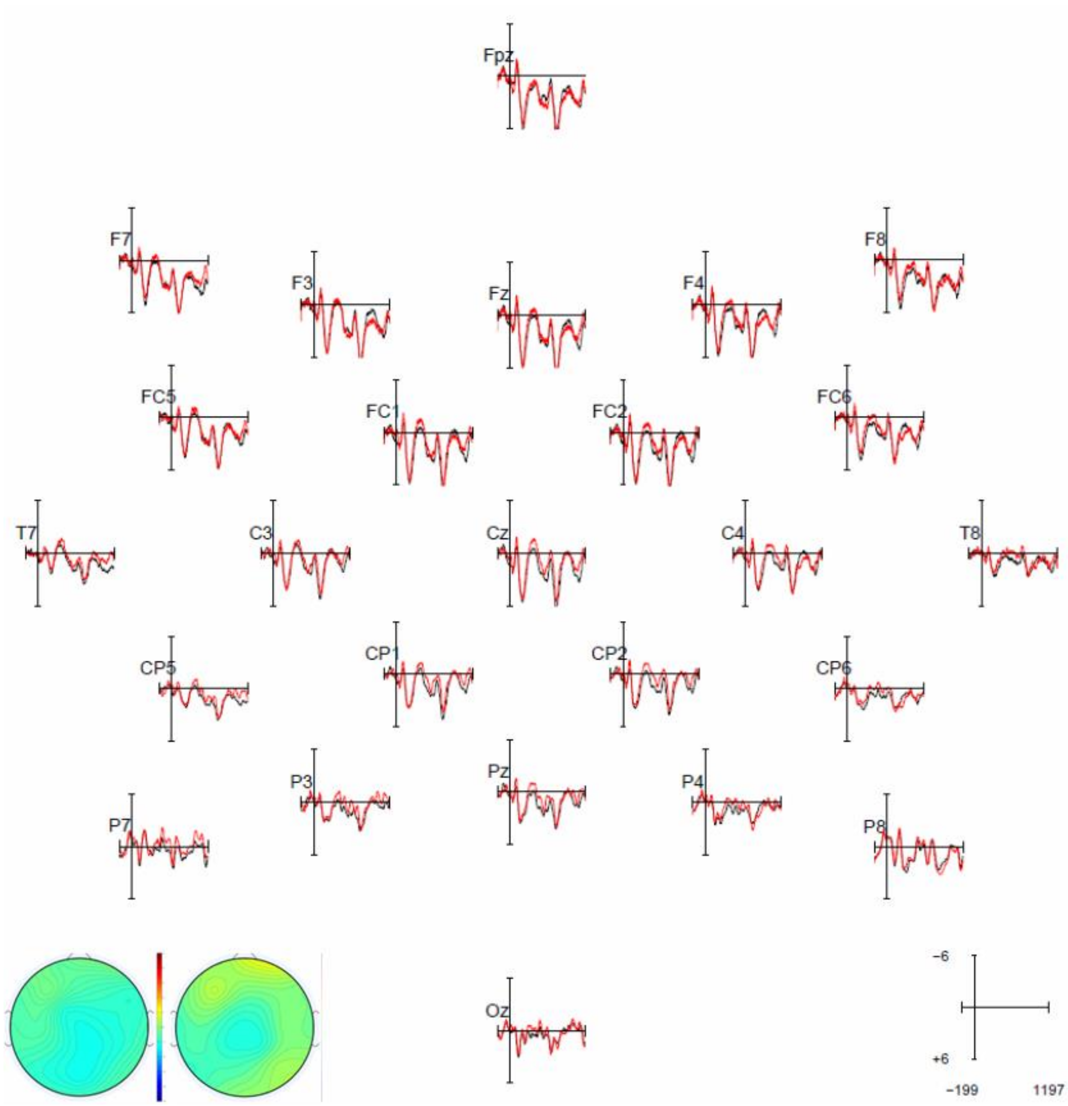
For Follow-Up data, visual inspection of grand average waveforms for the Article condition within the Study Abroad group suggested that violation waveforms were more negative than correct waveforms in frontal and fronto-central electrodes the earlier time window (300-500 ms) and again (a separate effect) in the later time window (600-900 ms). Additionally, a focal posterior positivity that was maximal around 750 ms was present. Statistical analyses confirmed the presence of both of these effects.

Statistical analyses in the 300-500 ms time window revealed a main effect of Violation was at midline ($F(1, 25) = 11.120$, $p = .013$) and medial electrodes ($F(1, 25) = 8.703$, $p = .021$), as well as a Violation by Electrode interaction at lateral sites ($F(4, 100) = 4.476$, $p = .006$). Follow-up analyses for lateral sites revealed a significant Violation by Electrode interaction in the left hemisphere ($F(2, 50) = 4.973$, $p = .004$), that was significant at the frontal electrode (F7). In the right hemisphere, there was a significant main effect of Violation ($F(1, 25) = 5.679$, $p = .049$), but no significant interaction with Electrode. These effects indicate the presence of a fronto-central negativity elicited by violations in the Article condition.

In the 600-900 ms time window, a Violation by Electrode interaction was found for ANOVAs over midline ($F(2, 50) = 8.728$, $p = .009$), medial ($F(4, 100) = 6.679$, $p = .027$), and lateral electrodes ($F(4, 100) = 9.032$, $p = .003$). Follow-ups revealed significant negativities at frontal and fronto-central electrodes along the midline (FPz, Fz), and also in medial-left (F3, FC1), medial-right (FC2), lateral-left (F7), and lateral-right (F8) sites, indicating the presence of a late, fronto-central negativity. The posterior positivity (focal P600 effect) reached significance only at the occipital midline electrode (Oz). Grand average ERPs for the Study Abroad group's processing of Article violations at Follow-Up are presented in Figure 20.

Visual inspection of grand average waveforms for the Adjective condition at Follow-Up within the Study Abroad group revealed a possible negativity, particularly over centro-parietal electrodes in the earlier time window. However, no statistically significant effects were found in either time window for this condition. Grand average ERPs for the Study Abroad group's processing of Adjective violations at Follow-Up are presented in Figure 21.

In summary, within the At Home group, no significant effects were found at either Baseline or Follow-Up testing for processing of grammatical gender agreement violations on articles or adjectives. Within the Study Abroad group, significant effects were found only for the Article condition at Follow-Up, where violations elicited a significant fronto-central negativity in both time windows, as well as a focal posterior positivity (P600) in the later time window. No significant effects were found for the Article condition at Baseline, or for the Adjective condition at Baseline or Follow-Up testing.



300-500 ms 600-900 ms

Figure 14. Grand average ERPs for At Home group's processing of agreement violations on Articles at Baseline.

Note. In waveforms, onset of the critical word (noun) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.

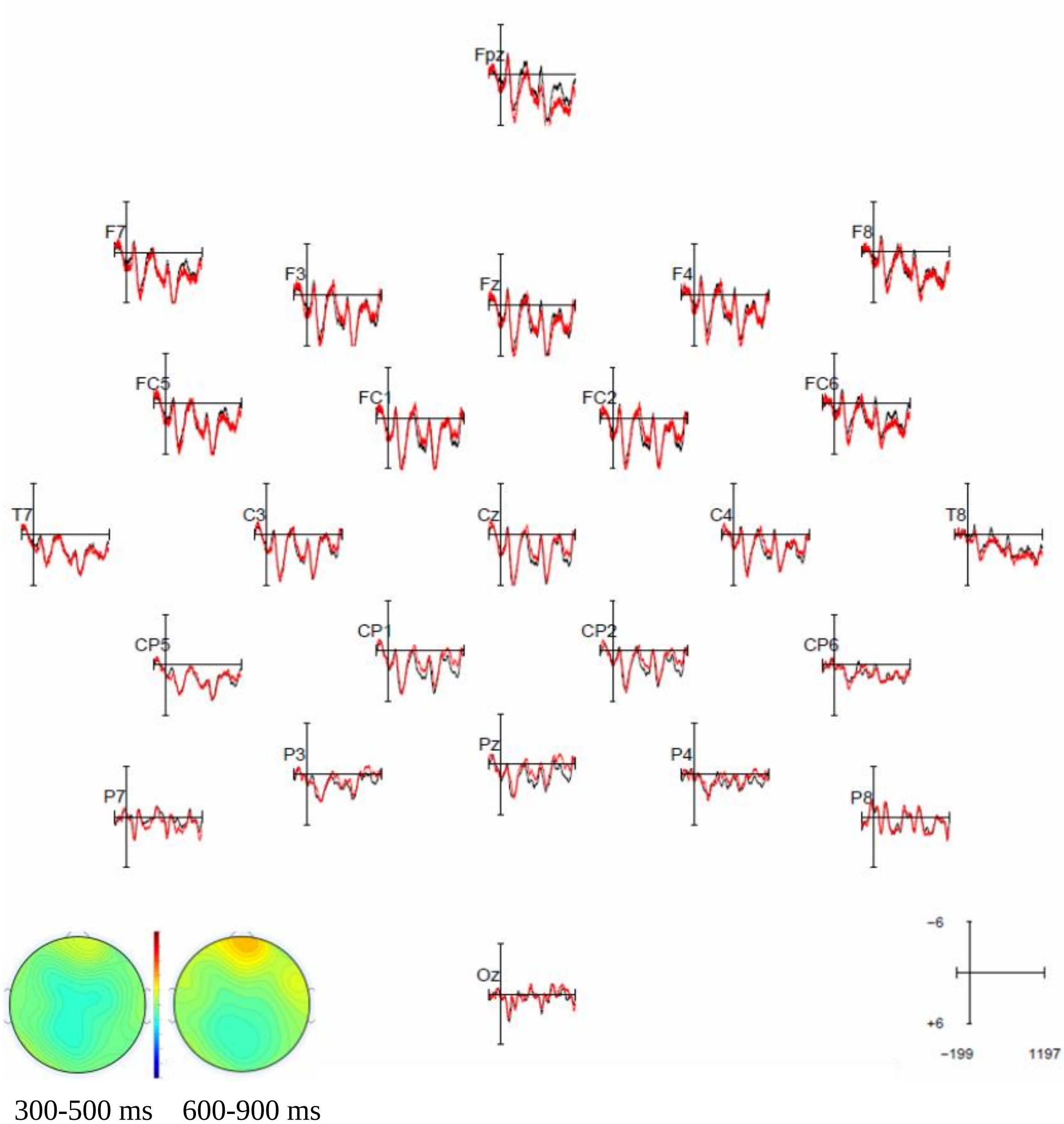
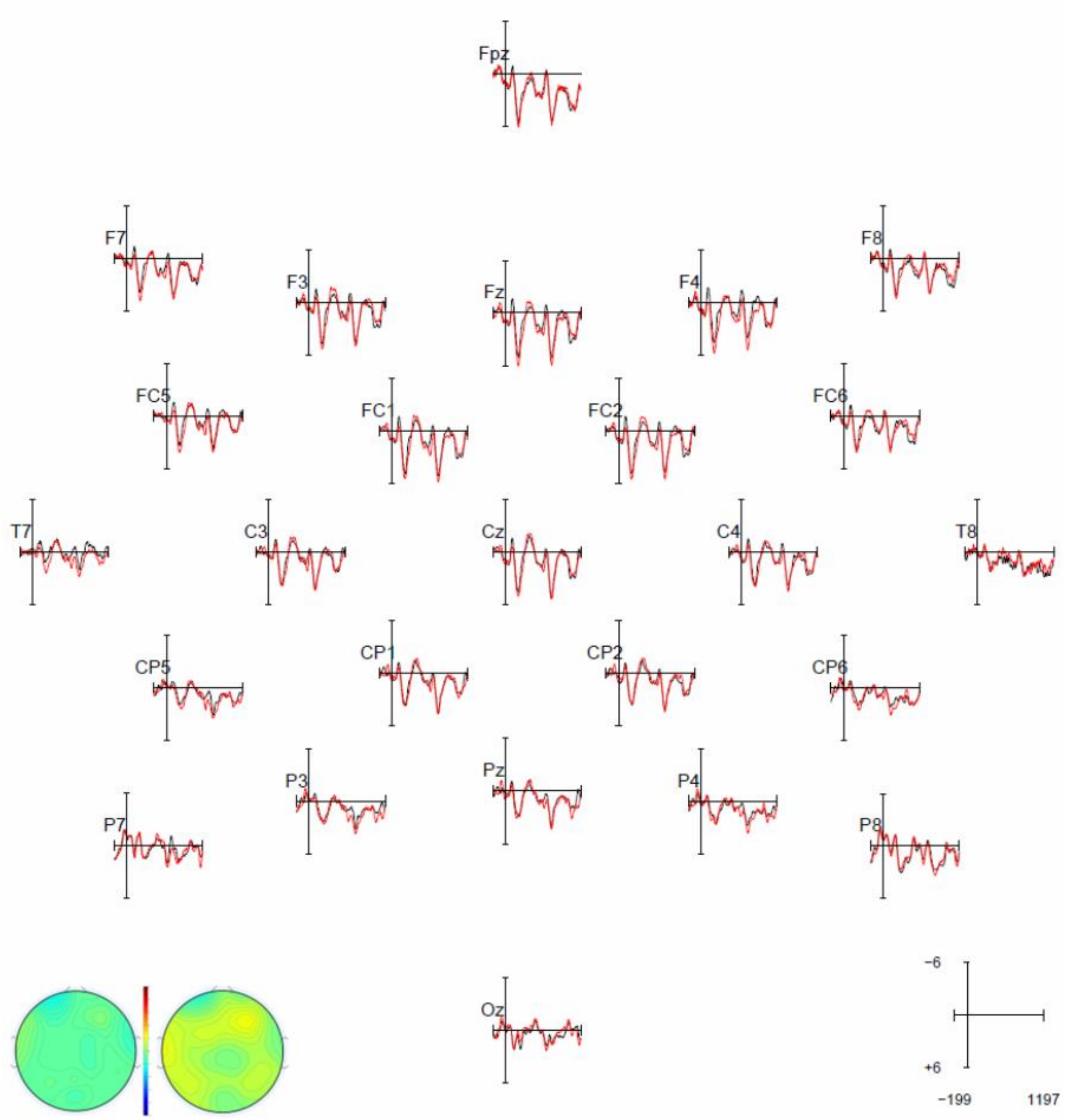


Figure 15. Grand average ERPs for At Home group's processing of agreement violations on Adjectives at Baseline.

Note. In waveforms, onset of the critical word (adjective) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.



300-500 ms   600-900 ms

Figure 16. Grand average ERPs for At Home group's processing of agreement violations on Articles at Follow-Up.

Note. In waveforms, onset of the critical word (noun) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (μV, positive voltage is plotted down). Voltage maps represent difference in voltage (μV) between correct and violation stimuli.

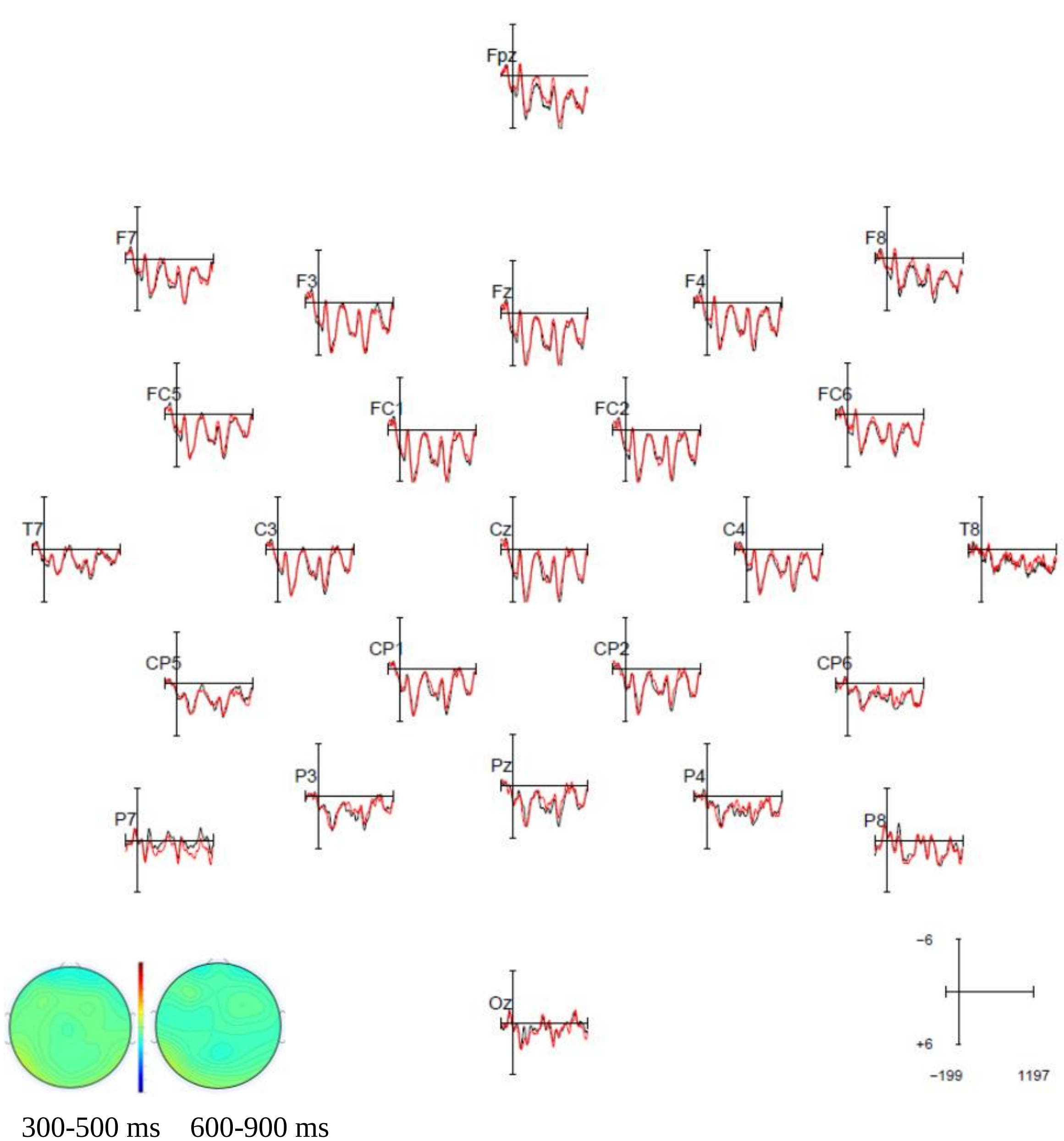


Figure 17. Grand average ERPs for At Home group's processing of agreement violations on Adjectives at Follow-Up.

Note. In waveforms, onset of the critical word (adjective) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.

