THE EFFECTS OF FORM-FOCUSED TRAINING AND WORKING MEMORY ON THE L2 PROCESSING AND LEARNING OF MORPHOLOGICAL CUES

A Dissertation in
Spanish and Language Science
by
Melisa J. Dracos

© 2013 Melisa J. Dracos

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

May 2013
The dissertation of Melisa J. Dracos was reviewed and approved* by the following:

Karen L. Miller  
Assistant Professor of Spanish Linguistics  
Dissertation Adviser  
Chair of Committee

Paola E. Dussias  
Head of the Department of Spanish, Italian, and Portuguese  
Associate Professor of Spanish, Linguistics, and Psychology

Robert W. Schrauf  
Head of the Department of Applied Linguistics  
Professor of Applied Linguistics

Ping Li  
Professor of Psychology, Linguistics, and Information Sciences & Technology

*Signatures are on file in the Graduate School.
ABSTRACT

Beyond the normal trials that confront an adult learner of a second language (L2), acquiring inflectional morphology presents peculiar and particularly daunting challenges. Beyond the linguistic factors (low salience, low reliability, redundancy) that render inflections less likely to be learned, the Associative-Cognitive model holds that language experience compounds this difficulty as attentional biases to cues learned in the first language (L1) can hinder the processing and in turn acquisition of important L2 grammatical cues (e.g., N. Ellis, 2006b, 2008b). Since learners’ working memory limitations often inhibit attention to multiple cues in L2 input, they frequently ignore morphological cues when they are redundant and compete with lexical cues (e.g., explicit subjects, adverbs) (e.g., VanPatten, 1996, 2006), especially if their L1 is morphologically impoverished compared to the L2 (e.g., English-Spanish learners) (e.g., N. Ellis & Sagarra, 2011). The difficulty in overcoming L1 cue biases can be exacerbated by the overuse of L2 lexical cues in the classroom (e.g., Dracos, 2010), which obviates the need to process inflections for meaning. LaBrozzi (2009) found that study abroad experience helped some learners, namely those with higher working memory, rely more on morphological cues. However, study abroad is not always an option, especially in the early stages of L2 learning. Moreover, exposure to intensive amounts of input containing the target form might not be enough to override learned attentional biases; some sort of form-focused intervention might be crucial, or at least play an important role, in helping learners attend to the L2 grammatical cues, and in turn promote their acquisition (e.g., N. Ellis, 2006b, 2008b). Under this motivation, the present dissertation evaluates the efficacy of longitudinal training that forces attention to verbal inflections, while also exploring whether the provision of feedback facilitates the processing and learning of L2 morphological cues. In addition to external instructional
variables, this study investigates the role of internal cognitive resources, namely working memory, in L2 learners’ processing of L2 morphological cues and their ability to learn from the training.

This dissertation used an innovative experimental paradigm that involved five computerized training sessions that recorded performance as well as offline and online assessment measures that were administered using a pretest/posttest procedure (with immediate and delayed posttests). Similar to previous Processing Instruction research (e.g., VanPatten, 2004b), the training involved exposure to input activities that required the processing of inflectional morphology for successful interpretation of meaning (determining the subject and tense), and the learners were assessed on their ability to correctly interpret the form as well as produce it in writing. In addition to these traditionally used assessments, a cue reliance task was employed to examine the learners’ biases toward lexical as opposed to morphological cues, and a self-paced reading task was used as a measure of online sensitivity to adverb-verb tense violations and subject-verb agreement violations.

A total of 264 second-semester L1 English learners of Spanish were included in the participant pool after exclusions. The participants were assigned to one of three experimental training groups differing only in the feedback provided (none, correct/incorrect, metalinguistic) or to the control group (no training). Results indicated that when focused on form, all three experimental groups, but not the control group, showed a robust increase in reliance on and accurate processing and production of verbal inflections, with metalinguistic feedback consistently enhancing the effectiveness of the training. However, results from the self-paced reading task revealed that the training had no effects on learners’ sensitivity to L2 morphological cues during more implicit, real-time processing while reading for overall meaning. With respect
to the role of working memory, the findings indicated that participants with higher working memory were more accurate overall in the aural processing of verbal inflections, and they were also able to learn more from the training and improve their accuracy in the processing and written production of inflections over time. In addition, self-paced reading task results revealed that learners with higher working memory capacity better processed verbal agreement online. In fact, comparison of low-span to high-span learners indicated that only learners with high working memory were sensitive to tense and subject-verb agreement violations.

The three central conclusions established through this research can be summarized as follows. First, the breadth and novelty of the employed experimental design offers significant support to previous research suggesting that native speakers of a morphologically weak L1 like English rely on meaningful lexical items over redundant inflections to derive meaning in a morphologically rich L2 such as Spanish, particularly at lower levels of proficiency (e.g., N. Ellis & Sagarr, 2010a, 2010b; Lee et al., 1997; Musumeci, 1989; VanPatten, 1996, 2004, 2006). Second and more importantly, the findings from the form-focused tasks indicate that refocusing learners’ attention through training intervention, particularly with metalinguistic feedback, can help learners overcome learned attentional biases and facilitate the processing and learning of L2 verbal morphology even at early stages of acquisition. This lends support to tenets of the Associative-Cognitive model (N. Ellis, 2006b, 2008b) and has important implications for instruction. Finally, this dissertation provides support for the small but growing body of research indicating effects of working memory on L2 grammar learning (e.g., Brooks et al., 2006; Kempe & Brooks, 2008; Kormos & Sáfar, 2008; Martin & N. Ellis, 2012), and, in response to an area of contention in the literature (e.g., Havik et al., 2009; Sagarra, 2007c; Sagarra & Herschensohn,
2010; cf. e.g., Juffs, 2004, 2005; Foote 2011), it also shows that working memory has a substantial influence on online L2 (morpho)syntactic processing.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................... ix

LIST OF TABLES ................................................................................................................ xi

ACKNOWLEDGEMENTS .............................................................................................. xii

CHAPTER 1: Introduction ................................................................................................. 1

CHAPTER 2: Literature Review ......................................................................................... 7

2.1 Introduction ................................................................................................................. 7

2.2 L2 Morphological Processing ..................................................................................... 7

2.2.1 Input Processing Model and Research .................................................................. 8

2.2.1.1 VanPatten’s Input Processing Model .......................................................... 8

2.2.1.2 Offline Research on Processing Redundant L2 Morphological Cues ........... 11

2.2.1.3 Summary and Limitations of Offline Research ........................................... 17

2.2.2 Online Processing of Agreement in an L2 ............................................................ 20

2.2.2.1 Studies on Processing of Nominal Agreement ......................................... 22

2.2.2.2 Studies on Processing of Verbal Agreement ............................................. 29

2.2.2.3 Summary of Online Research .................................................................. 42

2.3 Associative-Cognitive Model: Attentional Biases in L2 Processing and Acquisition .... 47

2.4 Instructional Factors and Strategies ......................................................................... 58

2.4.1 Structured Input Activities .................................................................................. 59

2.4.2 Corrective Feedback ........................................................................................... 63

2.5 Working Memory (WM) in L2 Processing and Learning .......................................... 67

2.5.1 Models of WM .................................................................................................... 69

2.5.1.1 The Original Model of WM and other Domain-Specific Models ............... 69

2.5.1.2 Domain-General Models of WM ............................................................... 78

2.5.2 WM Measures .................................................................................................... 94

2.5.2.1 Short-Term Memory Tests ........................................................................ 94

2.5.2.2 Complex WM Tests .................................................................................. 96

2.5.3 WM in L2 Processing and Learning .................................................................... 102

2.5.3.1 Studies on WM and L2 Processing of (Morpho)syntax .............................. 103

2.5.3.2 WM and Aspects of L2 Learning ................................................................ 116

2.6 The Present Study .................................................................................................... 128

CHAPTER 3: Experimental Study .................................................................................... 136

3.1 Overview of Method and Materials .......................................................................... 136

3.1.1 Participants ........................................................................................................ 137

3.1.2 Target structure: Spanish verbal inflectional morphology .................................. 138

3.1.3 Experimental Design ........................................................................................ 140

3.1.3.1 Screening and Proficiency Tasks .............................................................. 143

3.1.3.2 The Letter-Number Sequencing Working Memory Task ......................... 146

3.1.3.3 Training ...................................................................................................... 146

3.1.4 Descriptive Statistics for Screening, Proficiency, WM, and Training Measures .... 150

3.1.4.1 Means for Screening, Proficiency, and WM Measures ............................ 150

3.1.4.2 Training Accuracy Means ........................................................................ 152

3.2 Assessment 1: Aural Processing Task ..................................................................... 156

3.2.1 Participants ........................................................................................................ 156

3.2.2 Materials and Procedure .................................................................................... 158

3.2.3 Results ................................................................................................................ 159

3.2.4 Summary of Results .......................................................................................... 164
3.3 Assessment 2: Written Production Task ........................................................................ 165
  3.3.1 Participants.................................................................................................................. 165
  3.3.2 Materials and Procedure ................................................................................................. 167
  3.3.3 Results ............................................................................................................................ 168
  3.3.4 Summary of Results ....................................................................................................... 174
3.4 Assessment 3: Cue Reliance Task .................................................................................... 175
  3.4.1 Participants ..................................................................................................................... 175
  3.4.2 Materials and Procedure ............................................................................................... 177
  3.4.3 Results ............................................................................................................................ 180
  3.4.4 Summary of Results ....................................................................................................... 185
3.5 Assessment 4: Self-paced Reading Task ........................................................................... 186
  3.5.1 Participants ..................................................................................................................... 186
  3.5.2 Materials and Procedure ............................................................................................... 188
  3.5.3 Results ............................................................................................................................ 192
    3.5.3.1 Accuracy on the Picture Comprehension Measure .................................................. 193
    3.5.3.2 Reading Time Results Overall .................................................................................. 194
    3.5.3.3 Reading Time Results for Training Groups over Time .............................................. 204
    3.5.3.4 Reading Time Results for High vs. Low WM Groups ................................................. 206
  3.5.4 Summary of Results ....................................................................................................... 215
CHAPTER 4: Conclusions and Implications .............................................................................. 217
REFERENCES .......................................................................................................................... 228
APPENDIX A: Language Background Questionnaire .............................................................. 250
APPENDIX B: Vocabulary Test .................................................................................................. 252
APPENDIX C: Grammar Proficiency Test .................................................................................. 253
APPENDIX D: Written Production Experimental Stimuli .......................................................... 257
APPENDIX E: Letter-Number Sequencing Stimuli ................................................................. 259
APPENDIX F: Training Stimuli ................................................................................................. 260
APPENDIX G: Aural Processing Experimental Stimuli ............................................................... 262
APPENDIX H: Cue Reliance Experimental Item Sets ............................................................... 264
APPENDIX I: Self-paced Reading Experimental Stimuli ........................................................... 269
APPENDIX J: Examples of Self-Paced Reading Assessment Pictures ...................................... 273
LIST OF FIGURES

Figure 1. Baddeley & Hitch’s (1974) Three-Component WM Model ............................................. 70
Figure 2. Baddeley’s (2000, 2003, 2007) Current Multi-Component WM model ......................... 72
Figure 3. Cowan’s (1988) Embedded-Processes Model................................................................. 80
Figure 4. Schneider & Detweiler’s (1988) Connectionist Model of WM (system level) .................. 82
Figure 5. Kane et al.’s (2007) Measurement Model of the WM system (adapted from Engle, Kane et al., (1999) [version 1] and Engle & Kane (2004) [version 1.2]) ................................. 90
Figure 6. Experimental Design ........................................................................................................ 141
Figure 7. Example of a Training Visual Trial with a Tense Question in Feedback Conditions ........ 148
Figure 8. Example of Metalinguistic Feedback Following a Tense Question During Aural Training Practice ......................................................................................................................... 149
Figure 9. Example of Metalinguistic Feedback Following a Subject Question During Visual Training Practice .......................................................................................................................... 150
Figure 10. Training Accuracy for the Tense Condition by Group .................................................... 154
Figure 11. Training Accuracy for the Subject Condition by Group .................................................. 155
Figure 12. Example of a Trial for the Aural Processing Task .......................................................... 159
Figure 13. Accuracy in Aural Processing of Verbal Inflections for Tense ........................................ 162
Figure 14. Accuracy in Aural Processing of Verbal Inflections for Subject ..................................... 163
Figure 15. Accuracy in Written Production of Present Tense Verbal Inflections ............................. 170
Figure 16. Accuracy in Written Production of Preterite Tense Verbal Inflections .......................... 171
Figure 17. Accuracy in Written Production of Future Tense Verbal Inflections ............................. 172
Figure 18. Example of a Trial for the Cue Reliance Task ............................................................... 179
Figure 19. Percentage of Reliance on Verbal Inflections for Tense Condition .............................. 182
Figure 20. Percentage of Reliance on Verbal Inflections for Subject Condition ............................ 182
Figure 21. Example of a Trial for the Self-Paced Reading Task ..................................................... 190
Figure 22. Example of the Picture-Based Comprehension Measure in the Self-Paced Reading Task. ........................................................................................................................................... 191
Figure 23. Mean Reading Times for Tense Agreement and Violations Overall (All Groups Combined) ......................................................................................................................................................... 195

Figure 24. Mean Reading Times for S-V Agreement and Violations Overall (All Groups Combined) .......................................................................................................................................................... 196

Figure 25. Mean Reading Times for Preterite Tense Agreement and Violations (All Groups Combined) ........................................................................................................................................................................ 199

Figure 26. Mean Reading Times for Future Tense Agreement and Violations (All Groups Combined) ........................................................................................................................................................................ 200

Figure 27. Mean Reading Times for Preterite S-V Agreement and Violations (All Groups Combined) ........................................................................................................................................................................ 202

Figure 28. Mean Reading Times for Future S-V Agreement and Violations (All Groups Combined) ........................................................................................................................................................................ 203

Figure 29. Mean Reading Times on Critical Region for Low WM Group: Overall Tense and Subject-Verb Conditions.......................................................................................................................................................... 212

Figure 30. Mean Reading Times on Critical Region for High WM Group: Overall Tense and Subject-Verb Conditions.......................................................................................................................................................... 213

Figure 31. Mean Reading Times on Critical Region for Low WM Group: Preterite and Future Tense and Subject-Verb Conditions.......................................................................................................................................................... 213

Figure 32. Mean Reading Times on Critical Region for High WM Group: Preterite and Future Tense and Subject-Verb Conditions.......................................................................................................................................................... 214
LIST OF TABLES

Table 1. Verbal Paradigm for Spanish Present, Past (Preterite), and Future Tenses .......... 138

Table 2. Means and Standard Deviations for Screening, Proficiency, and WM Measures by Group ......................................................................................................................................................... 151

Table 3. Means for Training Accuracy in the Tense Condition by Group (SD in parentheses)..153

Table 4. Means for Training Accuracy in the Subject Condition by Group (SD in parentheses) .................................................................................................................................................................................. 153

Table 5. Descriptive Statistics for the Subset of the Participant Pool Used in the Aural Processing Task Analyses on Screening, Proficiency, and WM Measures ............... 157

Table 6. Descriptive Statistics for the Aural Processing Assessment (SD in parentheses) .... 160

Table 7. Descriptive Statistics for the Subset of the Participant Pool Used in the Production Task Analyses on Screening, Proficiency, and WM Measures ............................................................... 166

Table 8. Descriptive Statistics for the Written Production Assessment (SD in parentheses) .... 169

Table 9. Descriptive Statistics for the Subset of the Participant Pool Used in the Cue Reliance Task Analyses on Screening, Proficiency, and WM Measures ................................................................. 176

Table 10. Examples of Experimental Conditions for the Cue Reliance Task ....................... 178

Table 11. Descriptive Statistics for the Cue Reliance Assessment: Reliance on Verbal Morphology ........................................................................................................................................................................ 181

Table 12. Descriptive Statistics for the Subset of the Participant Pool Used in the Self-Paced Reading Analyses on Screening, Proficiency, and WM Measures ............................... 187

Table 13. Examples of Experimental Conditions for the Self-Paced Reading Task ............. 189

Table 14. Comprehension Accuracy for Experimental Trials on Self-Paced Reading Task at Pretest, Posttest, and Delayed Posttest (SD in parentheses) ......................................................... 193

Table 15. Relevant Regions for Reading Time Analyses .......................................................... 194

Table 16. Group Mean RTs on Critical Region (Verb) in the Tense and S-V Agreement Conditions (SD in parentheses) ........................................................................................................ 205

Table 17. Between-Group Results of Statistical Analyses (One-way ANOVAs) on Difference Scores (Agreement Violation – Agreement RTs) for Critical Word 5 (Verb) ............... 206

Table 18. Mean WM Span Scores and Mean RTs (SDs) on the Verb in Tense and S-V Agreement Conditions for High and Low WM Groups ........................................................................ 208
ACKNOWLEDGEMENTS

I would like to acknowledge my deep gratitude to the people who have been a constant support and help to me during the dissertation journey. First and foremost, I want to thank my husband, Marcus, for his constant support, love, encouragement, sacrifice, and selfless sharing of time to help me in every way he could. I would not have been able to do this if it were not for you. You sacrificially moved to Penn State with me so I could pursue my dream, and you have supported me every day along the way. You are also the only one who has really seen me throughout the entire dissertation process – the ups, the downs, the exciting moments of discovery, the disheartening obstacles, the late nights and early mornings. When I was discouraged, stressed, or questioning my work, I could always count on you to calm me down, cheer me up, and push me to persevere. Thank you for always listening to my ideas, helping me with editing, and offering words of advice. Thank you for always believing in me and my abilities. There are not words that can adequately express how grateful I am to you and how blessed I feel to be married to you.

The other person that is integral to this dissertation is my advisor, Karen Miller. I am indebted to you for your constant support, encouragement, and your selfless sharing of time, experience, and words of wisdom. You are much more to me than a wonderful advisor; you are a truly valued mentor, friend, and now collaborator, and I look forward to maintaining our relationship and working together on many future projects.

Next I want to sincerely thank the other members of my dissertation committee, Giuli Dussias, Robert Schrauf, and Ping Li. Each of you played a crucial role in helping me develop and carry out this dissertation research, and your guidance and support was important in me
succesfully completing the project. I am very appreciative. Other members of the academic community that I would like to thank for their support and valuable contributions to my dissertation research are Nuria Sagarra, Chip Gerfen, Jing Yang, Alvaro Villegas, Cari Bogulski, Astrid Román-Hernandez, and Aroline Hanson. I want to specifically thank Nuria Sagarra for her theoretical insights, her advice with respect to experimental design, and her help in organizing data collection. I want to give a special thanks to Jing Yang, who dedicated many hours to helping me with experimental programming. You taught me so much, and I am very grateful. I would also like to thank my wonderful friend and fellow graduate student peer, Aroline, for her constant encouragement, insights, and help in numerous ways. I am so grateful to have had you as a close companion throughout our graduate school journey; it would have been a much tougher road, and not near as fun, without your support and friendship. I have learned so much through our many friendly academic discussions, and it has been fun to be able to enjoy stress-relieving athletic and social adventures together as well. I also owe a sincere thanks to my research assistant, Michele DeMuth, who dedicated long, tiring hours to helping me collect and extract data. You were an invaluable asset to this project. The final members of the academic community that I would like to thank are the undergraduate student participants who took part in my training study.

Finally, I want to thank my family. In addition to my wonderful husband, who I have already thanked, I am blessed with amazing parents and grandparents who have never ceased to provide their endless love, support, and encouragement throughout my long academic journey. Thank you for always believing in me and lifting up my spirits and cheering me on at the most difficult times. I would like to thank in particular my mother. Thank you, mom, for always being available and willing to patiently listen to the ins and outs of my studies and research as well as
all of my trials and triumphs. You always offered thoughtful advice when appropriate and reminded me to breathe, relax, and take care of myself when necessary. Finally, I would like to thank my precious, three-month-old daughter, Romina, for being a true joy and giving me heartwarming smiles that lifted my spirits and helped me get through the final writing and revision stages of this dissertation.
CHAPTER 1: Introduction

Beyond the normal trials that confront an adult learner of a second or foreign language (L2), acquiring inflectional morphology appears to present peculiar and particularly daunting challenges. Many late learners exhibit persistent difficulty with inflections in both production (e.g., Guillelmon & Grosjean, 2001; Hawkins & Liszka, 2003; Lardiere, 1998a) and processing (e.g., Jiang, 2004; Keating, 2009), and problems in this domain extend even to learners who are otherwise highly proficient in the L2.

Consider specifically verbal inflections, which many languages rely on heavily to express tense and person, amongst other information such as aspect and mood. Classroom learners of such languages with rich inflection (e.g., Spanish) face the task of memorizing complex verbal paradigms, and even after they do show knowledge of inflection, they still struggle to comprehend and produce these endings in communicative contexts (e.g., Larsen-Freeman, 2010).

Adults who attempt to learn an L2 naturally through immersion, but without instruction, often fare no better. Naturalistic second language acquisition often stabilizes at a level far short of nativelike ability, which, although sufficient for everyday communication, often lacks functional inflection, such as missing or inconsistent use of tense or subject-verb agreement (Bardovi-Harlig, 2000; N. Ellis, 2008; Klein, 1986, 1998). Rather, both instructed and non-instructed learners have been shown to rely on content words such as temporal adverbs, explicit subjects, and the verb root (i.e., the lexical meaning of the verb and not the inflectional suffix) in both processing and production (e.g., Clahsen & Felser, 2006; VanPatten, 1996, 2004). In contrast, children acquiring their first language learn to process and produce inflections relatively quickly, and this is especially so in languages with complex morphology because they rely heavily on
grammatical morphemes to mark meaning (e.g., Austin, 2010). Why, then, do so many late learners experience such persistent problems with inflections, and is there anything that can be done to mitigate this difficulty?

This dissertation will operate specifically within the environs of these questions, aiming both (1) to offer theoretical insights into late learners’ difficulty with L2 inflectional morphology, and (2) to investigate potential ways to overcome this challenge through pedagogical intervention. I will aim to fulfill the first of these objectives by examining the processing behavior, learning, and the effects of cognitive individual differences in a language training study that focuses on learners attempting to acquire verbal inflections in a L2 (Spanish) that is morphologically richer than their L1 (English). The second goal will be achieved by evaluating the efficacy of the instructional practices employed in the training for facilitating the processing and learning of verbal inflections.

Various explanations have been posited to account for the challenges associated with adult second language acquisition of grammatical structures such as inflectional morphology. Some focus on a biological critical period for language acquisition, some on social interactional factors, and others on motivational differences. The Associative-Cognitive model (N. Ellis, 2006a, 2006b) offers a viable, alternative explanation in terms of general cognitive and associative principles that are also operative in other types of learning, specifically attentional processes in the associative learning of language structures. The current dissertation will work within this framework given that it provides a comprehensive and convincing language-experienced based account for adults’ difficulty acquiring an L2 that is in line with empirical evidence on L2 morphological processing (reviewed in Chapter 2).
In particular, according to the Associative-Cognitive model it is processing limitations, linguistic factors, and language experience that explain why certain L2 grammatical features are particularly difficult to learn, however available they might be in the input. Since processing limitations often inhibit attention to multiple cues in the L2 input, learners have to choose which aspects of the input they will process. The low salience, low-reliability, and frequent redundancy of inflections partially explain why they are less likely to be processed and learned. Yet, it is late learners’ prior language experience as it affects their attention to language that makes the acquisition of features such as inflectional morphology particularly difficult for L2 learners (N. Ellis, 2006), a thesis that is in line with other usage-based approaches such as the Competition Model (MacWhinney, 1987, 2001). As a result of L1 experience, regardless of the language, learners know that there are reliable and salient lexical cues to express information such as time (adverbs) and person (explicit subjects); so if frequently present in the L2 input, these known cues might block the processing and acquisition of inflections marking this same information. Moreover, this approach predicts that inflectional morphology is particularly difficult if the learner’s first language (L1) makes no, less extensive, or different use of the target morphology, which is consistent with other research (e.g., N. Ellis & Sagarra, 2011; Jiang, 2011; Tolentino & Tokowicz, 2011). In short, following this model, it is the conjunction of unfavorable linguistic factors with strong attentional biases based on a life-time of prior L1 usage that render verbal inflections especially difficult for English-Spanish learners.

The already challenging task of overcoming L1 cue biases can be compounded by factors inherent to the learning context, such as the overrepresentation, underrepresentation, and altered patterns of use of L2 cues in the classroom (e.g., Dracos, 2010; Gass, 1997; Goodall, 2008; Santilli, 1996; Sanz, 1999). Particularly problematic for the processing of L2 verbal morphology
is the overuse of lexical cues in ‘teacher talk’ as well as in other learners’ speech, which obviates
the necessity to process verbal inflections for meaning. LaBrozzi (2009) found that exposure to
intensive amounts of morphological cues during study abroad helped some intermediate English-
Spanish classroom learners, particularly those with higher working memory, rely more on verbal
inflections in a nativelike way. Yet, study abroad is not always an option, especially in the early
stages of L2 learning. Therefore, it is important to discover other ways to help classroom learners
override L1 cue preferences and attend to L2 morphological cues. Moreover, although exposure
to intensive amounts of input containing the target form is critical, it may not be enough to
override learned attentional biases and successfully acquire certain grammatical cues (e.g., N.
Ellis, 2008). Some sort of form-focused intervention, then, might be crucial, or at least play an
important role, in helping learners attend to the L2 grammatical cues, and in turn promote their
acquisition (e.g., DeKeyser, 2003; N. Ellis, 2006b, 2008b; Gass, 1997; MacWhinney, 2011).
Under this motivation, the present dissertation evaluates the efficacy of longitudinal training that
forces attention to verbal inflections, while also exploring whether the provision of feedback
facilitates the processing and learning of L2 morphological cues.

In addition to external instructional variables, this study investigates the role of internal
cognitive resources, namely working memory, on L2 learners’ processing of L2 morphological
cues and their ability to learn from the training. Previous research examining whether individual
differences in working memory capacity might account for variation in online (morpho)syntactic
processing has led to conflicting results (e.g., Havik et al., 2009; Juffs, 2004, 2005; Sagarra &
Herschenshohn, 2010). On the other hand, there is limited but encouraging evidence that
working memory capacity is positively related to explicit L2 grammar learning, presumably
because such processes draw heavily on the executive attention component of working memory
(Roehr, 2008; Williams, 2012). This dissertation will inform these bodies of research by exploring the possible influence of working memory on learners’ cue biases, their online processing of verbal agreement violations, and their ability to learn from form-focused training.

In short, the overarching goal of the current dissertation is to elucidate how certain external (form-focused instructional strategies) and internal (working memory) factors modulate learners’ processing and learning of notoriously difficult L2 morphological cues. To examine this, beginning late English learners of Spanish received intensive training on both familiar and novel L2 verbal inflections, which are non-salient, often redundant, and considerably less present and important in their native language. Following the suggestion of Larsen-Freeman (2010), this study, rather than merely examining learners’ performance at one point in time (as most previous research has done), adopts a more developmental perspective. It employs a within-subjects design to compare learners’ processing and production of L2 verbal morphology before and after they underwent training. The results of this study will inform both theory and pedagogy. On the theoretical side, this investigation will contribute to a better grasp of the cognitive processes of learned attention in L2 acquisition, specifically shedding light on whether, and the degree to which, learners’ prior attentional biases (and related processing strategies) can be altered through form-focused training. The results will elucidate the role of feedback, and specifically the provision of metalinguistic information, in facilitating processing and learning of L2 grammar. Moreover, the study will enhance our understanding of working memory capacity as a predictor of certain aspects of second language acquisition. From a pedagogical perspective, the importance of this study is rather clear. The persistent difficulties learners face with grammatical cues such as inflectional morphology makes research into effective instructional methods for improving grammar acquisition all the more important for second language education. This
dissertation will provide insights into whether form-focused training is an effective form of pedagogical intervention.

This dissertation is organized as follows. In order to contextualize and motivate the study, in chapter 2, I review the relevant background literature on L2 morphological processing and discuss how the Associative-Cognitive model offers a comprehensive usage-based approach that accounts for learners’ difficulty in this domain and thus subsequently in acquiring L2 grammatical morphemes. I also present research on external instructional factors as well as working memory that may have an effect on the processing and learning of grammatical morphemes. In chapter 3, I describe the experimental design and outline the specific method and materials for the training and the four assessment measures (aural processing task, written production task, cue reliance task, non-cumulative self-paced reading). For the ease of the reader, I also present the results in this chapter, directly following the description of the corresponding task. Finally, in chapter 4, I draw the relevant conclusions, relate them back to the literature, and then pose questions for future research.
CHAPTER 2: Literature Review

2.1 Introduction

The aim of this chapter is to provide an in-depth review of the literature relevant to the central questions and theoretical background of this dissertation. The chapter will begin with a review of research on L2 morphological processing (Section 2.2) that provides insights into the strategies late learners employ when processing L2 inflectional morphology, in particular (morpho)syntactic agreement, and the main factors that have been shown to modulate processing. Section 2.3 discusses how usage-based models such as the Associative-Cognitive model, the framework adopted for this dissertation, takes into account the evidence from the previously discussed research and offers a complete and convincing explanation for the difficulty late learners face in processing and acquiring L2 morphology. Section 2.4 presents an overview of research highlighting instructional factors and practices that may have an effect on the processing and learning of difficult grammatical structures such as inflectional morphemes. Section 2.5 provides background on the cognitive individual difference of working memory and how this factor has been examined in relation to L2 processing and grammar learning. Finally, Section 2.6 outlines the general aims of this dissertation, framing them within the literature discussed in Sections 2.2-2.5.