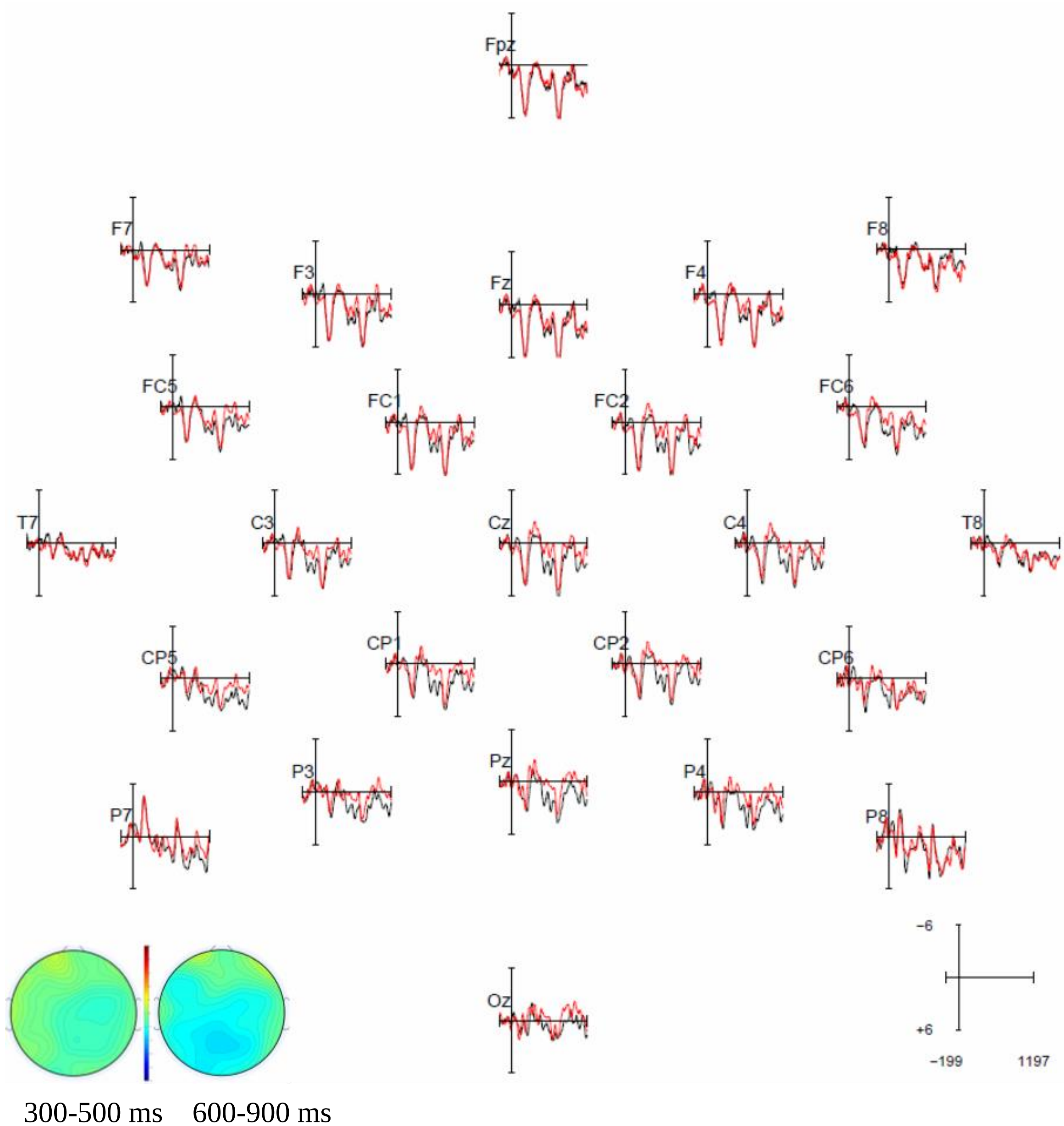
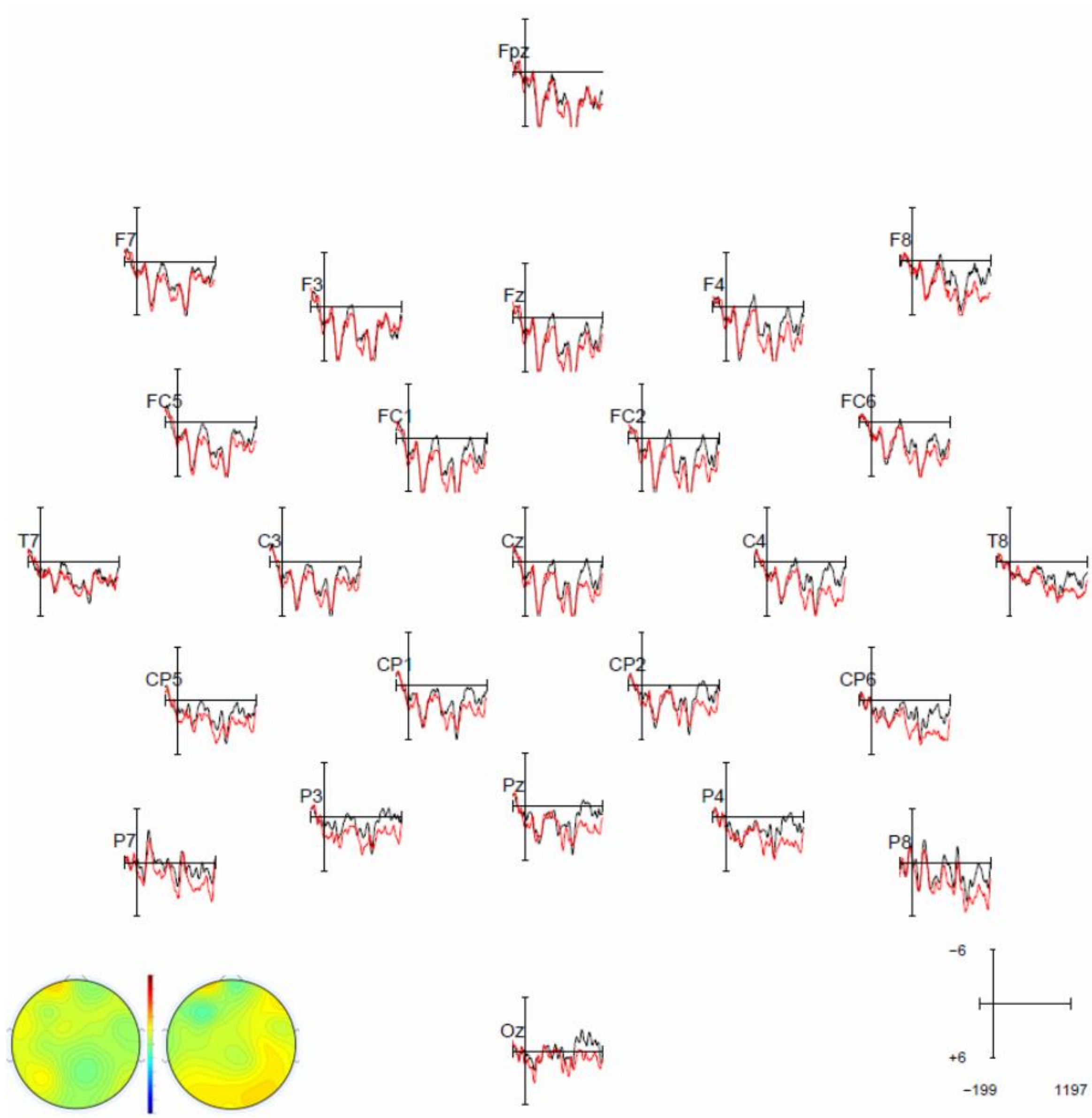


Figure 18. Grand average ERPs for Study Abroad group's processing of agreement violations on Articles at Baseline.

Note. In waveforms, onset of the critical word (noun) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.



300-500 ms   600-900 ms

Figure 19. Grand average ERPs for Study Abroad group's processing of agreement violations on Adjectives at Baseline.

Note. In waveforms, onset of the critical word (adjective) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.

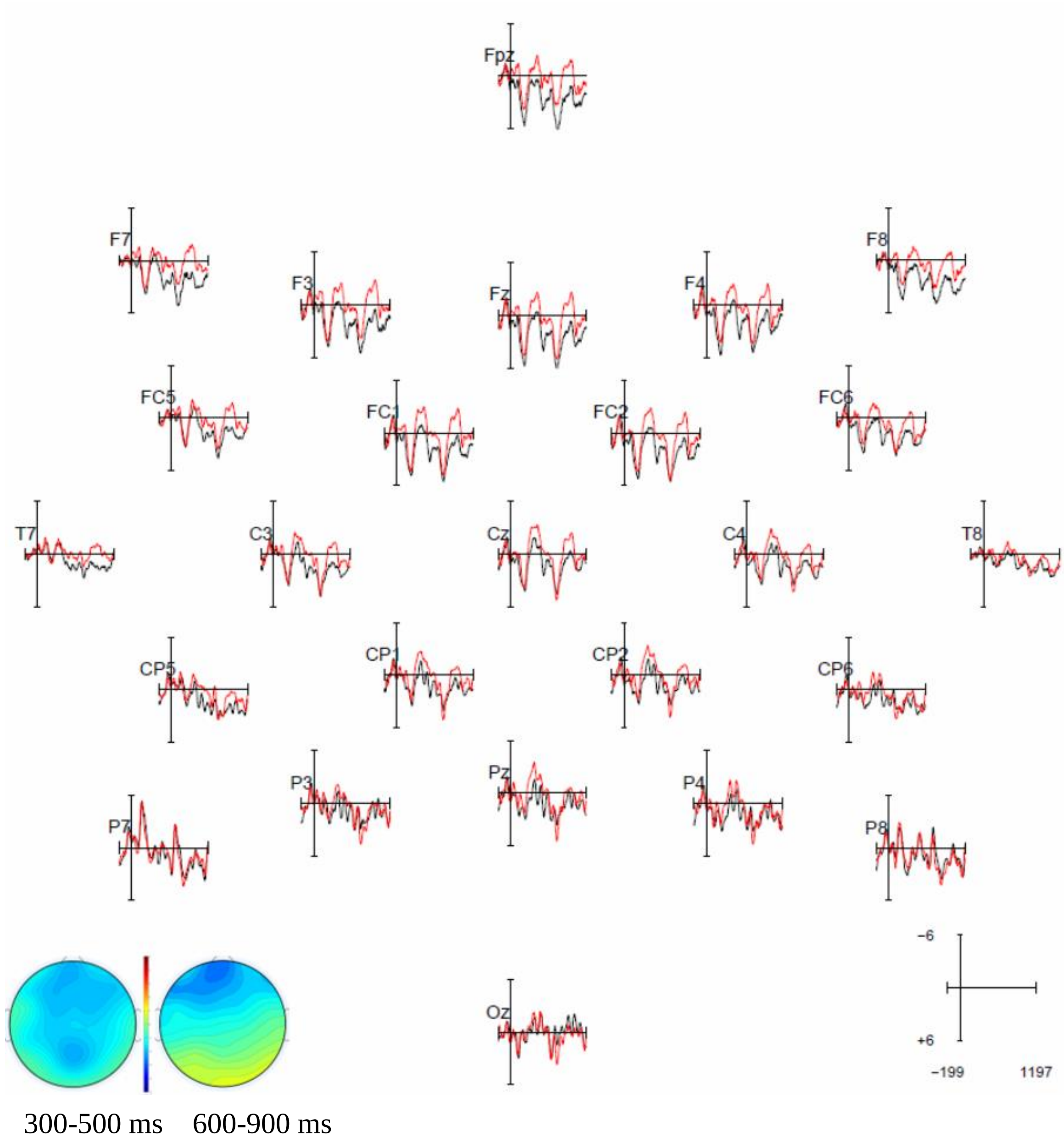
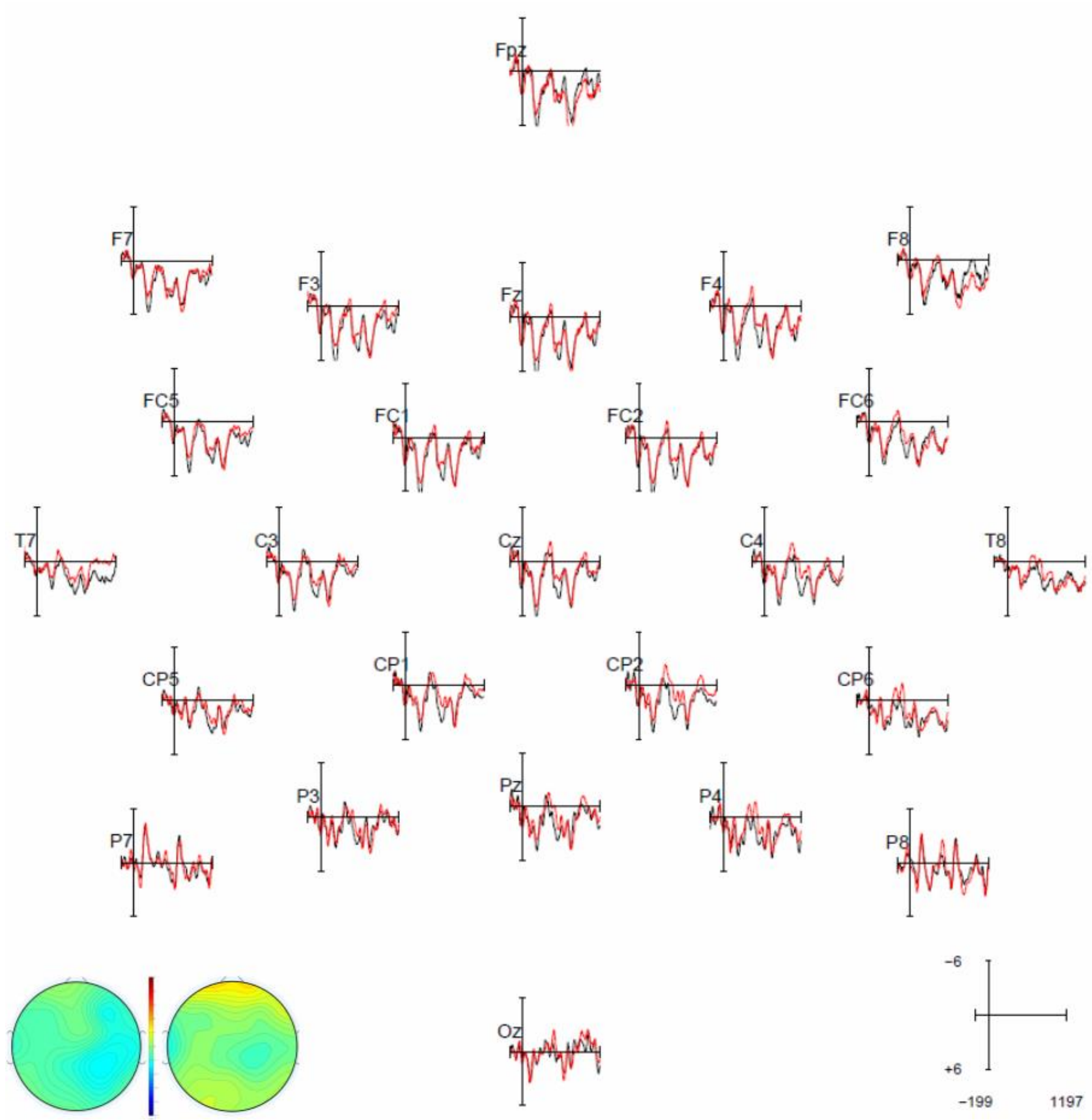


Figure 20. Grand average ERPs for Study Abroad group's processing of agreement violations on Articles at Follow-Up.

Note. In waveforms, onset of the critical word (noun) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.



300-500 ms 600-900 ms

Figure 21. Grand average ERPs for Study Abroad group's processing of agreement violations on Adjectives at Follow-Up.

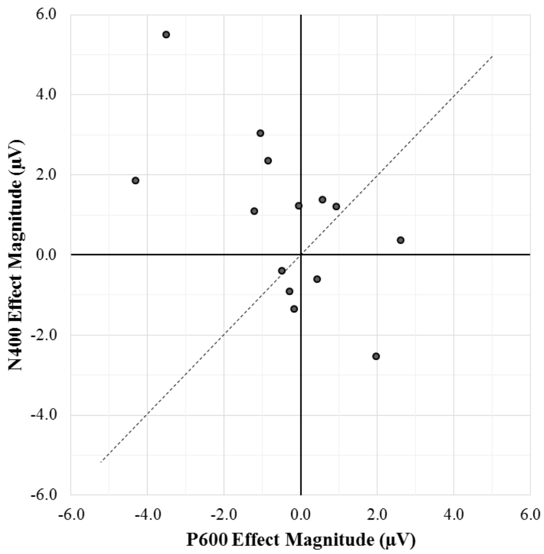
Note. In waveforms, onset of the critical word (adjective) is indicated by the vertical bar; x-axis: time (ms); y-axis: voltage (µV, positive voltage is plotted down). Voltage maps represent difference in voltage (µV) between correct and violation stimuli.

4.2.2 Individual-Level Processing Results

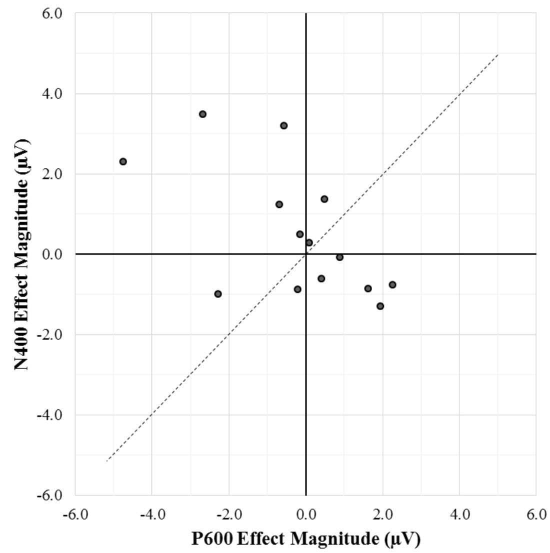
Despite a general lack of significant effects in the group-level ERP waveform analyses, inspection of individual participants' waveforms at Baseline and Follow-Up within both groups revealed that a number of participants did show sensitivities to violations in both conditions. In order to explore variability within each group, individual N400 and P600 effect magnitudes[13] were calculated for participants within each group for both conditions at Baseline and Follow-Up testing.

Figure 22 through Figure 25 provide scatterplots showing the distribution of individual N400 and each group (At Home and Study Abroad) at Baseline and Follow-Up for each condition (Article and Adjective). As can be seen in these figures, individual processing signatures for each condition are distributed along a continuum from negative to positive response dominance, with few participants showing equal response dominance, as indicated by points that fall on the dashed line). A great amount of variability in neural processing exists within each group.

[13] The term "N400 effect magnitude" is used here to describe a centro-parietally distributed negativity in the 300-500 ms time window. The term "P600 effect magnitude" is used here to describe to a centro-parietally distributed positivity in the 600-900 ms time window.