2.2 L2 Morphological Processing

Inflectional morphology presents significant obstacles for many late second language (L2) learners. In recent years, a growing body of research has emerged to explore learners’ problems in this domain from a processing perspective. The poignancy of this research vein follows from the fact that learners’ difficulties clearly extend beyond production to
comprehension, and acquisition is dependent upon making appropriate form-meaning/function connections during processing (e.g., N. Ellis & Collins, 2009; VanPatten, 2006). Moreover, processing research that limits the use of explicit knowledge by employing online techniques can shed light on learners’ underlying or integrated knowledge of the morphological form. Much of the research has one of two aims. First, some aims to elucidate the strategies and cues learners rely on, particularly during early and intermediate learning stages, and how these may affect acquisition of L2 morphology. Second, other research aims to ascertain whether native-like processing of L2 morphology is possible, and if not, offer an account for this. The goal of this section is to provide a critical review of this body of literature in an effort to shed light on how late learners process L2 morphology, in particular morphosyntactic agreement, and identify factors that likely modulate morphological processing.

This section is organized as follows. Section 2.2.1 presents the fundamentals of VanPatten’s (1996, 2004, 2006) Input Processing Model and the offline research associated with it in order to highlight the strategies L2 learners use to comprehend L2 input, and their effects on L2 morphological processing. Section 2.2.2 reviews research examining how learners at various levels of proficiency process verbal and nominal agreement online, highlighting the main factors found to affect processing.

2.2.1 Input Processing Model and Research

2.2.1.1 VanPatten’s Input Processing Model

The well-known psycholinguistic model of Input Processing (VanPatten, 1996, 2004, 2006) is concerned with the internal strategies learners use to process input and make form-meaning connections, and how these strategies might affect the acquisition of grammar. Based on prior research in cognitive psychology and second language acquisition (SLA), the model
makes four basic claims about what guides learners’ processing as they are engaged in comprehension: 1) learners are focused primarily on extracting meaning from the input, 2) learners must pay attention to the form, while simultaneously attending to meaning, in order for acquisition to occur, 3) making form-meaning connections is cognitively taxing, and 4) since working memory limitations restrict the amount of information learners can process and store during online comprehension, they have to choose which elements of the input they will process. Following from this, VanPatten proposed several principles that yield important insights into the processing and subsequent acquisition of nominal and verbal inflectional morphology. These include the following principles, which will be discussed in turn:

(a) Primacy of Content Words Principle  
(b) Lexical Preference Principle  
(c) Preference for Nonredundancy Principle  
(d) Meaning before Nonmeaning Principle

The Primacy of Content Words Principle claims that learners seek to grasp meaning by processing semantically rich content words (nouns, verbs, adverbs) before noncontent words and inflections. The primacy of processing content words is aided by prosodic factors in aural input, as content items, particularly the root of the words, tend to receive more stress than noncontent items and inflections (VanPatten, 2004). Consequently, “inflections on verbs and nouns may be skipped over or only partially processed and then dumped from working memory as the processing resources in working memory are exhausted by the efforts required to process lexical items” (8). This is especially the case if the semantic notion encoded by the inflection is redundant, which is common in natural languages. The Lexical Preference Principle encompasses this idea that learners will process lexical items for meaning over grammatical markers when they both provide the same semantic information. This implies that in a sentence
such as *Yesterday I played soccer*, learners will tend to comprehend the past nature of the event by processing the adverb *yesterday* rather than the verbal inflection *–ed*.

Yet, not all grammatical markers are redundant, and the *Preference for Nonredundancy Principle* proposes that learners are more likely to process non-redundant meaningful forms than redundant ones. This implies, for example, that learners of English would process, and thus likely acquire, the grammatical marker *–ing* before the *–s* used to mark third person singular verbs, which is consistent with morpheme acquisition studies (e.g., L1A: Brown, 1973; L2A: Goldschneider & DeKeyser, 2001; Johnson & Newport, 1989; Krashen, 1982). While *–ing* tends to be the only marker indicating the progressive nature of an event (e.g., *The girl is running*), third person *–s* is almost always redundant in English since explicit subjects are required (e.g., *The girl runs with friends*). Nevertheless, these verbal inflections carry meaning (i.e., “past”, “in progress”, “third-person singular”), and the *Meaning before Nonmeaning Principle* claims that learners are more likely to process these sorts of grammatical markers than those that VanPatten describes as nonmeaningful, such as inflections marking grammatical gender on adjectives in languages such as Spanish (e.g., *la casa blanca* ‘the white house*'). The gender marking on the adjective ‘white’ only informs us about its relation to the noun; moreover, it is redundant, further reducing the chances that it will be processed. Instead, learners will derive grammatical gender from the noun or the article, if they process it at all, just as they derive tense from adverbs of time instead of from verbal inflections. According to the Input Processing model, redundant verbal and nominal inflections resist processing and acquisition as long as comprehension remains effortful. Learners, rather, will attend to lexical items to the detriment of processing the morphemes since, from their perspective, “lexical items maximize the extraction of meaning” (VanPatten, 2006, p. 119). Inflections will not be processed until much later, if ever.
These claims regarding the processing of redundant grammatical forms and the primacy of lexical items stem primarily from early L2 processing studies employing offline experimental techniques in classroom settings, which will be reviewed in turn (but see also earlier work by Bardovi-Harlig, 1992; Klein, 1986, for supporting evidence from production-based research).

### 2.2.1.2 Offline Research on Processing Redundant L2 Morphological Cues

Musumeci (1989) was one of the first empirical studies to examine how L2 learners process input containing both lexical and morphological cues encoding the same semantic notion. Beginning learners of Italian, French, and Spanish (all L1 English) were divided into groups and asked to assign tense to sentences delivered under one of four conditions: 1) verb morphology and temporal adverbs, 2) verb morphology and physical gestures indicating time, 3) verb morphology with temporal adverbs and physical gestures, or 4) verb morphology as the only marker of tense. Results from a multiple-choice tense identification task clearly indicated that the presence of temporal adverbs was the significant factor determining correct tense assignment. This finding provides evidence for L2 learners’ preference toward processing lexical items over verbal inflections.

Mangubhai (1991) offered additional evidence for the primacy of lexical items in a longitudinal study examining the processing strategies of English speakers who were beginning to learn Hindi through the Total Physical Response Method. Although not specifically examining the processing of redundant morphology, results from concurrent think-aloud protocols, observations, and retrospective reports provide compelling evidence for the primacy of content words over grammatical forms, even at beginning stages of acquisition. The learners relied heavily on lexical items, pragmatic information when available, and chunking (processing whole phrases or expressions as one content word) to get meaning from the input, while neglecting to
devote attention to grammatical cues. Only on few occasions when certain learners could process meaning relatively automatically did they pay attention to grammatical cues, a finding that supports VanPatten’s (2004) claim that comprehension of meaning needs to be relatively effortless in order free up the necessary cognitive resources to process grammatical forms.

VanPatten (1990) further explored whether learners can consciously attend to both form and meaning when processing input at the discourse level. To address this question, he used an experimental paradigm that involved orienting learners’ attention toward detecting lexical or grammatical forms in the input while simultaneously processing for meaning. Learners of different levels of proficiency (first-semester, fourth-semester, and third-year conversation) were assigned to groups and instructed to listen to a passage either for content only, for content while simultaneously noting a lexical item, or for content while simultaneously noting a grammatical marker (either an article or a verbal inflection). The results demonstrated that directing learners to attend to lexical items did not inhibit the comprehension of content, but conscious attention to grammatical markers in the input did negatively affect overall comprehension for all groups. This negative effect on comprehension was especially striking when the grammatical feature was the less salient and frequently redundant bound morpheme marking person. In all of the other conditions proficiency level significantly enhanced recall, but when learners were instructed to attend to the verbal inflection, the most advanced group performed almost as poorly as the lower-level learners. These results suggest that learners experience great difficulty attending to both form and content in the input, at least at beginning and intermediate stages. The fact that conscious attention to content lexical items did not interfere with comprehension, though, is in accordance with the notion that lexical items are of primary importance during L2 input processing.
Lee, Glass, Cadierno, and VanPatten (1997) further investigated the hypothesis that learners focus on lexical items as opposed to grammatical cues when processing discourse-level input for semantic information. Like Musumeci (1989), the authors examined how learners use lexical items and verbal inflections when processing past temporal reference. Three different levels of L1 English learners of Spanish (first-, third-, and fifth-semester) listened to a passage containing regular past-tense (preterite) verbs in the third person singular form under one of two conditions: adverbs plus verbal inflections or verbal inflections only. Findings from a free-recall task indicated that lexical cues significantly enhanced third and fifth-semester learners’ ability to reconstruct past temporal references, whereas first-semester learners were not able to detect tense from the adverbs or the inflections. Beginning learners were, however, able to identify some past temporal references on the multiple-choice tense identification task, suggesting that they are detecting some aspects of the input even though they may not yet be able to use what they notice to express meaning. Independent of proficiency level, the use of lexical cues did not significantly improve tense recognition, although the data did reveal a trend toward this pattern. The findings indicate that language experience (i.e., proficiency) had a significant effect on learners’ ability to make use of lexical items and detect temporal reference. Beginning learners may not yet have enough previous knowledge of the target language to take advantage of content words as the more advanced learners do, a speculation that contradicts Mangubhai (1991) but is consistent with Han and Peverly (2007), as described below. Overall, the results of Lee et al. (1997) suggest that learners, at least beyond the initial learning stages, depend heavily on lexical cues to assign temporal reference, relying little, if at all, on grammatical cues, particularly when reconstructing content.
Lee (1999) replicated this general finding for the written mode. Learners in first or fourth-semester Spanish courses read and recalled the content from narrative passages with or without temporal adverbs. One week later, the participants met individually with the researcher and re-read the passages using a think-aloud procedure. Analyses of learners’ verbalizations revealed that regardless of level, the participants who had adverbs used them heavily to determine temporal reference, suggesting that relative beginners (first semester) can and do make use of these lexical cues. While the more advanced learners who saw adverbs were usually able to identify preterite verbal inflections when they were probed, they consistently relied on adverbs for tense. On the other hand, the advanced learners who were not exposed to adverbs more consistently and spontaneously identified the inflections. These findings suggest that L2 learners can use the L2 morphology when it is non-redundant, but when lexical cues are present, learners will rely on these over inflections to abstract meaning.

While the learners in Lee (1999) had been exposed to the target grammatical form prior to the study, Lee (2002) investigated how lexical cues and target verb-form frequency affects comprehension as well as the incidental processing of a new morphological form while reading in L2 Spanish. The target form was a perceptually salient morpheme of high communicative value, the third-person singular future tense (-á added to the end of the infinitive: cantará “will speak”). The second-semester learners read the passage under one of three conditions: pre-reading exercises orienting learners toward the meaning of the passage, pre-reading exercises directing learners to attend to the new form, or no pre-reading orientation (neutral condition). The passage contained 6, 10, or 16 future tense verb forms, either with or without a title and paragraph-initial adverbial phrases denoting temporal reference (6 passage versions). The frequency of occurrence of the verb form was found to be the most consistent variable positively affecting comprehension.
and processing. Exposure to temporal adverbs greatly enhanced learners’ comprehension and ability to recall target verbs, but the presence of these lexical items had no effect for the tense identification task, the form recognition task, or the form production task. This surprisingly limited effect of temporal adverbs and the facilitative role of frequent target verb forms may be explained by one or both of the following factors: the exceptional salience and communicative value of the future tense morpheme, and the restricted use of adverbial phrases (only paragraph initial) compared to previous studies (e.g., Lee et al., 1997; Lee, 1999) where the lexical items were employed in every sentence containing the target verb form. Nevertheless, the findings suggest that L2 learners can process a new, salient morphological form as a cue to temporal reference in a written passage.

Leeser (2004) further examined learners’ processing of the same unknown morphological form (the Spanish future tense). The study explored whether topic familiarity, mode (aural versus written), or pausing facilitates beginners’ overall comprehension and processing of the target verbal inflection in a passage also containing a title and lexical expressions indicating futurity. Results from the form-recognition and tense identification tasks revealed that only mode affected learners’ processing of the target form, where participants who read the passage were able to recognize and assign meaning to more future forms. These results indicate that if learners at beginning levels are going to make any form-meaning connections for morphological cues, they are more likely to do so while processing written as opposed to aural input. Yet, Leeser suggests that the lack of effect for the other factors may be due to the presence of the future discourse markers (lexical items), which may have obviated learners’ need to process the morphemes denoting future tense.
Rossomondo (2003, 2007) also investigated the incidental processing of the future tense in Spanish, specifically focusing on the variable of the presence or absence of lexical items to better understand their role in enhancing or impeding overall comprehension and processing of the target form. The author suggested that the limited number and strategic placement of lexical items in previous studies on the processing of a new morphological form, specifically in Lee (2002), disallow solid conclusions regarding the role of lexical cues in this process. Thus, in Rossomondo’s study, she designed two texts, one without a title or lexical items indicating future and one with a title and lexical items accompanying each target form, of which there were 13 total. The beginning learners were assigned to one of these passage versions and instructed to read under one of two text interaction formats: silent reading or a think-aloud protocol. Results from the multiple-choice form-recognition task and the cloze passage production task indicated that the presence of temporal adverbs enhanced learners’ comprehension of content without impeding processing of the new morphological form.

The studies reviewed thus far indicate that beginning up to at least intermediate learners tend to rely on content words over redundant morphology to extract meaning, a finding that supports VanPatten’s Input Processing model (1996, 2004, 2006). This conclusion with respect to beginning learners is based on examining the input processing strategies of learners in first-semester language classes (Lee, 1999; Lee et al., 1997; Leeser, 2004; Mangubhai, 1991; Musumeci, 1989; Rossomondo, 2003, 2007). Han and Peverly (2007) argued, though, that these learners – who were tested after the course had begun (with the exception of Mangubhai (1991)) – may have been able to utilize existing (even if limited) linguistic and non-linguistic knowledge when processing the input. Thus, to explore whether absolute beginners do in fact use a meaning-based approach (relying on content items) as the ‘default’ when processing input, Han
and Peverly examined how novice learners process L2 input when exposed to it for the very first time. The participants were 12 graduate students who had no knowledge of the target language, Norwegian, but all of whom knew two or more other languages, at least one of which they had learned through formal instruction. Rather than examining the processing of specific linguistic structures in line with previous research, this study looked at the processing of grammatical forms in general. Results from a recall and fill-in-the-blank task indicated that contrary to what previous research suggested, these absolute learners did not initially rely primarily on content words, but, rather, they adopted a form-based approach to input processing. The authors concluded that while learners with some prior L2 knowledge may rely predominantly on content words and non-redundant forms, those with no prior language experience show no preferences, processing content words and grammatical forms indiscriminately. While this may be the case, the conclusion is only suggestive given that there appears to be a confound introduced by the language background of the sample pool. Not only did the learners know multiple other languages, they had all learned one or more of the languages through formal instruction. This experience may have influenced their input-processing behavior, particularly if they were accustomed to focusing on grammatical forms as learners instructed through traditional methods often are.

2.2.1.3 Summary and Limitations of Offline Research

The research presented in the above section provides convincing evidence that learners, at least with some minimal knowledge of the language, rely on meaningful lexical items over redundant morphology to derive meaning during input processing. While content words were rather consistently found to enhance overall comprehension, conflicting findings and methodological differences across the studies (e.g., mode of input, target form, measurement
means) precludes a firm confirmation as to whether the presence of lexical cues that render morphological forms redundant are detrimental for the processing and acquisition of these grammatical items, as VanPatten’s Input Processing Model claims (1996, 2004, 2006). Albeit limited, this research also provided evidence pointing to other factors, in addition to lexical cues, that affected morphological processing, such as proficiency level (Han & Peverly, 2007; Lee, 1999; Lee et al., 1997; VanPatten, 1990), frequency of occurrence of the form (Lee, 2002), and mode (Leeser, 2004).

The insight this research offers into learners’ processing of redundant morphological forms is limited, though, due to the offline measures employed (e.g., think-aloud protocols, written recalls in the L1, multiple-choice tense recognition tasks, form production tasks). Offline measures such as these have two main drawbacks. First, they can only draw conclusions regarding the act of processing based on performance on tasks after processing is complete. Thus, offline measures cannot speak to how learners actually attend to the input during real-time processing. Second, offline techniques generally allow learners to draw upon explicit knowledge, and are thus inadequate for capturing underlying implicit knowledge (see e.g., Jiang, 2004; Tokowicz & MacWhinney, 2005). For example, learners’ verbalizations during think-aloud procedures reflect what they are consciously processing and fail to tap unconscious or implicit processing (Sagarra, 2007b). Thus, online tasks that are sensitive to implicit knowledge and attention as processing unfolds are necessary to elucidate how learners process redundant morphological forms in the L2. To this end, the following section will review research employing online techniques, including non-cumulative self-paced reading, eye-tracking, and event-related brain potentials, to examine learners’ sensitivity to redundant forms when processing agreement.
In addition to employing offline measures, there are three other crucial limitations to the aforementioned empirical research used to support VanPatten’s Input Processing Model. First, with the exception of Mangubhai (1991) and Han and Peverly (2007), which looked generically at the nature of processing with respect to all forms in the input, the remaining studies focused only on the processing of one type of redundant morphological form: verbal inflections marking tense. These grammatical markers encode important temporal information. What about other meaningful verbal and nominal morphemes that are often redundant, such as those denoting person on verbs, or plurality on nouns? It is unknown whether such forms will cause the same level of difficulty for learners, and we can only speculate that learners will also tend to rely on lexical cues over the grammatical markers to extract meaning in these cases. Moreover, what about redundant morphological forms that do not carry meaning, such as grammatical gender agreement marking on adjectives in languages such as Spanish? According to VanPatten’s *Meaning before Nonmeaning* principle, nonmeaningful forms such as these will cause more problems for learners and be processed later than meaningful forms. The above research provides no experimental evidence, though, that directly supports this principle and suggests that learners will always experience greater difficulty processing nonmeaningful redundant morphological forms as opposed to meaningful ones.

The second critical limitation is that the reviewed research failed to explore the possible influence of the learners’ native language on input processing. Research suggests that learners transfer processing strategies or cue settings from the L1 to the L2 (MacWhinney, 1997). With the exception of Han and Peverly (2007), whose participants had diverse L1s, the learners in all of the other studies were L1-English learners (and one L1-Chinese learner in Mangahai, 1991) of a morphologically richer L2 such as Hindi, Spanish, or Italian. Thus, it is impossible to
determine whether reliance on lexical cues and difficulty processing morphological cues is due to a universal strategy, linguistic characteristics of the form, or to L1 transfer.

Finally, the third limitation stems from the crucial role that the Input Processing model posits for limited cognitive abilities and working memory in L2 processing of lexical and grammatical forms (VanPatten, 1996, 2004, 2006). In spite of the fact that cognitive abilities are central to the model, none of the research used to support the model examined how this factor, specifically individual differences in working memory capacity, modulates processing of redundant morphological forms. Although not from the same theoretical framework, online research examining the processing of nominal and verbal agreement can inform these issues as well as our overall understanding of L2 processing of redundant morphological forms. This will be evidenced in subsequent sections.

2.2.2 Online Processing of Agreement in an L2

In recent years there has been a surge of online research investigating the L2 processing of nominal and verbal agreement in order to help account for learners’ difficulties with inflectional morphology, which have long been reported in offline studies such as those reviewed above as well as research informed by naturalistic or elicited production data (e.g., Franceschina, 2001; Lardiere, 1998a, b; Long, 2003; White, 2003) and grammaticality judgment tasks (e.g., Bialystok, 1997; Johnson & Newport, 1989). Using a variety of online psycholinguistic methods, a principle aim of the processing research of late has been to determine to what extent learners are able to process morphological forms in a native-like way, showing sensitivity to agreement violations, during real-time processing that invokes automatic, implicit grammatical knowledge and limits the use conscious, explicit knowledge. Evidence from this research generally concurs that there are apparent differences between native and nonnative processing; however,
researchers disagree as to how to account for and theoretically explain these differences. For instance, the shallow structure hypothesis proposed by Clahsen and Felser (2006b) claims that L2 processing is fundamentally different from native processing in that L2 learners compute shallower and less detailed syntactic representations as compared to native speakers. While native speakers compute full syntactic analyses, guided by the grammar (a system of symbolic rules and principles of structure building), late L2 learners rely primarily on lexical, semantic, and pragmatic information. Clahsen and Felser (2006a) argue that the reason L2 learners predominately adopt this shallow processing route is because the L2 grammar that feeds the parser is somehow incomplete or divergent as compared to the native grammar. This position is in accord with other representational deficit approaches that posit deficiencies in the underlying competence of late L2 learners, whereby certain grammatical features are not acquirable (e.g., Franceschina, 2001; Hawkins & Liszka, 2003). In contrast, other perspectives maintain that late L2 learners can achieve nativelike grammatical representations, and that any divergent L2 performance can be attributed to processing difficulties involving problems with morphological mapping (e.g., Lardiere, 2000), cognitive limitations in the L2 (McDonald, 2006; Hopp, 2007), and the use of inappropriate processing strategies (VanPatten, 1996, 2004, 2006). In spite of conflicting theoretical perspectives, most researchers concur that L1 transfer can also affect L2 processing; yet, they differ regarding the extent to which this factor can account for L1/L2 processing differences across diverse morphological features (Clahsen et al., 2010).

The debate over whether L2 learners have a grammatical deficit is peripheral to the main goals of this literature review, which are to elucidate how L2 learners actually process morphological cues and highlight the specific variables that have been empirically shown to influence such processing. Nevertheless, as will be illustrated in the following evaluation of
recent online studies, evidence of nativelike L2 processing, when critical factors align, discredits at least a strong view of the representational deficit approach. The first section below (2.2.2.1) will examine the processing of nominal agreement, the area in which the majority of the research on L2 morphological processing of agreement has been conducted. The subsequent section (2.2.2.2) will look at verbal agreement, including subject-verb agreement and adverb-verb tense agreement, which are the specific types of agreement examined in the present dissertation.

2.2.2.1 Studies on Processing of Nominal Agreement

Jiang (2004) was one of the first studies to examine whether L2 learners can process agreement information automatically during online processing as native speakers do (e.g., see Pearlmutter, Garnsey, & Bock, 1999). With the aim of extending the findings of Jiang (2004), which demonstrated that advanced Chinese-English bilinguals were not sensitive to number agreement between the subject and the verb (see following Section 2.2.2.2), Jiang (2007) employed the same non-cumulative self-paced reading paradigm to examine the processing of plural –s within complex NPs, as in The visitor took several of the rare coin/*s in the cabinet. The Chinese-English learners and English native-speaking participants were instructed to read the sentences word by word and then answer comprehension questions, focused on meaning, as quickly as possible. Results for the L2 speakers revealed no significant differences between reading times for grammatical versus ungrammatical sentences, indicating that these advanced learners were not sensitive to the morphosyntactic plurality cues. They were, however, sensitive to verb subcategorization rules. While Jiang (2007) interpreted these results as indicating that L2 knowledge is selectively integrated and the nominal plural morpheme is a ‘nonintegratable’ feature, subsequent work by Jiang and others suggest that this insensitivity is likely due to
properties of the learners’ native language Chinese, which does not mark plurality morphologically.

Jiang, Novokshanova, Masuda, and Wang (2011) examined L2 processing of the same structures using the test materials from Jiang (2007), but in this study he incorporated two language groups in order to test the morphological congruency hypothesis, which proposes that the presence of a similar morpheme in the L1 and L2 facilitates the processing and acquisition of that feature. The participants included native English speakers and advanced English learners with L1 Russian, which marks plurality like English, and L1 Japanese, which does not mark plurals obligatorily. Results from the self-paced reading task indicated that Russian learners were sensitive to plural errors, whereas Japanese learners were not. Jiang and colleagues concluded that the results support the morphological congruency hypothesis, suggesting that nativelike processing is exceptionally difficult, if not impossible, if the target grammatical morpheme is not instantiated in the learner’s L1.

Tokowicz and MacWhinney (2005) further explored how similarities and differences between the L1 and L2 affect online sensitivity to L2 morphosyntactic violations by testing beginning L2 learners of Spanish with L1 English on three constructions: 1) auxiliary marking, which is formed similarly in the L1 and L2, 2) determiner-number agreement, which is formed differently in the L1 and L2, and 3) determiner-gender agreement, which is a construction unique to the L2. Event-related brain potentials (ERPs) were used to measure implicit processing while participants read sentences word by word on the monitor, and grammaticality judgment responses after each sentence were used to measure explicit processing. The ERP results revealed that violations for the feature formed similarly in Spanish and English (auxiliary omission) elicited a P600 effect, which has been shown to be indicative of sensitivity to syntactic
violations (e.g., Osterhout & Mobley, 1995). However, there was no such effect for determiner-number agreement, signifying that the learners were not sensitive to violations for features that are formed differently between the L1 and L2. Moreover, learners demonstrated robust online sensitivity to violations for the feature unique to L2 Spanish (determiner-gender agreement), even though they did not show explicit knowledge of this feature on the grammaticality judgment measure. This suggests that beginning learners are sensitive to grammatical violations that they are not able to consciously identify using explicit knowledge. Consistent with the Competition Model (MacWhinney, 1987, 1997, 2005), Tokowicz and MacWhinney claim that learners experience more difficulty processing features that differ between the L1 and L2, due to conflicting cues and a tendency to rely on L1 processing strategies, as compared to features unique to the L2, which do not invoke competition, or features that are cross-linguistically similar and thus benefit from positive transfer. While the findings support this general conclusion, in the absence of ERP data from a native-speaking control group for comparison, it may be premature to consider this strong evidence that these early learners had “sufficiently learned” gender agreement (a feature not instantiated in their L1) and processed it in a native-like way.

Sabourin and Stowe (2008) also used ERPs to examine the effects of the L1 on L2 implicit processing of different grammatical constructions, yet they adopted a slightly different approach. The authors compared the processing of verbal domain dependency and determiner-noun gender agreement (with an intervening adjective not marked for gender) in L2 Dutch across two advanced-learner groups: L1 German speakers and L1 speakers of a Romance language. While agreement within the verbal domain (the use of participial versus infinitive morphological forms) is comparable across Dutch, German, and the Romance languages, only Dutch and German express gender similarly, as the Romance system is completely different at the lexical
level. Results revealed that while native Dutch speakers exhibited a P600 effect for both constructions, a P600 effect in the L2 Dutch learners was contingent upon L1-L2 similarity. The L1 German speakers demonstrated implicit sensitivity to both constructions like the native speakers, whereas the L2 Romance speakers were only sensitive to violations in the verbal domain. The findings provide support for Tokowicz and MacWhinney’s (2005) conclusion that implicit L2 processing depends on whether the target construction is similarly instantiated in the L1, and dissimilarity between the L2 and L1 construction can impede appropriate processing.

Sabourin and Stowe conclude that L2 learners can achieve native-like processing, at least in situations where there is substantial L1-L2 overlap and learners can employ L1 processing strategies in their L2. Specifically with respect to grammatical gender, the authors suggest that the presence of a gender system in the L1 is not enough for learners to achieve automatic processing of gender agreement, and overlapping lexical gender may be essential. This does not jibe, though, with the results reported by Tokowicz and MacWhinney (2005), where early English-Spanish learners were sensitive to gender violations. This discrepancy could be due to the generally more complex gender system in Dutch as compared to Spanish, the lack of adjacency between the determiner and the noun in Sabourin and Stowe (2008), or to the fact that this study did not examine a feature completely unique to the L2 as Tokowicz and MacWhinney did.

Frenck-Mestre, Foucart, Carrasco, and Herschensohn (2009) shed more light on the role of the properties of the native language in learners’ ability to implicitly process a redundant feature of the L2. They investigated the processing of noun-adjective gender agreement in L2 French by native speakers of German (a gendered language) or English (an ungendered language), all of whom were advanced learners. Results revealed that gender violations resulted
in a P600 effect for native French speakers and L1 English learners, but not for L1 German learners. The authors speculated that this lack of sensitivity for the L1 German group is the result of a cross-linguistic conflict between the way their native language and L2 French mark gender on plurals. Thus, in line with Tokowicz and MacWhinney, the authors proposed that processing is more difficult when features of the L1 and L2 are in competition as opposed to when the feature is unique to the L2, meaning no competition can occur. In a subsequent experiment, the authors also examined the role of input cues to gender agreement on processing. The findings indicated that both the native French control group and the advanced Spanish-French learners showed a robust P600 effect to all gender violations, but the effect was statistically larger for violations involving phonetically variable adjectives (i.e., adjectives marked by overt phonetic cues that distinguish between the masculine and feminine forms) as compared to those involving invariable adjectives without such phonetic cues. Thus, the authors concluded that native-like L2 processing is not limited to grammatical features present in the L1, but that the native language and salient cues in the input influence processing and acquisition.

Keating (2009) further explored the L2 processing of gender agreement in learners whose L1 lacks grammatical gender. Using eye-tracking, the study examined beginning, intermediate, and advanced English-Spanish learners’ ability to process noun-adjective agreement violations across three structural distances: adjectives within the DP (Un libro largo/*a ‘A large book’), adjectives in the VP of the matrix clause (Una película es bastante larga/*o ‘A rather long book’), and adjectives in the subordinate clause (Un refresco tiene muy buen sabor cuando está frío/*a y no caliente ‘A soft drink has a good flavor when it’s cold’). Results indicated that advanced learners, but not beginning or intermediate learners, were sensitive to gender agreement violations within the DP, as evidenced by longer reading times and more total
regressions for ungrammatical adjectives relative to grammatical adjectives. However, unlike the native speakers, the advanced learners were not sensitive to the violations at greater structural distances. Based on these findings, Keating argued that gender agreement is acquirable even for late learners whose L1 lacks this feature, but that it is acquired late and processing limitations inhibit nativelike processing of nonadjacent agreement elements across phrases. While the first part of this conclusion is sound and consistent with previous research, Keating acknowledged that the claim that nativelike processing of gender agreement may not be achievable in the L2 is premature given that individual differences amongst the advanced group suggested that some learners were in fact sensitive to agreement violations in nonlocal domains, possibly the result of higher working memory capacity given that long-distance dependencies are assumed to tax memory resources (e.g., Gibson, 1998).