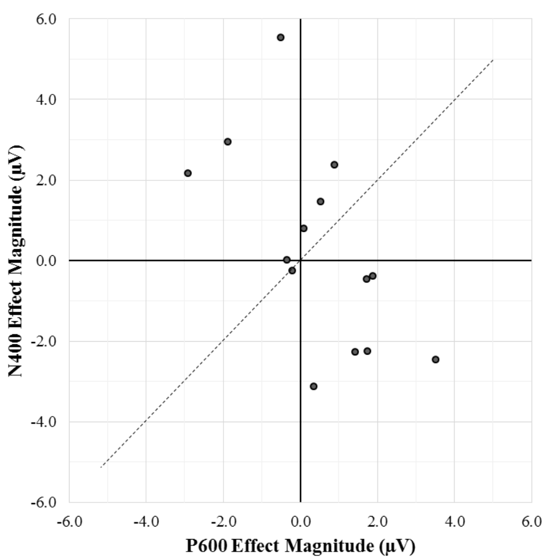


(a) Article

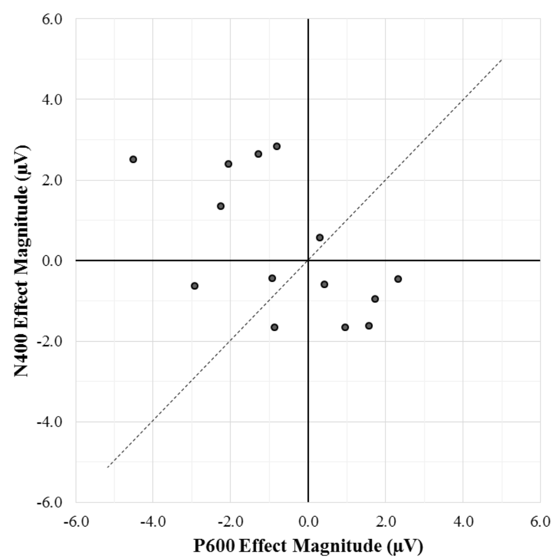


(b) Adjective

Figure 22. At Home Group, Baseline: Scatterplot showing the distribution of N400 and P600 effect magnitudes across participants. The dashed line represents equal N400 and P600 effect magnitudes. Each point represents one participant.

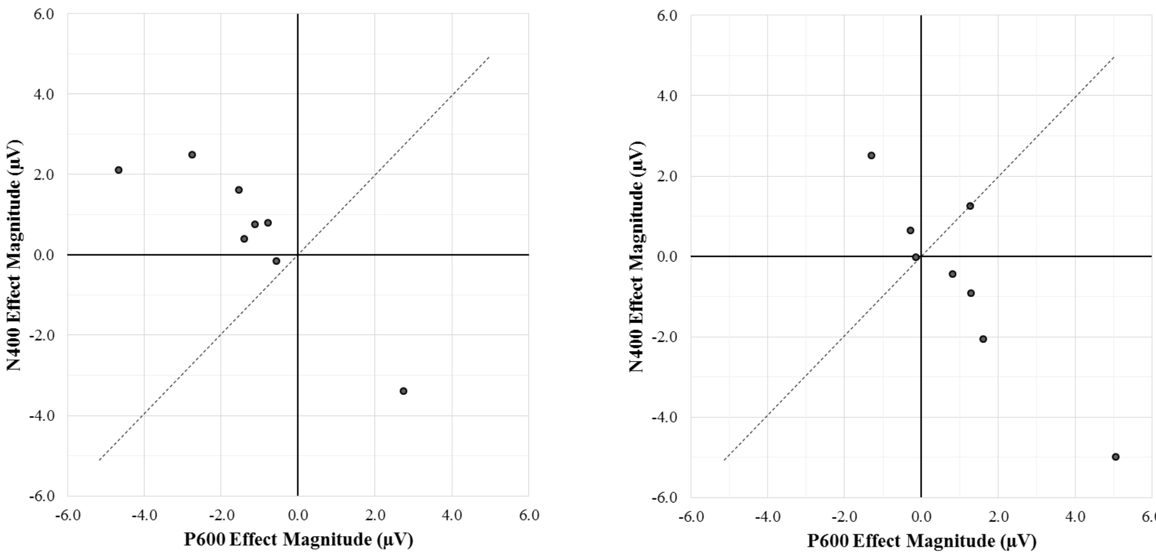


(a) Article



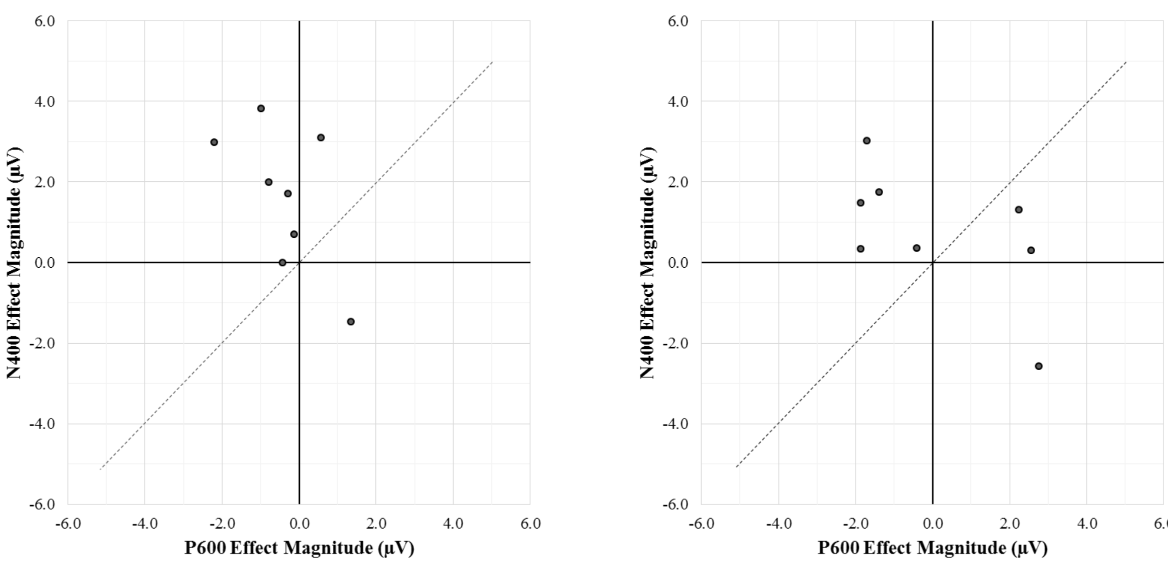
(b) Adjective

Figure 23. At Home Group, Follow-Up: Scatterplot showing the distribution of N400 and P600 effect magnitudes across participants. The dashed line represents equal N400 and P600 effect magnitudes. Each point represents one participant.



(a) Article (b) Adjective

Figure 24. Study Abroad Group, Baseline: Scatterplot showing the distribution of N400 and P600 effect magnitudes across participants. The dashed line represents equal N400 and P600 effect magnitudes. Each point represents one participant.



(a) Article (b) Adjective

Figure 25. Study Abroad Group, Follow-Up: Scatterplot showing the distribution of N400 and P600 effect magnitudes across participants. The dashed line represents equal N400 and P600 effect magnitudes. Each point represents one participant.

In order to further explore within-group variability, RMI (measure of overall sensitivity to agreement violations within the N400 and P600 time windows) and RDI (relative response dominance – negativity in the N400 time window versus positivity in the P600 time window) were calculated for each participant. Participants' RMI values for each condition at Baseline and Follow-Up are presented in Figure 26 and Figure 27 (At Home group) and Figure 28 and Figure 29 (Study Abroad group). Participants' RDI values for each condition at Baseline and Follow-Up are presented in Figure 30 and Figure 31 (At Home group) and Figure 32 and Figure 33 (Study Abroad group).

Descriptively, participants within each group showed different processing of violations within each condition at Baseline and Follow-Up testing. Furthermore, extensive variability in change in RMI and RDI from Baseline to Follow-Up was evidenced among participants in both groups.

In terms of RMI, many participants showed increased RMI (that is, greater neural response to violations) at Follow-Up as compared to Baseline, but some showed no change, or even a decrease in RMI. Comparisons between Baseline and Follow-Up RDI for both conditions showed similar variability: many participants' RDI shifted from negative to positive (or neutral) from Baseline to Follow-Up (a neutral RDI reflects relatively equal-sized effects in the N400 and P600 time windows, regardless of the size of these effects), but individuals also showed (1) more negative RDI at Follow-Up than Baseline, (2) positive RDI at Baseline and neutral (or slightly negative) RDI at Follow-Up.

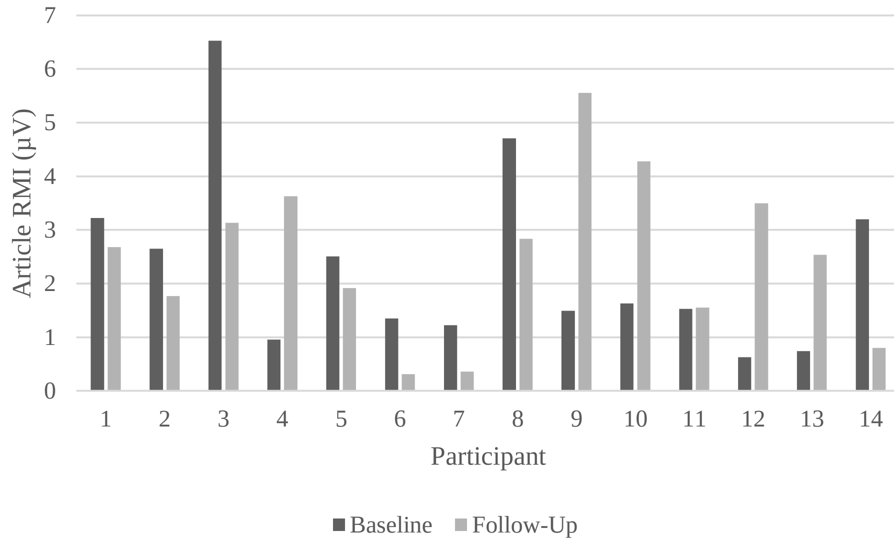


Figure 26. At Home group, Article RMI: Bar chart showing Article RMI at Baseline and Follow-Up for each participant.

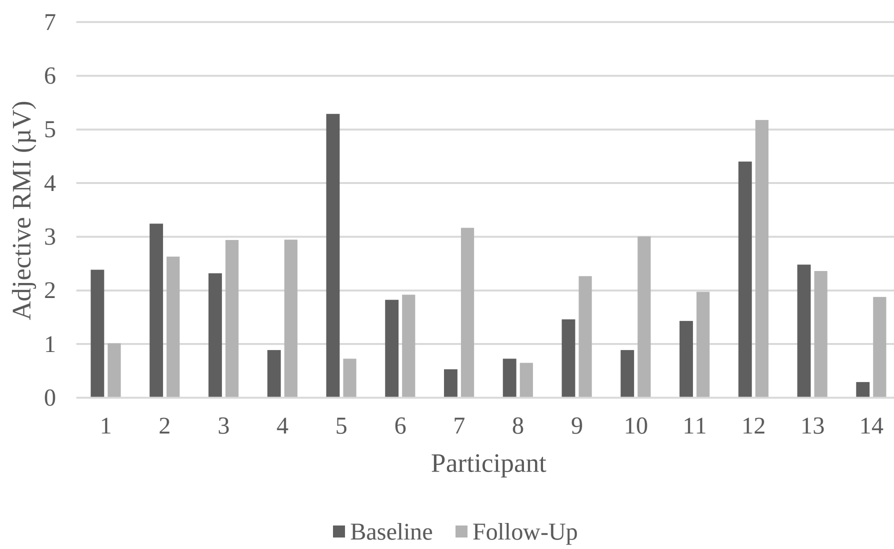


Figure 27. At Home group, Adjective RMI: Bar chart showing Adjective RMI at Baseline and Follow-Up for each participant.

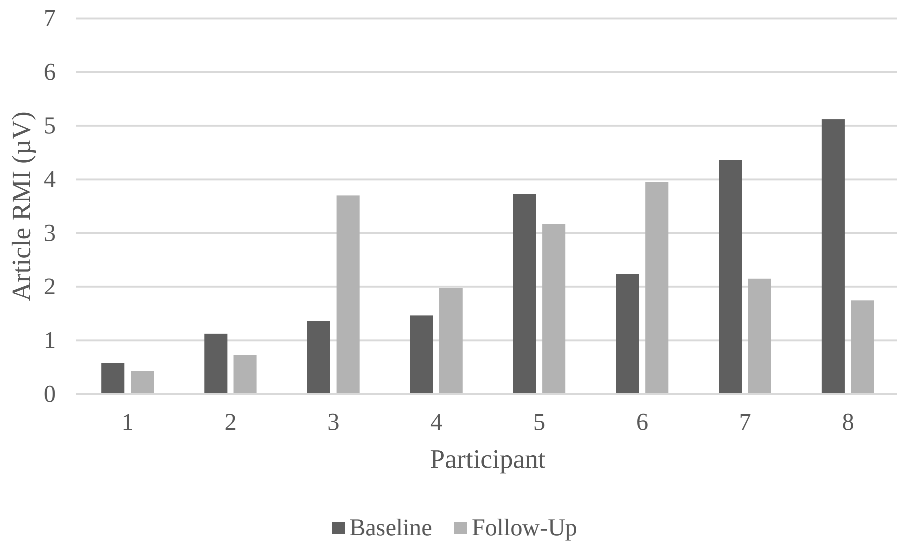


Figure 28. Study Abroad group, Article RMI: Bar chart showing Article RMI at Baseline and Follow-Up for each participant.

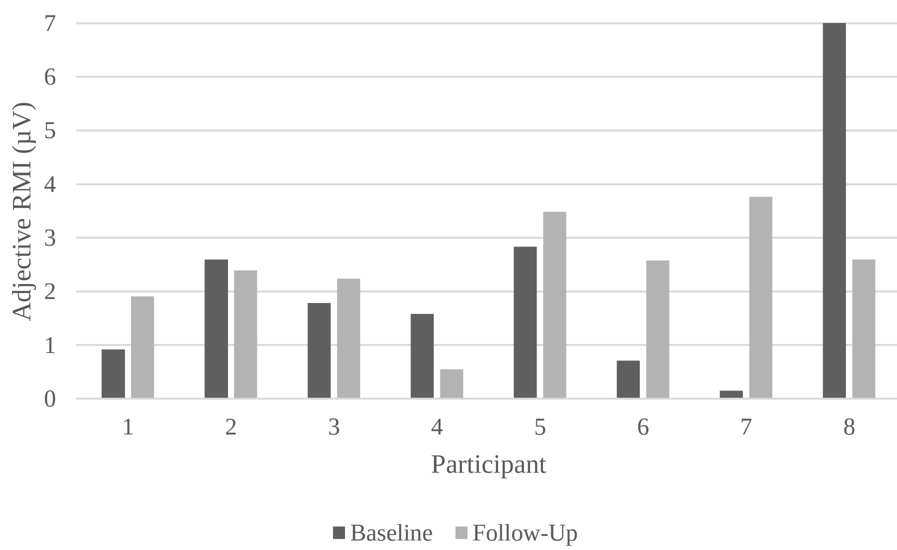


Figure 29. Study Abroad group, Adjective RMI: Bar chart showing Adjective RMI at Baseline and Follow-Up for each participant.

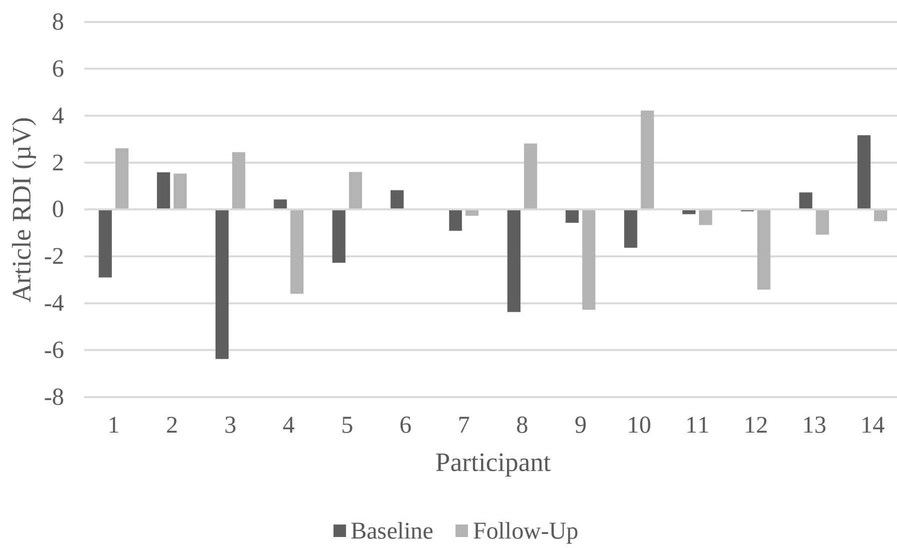


Figure 30. At Home group, Article RDI: Bar chart showing Article RDI at Baseline and Follow-Up for each participant.

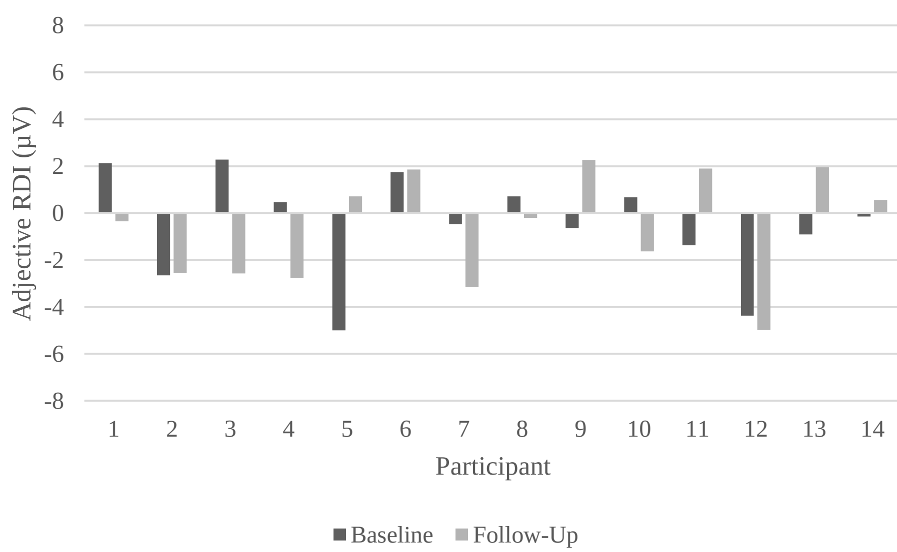


Figure 31. At Home group, Adjective RDI: Bar chart showing Adjective RDI at Baseline and Follow-Up for each participant.