Wen, Miyao, Takeda, Chu, and Schwartz (2010) aimed to better understand the role that distance effects and proficiency play in explaining nonnative-like processing of nominal agreement. Using self-paced reading, the study looked at intermediate and advanced Chinese-English and Japanese-English learners’ processing of demonstrative-noun agreement violations within simple NPs. In order to disentangle the effects of mere linear distance versus structural (interphrasal) distance, a confounding factor in Keating (2009), the authors manipulated distance by inserting an adjective between the demonstrative and the head noun within a single NP, as in this beautiful house/*houses. The results showed that advanced learners, but not intermediate learners, were sensitive to number agreement violations, and that non-local linear distance between the (dis)agreement elements did not encumber processing. In light of the findings, the authors argue that L2 learners at a certain level of proficiency can show nativelike sensitivity to number agreement, even if it is not instantiated in their L1, and that insensitivity to disagreement
in previous studies is likely due to structural distance and/or complex syntactic structures, both of which the authors claim are more taxing for working memory than non-adjacent items within the same NP. Despite speculations by Wen and colleagues as well as Keating (2009) that working memory affects L2 processing of nominal agreement, neither of these studies empirically tested this.

The role of individual differences in working memory on the processing of noun-adjective gender and number agreement was specifically examined in Sagarra and Herschensohn (2010), which also further explored the effects of proficiency in the L2. The participant pool comprised beginning and intermediate English-Spanish learners as well as a Spanish monolingual control group. The study employed an online non-cumulative self-paced reading task in addition to offline grammaticality judgments to examine the learners’ ability to detect number and gender agreement violations on adjectives adjacent to the nouns they modify. Results indicated that beginning learners were not sensitive to gender or number agreement errors, whereas intermediate learners performed like native speakers (albeit slower) showing sensitivity to both types of errors, evidenced by longer reading times for sentences with disagreement as compared to agreement during self-paced reading. Offline data, but not online data, also demonstrated that learners process number agreement more easily than gender disagreement, suggesting L1 influence since number agreement is instantiated in English whereas gender agreement is not. Furthermore, the working memory results revealed that intermediate learners with higher working memory capacity were more sensitive to gender agreement violations than learners with lower working memory. The authors concluded that late learners can exhibit nativelike processing of gender and number agreement once they attain a certain level of proficiency, but that individual differences in working memory modulate
processing ability. This finding is suggestive that working memory may also account for the individual differences evidenced in Keating (2009) for the processing of nonlocal agreement, which, if found to be the case, could reveal that nativelike processing across all domains is possible for highly proficient L2 learners with sufficient working memory capacity. The potential role of working memory in L2 online processing will be addressed in more detail in Section 2.5.3.1, which provides a review of the research examining the effects of WM on the processing of L2 morphology and syntax.

In sum, the research reviewed in this section suggests that nativelike processing of nominal agreement in an L2 is possible, but that such processing is conditioned by various factors, namely the first language, proficiency, properties of the input, and cognitive individual differences. Evidence is conflicting, though, as to how decisive these factors are in achieving nativelike processing of redundant morphological forms marking number or gender agreement. Before further evaluating the role of the factors highlighted in the current section, it is important to also review studies on the processing of verbal agreement, which is the type of agreement examined in this dissertation.

2.2.2.2 Studies on Processing of Verbal Agreement

2.2.2.2.1 Subject-Verb Agreement

As previously mentioned in Section 2.2.2.1, Jiang (2004) is one of the first online studies to examine late L2 learners’ sensitivity to morphological agreement features during L2 processing. Using a non-cumulative self-paced reading task, the study investigated advanced Chinese-English learners’ ability to process number agreement violations between the subject and the verb in sentences where the head noun and verb were split up by a prepositional phrase, as in ‘The bridge/*s to the island were about ten miles away. Results revealed that while the
native speaker control group was sensitive to the agreement violations, exhibiting significantly longer reading times on ungrammatical constructions, the learners were not sensitive to subject-verb disagreement of this nature. Yet, the learners were sensitive to violations of subcategorization and pronoun-verb agreement with the copular verb as in ‘I told you I/*she am a professor of psychology. Based on these findings, Jiang concluded that the plural –s is a nonintegratable inflectional morpheme that can thus not be implicitly activated during processing, which may be because grammatical number is rarely encoded in the learners’ L1, Chinese. While Jiang’s subsequent research examining nominal number agreement provides support for L1 influence, it is important to also consider the syntactic complexity of the structure used and the role of proficiency. All of the sentences contained nonadjacent (dis)agreement in complex NPs, which, following Keating (2009) and Wen et al. (2010), seem to impose greater processing demands on the learner and may thus explain the learners’ lack of sensitivity. Furthermore, Jiang classified his participants as ‘advanced’ learners, yet their self-ratings were low (scores as low as 2, with averages between 5 and 7 on a 10-point scale), suggesting that at least some of these learners had likely not reached their ‘end state’, and more proficient learners may show sensitivity. Thus, Jiang’s conclusion seems premature in that nativelike processing for Chinese-English bilinguals may be possible, although modulated by properties of the input (distance between agreement elements and syntactic complexity) and proficiency in addition to the role of their L1.

Further exploring these factors, Ojima, Nakata, and Kakigi (2005) used ERPs to investigate intermediate and advanced Japanese-English learners’ processing of subject-verb agreement involving adjacent elements, as in Turtles move/*s slowly. Like the native speakers, both learner groups showed large N400 responses to semantic violations in another condition, yet
the learners differed from the native speakers in their processing of subject-verb agreement errors. For the native English speakers, these agreement violations elicited both a P600 response and a left-lateralized negativity similar to a LAN effect, which has been shown to be another ERP component often elicited by syntactic violations (e.g., Osterhout & Holcomb, 1992), although to a lesser degree than the P600. However, the advanced learners only showed the left-lateralized negativity, and the intermediate learners showed neither effect. The authors interpret the nativelike negativity effect in the advanced learners as indicating sensitivity to agreement violations that approximates nativelike processing. They attribute the lack of a P600 effect in advanced learners to the absence of obligatory number agreement in the learners’ L1, Japanese. Based on the results, the authors underscore the role of proficiency in the processing of verbal agreement, proposing that L2 learners can develop sensitivity to verbal disagreement as proficiency increases. Yet, without evidencing a P600 effect, typically considered critical in determining sensitivity to syntactic violations (Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995), it is debatable to what degree the authors can claim that the advanced learners’ processing is qualitatively similar to that of native speakers.

Chen, Shu, Liu, Zhao, and Li (2007) also used ERP’s to examine sensitivity to subject-verb agreement violations in L2 learners whose L1, Chinese, does not mark subject-verb person agreement. The ERPs of a native English-speaking control group and the advanced L2 learners (classified as “proficient” by the researchers based on standardized tests) were recorded while they read sentences presented word-by-word and then made grammaticality judgments. The sentences comprised a subject noun modified by a prepositional phrase, followed by a verb phrase, as in The price of the car was too high. Only the number of the noun in the prepositional phrase (e.g., car or cars) and the number of the copular verb (e.g., was or were) were
manipulated to create four conditions involving agreement or disagreement between the head noun (*price*) and *was/were* as well as a local match or mismatch (e.g., between *car(s)* and *was/were*). The learners’ achieved high accuracy in their grammaticality judgments (scores between 86% and 90% on the four conditions), indicating that they could detect the subject-verb agreement violations. However, ERP results revealed that these violations did not elicit a P600 or LAN response for the L2 learners, as they clearly did for the native speakers. Yet, unlike the native speakers, the learners showed an N400 response followed by a P600-like effect when processing grammatical sentences that contain a local mismatch (e.g., *The price of the cars was too high*). The authors claim that this effect likely reflects a “limited scope semantic-syntactic analysis” (p. 171), which explains why the learners did not show sensitivity to the long-distance grammatical or ungrammatical agreement relationship between the noun and the verb. Chen et al. attribute the learners’ lack of sensitivity, or delayed sensitivity, to the absence of subject-verb person agreement in the L1, Chinese, concluding that “language-specific experiences help to shape the functional and neural structures of the brain” (171). While L1 influence clearly seems to be at play here, a finding consistent with previous research, it is again important to consider how the syntactic complexity of the NP (shown to present challenges even for native speakers (e.g., Bock & Miller, 1991)) as well as individual differences in cognitive abilities may have affected processing, an issue that will be examined in detail below.

Sato and Felser (2006, 2008) examined processing of subject-verb agreement in advanced L2 learners of English with three different L1 backgrounds (German, Japanese, and Chinese), of which only one of them (German) has subject-verb agreement. To tap the learners’ as well as native English speakers’ online processing, the researches selected a speeded grammaticality judgment task involving word-by-word rapid serial visual presentation of the stimuli. The
sentences were simple, three-word active sentences comprising a subject pronoun, an adverb, and an intransitive verb, as in *She rarely flirt/*s. All of the learners’ grammaticality judgment accuracy rates were well above chance (Germans: 93.4%; Japanese: 83.8%; Chinese: 62.8%), yet their reaction times were significantly slower than native speakers’ to accept grammatical sentences and reject ungrammatical ones, suggesting that they are slower processors, which is in line with previous studies. Moreover, L2 proficiency (measured by the Oxford Placement Test) correlated significantly with overall performance, where the high-proficiency Germans patterned with natives with their near-ceiling accuracy for their judgments. Yet, because all learner groups identified case violations (another condition) more accurately and faster than agreement violations in spite of the fact that case is instantiated in German and Japanese, but not Chinese, and subject-verb agreement is only present in German, the researchers concluded that the role of L1 transfer for L2 morphosyntactic processing is more limited than presumed. They speculated that bound verbal inflections are generally more difficult to process than free morphemes. There are methodological issues with the study, though, that make it difficult to entirely dismiss L1 influence and the possibility of nativelike implicit processing. First, unlike the previous studies reviewed, the task used here does not provide information regarding how the participants responded to individual words and violations during the act of processing, which renders it impossible to compare how the learners and native speakers responded to the critical word containing the agreement violation (which was also the final word, thus inhibiting natural delayed processing effects). Moreover, although the researchers incorporated a time constraint and word-by-word presentation of the stimuli to “maximize the chance of obtaining participants’ unreflected judgments” (p. 6), participants may still have been able to use conscious knowledge
of agreement for their responses, particularly given that the task focus directed them to attend to form as opposed to meaning.

In a longitudinal study, Osterhout, Poliakov, Inoue, McLaughlin, Valentine, Pitkanen, Frenck-Mestre, and Herschensohn (2008) used ERPs to examine novice English-French learners’ processing of verbal and nominal inflection during their first year of formal French instruction. They specifically investigated how L1-L2 similarity and the phonological realization of inflectional morphemes (versus morphemes marked orthographically but phonologically silent) affected learners’ processing after one, four, and eight months of instruction. In addition to a semantic condition, their stimuli included two morphosyntactic conditions: 1) subject-verb agreement (present in both L1 English and L2 French) with a phonologically realized verbal inflection, as in *Tu adores/*adorez le français. ‘You[2-sg] adore[2-sg]/*adore [2-pl] French’, and 2) determiner-number agreement (present in L2 French but not L1 English) with a phonologically silent nominal inflection, as in *Tu manges des hamburgers/*hamburger pour diner. ‘You[2-sg] eat[2-sg] some[pl] hamburgers[pl]/*hamburger[sg] for dinner’. After reading the sentences, presented word by word, the participants were instructed to make a grammaticality judgment. In that the results revealed striking individual differences across the learners for the judgment-accuracy data as well as the ERP data, the researchers divided the learners into two groups, “fast learners” and “slow learners”, and then just reported the results for the “fast learners”. Like the native speakers, these learners showed a robust N400 effect to semantically anomalous words at all three testing sessions. With respect to the subject-verb condition, the learners exhibited a response to agreement violations after just one month of instruction. However, these morphosyntactic violations elicited an N400-like effect rather than the P600 effect seen for the native French speakers. After four months of instruction, the N400 response
was replaced with a P600-like effect for subject-verb agreement violations, suggesting that the learners could implicitly process this type of disagreement. In contrast, no such effect was found, even after eight months of instruction, for determiner-noun disagreement. For this condition, there were no significant differences in the ERP responses between grammatical and ungrammatical sentences, indicating that the L2 learners did not develop nativelike sensitivity to nominal number disagreement. The researchers concluded that L1-L2 similarity in conjunction with phonological realization of the target inflectional morpheme led to fast L2 morphosyntactic learning, resulting in nativelike processing of subject-verb agreement. However, the opposite scenario led to poor learning and no evidence of sensitivity to nominal number agreement. As Osterhout et al. recognize, though, it is impossible to disentangle the two factors, L1 influence and phonological realization of inflections, to establish their relative impact on processing and acquisition. Furthermore, these learners were clearly not at an end state of learning, so they may still develop sensitivity to nominal number agreement.

2.2.2.2 Adverb-Verb Tense Agreement

Using a non-cumulative self-paced moving window task, Sagarra (2007b) examined third-semester English-Spanish learners’ processing of redundant verbal morphology for temporal reference, including the potential role of working memory. Specifically, Sagarra compared learners’ processing of adverb-verb tense agreement and adverb-verb tense disagreement with experimental sentences containing a temporal adverb denoting past as well as a regular Spanish verb conjugated in either the past (preterite) or present tense of the third person singular. After reading each sentence word by word, the participants responded to a yes-no comprehension question to ensure they were attending to meaning. The results revealed no significant differences between mean reading times on verbs in the tense agreement condition.
and verbs in the tense disagreement condition, indicating that these L2 learners were not sensitive to the tense violations. However, working memory results showed significant differences between the processing ability of high-span learners and low-span learners, where high span learners were sensitive to tense disagreement more often than the low span learners. These findings suggest that while learners at this low level of proficiency (third semester of formal instruction) generally do not appear to process redundant morphology, learners with higher working memory capacity are more successful at processing these grammatical morphemes. This research underscores the importance of working memory in facilitating simultaneous processing of sentential meaning and redundant grammatical forms, and it also points to the potential obstacle the overuse of lexical cues may have for the processing of grammatical cues, two findings that support VanPatten’s Input processing model (1996, 2004, 2006). However, two limitations of Sagarra’s experimental design prevent firm conclusions with respect to learners’ reliance on lexical cues (adverbs) as opposed to redundant morphological cues (inflections) in processing. First, the temporal adverbs were located in the most salient position in the sentence (sentence-initial) whereas the verbs were in the least salient place (mid-sentence) (see, e.g., VanPatten, 2004), which leaves open the possibility that learners may show more sensitivity to tense disagreement and enhanced processing of morphological cues if they are in sentence-initial position. Moreover, because self-paced reading prohibits regressions to previous words in the sentence, it is not possible to determine what cue the learner relies on when they do notice a tense conflict between the lexical item and the inflectional morphology. Both of these issues are addressed in the following studies, all of which control for the position of the cues and measure cue reliance through online processing with eyetracking.
N. Ellis and Sagarra (2010b) used eyetracking to explore how English-Spanish learners, English monolinguals, and Spanish monolinguals process lexical cues (adverbs) and morphological cues (verbal inflections) to temporal reference. The beginning learners (third-semester), intermediate learners (eighth-semester), and Spanish monolinguals read sentences in Spanish while their eyes were being tracked, whereas the English monolinguals read them in the English translation. After each sentence, the learners answered a yes-no comprehension question and received feedback on their responses. The experimental sentences included the following four conditions: past adverb-past verb, *past adverb-present verb, past verb-past adverb, *present verb-past adverb. Analyses were based on a comparison across the conditions of mean reading times for gaze duration (sum of all fixations on a word before moving on) and total time (sum of all fixations on a word, including regressions) on the verb and adverb. Results revealed that the beginning learners showed no differences in reading times between the congruent and incongruent condition whereas the intermediates did, indicating that only the intermediates realized (or were implicitly sensitive to) the tense disagreement, a finding consistent with previous studies. More interestingly, for incongruent sentences as compared to congruent sentences, Spanish monolinguals showed longer reading times at the verb, whereas both English monolinguals and intermediate learners showed longer reading times at the adverb. These findings illustrate that when there is an adverb-verb or verb-adverb tense conflict, native speakers of a morphologically rich language (Spanish) rely more on verbal inflections to resolve the conflict, whereas native speakers of a morphologically impoverished language (both English monolinguals and intermediate English-Spanish learners) clearly prefer to use the adverb.