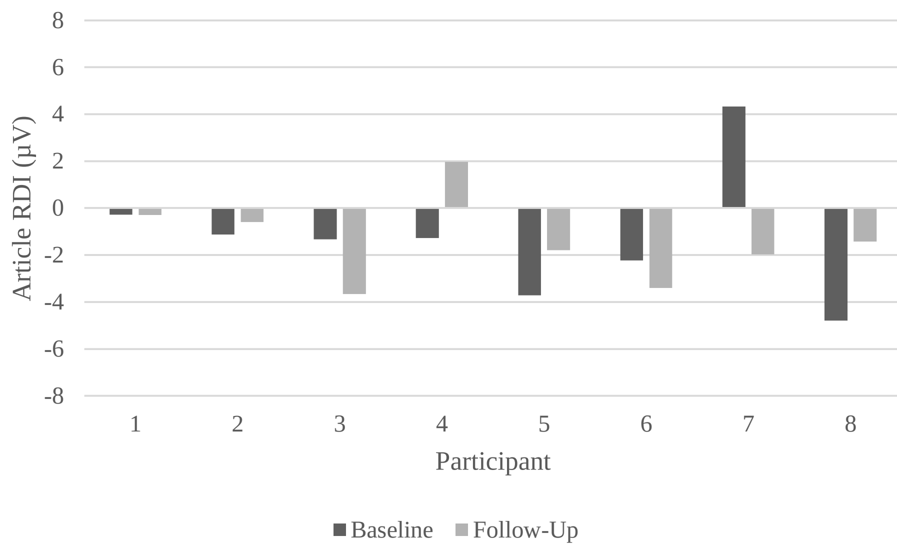


Figure 32. Study Abroad group, Article RDI: Bar chart showing Article RDI at Baseline and Follow-Up for each participant.

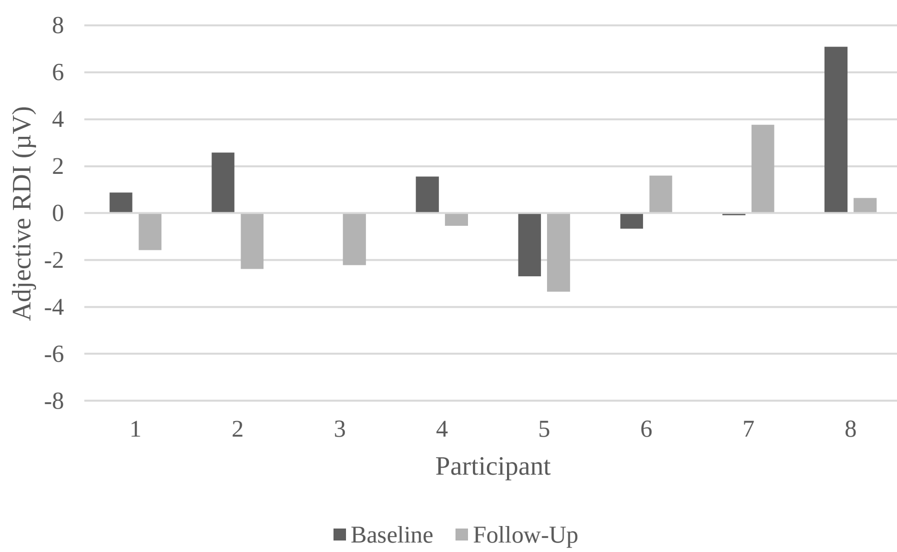


Figure 33. Study Abroad group, Adjective RDI: Bar chart showing Adjective RDI at Baseline and Follow-Up for each participant.

Given the variability present (1) within groups at each testing session (Figure 22 through Figure 25) and (2) in processing changes from Baseline to Follow-Up testing (Figure 26 through Figure 33), regression analyses were conducted to examine whether individual differences in L2 Contact (RQ2a) and/or WM abilities (RQ2b) could account for changes in individual processing signatures. Specifically, simple linear regressions were conducted in which each of the metrics of processing change (RMI and RDI change for Articles and Adjectives) were regressed onto the predictor variables L2 Contact and WM.

At Home Group

Within the At Home group, changes in individual processing signatures in the Article condition did not appear to be related to individual differences in L2 Contact or WM: neither L2 Contact nor WM predicted changes in processing of Articles, as indexed by RMI or RDI (Table XVII).

In terms of Adjective processing, regressions revealed relationships between individual difference factors and RMI change (change in the size of the neural response to violations over both the N400 and P600 time windows) as well as RDI change (change in relative response dominance – negativity in the N400 time window versus positivity in the P600 time window). An increase in RMI from Baseline to Follow-Up indicates an increased level of sensitivity to agreement violations, regardless of type of response; an increase in RDI from Baseline to Follow-Up testing indicates that a participant is becoming less negative-dominant / more positive-dominant.

Results of the linear regression analysis with WM as the predictor variable and Adjective RMI change as the dependent variable revealed that WM was a significant, positive predictor of

Adjective RMI change, indicating that increased RMI for adjective violations was associated with higher WM abilities. Specifically, for every additional point in WM composite (one standard deviation from the group mean), Adjective RMI change increased by approximately 1.5 μV (*B* = 1.479). WM accounted for nearly 44% of variance in Adjective RMI change (R^2 = .438). The relationship between WM and Adjective RMI change is represented graphically in Figure 34.

Results of the linear regression analysis with L2 Contact as the predictor variable and Adjective RDI change as the dependent variable revealed that L2 Contact was a significant, positive predictor of Adjective RDI change, indicating that positive RDI change for adjective violations was associated with higher levels of L2 Contact. Specifically, with every additional hour of average weekly L2 Contact reported, Adjective RDI increased by .157 μV (*B* = .157). L2 Contact accounted for approximately 25% of variance in Adjective RDI change (R^2 = .255). The relationship between L2 Contact and Adjective RDI change is represented graphically in Figure 34. Regressions for both Adjective processing metrics are presented in Table XVIII.

Table XVII

At Home Group: Change in Article Processing, Regressions

	Article RMI Change			Article RDI Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	-.269	1.110		1.566	2.193	
L2 Contact	.033	.067	.139	-.041	.133	-.089
R^2		.019			.008	
F		.237			.097	
			Working Memory			
Constant	.179	.624		1.001	1.190	
Working Memory	.214	.807	.076	-1.385	1.540	-.251
R^2		.006			.063	
F		.071			.808	

Note. RMI Change = change in RMI from Baseline to Follow-Up; RDI Change = change in RDI from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.

Table XVIII

At Home Group: Change in Adjective Processing, Regressions

	Adjective RMI Change			Adjective RDI Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	1.137	.847		-2.249	1.275	
L2 Contact	-.060	.051	-.318	.157	.077	.505
R^2		.101			.255	
F		1.349			4.116^	
			Working Memory			
Constant	.321	.374		-.102	.763	
Working Memory	1.479	.484		-1.403	.988	-.379
R^2		.438			.144	
F		9.344*			2.016	

Note. RMI Change = change in RMI from Baseline to Follow-Up; RDI Change = change in RDI from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.
^ $p < .1$; * $p < .05$

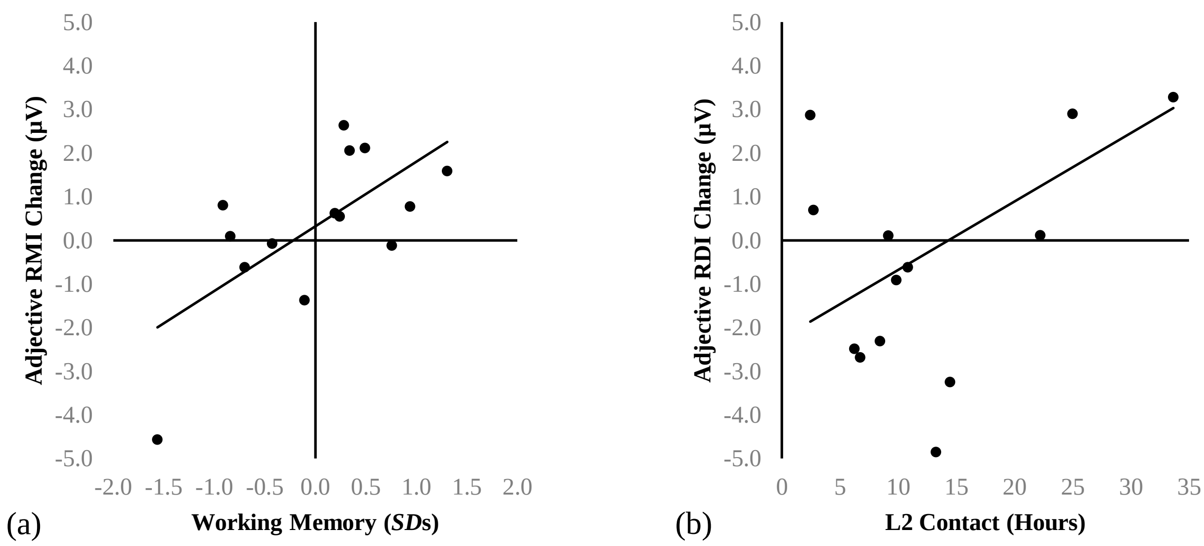


Figure 34. At Home Group: Adjective processing scatterplots illustrating the relationship between (a) WM and Adjective RMI change and (b) L2 Contact and Adjective RDI change. Each point represents one participant. Solid line shows the best-fit line for the data from the regression analysis.

Study Abroad Group

Within the Study Abroad group, changes in individual processing signatures did not appear to be related to individual differences in L2 Contact or WM. Indeed, results of simple linear regression analyses revealed that neither L2 Contact nor WM predicted changes in processing of Articles (Table XIX) or Adjectives (Table XX).

Table XIX

Study Abroad Group: Change in Article Processing, Regressions

	Article RMI Change			Article RDI Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	1.762	1.421		-2.177	2.252	
L2 Contact	-.020	.023	-.406	.043	.036	.516
R^2		.165			.266	
F		.791			1.452	
			Working Memory			
Constant	-.264	.618		-.095	1.226	
Working Memory	-1.402	.958	-.513	.237	1.900	.051
R^2		.263			.003	
F		2.140			.016	

Note. RMI Change = change in RMI from Baseline to Follow-Up; RDI Change = change in RDI from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.

Table XX

Study Abroad Group: Change in Adjective Processing, Regressions

	Adjective RMI Change			Adjective RDI Change		
Variable	*B*	*SEB*	*β*	*B*	*SEB*	*β*
			L2 Contact			
Constant	1.033	1.257		-1.277	3.079	
L2 Contact	-.010	.020	-.238	-.007	.049	-.072
R^2		.057			.005	
F		.241			.021	
			Working Memory			
Constant	.231	.871		-1.593	1.299	
Working Memory	-.907	1.350	-.265	-.454	2.013	-.092
R^2		.070			.008	
F		.452			.051	

Note. RMI Change = change in RMI from Baseline to Follow-Up; RDI Change = change in RDI from Baseline to Follow-Up; *B* = unstandardized regression coefficient; *SEB* = standard error of *B*; *β* = standardized regression coefficient.

To summarize the results from individual-level processing analyses, within the At Home group, variability in processing signatures was evidenced for both conditions at both testing sessions. Also evidenced was variability in changes in individual processing signatures (RMI and RDI change) from Baseline to Follow-Up testing. Within this group, neither L2 Contact nor WM accounted for changes in processing of grammatical gender agreement violations on articles. For the Adjective condition, WM was a significant, positive predictor of change in RMI, such that higher WM was associated with greater increase in Adjective RMI from Baseline to Follow-Up. L2 Contact was found to be a significant, positive predictor of change in RDI, such that increased L2 Contact was associated with a more positive shift in Adjective RDI from Baseline to Follow-Up testing.

Within the Study Abroad group, variability in processing signatures was evidenced for both conditions at both testing sessions. Also evidenced was variability in changes in individual processing signatures (RMI and RDI change) from Baseline to Follow-Up testing. Neither L2 Contact nor WM accounted for changes in processing of grammatical gender agreement violations on articles or adjectives within this group.

5 DISCUSSION AND CONCLUSIONS

This chapter reviews the predictions put forth in Chapter 2 and situates the findings of the study within the broader context of SLA and neurocognitive research. Results that speak to each particular hypothesis are addressed first, followed by limitations and future research directions, and general conclusions drawn from the present study.

5.1 <u>Discussion</u>

Behavioral Results

Hypothesis 1a predicted that L2 contact would account for behavioral gains in both At Home and Study Abroad learner groups. This prediction was partially supported within the Study Abroad group; within the At Home group, however, a negative relationship was revealed between L2 contact and behavioral change.

Among At Home learners, higher levels of L2 contact were associated with smaller, and even negative, changes in overall proficiency. No relationship was found between reported L2 contact and changes in sensitivity to or production of grammatical gender agreement on articles or adjectives. Among the Study Abroad learners, individual differences in L2 contact predicted gains in sensitivity to grammatical gender agreement violations on adjectives, as well as accurate production of gender-marked adjectives. No relationship was found between L2 contact and changes in sensitivity to or production of grammatical gender agreement on articles or in overall proficiency.

This set of outcomes, like the larger body of research addressing the role of L2 contact in linguistic development, provides ambiguous evidence regarding the role of L2 contact. Within the Study Abroad group, L2 contact hours served as a significant, positive predictor of gains in judgment (GJT) and production (Info Gap Task) accuracy for gender agreement on adjectives, but did not predict changes in overall proficiency as measured by the DELE. This pattern of results partially corroborates findings from Magnan and Back (2007), suggesting that L2 contact accounts for more change in production abilities than overall proficiency.

In terms of the At Home group, this study is the first to provide evidence of a role for L2 contact among learners in an at home setting. The somewhat surprising negative relationship between L2 contact and overall proficiency changes in the At Home group merits further consideration. Previous research has indicated that type of L2 contact, as opposed to overall amount, may be relevant in determining relationships with linguistic gains (e.g., Isabelli-García, 2010; Magnan & Back, 2007). A median split was conducted within the At Home learners based on average L2 contact hours in order to determine if a subgroup of learners were driving relationship with L2 contact. Results of the median split revealed that learners with "High" levels of L2 contact (n = 7) reported approximately double the average weekly hours as reported by the "Low" L2 contact group (n = 7) for each of the four language activities (speak, read, listen, write). Strikingly, "High" L2 contact learners reported an average of 5.4 hours per week listening to Spanish, compared to just 0.5 hours in the "Low" L2 contact group. Further investigation revealed that the majority of listening hours reported in the "High" L2 contact group represented hours spent listening to music in the L2; it is possible that this receptive form of L2 contact did not drive learners to depths of L2 processing that would lead to gains in overall proficiency (Freed et al., 2004).

Overall, these results indicate that L2 contact plays a different role in at home and study abroad contexts. In providing the first empirical evidence to indicate a role for L2 contact among at home learners, these results underscore the importance of including an empirical assessment of the amount of L2 contact among at home learners.

Hypothesis 1b predicated that WM would account for behavioral gains for learners in the At Home and Study Abroad groups, and that individual differences in WM ability would account for more variation among learners in the Study Abroad group. This hypothesis was partially supported within the At Home group, but was not supported within the Study Abroad group.

Within the At Home group, higher WM abilities were associated with greater changes in overall proficiency measure. WM did not account for changes in sensitivity to or production of grammatical gender agreement on articles or adjectives within the At Home group; within the Study Abroad group, WM did not account for behavioral change on any metric.

The positive relationship between WM and behavioral gains found for the At Home group is consistent with previous research (e.g., Sagarra, 2007; Sagarra & Herschensohn, 2010) as well as with theoretical claims regarding a role for WM in L2 acquisition (e.g., McDonald, 2006; Williams, 2012). This relationship, however, is limited to gains in overall proficiency. The lack of relationship with gains in grammatical gender agreement production and judgment accuracy may indicate that the role of WM varies by structure (e.g., McDonald, 2006; Sagarra & Herschensohn, 2010), and that at this proficiency level grammatical gender agreement is not a structure whose acquisition is supported by WM.

The dissociation between WM and behavioral gains within the Study Abroad group runs contrary to suggestions made by previous research and theories that WM plays a particularly important role in linguistic development in immersion settings (e.g., LaBrozzi, 2009; Sunderman & Kroll, 2009; Tokowicz et al., 2004). One explanation for the apparent disparity between results for the present group of Study Abroad learners and previous studies lies in research methodology differences: whereas the present study assessed WM abilities prior to immersion experience and analyzed the relationship between pre-study abroad WM and linguistic change that occurred during the semester of study, previous work has examined the relationship between WM and L2 abilities at a single time point after SAE has occurred. This difference in outcomes suggests that other factors may be at play in analyses that draw conclusions about interactions between WM and SAE.

In interpreting relationships between internal factors and behavioral gains for the two participant groups in the present study, it is important to note the limited behavioral "gains" observed from Baseline to Follow-Up testing. Within the At Home group, significant change was evidenced only for accuracy in production of gender-marked adjectives; within the Study Abroad group, significant changes were evidenced for overall proficiency and sensitivity to grammatical gender violations on articles. These results corroborate findings from the at home and study abroad groups studied by Isabelli-García (2010), who also failed to show significant gains from pre- to post-semester in terms of accuracy on grammatical gender agreement violation judgments, but also pose a challenge for individual differences to account for "gains." It is also important to note the small group sizes in the present study, which, although not strikingly different from other context-based studies, certainly impact statistical power.

Processing Results

Hypothesis 2a stated that L2 contact would account for positive change in RDI for learners in both groups. This prediction supported within the At Home group, but was not supported within the Study Abroad group.

Within the At Home group, L2 contact accounted for changes in RDI for grammatical gender agreement violations on adjectives. That is, learners who reported greater levels of L2 contact showed a greater positive change in response type from Baseline to Follow-Up testing. No relationship was found between L2 contact and changes in ERP processing signatures within the Study Abroad group.

In terms of situating these findings within the broader literature, the use of ERPs as a measure of individual differences is an emerging analysis method, and previous studies have not examined within-subjects changes in neural responses using these metrics. These results, however, are in line with longitudinal studies that show group-level changes from no ERP effect or N400 effects to P600 effects for morphosyntactic violations with increased L2 exposure (McLaughlin et al., 2010; Morgan-Short et al., 2010). These results can also be compared with previous work that has examined qualitative (response type) differences between subjects. Specifically, the background variables age of arrival in an L2 environment and motivation to sound like a native speaker were associated with relative brain response type (RDI; N400 or P600) among highly proficiency, immersed learners (Tanner et al., 2014). Previous research has found a relationship between grammatical sensitivity (*d'* score) and RDI, such that learners with higher *d'* scores showed a more positive (P600-like) response to morphosyntactic violations (Tanner et al., 2013), an effect that was absent in both the At Home and Study Abroad learner groups in the present study.[14]

Hypothesis 2b predicted that WM ability would account for positive change in RMI and RDI for learners in both groups, and that WM ability would account for more variation in

[14] Post-hoc analyses within the present participant groups revealed no significant relationships between *d'* score at Baseline or Follow-Up testing and processing signatures at either session, revealing that within the present groups of learners, neither response magnitude nor response type were predicted by behavioral sensitivity to agreement violations.

processing changes among the Study Abroad learners. This prediction was supported within the At Home group, but was not supported within the Study Abroad group.

Within the At Home group, WM accounted for changes in RMI for grammatical gender agreement violations on adjectives. That is, learners with higher WM abilities showed greater increases in neural sensitivity to these violations (RMI), but did not show systematically different changes in response type (RDI). The relationship between WM and processing of adjectives, but not articles, may reflect greater processing demands for adjectives than for articles. Indeed, behavioral scores for article violations and production were consistently higher than for adjectives within the At Home group, suggesting that grammatical gender agreement on adjectives posed a greater processing demand for this group. These results are largely in line with Sagarra and Herschensohn (2010), where WM played a role in gender – but not number – agreement processing in a self-paced reading task, where behavioral scores revealed that gender agreement was more difficult for learners and therefore constituted greater processing demands.

The lack of relationship between neural sensitivity and WM in the Study Abroad group is not consistent with previous research, which has found that higher WM facilitates L2 processing for learners with SAE (LaBrozzi, 2009; Lafford, 2006; Tokowicz et al., 2004; Sunderman & Kroll, 2009). In addition to differences in overall research design between the present study and this previous work, the present study also utilized a different measure of online processing (ERPs, as opposed to eye-tracking or self-paced reading). These significant differences in research methodology may partially explain the disparity in results. The limited number of participants within the Study Abroad group may also have contributed to the lack of significant relationships between internal factors and processing changes.

Taken together, these results contribute to the discussion of individual variability in morphosyntactic processing signatures, and the potential use of ERPs to quantify these differences. In addition to supporting the use of ERPs as a measure of individual differences in linguistic processing across subjects, the present findings suggest that experiential (L2 contact) and cognitive (WM) factors can account for *changes* in processing signatures within subjects. Additional research that measures cognitive, background, and behavioral factors in conjunction with individual measures of ERP response size and type is vital to understanding these relationships.

5.2 Limitations and Future Directions

Although this work has attempted to address numerous issues related to experimentally assessing the role of individual differences in L2 development in natural settings, a number of limitations remain.

First, the number of participants in the present study is quite low, especially given the research goals of exploring individual differences. Low participant numbers likely impacted statistical power in the present study; future research will benefit from including more participants, with continued careful assessment of learner characteristics such as previous L2 experience and L2 proficiency level.

Another limitation is the short duration of immersion in the L2 environment. The results of analyses with Study Abroad learners lead to limited significant findings. Longer stays in the L2 environment may have a stronger impact on morphosyntactic processing.

Variability in L2 experiences within both the At Home and Study Abroad groups during the semester of participation constitutes another major limitation of this research. Within the At

Home group, for example, learners were enrolled in one to three Spanish courses, including diverse topics such as advanced grammar review, introduction to Spanish linguistics, and literary analysis. The variability in coursework completed during the semester of participation may have impacted learning outcomes within the At Home group. Due to logistic limitations, learners within the Study Abroad group participated in a number of different study abroad programs in different Spanish-speaking countries. Although participants in the Study Abroad group completed all of their coursework in Spanish, a number of variables were uncontrolled. As in the At Home group, Study Abroad participants were enrolled in a range of different courses: participants completed between three and five courses taught in Spanish that covered a broad range of topics (e.g., Spanish grammar, linguistics, literary analysis, history, culture, and gastronomy). Participants also reported different living situations: home stay and private apartment, and different program organization: direct-enroll with native Spanish-speakers and foreign-student only programs.

Given this intra-group variability, the present study is unable to draw conclusions directly related to the role of context of learning in L2 development. It is important to understand and quantify aspects of the at home and study abroad context that may interact with individual differences in experiential and cognitive factors to impact linguistic development. A systematic analysis of learning context, including different study abroad programs, is crucial in addressing practical, applied questions related to the context – and specific program type – that may be most beneficial for a particular learner.

An additional limitation to the present study is a lack of longitudinal data aimed at examining whether behavioral and processing changes are durable. Future research may benefit from including a delayed post-test in order to test the durability of behavioral and processing changes.

Finally, this study makes a novel contribution to the field by measuring changes in online processing. Given the lack of a control group (e.g., learners at a similar proficiency level who are no longer engaged in L2 study), it is not possible to guarantee that changes in ERP signatures from Baseline to Follow-Up reflect changes in morphosyntactic processing related to L2 development and not merely maturational effects. Within-subjects changes in processing in both L1 and L2 need to be empirically examined by future work.

5.3 Conclusions

The present study investigated whether individual differences in language contact and WM may explain variability in behavioral and processing changes among L2 learners in at home and study abroad settings. The results suggest that both language contact and WM account for variability in behavioral and processing changes, but that these factors play out differently for learners in different contexts, and that the impact of these internal factors is not consistent for all grammatical structures or linguistic abilities.

Specifically, WM accounted for gains in overall proficiency within the At Home group, as well as increase in neural response to grammatical gender agreement violations on adjectives. L2 contact predicted qualitative changes in neural response to grammatical gender agreement violations on adjectives within the At Home group, and was also associated with suppressed improvement on the overall proficiency measure. Within the Study Abroad group, L2 contact predicted improvement in grammatical sensitivity to and accurate production of gender-marked adjectives. The lack of relationship between WM and behavioral and processing changes within the Study Abroad group indicate that the role of working memory in immersion settings may not be as significant as posited based on non-experimental assessments of immersion experience (e.g., LaBrozzi, 2009; Sunderman & Kroll, 2009; Tokowicz et al., 2004) and highlight the importance of examining this relationship using a variety of metrics within a pre-test/post-test design.

Importantly, the results reveal gains in grammatical abilities over the course of one semester of study for both At Home and Study Abroad learners at a low to intermediate level of proficiency. ERP data reveal both quantitative and qualitative changes in processing of grammatical gender agreement violations on articles and adjectives for individual participants within both groups – changes that are mostly obscured in group-level analyses. These findings provide support for models of SLA that implicate changes in the neural substrates of L2 grammatical processing with increased proficiency and exposure, and underscore the potential for obtaining rich, informative data via individual-level ERP analysis.

Taken together, the results of this study provide support for Collentine and Freed's (2004) call for theories of language acquisition and processing that incorporate consideration of cognitive abilities and context of learning. Future research that utilizes a multidimensional approach informed by SLA and cognitive neuroscience is likely to provide further insights into the relationships between external and internal factors in L2 development and have significant implications for identifying the predictors of successful L2 acquisition among adult learners.

CITED LITERATURE

Baddeley, A. D. (2007). *Working memory, thought, and action*. New York, NY, US: Oxford University Press, New York, NY.

Baddeley, A., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137–153.

Berquist, B. (1997). Individual differences in working memory span and L2 proficiency: Capacity or processing efficiency? In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition,* Edinburgh, UK.

Bialystok, E., & Frölich, M. (1978). Variables of classroom achievement in second language learning. *The Modern Language Journal*, *62*, 327-336.

Bond, K., Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In D. Tanner & J. Herschensohn (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, pp. 17–29. Somerville, MA: Cascadilla Proceedings Project.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia, 51*, 2492-2511.

Brecht, R. D., Davidson, J. L., & Ginsberg, R. B. (1995). Predictors of foreign language gain during study abroad. In B. Freed (Ed.), *Second Language Acquisition in a study abroad context* (pp. 37-66). Amsterdam: John Benjamins.

Carpenter, H. (2008). *A behavioral and electrophysiological investigation of different aptitudes for L2 grammar in learners equated for proficiency level*. (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Carroll, J. B. (1958). A factor analysis of two foreign language aptitude batteries. *Journal of General Psychology, 59(1),* 3-19.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87-136). Pittsburgh: University of Pittsburgh Press.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in foreign language aptitude.* (pp. 83-118). Rowley, MA: Newbury House.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. San Antonio, TX, US: Psychological Corporation, San Antonio, TX.

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

Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry, 31*(2), 319-355.

Cheng, A. C., & Mojica-Diaz, C. (2006). The Effects of Formal Instruction and Study Abroad on Improving Proficiency: The Case of the Spanish Subjunctive. *Applied Language Learning, 16*(1), 17-36.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: Massachusetts Institute of Technology.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*(1), 3–42.

Collentine, J. & Freed, B.F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition, 26*(2), 153-171.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, *26*(2), 227-248.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21-58.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

Davidson, D. E. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals, 43*(1), 6-26.

DeKeyser, R. (1986). From learning to acquisition? Foreign language development in a U. S. classroom and during a semester abroad. (Unpublished doctoral dissertation). Stanford University, Stanford, California.

DeKeyser, R. (1990). From learning to acquisition? Monitoring in the classroom and abroad. *Hispania*, *73*(1), 238-247.

DeKeyser, R. M. (1991). Foreign language development during a semester abroad. In B. F. Freed (Ed.), *Foreign language acquisition and the classroom* (pp. 104-118). Lexington, MA: D. C. Heath.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22,* 499-533.

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

Dewaele, J. M., & Veronique, D. (2001). Gender assignment and gender agreement in advanced French inter-language: A cross-sectional study. *Bilingualism: Language and Cognition, 4*(3), 275–297.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Doughty, C. & Williams, J. (1998). In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.

Ehrman, M. E., & Oxford, R. L. (1995). (1995), Cognition Plus: Correlates Of Language Learning Success. *The Modern Language Journal*, 79(1), 67–89.

Ellis, N. C. (1998). Emergentism, connectionisn, and language learning. *Language learning 48*(4), 631-664.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics - Introduction to the special issue. *Applied Linguistics*, *27*(4), 558–589

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning, 59*(1), 90-125.

Ely, C. (1986). An analysis of discomfort, risk-taking, sociability, and motivation in the L2 classroom. *Language Learning, 36*(1), 1-25.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23.

Engle, R.W., & Kane, M.J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York: Elsevier.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*(1), 250-272.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379-399.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language, 66*(1), 226-248.

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amsterdam: John Benjamins.

Freed, B. F. (1990). Language learning in a study abroad context: The effects of interactive and non-interactive out-of-class contact on grammatical achievement and oral proficiency. In J. Atlatis (Ed.), *Selected Proceedings of the Georgetown University Round Table on Languages and Linguistics: Linguistics, language teaching and language acquisition: the interdependence of theory, practice and research.* (pp. 459-477). Washington, DC: Georgetown University Press.

Freed, B. F. (1995a). Introduction. In B. F. Freed (Ed.), *Second Language Acquisition in a study abroad context* (pp. 3-34). Amsterdam: John Benjamins.

Freed, B. F. (1995b). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123-148). Amsterdam: John Benjamins.

Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. *Frontiers 4*(1), 31-60.

Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition, 26*(2), 349-356.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition, 26*(2), 275-301.

Frenck-Mestre, C., Foucart, A., Carrasco, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language: Evidence from ERPs. *EUROSLA Yearbook, 9*(1), 76-106.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica, 128*(3), 528-536.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second pass parsing processes. *Journal of Psycholinguistic Research, 25*(1), 157-176.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, 37*(4), 581-590.

Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2013). Examining second language development using event-related potentials: a cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism, 3*(2), 213-232.

Gardner, R. C. (2004). *Attitude/Motivation test battery: International AMTB research project.* Canada: The University of Western Ontario.

Gillon Dowens, M., Barber, H., Guo, T., Guo, J., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from event related potentials. *Neuropsychologia 49*(7):1651-1659.

Gillon Dowens, M., Barber, H., Vergara, M., & Carreiras, M. (2010). Morpho-syntactic processing in late L2 learners. *Journal of Cognitive Neuroscience, 22*(8), 1870-1887.

Golonka, E. M. (2006). Predictors revised: Linguistic knowledge and metalinguistic awareness in second language gain in Russian. *Modern Language Journal, 90*(4), 496-505.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics, 22*(1), 123–134.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883–899.

Hagoort, P., Brown, C. M., & van Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*(1). 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(2), 123-141.

Hahne, A., Mueller, J., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121-134.

Harley, B. & Hart, D. (1997). Language aptitude and language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*(1), 379-400.

Harley, B. & Hart, D. (2002). Age, aptitude, and second language learning on a bilingual exchange. In P. Robinson (Ed.), *Individual differences in instructed language learning* (pp. 301-330). Philadelphia: John Benjamins Publishing Company.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*(1), 25-38.

Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage, 15*(3), 647-60.

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning, 59*(1), 73-112.

Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *Modern Language Journal, 94*(4), 600-617.

Hinojosa, J. A., Martín-Loeches, M., & Rubia, F. J. (2001). Event-related potentials and semantics: An overview and an integrative proposal. *Brain and Language, 78*, 128-139.

Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing.* (Unpublished doctoroal dissertation). University of Groningen, Groningen, The Netherlands.

Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal, 70*(2), 125-132.

Isabelli, C.A. (2001). The impact of a study-abroad experience on the acquisition of L2 Spanish syntax: The null subject parameter. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Isabelli, C.A. (2004). The acquisition of NSP properties in SLA: Some effects of positive evidence in a natural learning setting. *Hispania, 87*, 150-162.

Isabelli, C.A. (2007). Development of the Spanish subjunctive by advanced learners: Study abroad followed by at-home instruction. *Foreign Language Annals, 40*(2), 330-341.

Isabelli, C. A., & Nishida, C. (2005). Development of the Spanish Subjunctive in a Nine-Month Study-Abroad Setting. In D. Eddington (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 78-91). Somerville, MA: Cascadilla Proceedings Project.

Isabelli-Garcia, C. (2010). Acquisition of Spanish gender agreement in two learning contexts: Study abroad and at home. *Foreign Language Annals, 43*(2), 289-303.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199-225.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159-201.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169-183.

Kane, M. J., Conway, A.R.A., Hambrick, D.Z., & Engle, R.W. (2007). Variation in working-memory capacity as variation in executive attention and control. In A.R.A. Conway, C. Jarrold, M.J.Kane, A. Miyake, & J.N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217.

Kaufman, A.S., & Kaufman, N.L. (2004). Kaufman Assessment Battery for Children Second Edition. Circle Pines, MN: American Guidance Service.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*(3), 376-395.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137-171.

Kutas, M. & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, *11*, 539-550.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential *Annual Review of Psychology, 62(1),* 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(1), 203-205.

LaBrozzi, R. M. (2009). *Processing of lexical and morphological cues in a study abroad context.* (Unpublished doctoral dissertation). Pennsylvania State University, State College, P.A.

LaBrozzi, R. M. (2012). The role of study abroad and inhibitory control on processing redundant cues. *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, ed. Kimberly Geeslin and Manuel Díaz-Campos, 228-241. Somerville, MA: Cascadilla Proceedings Project.

Lafford, B. A. (1995). Getting into, through, and out of a survival situation: A comparison of communicative strategies used by students studying Spanish abroad and “at home.” In B. F. Freed (Ed.), *Second language acquisition in the study abroad context* (pp. 97-121). Amsterdam: John Benjamins.

Lafford, B. A. (2004). The effect of the context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition*, *26*(2), 201-225.

Lafford, B. A. (2006). The effects of study abroad vs. classroom contexts on Spanish SLA: Old assumptions, new insights and future research directions. In C.A. Klee & T.L. Face (Eds.), *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages* (pp. 1-25). Somerville, MA: Cascadilla Proceedings Project.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B.F. Freed (Ed.), *Second language acquisition in the study abroad context* (pp. 67-94). Amsterdam: John Benjamins.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: [de]constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition, 25*(1), 37-63.

Leeser, M. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: topic familiarity and working memory. *Language Learning*, *57*(2), 229-270.

Lightbown, P. M., & Spada, N. (1990). Focus on form and corrective feedback in communicative language teaching: effects on second language learning. *Studies in Second Language Acquisition, 12*(1), 429-48.

Long, M. H. 1997. ‘Construct validity in SLA: A response to Firth and Wagner,’ The Modern Language Journal 81(3), 18–23.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35-54.

Mackey, A., Adams, R., Stafford, C., Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*(3), 501- 533.

Mackey, A., Philp, J., Fujii, A., Egi, T., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson & P. Skehan (Eds.), *Individual differences in L2 learning* (pp. 181-208). Philadelphia: John Benjamins.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide, 2nd Edition*. New York, NY: Cambridge University Press.

Macready, C., & Tucker, C. (2011). *Who goes where and why? An overwiew and analysis of global educational mobility.* (Global Education Research Reports).IIE and the AIFS Foundation.

Magnan, S. S., & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals, 40*(1), 43-61.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*(3), 381-401.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*(1), 703–704.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G. & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning,* 60, 123-150.

Meuter, R. (2005). Language selection in bilinguals: Mechanisms and processes. In: J.F. Kroll & A.M.B. De Groot (eds), *Handbook of Bilingualism: Psycholinguistics approaches* (pp. 349–370). New York: Oxford University Press.

Michael, E. B., & Gollan, T. H. (2005). *Being and becoming bilingual: Individual differences and consequences for language production*. New York, NY, US: Oxford University Press, New York, NY.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. J. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*(8), 908-930.

Molinaro, N., Vespignani, F., & Job, R. (2008). Deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research, 1228*(1), 161– 176.

Montrul, S. A. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. Amsterdam: John Benjamins.

Montrul, S. A. (2005). Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research, 21*(3), 199-249.

Montrul, S. A., Foote, R., Perpinan, S., Thornhill, D., & Vidal, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning, 58*(3), 503–553.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish-English bilinguals. *Cognitive Brain Research, 22*(2), 205-220.

Morgan-Short, K. (2007). *A neurolinguistic investigation of late-learned second language knowledge: The effects of explicit and implicit conditions.* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition, 17*(1), 56-72.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *Plos One, 7*(3), e32974.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60*(1), 154-193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933-947.

Morgan-Short, K. & Tanner, D. (2014). Event-related potentials (ERPs). In J. Jegerski & B. VanPatten (Eds.), *Research Methods in Second Language Psycholinguistics* (pp. 127-152). New York: Routledge.

Mueller, J. L. (2009). The influence of lexical familiarity on ERP responses during sentence comprehension in language learners. *Second Language Research, 25*(1), 43-76.

Mueller, J. L., Hahne, A., Fujjj, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience, 17*(8), 1229-1244.

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure an ERP study in adult learners. *BMC Neuroscience, 10(1)*, 89.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience, 22*(12), 2886-2898.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes - evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151-165.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of english. *Journal of Cognitive Neuroscience, 24*(5), 1205-1223.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphor noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*(1), 119–131.

Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature, 372*(6503), 260-263.

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning, 51*(1), 157-213.

Ogilvie, S. (2009). Concise Encyclopedia of Languages of the World, in K. Brown (Ed.), *Concise Encyclopedia of Languages of the World* (1st Edition). Amsterdam, Netherlands: Elsevier.

Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience Research, 70*(2), 197-205.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *The Bulletin of the Psychonomic Society, 30*(4), 287-289.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739-773.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*(3), 283-317.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786-803.

Osterhout, L., McLaughlin, J., Kim, A., Greewald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (Eds.). *The on-line study of sentence comprehension: Eye-tracking, ERPs, and beyond*, pp. 271–308. New York: Psychology Press.

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728-2744.

Ranjan, R. (2013). Teaching strategies of grammatical gender in L2 Hindi/Urdu. *Arizona Working Papers in SLA & Teaching*, *20*, 1-19.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, K. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164-171.

Rees, J., & Klapper, J. (2008). Issues in the quantitative longitudinal measurement of second language progress in the study abroad context. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 89-105). London: Routledge.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*(1), 45-99.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*(4), 368-392.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition.* (pp. 631-678). Malden, MA: Blackwell Publishing.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27*(2), 235-268.

Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching* (pp. 57-76). Springer: Oxford.

Robinson, P. (Ed.). (2002). *Individual differences and instructed language learning*. Philadelphia: John Benjamins Publishing Company.

Roehr, K. (2008). Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics*, *19*(1), 67-106.

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030-2048.