The results also showed, though, that the intermediate learners regressed less overall to the adverb as compared to the beginners, suggesting that with more experience in the target
language, the learners began to adjust their processing strategies to rely less on lexical cues and more on morphological ones. The researchers conclude that the findings demonstrate long-term learned attention effects, where L1 reliance on lexical cues in a morphologically impoverished language biases attention to these same cues in the L2, thus blocking the acquisition of morphological cues in an L2 that more extensively uses these as compared to lexical cues. Although the findings are consistent with previous research demonstrating the crucial role of L1 influence and proficiency in L2 processing, there were two limitations to this study that leave some unanswered questions. First, without groups of diverse L1 backgrounds (differing in their morphological richness), it is not possible to ensure that the L1 English participants’ behavior was not due to a general L2 tendency to rely on lexical cues. A subsequent study by the same researchers and colleagues (reviewed below) directly addresses this limitation. Second, the decreased reliance on lexical cues by the intermediate learners (eighth-semester), as compared to the beginners (third-semester) suggest that learners can begin to overcome their L1 processing strategies with a sufficient amount of L2 experience, yet, since many of the intermediate learners had studied abroad, it is unclear whether formal classroom instruction is ‘sufficient’, or if intensive exposure to the L2 in an immersion setting is critical. The next two studies reviewed each tackle one of these limitations.

Using a similar experimental procedure with eye-tracking, Sagarra, N. Ellis, Hanson, and Gauthier (2010) compared the L2 processing of lexical and morphological cues in a morphologically rich language (Spanish) by late learners whose L1 is morphologically weak (English) or morphologically rich (Romanian). The two intermediate-level learner groups, who were matched for proficiency, read sentences in Spanish containing an adverb and a morphological cue to temporal reference, and then they chose the picture (out of four) that best
depicted the meaning of the sentence. Like the previous study, the experimental sentences were either grammatically congruent or incongruent, and the adverb was located either pre- or post-verbally. The main difference with respect to the stimuli is that this study avoided sentence-initial adverbs by incorporating a present-tense verbal phrase such as *Dicen que* ‘They say’ prior to the first target adverbial or verbal cue. The results showed no significant differences for reading times on the verb between the congruent and incongruent condition, yet total time on the adverb was significantly greater for incongruent sentences when the adverb was preverbal. This indicates both learner groups relied more on the lexical cue to resolve tense conflicts when it appeared before the morphological cue. However, analyses of the total number of regressions reveal different processing behaviors between the two groups. Similar to the Spanish monolinguals in N. Ellis and Sagarra (2010b), in the verb-adverb condition the Romanian-Spanish learners regressed more to the verb in the incongruent condition as compared to the congruent condition. Contrariwise, the English-Spanish learners regressed more to the adverb in this condition, indicating that this is clearly their preferred cue. In short, while the learners coming from a morphologically poor L1 always relied on the adverb when there was a tense conflict, learners coming from a morphologically rich L1 only relied on the adverb when it was the first cue they came across, likely due to its salience. Thus, the findings provide further support for L1 transfer effects in L2 processing of lexical and morphological cues, yet they also underscore the influence of linguistic factors, namely cue salience.

LaBrozzi (2009, 2011) also carried out an eye-tracking experiment to investigate the L2 processing of lexical and morphological cues for temporal reference, with the specific aim of exploring how immersion experience and cognitive abilities influence English-Spanish learners’ processing of redundant morphological cues in L2 Spanish. Thus, the sample pool consisted of
two groups of intermediate-level English-Spanish learners, matched in proficiency, who began learning Spanish after puberty: classroom learners with and without study-abroad experience. The same four experimental conditions as the previous studies were used, yet, the stimuli and procedure replicated that of N. Ellis and Sagarra (2010b) in that the adverb was in sentence-initial position in the two conditions were it preceded the verb, and the sentences were followed by yes-no comprehension questions as opposed to pictures. Reading times for first gaze duration on the adverb and verb across all conditions indicated that the learners were not initially sensitive to a tense incongruency between the cues. However, a comparison of the reading times for the adverb + 1 (i.e., the adverb and the following word) and the verb + 1 in the congruent and incongruent conditions showed evidence of delayed processing, suggesting that the L2 learners did notice the error, but often not until after leaving it. These first two analyses revealed similar processing behavior for both learner groups, yet total reading times on the adverb and verb uncovered differences between them. While the classroom learners attended to the adverb to resolve a tense conflict regardless of whether it was pre- or post-verbal, the position of the adverb was important for the study abroad learners, as they spent significantly less time on the adverb when it preceded the verb, thus rendering it the farthest cue to return to when they encountered an incongruency. Total reading times on the verb revealed that only the study abroad learners spent more time on the verb in the incongruent conditions as compared to the congruent ones. These findings illustrate that while the classroom learners unconditionally relied on the lexical cue, the study abroad group relied on both cues. LaBrozzi proposes that the study abroad learners are in a transitional phase moving away from relying primarily on lexical cues, as they do in their L1, and beginning to develop native-like L2 processing strategies. Working memory was also found to enhance processing of the redundant morphological cues for the study
abroad group, but this cognitive individual difference had no effect on the classroom learners. LaBrozzi concludes that immersion experience alone, and particularly in conjunction with high working memory capacity, can facilitate L2 processing of morphological cues. While clearly important, the extent to which these factors are beneficial cannot be determined since the researcher used a between-subjects design with two separate groups of learners, rather than examining the same group of participants before and after immersion experience.

The studies in this section have shown that late L2 learners can exhibit online sensitivity to tense conflicts, suggesting that they are able to process redundant verbal morphology for temporal reference. However, such processing is qualified by proficiency, as only rather proficient intermediate learners (seventh- to eighth-semester), and not relative beginners (third-semester), realized that there was a conflict between the adverb and the verb. Nevertheless, the eye-tracking studies revealed that although the intermediate learners were sensitive to agreement violations, they differed from the native speakers with respect to the cues, or processing strategies, they used to resolve the conflict. While the native speakers of a morphologically rich language (Spanish monolinguals and Romanian-Spanish learners) relied more on verbal inflections, the adverb was clearly the preferred cue of the native speakers of a morphologically poor language (English monolinguals and late English-Spanish learners). Only LaBrozzi’s intermediate learners who had studied abroad demonstrated more native-like cue reliance, evidenced by increased attention to inflections and less dependence on adverbs, although this processing behavior was also affected by working memory. These findings provide evidence, contributing to VanPatten’s Input Processing model, that late learners’ overreliance on lexical cues and decreased attention to morphological cues may be influenced by L1 transfer and
modulated by working memory. The role of working memory in L2 processing and learning will be addressed in detail in Section 2.5.3.

2.2.2.3 Summary of Online Research

Before any strong conclusions can be drawn, clearly more research on L2 morphological processing in late learners is necessary. However, taken as a whole, the results from the above online studies examining nominal and verbal agreement strongly suggest that adult L2 learners can process redundant L2 morphology in a way that is qualitatively similar to native speakers, albeit likely slower, as reduced processing speed was widely reported even when learners exhibited nativelike sensitivity (e.g., Sagarra & Herschensohn, 2010; Sato & Felser, 2006, 2010; Tokowicz & MacWhinney, 2005). Crucially, though, achieving such nativelike processing is conditioned by certain factors, namely L1 influence, proficiency, linguistic factors, and cognitive abilities. With respect to L1 influence, evidence from the reviewed studies (Frenck-Mestre et al., 2009; Jiang et al., 2011; Osterhout et al., 2008; Sabourin & Stowe, 2008; Tokowicz and MacWhinney, 2005) indicates that L1-L2 morphological similarity facilitates L2 processing for late learners, a finding consistent with other types of research pointing to a morphological congruency effect (grammaticality judgment studies: e.g., Bialstock, 1997; Sabourin, Stowe, & de Haan, 2006; production studies: e.g., Guillelmon & Grosjean, 2001; Hawkins & Liszka, 2003). The research does not, however, provide uniform consensus on whether late learners can show nativelike sensitivity to L2 features that are either absent or formed differently in their L1. Jiang (2004, 2007), Jiang et al. (2011), and Chen et al. (2007) suggest that learners may not be able to achieve nativelike abilities for processing morphemes that are absent in their L1, whereas Tokowicz and MacWhinney (2005) and Frenck-Mestre et al. (2009) indicate that they can, but only if the feature is unique to the L2 and will therefore not invoke detrimental L1-L2
competition (also shown to impede processing in Sabourin and Stowe (2008) and Osterhout et al., (2008)). Keating (2009), Sagarra and Herschensohn (2010), and Wen et al. (2010) provide further evidence that L2 learners can process L2 features absent in the L1, but only at higher levels of proficiency, and even then processing was qualified by linguistic and cognitive factors.

Proficiency level is shown to be an important factor determining implicit sensitivity to agreement violations in various studies reviewed (N. Ellis & Sagarra, 2010b; Keating, 2009; Ojima et al., 2005; Osterhout et al., 2008; Sagarra & Herschensohn, 2010; Sato & Felser, 2006, 2010; Wen et al., 2010). By assessing L2 proficiency independently or testing different proficiency groups (using a within-subject longitudinal design in the case of Osterhout et al., 2008), these studies illustrated that nativelike processing is contingent upon having reached a certain level of proficiency (intermediate to advanced), as less proficient learners were completely insensitive to violations of number agreement (Sagarra & Herschensohn, 2010; Wen et al., 2010), gender agreement (Keating, 2009; Sagarra and Herschensohn, 2010), subject-verb agreement (Ojima et al., 2005; Osterhout et al., 2008), and adverb-verb tense agreement (N. Ellis & Sagarra, 2010b). Moreover, while both groups of LaBrozzi’s (2009) intermediate learners showed sensitivity to tense agreement violations, experience with the second language in an immersion setting, entailing more extensive exposure to morphological cues, was the critical factor associated with more nativelike reliance on verbal morphology to resolve grammatical conflicts. This suggests that L2 processing of morphology may hinge more on language experience (particularly with the target feature), as opposed to on general proficiency per se. This factor could also potentially explain why some studies, such as Jiang (2004, 2007; Jiang et al., 2011), did not show learner sensitivity to agreement violations. As mentioned previously, Jiang classified his participants as ‘advanced’, yet their self-ratings were rather low (scores as low as 2,
with averages between 5 and 7.6 on a 10-point scale), suggesting that at least some of these learners had likely not reached their ‘end state’, and more proficient learners may in fact show sensitivity.

Linguistic or input characteristics were also found to have a robust effect on L2 processing of redundant morphology. Frenck-Mestre et al. (2009) and Osterhout et al. (2008) showed that overt phonetic cues enhanced processing of inflectional morphology, as evidenced by increased sensitivity to morphemes that are phonetically realized as opposed to those that are silent and vary only in orthography. This finding is likely due to the fact that phonetic cues, which are available in spoken input and can be activated during silent reading, make morphemes more salient and less ambiguous, rendering them a more reliable cue to agreement (Frenck-Mestre et al., 2009). The other linguistic factor shown to modulate processing is the distance between the (dis)agreement elements. Keating (2009) found that highly proficient learners only exhibited nativelike processing to adjacent (dis)agreement within the NP, but not across phrases, which he took as evidence that processing limitations restrict the use of implicit knowledge to either adjacent agreement or to local domains. By manipulating linear distance, Wen et al. (2010) demonstrated that advanced learners do exhibit sensitivity to nonadjacent items within simple NPs. This suggests that insensitivity to agreement violations in certain studies (Jiang, 2004, 2007; Jiang et al., 2011; Chen et al., 2007) may be due to structural distance or the use of syntactically complex NPs, which pose an additional processing load on learners’ limited cognitive resources and may thus overtax their working memory capacity.

While many studies speculated that working memory might have significant effects on the processing of redundant L2 morphology (e.g., Keating, 2009; Wen et al., 2010), few studies have empirically investigated this hypothesis. However, all three of the reviewed studies that did
examine the role of individual differences in working memory (Sagarra, 2007b; Sagarra and Herschensohn, 2010; LaBrozzi, 2009) found that this factor had a robust effect on processing. In fact, working memory capacity was shown to be the critical factor determining whether low-proficient learners exhibited sensitivity to adverb-verb tense disagreement (Sagarra, 2007b), a finding Sagarra replicated for gender disagreement in another study (2007a). In addition, for intermediate learners working memory capacity was shown to facilitate gender agreement processing (Sagarra & Herschensohn, 2010) as well as enhance processing of verbal morphology, while reducing overreliance on lexical cues (LaBrozzi, 2009). In light of the facilitative role higher working memory capacity had on L2 processing of redundant morphology in these studies, this variable may also help explain individual differences reported in studies indicating a lack of sensitivity to agreement violations (e.g., Chen et al., 2007; Keating, 2009), perhaps even uncovering exceptional high-span learners that exhibited nativelike processing.

The findings from this online research on L2 processing of nominal and verbal agreement can also make important contributions to VanPatten’s model of Input Processing (1996, 2004, 2006). First, the evidence demonstrating a role for working memory in the processing of redundant inflectional morphology (LaBrozzi, 2009; Sagarra, 2007a, 2007b; Sagarra & Herschensohn, 2010) provides empirical support for VanPatten’s claim, at the heart of the model, that L2 processing of redundant grammatical forms is constrained by working memory limitations. The benefits of high working memory capacity seen in these online studies accord with the model’s prediction that, for L2 learners to process redundant grammatical forms, the processing of overall sentential meaning must not tax their working memory and use up all of their available cognitive resources (2004, 2006). Nevertheless, more research is clearly needed on the role of working memory in this specific domain because some studies do not show
working memory to have an effect on L2 online processing, as will be evidenced in Section 2.5.3.1.