Sabourin, L., & Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 175–195). Amsterdam: John Benjamins.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1-29.

Sadowska, I. (2012). *Polish: A Comprehensive Grammar*. Oxford; New York: Routledge.

Sagarra, N. (2000). The longitudinal role of working memory on adult acquisition of L2 grammar. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Sagarra, N. (2008). Working memory and L2 processing of redundant grammatical forms. In Z. Han (Ed.), *Understanding second language process* (pp. 133-147). Clevedon, UK: Multilingual Matters.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 229-248). Oxford: Oxford University Press.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022-2039.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing. II. segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research, 15*(3), 214-227.

Sanz, C. (2000). Bilingual education enhances third language acquisition: Evidence from Catalonia. *Applied Psycholinguistics, 21*(issue), 23-44.

Sanz, C. (2005). *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice.* Washington, DC: Georgetown University Press.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*(1), 35-78.

Schwartz, B.D., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*(1), 40–72.

Segalowitz, N. & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory and Cognition, 33*(4), 644-653.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 173-199.

Segalowitz, N., Freed, B. F., Collentine, J., Lafford, B. A., Lazar, N., & Diaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10(1),* 1-18.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 13*, 275-298.

Skehan, P. (2008). Factors that affect test task difficulty in task-based assessment. *Foreign Language Education in China, 1*(1), 64-78.

Spada, N. & Tomita, Y. (2010). Interactions between type of instructions and type of language feature: a meta-analysis. *Language Learning, 60*(2), 263-308.

Spada, N. (1986). The interaction between type of contact and type of instruction: Some effects on the L2 proficiency of adult learners. *Studies in Second Language Acquisition, 8*(2), 181-199.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavioral Research Methods, Instruments, & Computers, 31*(1), 137-149.

Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer & H. A. Whitaker (Eds.) *Handbook of the neuroscience of language* (pp. 91-104). Oxford, UK: Elsevier.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13-41.

Stewart, J. A. (2010). Using e-journals to assess students' language awareness and social identity during study abroad. *Foreign Language Annals, 43*(1), 138-159.

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30*(1), 79-99.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition, 16,* 367-382..

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition, 16,* 367-382.

Tanner, D., Osterhout, L., & Herschensohn, J. (2009). Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. *Proceedings of the Annual Boston University Conference on Language Development, 33*(2), 528-539.

Tanner, D. & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, 56, 289-301.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition, 27*(2), 173-204.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*(3), 255-272.

Tsimpli, I., & Dimitrakopoulou, M. (2007). The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*(2), 215–242.

Tsimpli, I., & Mastropavlou, M. (2007). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. Liceras, H. Zobl, & H. Goodluck, (Eds.), *The Role of Formal Features in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-154.

Unsworth, N. & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex spans and their relation to higher-order abilities. *Psychological Bulletin*, *133*, 1038-1066.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505.

Van Petten, C. (1995). Words and sentences: Event-related brain potentials measures. *Psychophysiology*, *32*(1), 511-525.

Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, *38*(1), 41-63.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *49*(1), 51-79.

Weber, K., & Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology*, *45*(6), 920-925.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231-256.

White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, *6*(2), 129-141.

Wickens, T. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.

Williams, J. N. (2012). Working memory and SLA in S. Gass, & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 427-447). London: Routledge.

APPENDICES

APPENDIX A

Language Background and Experience Questionnaire

(modified from Language Experience and Proficiency Questionnaire; Marian, Blumenfeld & Kaushanskaya, 2007)

Researcher completes this form with participant.

Background Information

Participant number:
Date of testing:
Date of Birth:
Gender:
Current year in school:

How many years of formal education do you have?
(include kindergarten through college/graduate school)

Please indicate your highest level of education (or the approximate US equivalent to a degree obtained in another country):

- ☐ Less than high school
- ☐ High school
- ☐ Professional training
- ☐ Some college
- ☐ College
- ☐ Some graduate school
- ☐ Masters
- ☐ Ph.D. / M.D. / J.D.
- ☐ Other: __________________

Please check your FATHER'S highest education level (or the approximate US equivalent to a degree obtained in another country) [same options listed as above]

Please check your MOTHER'S highest education level (or the approximate US equivalent to a degree obtained in another country) [same options listed as above]

Date of immigration to the US, if applicable:
If you have ever immigrated to another country, please provide name of country and date of immigration here:

APPENDIX A (continued)

Have you ever had any of the following?

- ☐ Vision problem
- ☐ Hearing impairment
- ☐ Language disability
- ☐ Learning disability
- ☐ None

If you checked any of the boxes on the previous question, please explain (including any corrections).

Have you ever had a concussion or other head trauma? If yes, please describe, include AGE at incident, treatment (if applicable), and duration of symptoms (if applicable). *Enter YES or NO, and describe each incident.

The NIH (National Institutes of Health) requires us to collect information about the ethnic origin of our participants. Please mark the category/categories that apply to you. This information will be reported to them anonymously. [NIH-provided categories listed]

Please list all the languages you know in order of acquisition (Language 1, Language 2, etc)

Please list all the languages you know in order of dominance (Language 1, Language 2, etc)

Have you ever traveled to a Spanish-speaking country prior to this study?

Duration of Travel: If yes, when (what age) and for how long?

Have you ever held residence in a Spanish-speaking country prior to this study?

Duration of Residence: If yes, when (what age) and for how long?

APPENDIX A (continued)

Native Language(s)

Native language refers to any language you were completely fluent in as a child and are still completely fluent in now. It can also refer to a language that you have been exposed to since birth or early childhood.

NATIVE LANGUAGE #1: __________________
Age of Exposure to Native Language #1: _____________
Place of exposure to Native Language #1 (please mark all that apply):

- ☐ Home
- ☐ Other family
- ☐ School
- ☐ Work
- ☐ Other: _________________

Proficiency in Native Language #1
Please rate your proficiency of the following aspects of NATIVE LANGUAGE #1 on a 1-5 scale, where 1 indicates low proficiency and 5 indicates high, native- like proficiency.

Speaking	1	2	3	4	5
Understanding	1	2	3	4	5
Reading	1	2	3	4	5
Writing	1	2	3	4	5

For the following 3 items, please list the number of years and months you spent in each language environment:
A COUNTRY where Native Language #1 is spoken (number of years and months):
A family where Native Language #1 is spoken (number of years and months):
A SCHOOL and/ or WORK environment where Native Language #1 is spoken (number of years and months):

Additional comments about Native Language #1:

Note. “Native Language” section repeated for second and third additional languages, if applicable.

APPENDIX A (continued)

Other Languages

What other languages do you know? Please include any language(s) other than your native language(s) that you have studied or learned either formally or informally. For each language, please include the information requested.

OTHER LANGUAGE #1: ___________________
Age of Exposure to Other Language #1: ______________
Place of exposure to Other Language #1 (please mark all that apply):

- ☐ Home
- ☐ Other family
- ☐ School
- ☐ Work
- ☐ Other: __________________

Proficiency in Other Language #1
Please rate your proficiency of the following aspects of OTHER LANGUAGE #1 on a 1-5 scale, where 1 indicates low proficiency and 5 indicates high, native- like proficiency.

Speaking	1	2	3	4	5
Understanding	1	2	3	4	5
Reading	1	2	3	4	5
Writing	1	2	3	4	5

For the following 3 items, please list the number of years and months you spent in each language environment:

A COUNTRY where Other Language #1 is spoken (number of years and months):
A FAMILY where Other Language #1 is spoken (number of years and months):
A SCHOOL and/ or WORK environment where Other Language #1 is spoken (number of years and months):

Additional comments about Other Language #1:

Note. “Other Language” section repeated for second and third additional languages, if applicable.

APPENDIX B

Working memory tasks – example trials.

Table XXI

Example Trials from Automated Operation Span Task

Math Problem	Decision	Letter to Recall
(2*5)-6=	4 True / False	F
(5/1)+8=	13 True / False	L
(3/1)+2=	5 True / False	T
(3*5)-7=	9 True / False	R
(3/3)-1=	2 True / False	T
(8/4)+6=	8 True / False	P
(6*2)+2=	16 True / False	H
(1*7)-4=	7 True / False	F
(4*2)-2=	2 True / False	Q
(2/2)-1=	2 True / False	K
(2*3)+5=	13 True / False	H
(4*3)-8=	4 True / False	Q
(4*3)+7=	11 True / False	N
(8/8)-1=	0 True / False	J
(1*9)-6=	3 True / False	Q

Note. Task provided by R. Engle Laboratory; Unsworth, Heitz, Schrock & Engle, 2005

APPENDIX B (continued)

Table XXII

Example Trials from Automated Reading Span Task

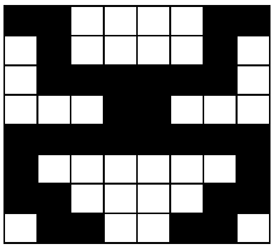
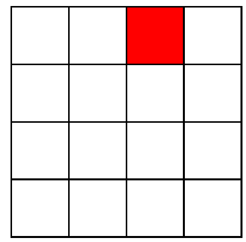
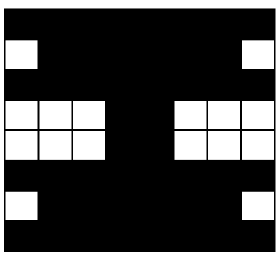
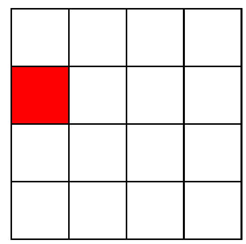
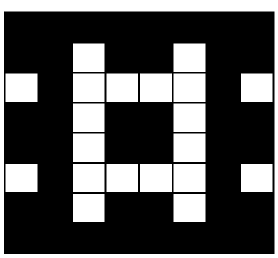
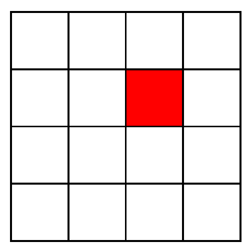
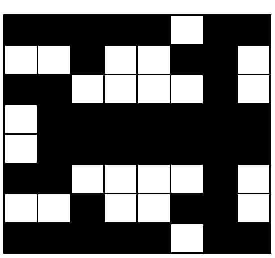
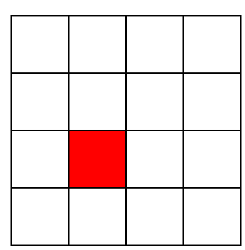
Sentence	Decision	Letter
Unaware of the hunter, the deer wandered into his shotgun range.	This sentence makes sense True / False	R
Raising children requires a lot of dust and the ability to be firm.	This sentence makes sense True / False	Q
My mother and father have always wanted to love near the cup.	This sentence makes sense True / False	Q
Stacey stopped dating the light when she found out he had a wife.	This sentence makes sense True / False	R
When John and Amy moved to Canada their wish had a huge garage sale.	This sentence makes sense True / False	F
He wrecked his car because he was going too fast in the rain.	This sentence makes sense True / False	N
Before Katie left for the city, she took a self-defense class at the gym.	This sentence makes sense True / False	K
John wants to be a football player when he gets older.	This sentence makes sense True / False	F
The seventh graders had to build a volcano for their science class.	This sentence makes sense True / False	N
The couple decided that they wanted to have a picnic in the park.	This sentence makes sense True / False	P

Note. Task provided by R. Engle Laboratory; Unsworth, Heitz, Schrock & Engle, 2005

APPENDIX B (continued)

Table XXIII

Example Trials from Automated Symmetry Span Task

Symmetry Image	Decision	Matrix
	Was image symmetrical? Yes / No	
	Was image symmetrical? Yes / No	
	Was image symmetrical? Yes / No	
	Was image symmetrical? Yes / No	

Note. Task provided by R. Engle Laboratory; Unsworth, Heitz, Schrock & Engle, 2005

APPENDIX C

Information Gap Activity

(adapted from Leeman, 2003)

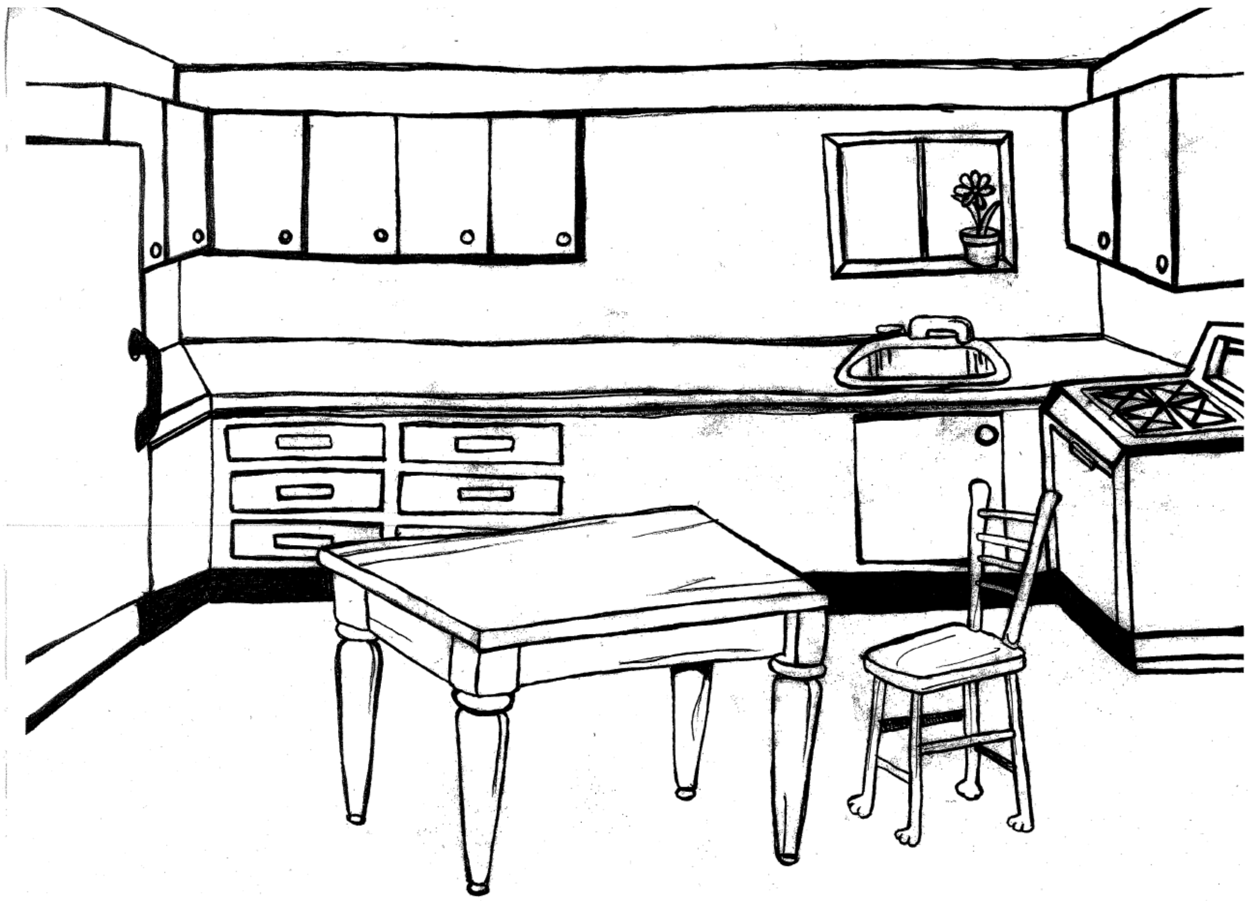


Figure 35. Blank kitchen used by researcher during Info Gap Activity.

APPENDIX C (continued)

Figure 36. Sample participant kitchen used during Info Gap Activity.

APPENDIX C (continued)

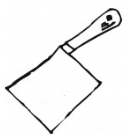
Vocabulary Sheet provided to participants during Info Gap Activity.

taza

fresa

gato

botella



cuchillo

perro

manzana

helado

piña

cebolla

huevo

sandía

pera

pollo

vaso

zapato

amarillo—yellow
morado—purple
negro—black
rojo—red
rosado—pink

pequeño—small
mediano—medium
grande—large

encimera—counter-top
horno—oven
lavamanos—sink
mesa—table
piso—floor
silla—chair
ventana—window

estar—to be located
ser—to be

abajo de—beneath
al lado de—at the side of
encima de—on top of
en—in/on
izquierda – left
derecha – right

APPENDIX D

Article Agreement Stimuli – Grammaticality Judgment Task

César dice que **la** mesa que está muy cerca es grande y le gusta mucho.

César dice que **el** mesa que está muy cerca es grande y le gusta mucho.

Andrés piensa que **el** escritorio es muy útil.

Andrés piensa que **la** escritorio es muy útil.

Luz dice que **la** casa donde vive está en la ciudad.

Luz dice que **el** casa donde vive está en la ciudad.

Lupe sabe que **el** plato que tiene que limpiar es muy grande.

Lupe sabe que **la** plato que tiene que limpiar es muy grande.

Felipe dice que **el** abrigo que lleva cuando hace frío es de cuero.

Felipe dice que **la** abrigo que lleva cuando hace frío es de cuero.

Según Maribel **la** blusa que está allí arriba es enorme.

Según Maribel **el** blusa que está allí arriba es enorme.

Tomás piensa que **la** corbata de rayas combina bien con su traje.

Tomás piensa que **el** corbata de rayas combina bien con su traje.

Según Irene **el** vestido que siempre lleva cuando viaja es de moda.

Según Irene **la** vestido que siempre lleva cuando viaja es de moda.

Maribel piensa que **la** falda que lleva cuando sale le queda bien.

Maribel piensa que **el** falda que lleva cuando sale le queda bien.

Inés dice que **el** escritorio es grande pero le gusta mucho.

Inés dice que **la** escritorio es grande pero le gusta mucho.

Víctor piensa que **el** contrato que tiene que explicar es comprensible.

Víctor piensa que **la** contrato que tiene que explicar es comprensible.

Juan dice que **la** mochila es muy útil.

Juan dice que **el** mochila es muy útil.

Según Carlos **el** regalo que está por allí abajo no es muy grande.

Según Carlos **la** regalo que está por allí abajo no es muy grande.

Noemí sabe que **la** bolsa cuesta mucho y es grande.

APPENDIX D (continued)

Noemí sabe que **el** bolsa cuesta mucho y es grande.

Irene se pone **la** blusa de lunares cuando visita a su abuela.

Irene se pone **el** blusa de lunares cuando visita a su abuela.

Pilar piensa que **el** vestido que quiere comprar no cuesta demasiado.

Pilar piensa que **la** vestido que quiere comprar no cuesta demasiado.

Según Enrique **el** barrio donde va para divertirse tiene casas fabulosas.

Según Enrique **la** barrio donde va para divertirse tiene casas fabulosas.

Jaime limpia **la** casa cada día antes de trabajar.

Jaime limpia **el** casa cada día antes de trabajar.

Según Carol **el** teléfono que usa para hablar con sus amigos funciona bien.

Según Carol **la** teléfono que usa para hablar con sus amigos funciona bien.

Carlos trabaja en **la** mesa azul todos los días.

Carlos trabaja en **el** mesa azul todos los días.

Raúl dice que **la** mochila que está allí abajo le sirve bien.

Raúl dice que **el** mochila que está allí abajo le sirve bien.

Jorge limpia **el** plato después de comer.

Jorge limpia **la** plato después de comer.

Beatriz piensa que **la** bolsa que lleva cuando llueve es impermeable.

Beatriz piensa que **el** bolsa que lleva cuando llueve es impermeable.

Según Montse **el** regalo para Juan es espectacular.

Según Montse **la** regalo para Juan es espectacular.

Según José **la** bufanda que va a llevar hoy es de cuadros.

Según José **el** bufanda que va a llevar hoy es de cuadros.

Javier cree que el contrato que usan es fácil de leer.

Javier cree que **la** contrato que usan es fácil de leer.

Maite dice que el abrigo que lleva cuando hace frío es elegante.

Maite dice que **la** abrigo que lleva cuando hace frío es elegante.

Según Montse **la** falda que lleva cuando hace frío es de lana.

Según Montse **el** falda que lleva cuando hace frío es de lana.

APPENDIX D (continued)

Pepe dice que **el** libro que lee es interesante.

Pepe dice que **la** libro que lee es interesante.

Víctor dice que **la** tarea es difícil y no le gusta.

Víctor dice que **el** tarea es difícil y no le gusta.

Según Lourdes **el** periódico contiene mucha información importante.

Según Lourdes **la** periódico contiene mucha información importante.

Según Inés **la** novela que lee cuando quiere descansar es muy grande.

Según Inés **el** novela que lee cuando quiere descansar es muy grande.

Dolores lee **el** periódico todos los días para informarse.

Dolores lee **la** periódico todos los días para informarse.

Inés dice que **la** novela es fascinante y le gusta mucho.

Inés dice que **el** novela es fascinante y le gusta mucho.

Esteban dice que **el** libro que lee antes de dormir es agradable.

Esteban dice que **la** libro que lee antes de dormir es agradable.

Rafael dice que no le molesta hacer **la** tarea antes de salir.

Rafael dice que no le molesta hacer **el** tarea antes de salir.

Ramón dice que no le gusta **la** corbata de lunares que siempre lleva.

Ramón dice que no le gusta **el** corbata de lunares que siempre lleva.

Miguel solamente usa **el** teléfono y no el correo electrónico.

Miguel solamente usa **la** teléfono y no el correo electrónico.

Según Noemí **el** barrio de su familia tiene muchos parques.

Según Noemí **la** barrio de su familia tiene muchos parques.

Trini dice que **la** bufanda que va a comprar le gusta mucho.

Trini dice que **el** bufanda que va a comprar le gusta mucho.

Maite usa **el** armario de su habitación para guardar su ropa.

Maite usa **la** armario de su habitación para guardar su ropa.

Según Raúl es importante tener **el** armario accesible para encontrar todo fácilmente.

Según Raúl es importante tener **la** armario accesible para encontrar todo fácilmente.

Raúl dice que **la** maleta que está allí abajo es bastante grande.

APPENDIX D (continued)

Raúl dice que **el** maleta que está allí abajo es bastante grande.

Según Pilar **la** maleta que usa para viajes cortos funciona bien.

Según Pilar **el** maleta que usa para viajes cortos funciona bien.

Según Felipe **el** carro que conduce hace mucho ruido.

Según Felipe **la** carro que conduce hace mucho ruido.

Irene usa **el** carro para ir a trabajar cuando llueve.

Irene usa **la** carro para ir a trabajar cuando llueve.

A Javier le gusta mucho **la** bicicleta que usa para hacer turismo.

A Javier le gusta mucho **el** bicicleta que usa para hacer turismo.

Lourdes usa **la** bicicleta de su hermana cuando hace buen tiempo.

Lourdes usa **el** bicicleta de su hermana cuando hace buen tiempo.

A Ramón no le gusta almorzar en **la** oficina todos los días.

A Ramón no le gusta almorzar en **el** oficina todos los días.

Dolores trabaja mucho en **la** oficina por la tarde.

Dolores trabaja mucho en **el** oficina por la tarde.

Según Víctor es importante seguir **el** horario para evitar sorpresas.

Según Víctor es importante seguir **la** horario para evitar sorpresas.

Lourdes piensa que **el** horario que sigue le gusta mucho.

Lourdes piensa que **la** horario que sigue le gusta mucho.

A Juan le fascina **el** sombrero que lleva su abuelo.

A Juan le fascina **la** sombrero que lleva su abuelo.

Flor siempre lleva **el** sombrero de lunares cuando hace sol.

Flor siempre lleva **la** sombrero de lunares cuando hace sol.

Enrique no sabe usar **la** cafetera de sus padres.

Enrique no sabe usar **el** cafetera de sus padres.

Trini dice que **la** cafetera que usa cada mañana no funciona muy bien.

Trini dice que **el** cafetera que usa cada mañana no funciona muy bien.

Miguel quiere pintar **el** baño cuando tiene vacaciones.

APPENDIX D (continued)

Miguel quiere pintar **la** baño cuando tiene vacaciones.

Carmen siempre evita limpiar **el** baño porque no le gusta para nada.

Carmen siempre evita limpiar **la** baño porque no le gusta para nada.

Fidel nunca usa **la** calculadora para resolver problemas.

Fidel nunca usa **el** calculadora para resolver problemas.

A Irene le gusta mucho **la** calculadora que tiene en la oficina.

A Irene le gusta mucho **el** calculadora que tiene en la oficina.

Adjective Agreement Stimuli – Grammaticality Judgment Task

Felipe trabaja mucho en su mesa **alto** por la tarde.

Felipe trabaja mucho en su mesa **alta** por la tarde.

Juan usa su escritorio **alta** cuando estudia.

Juan usa su escritorio **alto** cuando estudia.

A Irene le gusta mucho su casa **organizado** porque sabe dónde está todo.

A Irene le gusta mucho su casa **organizada** porque sabe dónde está todo.

Irene limpia su plato **blanca** después de la cena.

Irene limpia su plato **blanco** después de la cena.

César lleva su abrigo **oscura** cuando hace frío.

César lleva su abrigo **oscuro** cuando hace frío.

Lourdes lleva su blusa **oscuro** con su falda gris.

Lourdes lleva su blusa **oscura** con su falda gris.

Luis se pone su corbata **oscuro** cuando presenta un trabajo.

Luis se pone su corbata **oscura** cuando presenta un trabajo.

Carol se pone su vestido **oscura** cuando tiene que viajar.

Carol se pone su vestido **oscuro** cuando tiene que viajar.

Luz lleva su falda **sencillo** cuando sale con sus amigos.

Luz lleva su falda **sencilla** cuando sale con sus amigos.

Pilar trabaja en su escritorio **llena** cuando trabaja desde casa.

APPENDIX D (continued)

Pilar trabaja en su escritorio **lleno** cuando trabaja desde casa.

Pepe prefiere mantener su contrato **sencilla** para evitar problemas.

Pepe prefiere mantener su contrato **sencillo** para evitar problemas.

Andrés prefiere usar su mochila **deportivo** cuando tiene un partido de fútbol.

Andrés prefiere usar su mochila **deportiva** cuando tiene un partido de fútbol.

Enrique tiene su regalo **bonita** en el coche.

Enrique tiene su regalo **bonito** en el coche.

Inés lleva su bolsa **bonito** a clase todos los días.

Inés lleva su bolsa **bonita** a clase todos los días.

Lupe se pone su blusa **bonito** cuando tiene una cita.

Lupe se pone su blusa **bonita** cuando tiene una cita.

Carmen se pone su vestido **bonita** cuando va a un restaurante elegante.

Carmen se pone su vestido **bonito** cuando va a un restaurante elegante.

Carlos visita su barrio **antigua** los fines de semana.

Carlos visita su barrio **antiguo** los fines de semana.

Soledad tiene que limpiar su casa **antiguo** antes de poder salir.

Soledad tiene que limpiar su casa **antigua** antes de poder salir.

Flor no puede usar su teléfono **rota** para hablar con sus amigos.

Flor no puede usar su teléfono **roto** para hablar con sus amigos.

Jaime tiene que usar su mesa **roto** para estudiar por la mañana.

Jaime tiene que usar su mesa **rota** para estudiar por la mañana.

Jorge prefiere usar su mochila **pequeño** cuando viaja.

Jorge prefiere usar su mochila **pequeña** cuando viaja.

Javier no quiere usar su plato **rota** para servir la comida.

Javier no quiere usar su plato **roto** para servir la comida.

Maite prefiere llevar su bolsa **pequeño** cuando va al cine.

Maite prefiere llevar su bolsa **pequeña** cuando va al cine.

Éster lleva su regalo **pequeña** a la fiesta de cumpleaños.

Éster lleva su regalo **pequeño** a la fiesta de cumpleaños.

APPENDIX D (continued)

Raúl lleva su bufanda **largo** a la fiesta de cumpleaños.

Raúl lleva su bufanda **larga** a la fiesta de cumpleaños.

José lee su contrato **larga** en su oficina.

José lee su contrato **largo** en su oficina.

Isabel prefiere ponerse su abrigo **larga** cuando hace frío.

Isabel prefiere ponerse su abrigo **largo** cuando hace frío.

Beatriz se pone su falda **largo** cuando hace frío.

Beatriz se pone su falda **larga** cuando hace frío.

Tomás escribe en su libro **aburrida** por una hora cada día.

Tomás escribe en su libro **aburrido** por una hora cada día.

Héctor no hace su tarea **aburrido** los fines de semana.

Héctor no hace su tarea **aburrida** los fines de semana.

Maribel lee su periódico **aburrida** cuando está en el tren.

Maribel lee su periódico **aburrido** cuando está en el tren.

Noemí lee su novela **aburrido** por la noche.

Noemí lee su novela **aburrida** por la noche.

Éster lee su periódico **divertida** todos los días.

Éster lee su periódico **divertido** todos los días.

Iris lee su novela **divertido** antes de clase.

Iris lee su novela **divertida** antes de clase.

Miguel lee su libro **divertida** antes de ir al gimnasio.

Miguel lee su libro **divertido** antes de ir al gimnasio.

Ramón hace su tarea **divertido** antes de limpiar su cuarto.

Ramón hace su tarea **divertida** antes de limpiar su cuarto.

Esteban busca su corbata **nuevo** por todas partes.

Esteban busca su corbata **nueva** por todas partes.

Rafael usa su teléfono **nueva** cuando debe hacer la tarea.

Rafael usa su teléfono **nuevo** cuando debe hacer la tarea.

A Luz le gusta mucho su barrio **nueva** porque tiene buenos vecinos.

APPENDIX D (continued)

A Luz le gusta mucho su barrio **nuevo** porque tiene buenos vecinos.

Trini lleva su bufanda **nuevo** a clase cuando tiene frío.

Trini lleva su bufanda **nueva** a clase cuando tiene frío.

Carmen guarda todo en su armario **alta** en su habitación.

Carmen guarda todo en su armario **alto** en su habitación.

Javier mantiene su armario **organizada** para evitar problemas.

Javier mantiene su armario **organizado** para evitar problemas.

Jorge prefiere usar su maleta **blanco** cuando viaja al extranjero.

Jorge prefiere usar su maleta **blanca** cuando viaja al extranjero.

Maribel tiene su maleta **lleno** y quiere ir al aeropuerto.

Maribel tiene su maleta **llena** y quiere ir al aeropuerto.

Miguel dice que su carro **antigua** todavía funciona muy bien.

Miguel dice que su carro **antiguo** todavía funciona muy bien.

Flor conduce su carro **deportiva** durante los fines de semana.

Flor conduce su carro **deportivo** durante los fines de semana.

Según Andrés su bicicleta **deportivo** le ayuda ganar carreras.

Según Andrés su bicicleta **deportiva** le ayuda ganar carreras.

Soledad tiene dificultad en subirse a su bicicleta **alto** cuando tiene que ir a trabajar.

Soledad tiene dificultad en subirse a su bicicleta **alta** cuando tiene que ir a trabajar.

Luis prefiere tener su oficina **organizado** porque puede concentrarse mejor.

Luis prefiere tener su oficina **organizada** porque puede concentrarse mejor.

A Montse le gusta mucho su oficina **blanco** porque los colores le distraen.

A Montse le gusta mucho su oficina **blanca** porque los colores le distraen.

Esteban tiene su horario **llena** para poder alcanzar mucho cada día.

Esteban tiene su horario **lleno** para poder alcanzar mucho cada día.

Lupe prefiere tener su horario **organizada** porque trabaja mucho.

Lupe prefiere tener su horario **organizado** porque trabaja mucho.

Tomás lleva su sombrero **deportiva** cuando juega fútbol.

Tomás lleva su sombrero **deportivo** cuando juega fútbol.

APPENDIX D (continued)

Maribel quiere llevar su sombrero **sencilla** para trabajar afuera.

Maribel quiere llevar su sombrero **sencillo** para trabajar afuera.

César trae su cafetera **lleno** a la mesa para servir el café.

César trae su cafetera **llena** a la mesa para servir el café.

Isabel prefiere usar su cafetera **sencillo** en casa cuando prepara el desayuno.

Isabel prefiere usar su cafetera **sencilla** en casa cuando prepara el desayuno.

Javier limpia su baño **pequeña** cada sábado porque vive con dos amigos.

Javier limpia su baño **pequeño** cada sábado porque vive con dos amigos.

Montse dice que su baño **blanca** es muy grande y tiene buena luz.

Montse dice que su baño **blanco** es muy grande y tiene buena luz.

Raúl no puede usar su calculadora **roto** para el examen de matemáticas.

Raúl no puede usar su calculadora **rota** para el examen de matemáticas.

Carol dice que su calculadora **antiguo** le sirve bien.

Carol dice que su calculadora **antigua** le sirve bien.

APPENDIX E

Motivation Questionnaire

(adapted from Gardner, 2004)

The purpose of this questionnaire is to determine your feelings about a number of things. We want you to rate each of the following items in terms of how you feel about it. Each item is followed by a scale that has a label on the left and another on the right, and the numbers 1 to 7 between the two ends. For each item, please choose the number that best describes you.

1. My motivation to learn Spanish in order to communicate with Spanish speaking people is:

   Weak 1 2 3 4 5 6 7 Strong

2. My attitude toward Spanish-speaking people is:

   Unfavorable 1 2 3 4 5 6 7 Favorable

3. My interest in foreign languages is:

   Very low 1 2 3 4 5 6 7 Very high

4. My desire to learn Spanish is:

   Weak 1 2 3 4 5 6 7 Strong

5. My attitude toward learning Spanish is:

   Unfavorable 1 2 3 4 5 6 7 Favorable

6. My attitude toward my Spanish teacher is:

   Unfavorable 1 2 3 4 5 6 7 Favorable

7. My motivation to learn Spanish for practical purposes (e.g., to get a good job) is:

   Weak 1 2 3 4 5 6 7 Strong

8. I worry about speaking Spanish outside of class:

   Very little 1 2 3 4 5 6 7 Very much

9. My attitude toward my Spanish course is:

   Unfavorable 1 2 3 4 5 6 7 Favorable

10. I worry about speaking in my Spanish class:

    Very little 1 2 3 4 5 6 7 Very much

11. My motivation to learn Spanish is:

    Very low 1 2 3 4 5 6 7 Very high

12. My parents encourage me to learn Spanish:

    Very little 1 2 3 4 5 6 7 Very much

APPENDIX F

Weekly Language Contact Questionnaire

(adapted from Language Contact Profile; Freed, Segalowitz, Dewey & Halter, 2004)

Study Abroad Version

Thank you for taking the time to fill out this survey. The information you provide will help us to better understand your learning experience. Your honest and detailed responses are greatly appreciated. ALL of the questions refer to your use of SPANISH during the past week unless otherwise noted. Please contact Mandy with any questions: cogsla.uic@gmail.com. Thank you!

Please enter your name and the city and country where you are studying.

Please enter the dates for which your survey responses apply. For example, if you are answering the questions on July 7 for the previous week, you would enter July 1 - July 7, 2012.

2. How much time did you spend speaking, in Spanish, outside of class with native or fluent Spanish speakers this week?
Please select (1) the number of DAYS this week, as well as (2) the typical number of HOURS per day.

- ☐ 0 days
- ☐ 1 day
- ☐ 2 days
- ☐ 3 days
- ☐ 4 days
- ☐ 5 days
- ☐ 6 days
- ☐ 7 days

- ☐ 0 hours
- ☐ 0-1 hour/day
- ☐ 1-2 hours/day
- ☐ 2-3 hours/day
- ☐ 3-4 hours/day
- ☐ 4-5 hours/day
- ☐ 5+ hours/day

[Note. Each item included the prompt *Please select (1) the number of DAYS this week, as well as (2) the typical number of HOURS per day* and also included the seven day/hour range options.]

3. This week, outside of class, I tried to speak Spanish to MY INSTRUCTOR...
4. This week, outside of class, I tried to speak Spanish to FRIENDS WHO ARE NATIVE/FLUENT SPEAKERS OF SPANISH...
5. This week, outside of class, I tried to speak Spanish to CLASSMATES...
6. This week, outside of class, I tried to speak Spanish to STRANGERS WHOM I THOUGHT COULD SPEAKER SPANISH...
7. This week, outside of class, I tried to speak Spanish to A HOST FAMILY / SPANISH ROOMMATE / SPANISH SPEAKERS IN MY DORMITORY...

APPENDIX F (continued)

8. This week, outside of class, I tried to speak Spanish to SERVICE PERSONNEL...
9. This week, outside of class, I tried to speak Spanish to OTHER PEOPLE (please specify below)...
10. This week, outside of class, I used SPANISH to clarify classroom-related work...
11. This week, outside of class, I used SPANISH to obtain directions or information (e.g., "Where is the post office?" "How much are stamps?" etc)...
12. This week, outside of class, I used SPANISH for superficial or brief exchanges (e.g., greetings, ordering in a restaurant, etc)...
13. This week, outside of class, I used SPANISH for extended conversations with host family, friends, etc...
14. This week, how often did you deliberately try to use things you were taught in the classroom (grammar, vocabulary, expressions) with native or fluent speakers of Spanish outside of the classroom?
15. This week, how often did you take things you learned outside of the classroom (grammar, vocabulary, expressions) back to class for questions or discussion?
16. This week, how much time did you spend speaking a language other than English or Spanish to speakers of that language (e.g., Chinese to a Chinese-speaking friend)? Include in class as well as outside of class. Please list language spoken in the "Other" section below.
17. This week, how much time did you spend speaking SPANISH to native or fluent speakers of Spanish? Include in class as well as outside of class.
18. This week, how much time did you spend speaking ENGLISH to native or fluent speakers of Spanish? Include in class as well as outside of class.
19. This week, how much time did you spend speaking SPANISH to nonnative speakers of Spanish (i.e., classmates)? Include in class as well as outside of class.
20. This week, how much time did you spend speaking ENGLISH to nonnative speakers of Spanish (i.e., classmates)? Include in class as well as outside of class.
21. This week, how much time did you spend READING in Spanish (overall) outside of class?
22. This week, how much time did you spend reading Spanish newspapers outside of class?
23. This week, how much time did you spend reading novels in Spanish outside of class?
24. This week, how much time did you spend reading Spanish language magazines outside of class?
25. This week, how much time did you spend reading schedules, announcements, menus, and the like in Spanish outside of class?
26. This week, how much time did you spend reading email or Internet web pages in Spanish outside of class?
27. This week, how much time did you spend LISTENING to Spanish (overall) outside of class?
28. This week, how much time did you spend listening to Spanish television and radio outside of class?
29. This week, how much time did you spend listening to Spanish movies or videos outside of class?
30. This week, how much time did you spend listening to Spanish songs outside of class?
31. This week, how much time did you spend trying to catch other people's conversations in Spanish outside of class?