Next, VanPatten’s model and related Input Processing research identify proficiency as a favorable factor for processing (Lee, 1999; Lee et al., 1997; VanPatten, 1990; VanPatten, 1996, 2004), as increased vocabulary size and ability to retrieve lexical items makes overall comprehension less laborious. As discussed above, the online research also pointed to favorable effects of proficiency (or language experience) for L2 morphological processing (e.g., Keating, 2009; Sagarra & Herschensohn, 2010; Wen et al., 2010). Although most of the online studies did not specifically investigate learners’ cue preferences, their lack of sensitivity to grammatical violations involving different morphological forms, particularly at beginning and low-intermediate stages (e.g., Keating, 2009; Sagarra & Herschensohn, 2010; Wen et al., 2010), is in line with the *Lexical Preference Principle* and offline research demonstrating learners’ reliance on lexical items over grammatical forms (e.g., Lee et al., 1997; Musumeci, 1989; Rossomondo, 2003; VanPatten, 1996, 2004, 2006). The strongest support for the *Lexical Preference Principle* comes from the eye-tracking studies (N. Ellis & Sagarra, 2010b; LaBrozzi, 2009) that demonstrated, by examining regressions, a preference for lexical items on the part of L2 learners. However, the fact that these learners all came from a morphologically weak L1 and other online studies point to a critical role for the L1—either facilitating or hindering L2 processing (e.g., Jiang et al., 2011; Sabourin & Stowe, 2008; Sagarra et al., 2010; Tokowicz & MacWhinney, 2005) – strongly suggests that learners do not all adopt the same universal strategies, as the Input Processing model originally proposed (VanPatten, 1996, 2004). Thus, to account for this recent online evidence, the model must adopt and expand upon the tentative *L1 Transfer Principle* that VanPatten included in his latest update on Input Processing (2006). The only specific Input
Processing principle discussed earlier that the online research does not support is the *Meaning before Nonmeaning Principle*. There was no direct evidence that inflectional morphology carrying meaning (e.g., morphemes marking plurality on nouns, morphemes marking number or tense on verbs) was universally processed sooner or easier as compared to nonmeaningful inflections (e.g., morphemes marking grammatical gender on nouns or adjectives). Yet, no study set out to directly explore this issue, and inconsistent as well as confounding factors across the online studies, including the presence or absence of the feature in the L1 and divergent proficiency levels, render it impossible to assess whether all learners tend to process meaningful morphological forms before nonmeaningful ones.

Taken together, the research reviewed indicates that L1 transfer, proficiency (language exposure), linguistic characteristics of the input, and cognitive resources modulate L2 morphological processing. These same factors are integral to usage-based perspectives to second language processing and acquisition, specifically the Associative-Cognitive model (N. Ellis, 2006a, 2006b, 2006c), which is the framework adopted for the current dissertation. The following section will illustrate how this model rather impressively accounts for the evidence from the above studies by drawing on linguistic characteristics, cognitive abilities, and language experience to explain late learners’ difficulty with L2 morphological processing.

### 2.3 Associative-Cognitive Model: Attentional Biases in L2 Processing and Acquisition

Usage-based models hold that language processing and acquisition are shaped by experiences with language input (e.g., Bybee, 2008; N. Ellis, 2002, 2008; Tomasello, 2003). In this perspective, the basic units of language representation are constructions that map linguistic form (morphological, syntactic, lexical) with particular functions (semantic, pragmatic, discourse). Thus, the acquisition of grammar entails the piecemeal learning of numerous
constructions, or form-meaning mappings, through actively engaging in communication (Barlow & Kemmer, 2000; N. Ellis, 2002, 2006b). This process is an “intuitive statistical learning problem” (N. Ellis, 2006, p. 26) governed by domain-general cognitive abilities and associative learning that reflects the probabilities of occurrence of form-function mappings. In particular, frequency (type and token), salience, contingency, and cue competition have a robust influence on processing and acquisition of constructions (Bybee, 2008; N. Ellis, 2002, 2006b, 2006c, 2008). Moreover, in that all late L2 learners have necessarily already learned a first language, usage-based models posit that prior experience with the L1 will have discernible effects on L2 acquisition (Bybee, 2008). On this issue, the Associative-Cognitive model (N. Ellis, 2006a, 2006b) has great explanatory power. It offers a thoroughgoing account of how a lifetime of prior L1 usage explains difficulties with aspects of L2 grammar, even if exposure and the linguistic characteristics of the input are identical for both L2 and L1 acquisition.

Drawing on the same associative and cognitive principles that govern other types of learning, the Associative-Cognitive model accounts for the challenges associated with adult L2 acquisition in terms of attentional processes in the associative learning of constructions (N. Ellis, 2006a, 2006b, 2006d). More specifically, this theory posits that the following factors modulate the processing and acquisition of grammatical items: limitations in cognitive resources, linguistic characteristics of the input, and language experience. These factors will be examined in turn, with particular attention to their role in the processing and acquisition of L2 inflectional morphology.

Successful acquisition of grammar, including morphological forms, requires attention to both meaning and form (Doughty, 2001; Doughty & Williams, 1998; N. Ellis, 2011; Long, 1991; Long & Robinson, 1998). However, making form-meaning connections is cognitively taxing,
and given working memory limitations (Baddeley, 2003; Just & Carpenter, 1992), learners have
to select which aspects of the input they will process (Gass, Svetics, & Lemelin, 2003;
MacWhinney, 2008; Sagarra, 2007; VanPatten, 2006). As mentioned above, the Associative-
Cognitive model maintains that L2 cue selection, or which forms learners process, is guided by
linguistic characteristics and language experience factors.

Beyond raw frequency of occurrence, the principle linguistic factors that affect the
processing of L2 forms are salience, reliability, and redundancy (N. Ellis, 2006a, 2006b, 2006d).
Salience refers to the perceived strength of the cue, or its “physical magnitude and psychological
significance” (N. Ellis & Sagarra, 2010b, p. 87). Reliability pertains to the power of a cue to
predict a specific outcome or the “contingency of form-function mapping” (p. 87), where a fully
reliable cue is one that always correctly predicts that outcome, although it may or may not
always be available (MacWhinney, 2008). Redundancy relates to the amount of similarity
between two cues within the discourse context, where a redundant cue is one that is accompanied
by another cue that expresses the same meaning.

These linguistic factors have a robust effect on the processing and acquisition of
inflectional morphology. Nominal and verbal inflections tend to be of low salience in the
language speech stream. Even in slow and careful speech, bound inflections tend to be short and
low in stress (Bates & Goodman, 1997; N. Ellis, 2006b, 2006d, 2008), rendering them difficult
to perceive in fluent speech, particularly for learners (N. Ellis, 2006d, 2008). Moreover, these
grammatical morphemes tend to become more phonologically fused with surrounding language,
especially in informal and rapid speech, because their frequent use leads to reduction processes
(Bybee, 2003). These processes also explain in part the low reliability of many morphological
cues, since shorter function words and inflections tend to be more homophonous and therefore
more ambiguous in their interpretation (consider the unreliable mappings between –s and its interpretations as plural, third person plural, or the contracted form of the copula ‘be’) (N. Ellis, 2006b, 2006d, 2008). In addition, the plurifunctionality of inflections (i.e., inflections can encode multiple types of information including person, mood, tense, aspect) renders them less reliable.

Reduced attention to morphological cues is exacerbated when they compete with more salient and reliable lexical cues that convey the same meaning (e.g., *yesterday I played*). When there is cue redundancy, learners will often only attend to one of the two cues, especially in the early stages of SLA when cognitive resources and knowledge of the L2 are more limited (MacWhinney, 2008; VanPatten, 2006). Since lexical items, such as adverbs and nouns, are much more salient and often more reliable than inflections, as discussed previously, learners are more likely to interpret the lexical cue for the grammatical meaning and ignore the morphological cue (N. Ellis, 2006a; VanPatten, 1994, 2004, 2006).

The low salience, low reliability, and frequent redundancy of many morphological cues make them more difficult to attend to and process, and thus likely contributes to the challenges late learners experience in acquiring L2 inflections (N. Ellis, 2006a, 2006b; Goldschneider & DeKeyser, 2001; Tomasello, 2003; Zobl & Liceras, 1994). However, since these linguistic factors affect both L1 and L2 acquisition, there must be something else that explains why children acquiring their L1 learn to process and produce inflections relatively quickly (particularly in morphologically rich languages such as Spanish (e.g., Austin, 2010)), whereas L2 learners typically do not, with L2 acquisition frequently stabilizing at an end state that falls far short of native-like ability in this domain (Bardovi-Harlig, 2000; N. Ellis, 2001; Klein, 1998). According to the Associative-Cognitive model, what makes the acquisition of L2 morphological
cues particularly difficult for adult learners is their prior language experience as it affects their attention to the second language (N. Ellis, 2006a, 2006b, 2006d).

The Associative-Cognitive model describes how the general associative learning phenomenon of ‘learned attention’, including blocking, overshadowing, and other effects of transfer and interference, applies to language learning and can account for adults’ difficulty acquiring foreign languages (N. Ellis, 2006a, 2006b, 2006d). Research demonstrates that once an individual has learned that a particular stimulus (A) is associated with a particular outcome (X), it becomes harder to later learn that a different stimulus (B) also predicts this same outcome because learners’ attentional biases adhere to the perfectly predictive earlier-learned stimulus (A) (N. Ellis, 2006c, 2008; Kruschke & Blair, 2000). In relation to language acquisition, N. Ellis and Sagarra (2010b) define ‘learned attention’ as “the shifting of a learner’s attention to certain aspects of the linguistic input as a result of language experience” (p. 87).

Learned attention can result in overshadowing and blocking, and these mechanisms play an important role in determining which cues L2 learners attend to in the input (N. Ellis, 2006d). The phenomenon of overshadowing comes about when two cues are presented together and they both predict the same outcome. The strength of conditioning to each cue depends upon their salience and relative reliability, with the more salient and reliable cue becoming associated with the outcome and the less salient one being overshadowed and thus going relatively unnoticed. Over time, overshadowing results in blocking, or learned inattention to certain elements in the input (N. Ellis, 2006d).

These phenomena of learned attention have a greater effect on L2 acquisition than L1 acquisition, whereby blocking in the L2 often results from redundant L2 cues being overshadowed due to prior L1 experience (N. Ellis, 2006b, 2006d). For instance, forms of low
salience in the L2 like inflections may be blocked because learners’ L1 experience guides them to look to other cues for interpretation. According to Kruschke and Blair (2000) and N. Ellis (2006d), blocking is due to an automatically learned inattention that can be pervasive and enduring, thus impairing further learning about the blocked cue. Once a learner knows from previous experience that cue (A) is a reliable predictor of (X), cue (B) is seen as merely a distraction from a perfectly predictive cue (A). To avoid potential error by attending to this distraction, the learners shift their attention away from cue (B) to cue (A), and as a result only develop a weak association from (B) to (X). In addition, they learn that when cue (B) and (A) appear together, they should ignore (B) and attend to (A). That is, they learn to attentionally block cue (B) (N. Ellis, 2006d).

In short, according to the Associative-Cognitive model, the phenomenon of blocking accounts for why some cues like typically redundant inflections are less readily acquired by L2 learners (N. Ellis, 2006a, 2006b, 2006d). L1 experience has led L2 learners to rely on certain cues, and they will likely use these cues whenever available in the L2 (i.e., initial L2 cue settings tend to closely match those of the L1) (N. Ellis, 2006a; MacWhinney, 2008). Reliance on these L1 cues can be to the detriment of learning other important cues in the L2, such as nominal and verbal morphology. As a result of their L1 experience, regardless of the language, adult learners know, for instance, that there are reliable and salient lexical cues to express time (adverbs like yesterday, ayer, hier, gestern, dün), and if frequently present in the L2 input, these known cues may block the acquisition of temporal morphology. For example, upon hearing *Yesterday they worked*, the morphological tense marker *-ed* is redundant, meaning that it does not need to be processed for successful interpretation of temporal reference, and persistent lack of processing means lack of acquisition (N. Ellis, 2006a, 2006b, 2006c). Similarly, verbal inflections for
person and number are often overshadowed by the more salient explicit subject of the verb (e.g., *The girl eats pizza on Fridays*) (N. Ellis, 2006c). This also applies to nominal inflections, where plurality and gender can be redundantly marked on many forms including articles, nouns, pronouns, and adjectives in languages like Spanish (e.g., *Esta alumna es muy aplicada* ‘This student is very diligent’; *Visitamos a los niños enfermos* ‘We visit the sick children’).

Despite the fact that inflections are the more available cue in many morphologically rich languages, considering that temporal adverbs are not always present and nor are explicit subjects in null subject languages (e.g., Spanish, Russian, Greek), many late L2 learners still initially rely predominately on the simpler, more salient lexical cues to interpret the message (whenever they are available) and to communicate their own message. This is not the case for children acquiring their native language, as they prefer available cues to reliable cues (MacWhinney, 2008). For example, children initially rely on verbal over adverbial cues to time, not acquiring the meanings of temporal adverbs until rather late in development (e.g., N. Ellis, 2006b; Smith, 1980; Valian, 2006). As children get older, their cue settings will adapt if necessary to match the adult pattern, where the strongest cue is the most valid one, which is based not only on its availability, but also, crucially, on its reliability in that language (MacWhinney, 2008).

Since L2 learners perceive the L2 through the L1-tuned processor, the degree of L1-L2 similarity can either aid or encumber SLA (N. Ellis, 2005, 2006c; MacWhinney, 2008), as clearly shown in the empirical research previously discussed (e.g., Jiang et al., 2011; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005). Thus, with respect to the acquisition of inflectional morphology, the degree to which learners’ native language makes extensive use of inflections (i.e., how available and reliable they are) affects sensitivity to these cues in the L2.
The eye-tracking study reviewed above (Sagarra et al., 2010) provided evidence of this phenomenon in intermediate English-Spanish and Romanian-Spanish learners, where native speakers of morphologically rich Romanian relied significantly more on verbal inflections to resolve tense conflicts than English speakers. Moreover, several recent laboratory-learning studies by N. Ellis and colleagues offer additional support for learned attention effects in the acquisition of L2 morphology.

N. Ellis (2007) and N. Ellis and Sagarra (2010a) investigated the short- and long-term effects of learned attention in the acquisition of temporal reference in a small subset of Latin. For the first experiments in both studies, the participants (native speakers of English, Chinese, and Korean in N. Ellis (2007); all native English speakers in N. Ellis & Sagarra (2010a)) were randomly assigned to one of the three groups: Adverb Pretraining, Verb Pretraining, and Control. In Phase 1 (pretraining), the Adverb Pretraining participants learned two Latin adverbs and their temporal reference referring to present or past while the Verb Pretraining group learned two verb forms, one marking present and one marking past. The control group received no pretraining. In Phase 2 (training), all participants were exposed to sentences with appropriate adverb-verb combinations (also including new future adverb and verb forms), and they had to decide whether the sentences referred to the past, present, or future tense. They received computerized feedback if incorrect. In Phase 3 (reception test), all combinations of the three adverbs and three verb forms were combined and the participants were instructed to decide whether each sentence referred to the past, present, or future on a 5-point scale ranging from extreme past to extreme future. In Phase 4 (only conducted in N. Ellis and Sagarra, 2010a), the participants translated sentences from English to Latin that required the production of the adverbs and verbs seen. The results revealed that when the adverbial and verbal cues conflicted in Phase 3, the Adverb-
Pretraining group relied on the adverb whereas the Verb-Pretraining group relied on the verbal cue. The control group did not favor either cue overall, but rather attended equally to both cues in these conflict situations. Phase 4 replicated these cue-reliance patterns. While the Adverb-Pretraining group exhibited high accuracy in adverb production and low accuracy in verb production, the Verb-Pretraining group showed the opposite pattern, and the control group produced both cues at similar accuracy levels. The authors interpreted these findings as support that early-learned cues in instruction block the acquisition of later-experienced cues conveying the same meaning.