APPENDIX F (continued)

32. This week, how much time did you spend WRITING in Spanish (overall) outside of class?
33. This week, how much time did you spend writing homework assignments in Spanish outside of class?
34. This week, how much time did you spend writing personal notes or letters in Spanish outside of class?
35. This week, how much time did you spend SPEAKING ENGLISH outside of class?
36. This week, how much time did you spend reading newspapers, magazines, or novels, or watching movies, television, or videos in ENGLISH outside of class?
37. This week, how much time did you spend reading email or Internet web pages in ENGLISH outside of class?
38. This week, how much time did you spend writing email in ENGLISH outside of class?
39. This week, how much time did you spend writing personal notes and letters in ENGLISH outside of class?

Thank you for taking the time to fill out this week's survey!

APPENDIX G

IRB Approval – Original Submission

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice

Initial Review (Response to Modifications)

July 7, 2008

Kara Morgan Short, Ph.D.
Spanish, French, Italian and Portuguese
601 S. Morgan St.
1718 UH, M/C 315
Chicago, IL 60607-7117
Phone: (312) 996-5215 / Fax: (312) 413-1044

RE: Protocol # 2008-0496
"The Cognition of Language Acquisition and Processing"

Dear Dr. Short:

Your Initial Review Application (Response to Modifications) was reviewed and approved by the Expedited review process on June 23, 2008. You may now begin your research.

Please note the following information about your approved research protocol:

Protocol Approval Period: June 23, 2008 - June 22, 2009
Approved Subject Enrollment #: 500
Additional Determinations for Research Involving Minors: These determinations have not been made for this study since it has not been approved for enrollment of minors.
Performance Sites: UIC
Sponsor: Office of Social Science Research (OSSR)
PAF#: Not applicable
Grant/Contract No: Not applicable
Grant/Contract Title: Understanding heritage language processing: A unique opportunity for research and community

APPENDIX G (continued)

Research Protocol(s):

a) The Cognition of Language Acquisition and Processing; Version 1; 06/03/2008

Recruitment Material(s):

a) The Cognition of Language Handout; Version 2; 06/12/2008
b) The Cognition of Language Message; Version 2; 06/18/2008
c) The Cognition of Language Research Form; Version 2; 06/18/2008
d) The Cognition of Language Research Flyer; Version 2; 06/18/2008

Informed Consent(s):

a) The Cognition of Language Research Consent Form; Version 2; 06/18/2008

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific categories:

(4) Collection of data through noninvasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving X-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. (Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for expedited review, including studies of cleared medical devices for new indications.) Examples: (a) physical sensors that are applied either to the surface of the body or at a distance and do not involve input of significant amounts of energy into the subject or an invasion of the subject's privacy; (b) weighing or testing sensory acuity; (c) magnetic resonance imaging; (d) electrocardiography, electroencephalography, thermography, detection of naturally occurring radioactivity, electroretinography, ultrasound, diagnostic infrared imaging, doppler blood flow, and echocardiography; (e) moderate exercise, muscular strength testing, body composition assessment, and flexibility testing where appropriate given the age, weight, and health of the individual;

(6) Collection of data from voice, video, digital, or image recordings made for research purposes; and

(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
06/04/2008	Initial Review	Expedited	06/08/2008	Modifications Required
06/18/2008	Response to Modifications	Expedited	06/23/2008	Approved

Please remember to:

→ Use your **research protocol number** (2008-0496) on any documents or correspondence with the IRB concerning your research protocol.

APPENDIX G (continued)

→ Review and comply with all requirements on the enclosure, "**UIC Investigator Responsibilities, Protection of Human Research Subjects**"

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 355-2908. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne
Assistant Director, IRB # 2
Office for the Protection of Research Subjects

Enclosure(s):

1. **UIC Investigator Responsibilities, Protection of Human Research Subjects**
2. **Informed Consent Document(s):**
   a) The Cognition of Language Research Consent Form; Version 2; 06/18/2008
3. **Recruiting Material(s):**
   a) The Cognition of Language Handout; Version 2; 06/12/2008
   b) The Cognition of Language Message; Version 2; 06/18/2008
   c) The Cognition of Language Research Form; Version 2; 06/18/2008
   d) The Cognition of Language Research Flyer; Version 2; 06/18/2008

cc: Dianna Niebylski, Spanish, French, Italian and Portuguese, M/C 315
OVCR Administration, M/C 672

APPENDIX H

IRB Approval – Amendment Approval

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice
Amendment to Research Protocol and/or Consent Document – Expedited Review
UIC Amendment # 8

November 16, 2012

Kara Morgan Short, Ph.D.
Spanish, French, Italian and Portuguese
601 S. Morgan St.
1718 UH, M/C 315
Chicago, IL 60607-7117
Phone: (312) 996-5743 / Fax: (312) 413-1044

RE: **Protocol # 2008-0496**
"The Cognition of Language Acquisition and Processing"

Dear Dr. Short:

Members of Institutional Review Board (IRB) #2 have reviewed this amendment to your research and/or consent form under expedited procedures for minor changes to previously approved research allowed by Federal regulations [45 CFR 46.110(b)(2)]. The amendment to your research was determined to be acceptable and may now be implemented.

Please note the following information about your approved amendment:

Amendment Approval Date: November 13, 2012
Amendment:

Summary: UIC Amendment #8 dated October 31, 2012 (received November 09, 2012) is an investigator-initiated amendment regarding the following: (1) Update the Protocol to include a sample of subject who will not be participating in the fMRI portion of the study (revised Protocol, v4, 11/14/2012); (2) Submit a new consent form for participants that do not participate in fMRI data collection (Consent Form: Behavioral, ERP, and Eye-Tracking, v1, 10/31/2012); (3) Submit new recruitment documents (The Cognition of Language Message, v1, 10/31/2012; The Cognition of Language Flyer, v1, 10/31/2012; The Cognition of

APPENDIX H (continued)

Language Handout, v1, 10/31/2012; The Cognition of Language Research Interest Form, v1, 10/31/2012); (4) Request a waiver of documentation for screening purposes only (revised initial review, pages 1, 24-26).

Approved Subject Enrollment #: 500
Performance Sites: UIC, Northwestern University
Sponsor: Office of Social Science Research (OSSR)
PAF#: Not available
Grant/Contract No: Not available
Grant/Contract Title: Understanding heritage language processing: A unique opportunity for research and community

Research Protocol(s):

a) The Cognition of Language Acquisition and Processing; Version 4; 11/14/2012

Recruiting Material(s):

a) Message: L2 Study Abroad/At Home; Version 1; 10/31/2012
b) Flyer: L2 Study Abroad/At Home; Version 1; 10/31/2012
c) Handout: L2 Study Abroad/At Home; Version 1; 10/31/2012
d) Interest Form: L2 Study Abroad/At Home; Version 1; 10/31/2012

Informed Consent(s):

a) Consent Form: Behavioral, ERP, and Eye-Tracking; Version 1; 10/31/2012

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
11/09/2012	Amendment	Expedited	11/13/2012	Approved

Please be sure to:

→ **Use only the IRB-approved and stamped consent document(s) and/or HIPAA Authorization form(s) enclosed with this letter when enrolling subjects.**

→ Use your research protocol number (2008-0496) on any documents or correspondence with the IRB concerning your research protocol.

→ Review and comply with all requirements on the enclosure,
"**UIC Investigator Responsibilities, Protection of Human Research Subjects**"

Please note that the UIC IRB #2 has the right to ask further questions, seek additional information, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

APPENDIX H (continued)

We wish you the best as you conduct your research. If you have any questions or need further help, please contact the OPRS at (312) 996-1711 or me at (312) 355-0816. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Alison Santiago, MSW, MJ
IRB Coordinator, IRB # 2
Office for the Protection of Research Subjects

Enclosure(s):

1. **UIC Investigator Responsibilities, Protection of Human Research Subjects**
2. **Data Security Enclosure**
3. **Informed Consent Document(s):**
   a) Consent Form: Behavioral, ERP, and Eye-Tracking; Version 1; 10/31/2012

4. **Recruiting Material(s):**
   a) Message: L2 Study Abroad/At Home; Version 1; 10/31/2012
   b) Flyer: L2 Study Abroad/At Home; Version 1; 10/31/2012
   c) Handout: L2 Study Abroad/At Home; Version 1; 10/31/2012
   d) Interest Form: L2 Study Abroad/At Home; Version 1; 10/31/2012

cc: Rosilie Hernandez-Pecoraro, Spanish, French, Italian and Portuguese, M/C 315

VITA

EDUCATION

PhD	Hispanic Linguistics with Doctoral Concentration in Neuroscience University of Illinois at Chicago	August 2014
MA	Hispanic Linguistics University of Illinois at Chicago	May 2009
BA	Spanish with Certificate in TESOL University of Wisconsin-Madison	May 2007

PUBLICATIONS

Morgan-Short, K., Faretta-Stutenberg, M., & Bartlett-Hsu, L. (forthcoming). Contributions of event-related potential research to issues in explicit and implicit second language acquisition. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages.* Amsterdam: John Benjamins.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong. P.C.M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*(1), 56-72.

Faretta-Stutenberg, M. & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams (2005). In G. Granena et al. (Eds.), S*elected Proceedings of the 2010 Second Language Research Forum: Reconsidering SLA Research, Dimensions, and Directions* (pp. 18-28). Somerville, MA: Cascadilla Proceedings Project.

AWARDS AND FELLOWSHIPS

2014 Language Learning Dissertation Grant

2013 UIC Provost's Award for Graduate Research

2013 Audrey Lumsden-Kouvel Dissertation Fellowship

2013 UIC Undergraduate Mentoring Award for Graduate Students

2013 NIH-funded Society for the Neurobiology of Language Conference Travel Award

2013 UIC Hispanic and Italian Studies Excellence in Teaching Award (also 2009)

VITA (continued)

PRESENTATIONS AND POSTERS: PEER-REVIEWED
(Note. * denotes undergraduate student)

Morgan-Short, K., Bartlett-Hsu, L., & Faretta-Stutenberg, M. (2014, March). *Understanding second language grammatical gender agreement: Relationships between language experience, proficiency, performance and neurocognitive processing.* Paper presented at the American Association for Applied Linguistics (AAAL) Conference, Portland, Oregon.

Faretta-Stutenberg, M., Tanner, D., & Morgan-Short, K. (2013, November). *Individual differences in declarative and procedural memory and changes in L2 ERP signatures over time.* Poster presented at the annual meeting of the Society for the Neurobiology of Language (SNL) Conference, San Diego, California.

Faretta-Stutenberg, M., Tanner, D., & Morgan-Short, K. (2013, October). *The relationship between working memory and second language development at home and abroad.* Paper presented at the Second Language Research Forum (SLRF), Provo, Utah.

Issa, B., Faretta-Stutenberg, M., Zalbidea, J., Bartlett, L., & Morgan-Short, K. (2013, October). *Musical background and second language acquisition: Moving beyond phonology.* Paper presented at the Second Language Research Forum (SLRF), Provo, Utah.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P.C.M. (2012, October). *Declarative and procedural memory as individual differences in second language acquisition.* Paper presented at the Second Language Research Forum Conference (SLRF), Pittsburgh, Pennsylvania.

Brill, K. A., Faretta-Stutenberg, M., Wong, F., Wong, P.C.M., & Morgan-Short, K. (2011, May). *Declarative and procedural memory abilities and successful adult language learning.* Poster presented at the annual meeting of the Midwestern Psychological Association (MPA), Chicago, Illinois.

*Karpouzian, T., Faretta-Stutenberg, M., Wong, F., Wong, P.C.M., & Morgan-Short, K. (2011, May). *The role of working memory in second language development.* Poster presented at the annual meeting of the Midwestern Psychological Association (MPA), Chicago, Illinois.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P.C.M. (2011, April). *The relationship between artificial and natural language learning.* Paper presented at the 2011 University of Illinois at Chicago Bilingualism Forum, Chicago, Illinois.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P.C.M. (2011, January). *The relationship between artificial and natural language learning.* Paper presented at the 85th meeting of the Linguistic Society of America, Pittsburgh, Pennsylvania.

VITA (continued)

Morgan-Short, K., Bartlett, L., Faretta-Stutenberg, M., & González-Vilbazo, K. (2011, April). *Local and distant morphosyntactic processing at early stages of second language acquisition: An event-related potential study.* Poster presented at the annual meeting of the Cognitive Neuroscience Society (CNS), San Francisco, California.

Faretta-Stutenberg, M., & Morgan-Short, K. (2010, October). *Learning without awareness revisited: Examining the role of prior knowledge in implicit learning.* Poster presented at the Second Language Research Forum Conference (SLRF), College Park, Maryland.

McCarthy, B., Faretta, M., Wong, F., Wong, P.C.M., & Morgan-Short, K. (2010, August). *Individual differences in successful second language learning: The roles of working memory and intelligence.* Poster presented at the annual meeting of the Cognitive Science Society (CSS), Portland, Oregon.

Brill, K. A., Faretta, M., Wong, F., Wong, P.C.M., Morgan-Short, K. (2010, August). *Declarative and procedural memory abilities as predictors of successful adult language learning.* Poster presented at the annual meeting of the Cognitive Science Society (CSS), Portland, Oregon.

Morgan-Short, K., Steinhauer, K., Sanz, C., Faretta, M., & Ullman, M. T. (2009, March). *The neurocognition of morpho-syntactic processing in second language: An artificial language study.* Poster presented at the annual meeting of the Cognitive Neuroscience Society (CNS), San Francisco, California.

Morgan-Short, K., Faretta, M., Sanz, C., Steinhauer, K., & Ullman, M. T. (2009, March) *Awareness and explicitness in second language development*. Paper presented at the Georgetown University Round Table (GURT), Washington, D.C.

<u>POSTERS: OTHER</u>

*Bautista, K., Issa, B., Faretta-Stutenberg, M., Zalbidea, J., Bartlett, L., & Morgan-Short, K. (2013, April). *The role of musical training in acquiring a second language.* Poster presented at the University of Illinois at Chicago Student Research Forum, Chicago, Illinois.

Awarded Second Place for Social Sciences at the 2013 UIC Student Research Forum

*Orozco, L.M., Faretta-Stutenberg, M., & Morgan-Short, K. (2013, April). *Working memory and adult second language acquisition*. Poster presented at the University of Illinois at Chicago Student Research Forum, Chicago, Illinois.

*Quadri, N., Faretta-Stutenberg, M., & Morgan-Short, K. (2013, April). *Declarative and procedural memory in second language processing.* Poster presented at the University of Illinois at Chicago Student Research Forum, Chicago, Illinois.

VITA (continued)

*Karpouzian, T., Faretta-Stutenberg, M., Wong, F., Wong, P.C.M., & Morgan-Short, K. (2011, April). *The role of working memory in second language development.* Poster presented at the University of Illinois at Chicago Student Research Forum, Chicago, Illinois.

Awarded Second Place for Social Sciences at the 2011 UIC Student Research Forum

*Lee, L., Faretta-Stutenberg, M., Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M.T. (2011, April). *The effect of language experience on subsequent language learning.* Poster presented at the University of Illinois at Chicago Student Research Forum, Chicago, Illinois.

RESEARCH AND MENTORING EXPERIENCE

Research Assistant to Dr. Kara Morgan-Short 2008-2014
Cognition of Second Language Acquisition Laboratory, UIC

Supervising Graduate Mentor to Undergraduate Research Assistants 2008-2014
Cognition of Second Language Acquisition Laboratory, UIC
Chancellor's Undergraduate Research Award (N = 2)
Letters and Sciences Undergraduate Research Initiative (N = 4)
Summer Research Opportunities Program (N = 2)
Honors College Capstone Projects (N = 6)
Volunteer and Course Credit only (N = 7)

TEACHING EXPERIENCE

Basic Language Program Coordinator 2010-2012
Department of Hispanic and Italian Studies, UIC
Elementary Spanish (SPAN 102)

Teaching Assistant 2007-2013
Department of Hispanic and Italian Studies, UIC
Instructor: Spanish Basic Language: Elementary, Intensive Elementary, Intermediate
Instructor: Spanish Grammar in Practice: Advanced
Assistant: Second Language Learning - Introduction to SLA Theories (graduate course)
Assistant: Introduction to Hispanic Linguistics (undergraduate course)

VITA (continued)

INVITED LECTURES

2014 Psycholinguistics and Second Language Acquisition,
Department of Foreign Languages and Literatures, Northern Illinois University

2013 Theoretical and Research Foundations of Communicative Language Teaching,
School of Literature, Cultural Studies, & Linguistics, UIC

2013 Graduate Seminar on Study Abroad,
Department of Spanish & Portuguese, Georgetown University

2013 Professional Development Workshop for Spanish Majors,
Department of Hispanic and Italian Studies, UIC

BROWN BAG TALKS

Working memory, linguistic development, and changes in ERP signatures over time. (2013, October). UIC Talks in Linguistics, Department of Hispanic and Italian studies, University of Illinois at Chicago.

Awareness and explicitness in second language acquisition, with Morgan-Short, K. (2009, March). UIC Talks in Linguistics, Department of Hispanic and Italian studies, University of Illinois at Chicago.

Electrophysiological investigation of bilingual processing: ERP methods for heritage and second language learning, with Morgan-Short, K. and Bartlett, L. (2008, November). UIC Talks in Linguistics, Department of Spanish, French, Italian & Portuguese, University of Illinois at Chicago.

SERVICE TO THE UNIVERSITY

AY2009, AY2014	UIC Talks in Linguistics (TiL) Organizing Committee
AY2009, AY2011	UIC Bilingualism Forum Organizing Committee
AY2010	Chancellor's Committee to Review Graduate Appointment Processing
AY2013	Graduate Student Council Representative

VITA (continued)

SERVICE TO THE PROFRESSION

Abstract Reviewer

UIC Bilingualism Forum, 2011, 2012
Society for the Neurobiology of Language Conference, 2010

Ad Hoc Reviewer

Journal of Cognitive Neuroscience
Selected Proceedings of the 2012 Second Language Research Forum
Learning and Individual Differences

TRAVEL AWARDS

2014 Student Presenter Award, Graduate College, UIC (also 2008, 2010, 2012)
2014 Presenter Award, Graduate Student Council, UIC (also 2012)
2013 Travel Award, School of Literatures, Cultural Studies & Linguistics, UIC (also 2012)
2013 Ph.D. Student Travel Award, College of Liberal Arts and Sciences, UIC (also 2012)

PROFESSIONAL AFFILIATIONS

American Association for Applied Linguistics
Linguistic Society of America
Society for the Neurobiology of Language