Having illustrated that short-term learned attention effects from instruction bias cue acquisition, N. Ellis (2007) and N. Ellis and Sagarra (2010b) conducted subsequent experiments to investigate whether long-term learned attention effects based on L1 experience also bias cue acquisition. The motivation for these latter experiments is the viewpoint of usage-based approaches that short-term effects lead to long-term effects (Barlow & Kemmer, 2000; N. Ellis & Sagarra, 2010b). All of the participants in these latter experiments were advanced L2 English learners with L1 Chinese (a language without any verb tense morphology). The participants followed the same procedure as the control groups in the previous experiments (no pretraining). These L1 Chinese participants showed preference for the adverbial cue, performing more similar to the original Adverb-Pretraining group than the L1 English control group. In additional experiments reported in N. Ellis and Sagarra (2010b), the researchers replicated this design with a more complex verbal paradigm including three different persons in the three tenses (nine verb forms total). They included four participant groups with different L1 backgrounds varying with respect to their use of inflectional morphology: Chinese (none), English (impoverished), Spanish and Russian (rich). The results indicated that with a more complicated verbal paradigm, all
participants increased their reliance on the more salient and simple adverbial cues to temporal reference. Nevertheless, the participants’ L1 biased their use of the cues. The greater the use of verbal morphology in the L1, the more the learners relied on and acquired the inflections. These findings provide experimental evidence for long-term language transfer effects.

The previous experiments illustrated that the participants coming from L1s that lack or have weak morphology (Chinese and English, respectively) relied the most on adverbial cues and experienced the most difficulty learning inflections. Thus, in two follow-up studies described in N. Ellis and Sagarra (2010b), the researchers explored whether such learners could overcome these long-term learned attention effects through instructional practices involving preexposure to verbal cues (Experiment 1) or textual enhancement of the inflections (Experiment 2). Both experiments included L1 Chinese and L1 English participants, and they followed the design of the prior studies using the more complex verbal paradigm, with the exception of the instructional modifications. Results of Experiment 1 revealed that verb pretraining increased both groups’ processing of verbal inflections in the sentence processing task that followed, but this practice had a stronger impact for participants who came from an L1 with impoverished, yet nevertheless present, morphology (English). Textual enhancement of the inflections (presented in bold and blue) in the second experiment also increased processing and learning of morphology, and this practice was equally effective for both L1 English and L1 Chinese participants. These two experiments illustrate that re-focusing learners’ attention through instructional practices can help them overcome long-term learned attention effects, at least in the case of laboratory learning of a subset of a language. By examining the effects of form-focused training in the current dissertation, we will be able to determine whether these effects obtained with a subset of Latin in
a single laboratory session extend to real classroom learners that are actually attempting to acquire a complex language.

Taken together, these laboratory-learning studies illustrate robust effects of attentional biases stemming from both short-term and long-term learning experiences. Cues learned in prior language experience (either early in instruction or in the L1) block the processing and acquisition of later learned cues. Based on the research presented here as well as the eye-tracking studies examining real classroom learners that were discussed earlier, N. Ellis and Sagarra (2010b) conclude that linguistic factors and language experience modulate learners’ attention to inflectional morphology, where acquisition is more successful in cases of less complex verbal paradigms, learners coming from morphologically rich native languages, and increased exposure due to longer time learning the language. Fortunately, though, the research also suggests that certain types of pedagogical intervention may help learners attend more to inflections and overcome learned attentional biases toward lexical cues. This will be examined in the present dissertation.

In sum, this experimental research provides support for the Associative-Cognitive model, and more generally for usage-based explanations for the typical limited end-state of L2 grammar. In addition to linguistic factors (frequency, salience, reliability) that affect both L1 and L2 acquisition and can render certain forms more difficult to learn, a lifetime of prior L1 usage strongly influences L2 learners’ attention to language, processing biases, and acquisition of constructions. In short, following the Associative-Cognitive model, it is the conjunction of unfavorable linguistic factors with strong learned attentional biases, based on L1 experience, that render grammatical morphemes such as inflectional morphology especially difficult for many late L2 learners to process and acquire (N. Ellis, 2006, 2008). Nevertheless, despite the
difficulties L2 learners face as a result of their previous language experience, unlike classic
critical-period approaches, usage-based perspectives generally predict that, although difficult,
any morphosyntactic pattern can be acquired with sufficient exposure, practice, motivation, and
ability (Bybee, 2008; N. Ellis, 2008), which leaves open hope for intervention methods. In this
vein, the current dissertation aims to examine whether form-focused training can help real
classroom learners at beginning stages of SLA overcome learned attentional biases and promote
the processing and learning of grammatical morphemes, thus extending the laboratory findings
(based on a small subset of Latin) reported in this section.

2.4 Instructional Factors and Strategies

The challenge of overcoming learned attentional biases is exacerbated by the
overrepresentation, underrepresentation, and altered pattern of use of certain L2 cues in the
classroom (e.g., Chaudron, 1988; Dracos, 2010; Goodall, 2008; Santilli, 1996; Sanz, 1999).
Particularly problematic for the processing of L2 verbal morphology is the overuse of lexical
cues in ‘teacher talk’ as well as in other learners’ speech. For instance, Dracos (2010) found that
L1 Spanish instructors produced substantially more overt subject pronouns in the classroom with
English-learners of Spanish as compared to when they interacted naturally with other native
Spanish speakers. Such redundant overuse of lexical cues – likely caused by the desire to
facilitate learners’ overall comprehension – obviates the necessity to process verbal inflections
for meaning. LaBrozzi (2009) found that studying abroad helped some intermediate English-
Spanish learners rely more on verbal inflections, particularly those with high working memory.
But since study abroad is not always an option, especially in the initial stages of L2 learning, it is
crucial to find alternative ways to help override L1-based cue biases and promote the processing
of fundamental grammatical cues like inflectional morphology. Moreover, even if study abroad
or immersion experience is an option, pedagogical intervention that explicitly draw learners’
attention to the grammatical cues at early stages of acquisition might assist students in overriding
biases early on and, in turn, help maximize the results of subsequent study abroad or immersion
experiences.

In light of the mounting evidence indicating that attention to both form and
meaning/function are necessary for successful L2 acquisition (e.g., N. Ellis, 2011; Doughty &
Williams, 1998; Gass, 1997; Long, 1991), various instructional practices have been developed to
help draw learners’ attention to form while also maintaining their focus on meaning and/or
function. Some of these include explicit grammar instruction in conjunction with an abundance
of meaningful input (Terrell, 1991), input enhancement (Sharwood Smith, 1993; see Han, Park,
& Combs, 2008, for a review), task-based instruction (Long & Crookes, 1992; see Van den
Branden, Bygate, & Norris, 2009, for a review), feedback on error (Long, 1991; see Li, 2010, for
a review), and structured input activities that require the processing of the grammatical form for
meaning (VanPatten, 1996, 2002). The language training employed in the current dissertation
involves processing activities designed in a similar vein to structured input activities as well as
corrective feedback, and thus both of these specific instructional techniques will be discussed
briefly in turn.

2.4.1 Structured Input Activities

Structured input activities are the key component of Processing Instruction, an approach
to grammar instruction that derives from VanPatten’s input processing model (1996, 2004a,
2006) described in Section 2.2.1.1. The aim of Processing Instruction is “to alter how learners
process input and to encourage better form-meaning mapping that results in grammatically richer
intake” (VanPatten, 1996, p. 6). While Processing Instruction involves both presenting explicit
information about the target form or structure (including information about processing strategies) and providing practice interpreting the form for meaning in structured input activities, research strongly suggests that the structured input practice component is the cause, or at least the principle cause, for Processing Instruction’s effectiveness in promoting learning (Benati, 2004a, 2004b; Fernández, 2008; Farley, 2004; Marsden, 2006; Sanz & Morgan-Short, 2004; VanPatten & Oikennon, 1996; Wong, 2004b; see Benati & Lee, 2010, for a summary of the research). In structured input activities the input is manipulated in specific ways in order to push learners to become dependent on the form and structure to get meaning. For instance, given that many L2 learners, especially those whose L1 is morphologically weak, tend to rely on temporal adverbs over verbal inflections to establish the time frame (a processing strategy specifically outlined in VanPatten’s (1996) Lexical Preference Principle), structured input activities addressing this processing problem would require learners’ to process and respond to sentences that lack temporal adverbs such as the following: John played baseball with his friends. In an activity requiring learners to determine whether the statement refers to an action in the past or present, for example, the learners must attend to the grammatical marking (in this case the verbal inflection –ed). Such structured input activities make interpretation of the form to get meaning necessary to the task, and this helps learners make the necessary form-meaning connections for acquisition (e.g., Benati & Lee, 2010; N. Ellis, 2005; VanPatten, 1996, 2002, 2004b). VanPatten (1996) outlined the following guidelines for the creation of structured input activities for classroom instruction: a) present one new form at a time, b) keep meaning in focus, c) move from sentences to connected discourse, d) use both oral and written input, e) have the learner do something with the input, and f) keep the learner’s processing strategies in mind.
There is an abundance of research examining the effectiveness of Processing Instruction (or specifically structured input activities) for learning grammar, and as aforementioned the results are positive. Numerous studies have compared Processing Instruction to other common approaches to grammar instruction, primarily to one or both of the following: a) traditional instruction involving explanation of the target item, the complete paradigm of the forms, and then drills (mainly mechanical) requiring attention to form but not to meaning, and b) meaning-based output instruction involving explanation of the target item followed by production-oriented language activities that maintained meaning and form in focus. This research has overwhelmingly shown that Processing Instruction is most effective for learning the target form, and the benefits of this approach are generalizable to various structures (e.g., verbal morphology, object pronouns, copular verbs, adjective agreement), languages (e.g., Spanish, French, Japanese, English), mediums of delivery (classroom and computer), and levels of proficiency (e.g., Benati, 2005; Cadierno, 1995; Cheng, 2004; Lee and Benati, 2007a, 2007b; VanPatten and Cadierno, 1993; VanPatten and Fernández, 2004; for recent comprehensive reviews of the research, see Lee & Benati, 2009; Benati & Lee, 2010).

Learners’ performance in these studies has mainly been measured by offline sentence-level production and interpretation tasks administered before and after the treatment. For the production task, learners must supply the correct form missing in a given sentence. In the interpretation task, learners hear sentences that require them to interpret the target grammar form for meaning (i.e., they are given structured input similar to activities in their treatment). Proponents of Processing Instruction claim that it is an effective form of intervention for altering learners’ inappropriate or inefficient processing strategies (e.g., processing temporal adverbs over grammatical cues) (e.g., Benati & Lee, 2010; Fernández, 2008; Wong, 2004a; VanPatten,
I argue, though, that this claim about the merits of Processing Instruction is without evidential support. While research clearly shows that Processing Instruction improves learners’ ability to produce the target form and process it for meaning when it is non-redundant and/or more salient (e.g., when it is the only available cue in a more salient position) in structured input, it is unclear whether it actually alters their strategies for subsequent processing of unstructured input, where the efficient and inefficient processing strategies are both available to them. If learners truly alter their processing strategies due to the treatment, then when they later encounter the form in unstructured input such as redundant contexts – which, as discussed previously, is a frequent occurrence in natural language and particularly in learner-directed speech (Dracos, 2010; N. Ellis, 2006c; VanPatten, 2003, 2004a) – they should continue to use the appropriate or efficient processing strategy and attend to the grammatical cue. But Processing Instruction research to date does not address this issue. For example, with respect to VanPatten’s Lexical Preference Principle (1996, 2006), the following questions remain unanswered. After exposure to structured input activities, will L1 English learners of a morphologically rich L2 exhibit increased reliance on and improved processing of verbal morphology even when it is redundant and accompanied by more salient lexical cues (temporal adverbs and subject pronouns)? Or will learners revert back to their L1-experience-based strategy of attending to the lexical cue to the detriment of processing the grammatical cue? While this dissertation does not aim to test the tenets of Processing Instruction, and cannot directly do so given the methodological differences in the design and implementation of the training used, the results of the current study will shed some light on these unanswered questions by examining learners’ cue reliance and online processing of agreement violations before and after extensive exposure to
processing-based training activities that share many commonalities with structured input activities (see Section 3.1.3.3 for training materials and procedure).

2.4.2 Corrective Feedback

There is substantial evidence that corrective feedback can facilitate the acquisition of targeted hard-to-learn grammatical forms, and this has been argued to be the case because corrective feedback can help learners validate or reject interlanguage hypotheses to make correct form-meaning/function connections (Doughty, 2001; Doughty & Williams, 1998b; N. Ellis, 2005, 2006b; R. Ellis, 2008; R. Ellis, Loewen, & Erlam, 2006; Gass, 1997; Long & Robinson, 1998; Tomasello & Herron, 1989). While there is little disagreement that corrective feedback contributes to SLA, empirical research remains inconclusive with respect to the role that different feedback techniques play in L2 grammatical development, particularly the relative efficacy of implicit and explicit types (for comprehensive reviews see R. Ellis et al., 2006; Li, 2010; Long, 2007). Many studies have shown advantages for more explicit feedback techniques over implicit types typically in the form of recasts or simple correct/incorrect responses (e.g., Ammar & Spada, 2006; Carroll & Swain, 1993; R. Ellis et al., 2006; R. Ellis, 2008; Heift, 2004, 2010; Lyster & Ranta, 1997; Lyster, 2004; Nagata, 1993; Nagata & Swisher, 1995; Nassaji, 2009; Rosa & Leow, 2004; Sauro, 2009), yet other studies have found no significant differences (e.g., Kim & Mathes, 2001; Loewen & Erlam, 2006; Lyster & Izquierdo, 2009; Loewn & Nabei, 2007; Sanz, 2004; Sanz & Morgan-Short, 2004).

In experimental classroom research, Carroll and Swain (1993) is one of the early studies that concluded that more explicit feedback techniques were more effective than implicit types. The learners in this study who were provided with the most explicit form of feedback, direct metalinguistic explanation, during training on English dative alternation significantly
outperformed the groups receiving no feedback and other types of feedback (recasts, utterance rejections, or prompts) on production tasks. R. Ellis et al. (2006) is an example of a recent study that provides additional support for the positive effects of explicit feedback. R. Ellis and colleagues found that metalinguistic information led to greater improvement as compared to recasts on different testing measures specifically designed to examine implicit and explicit knowledge (oral imitation task and grammaticality judgment task, respectively) of past tense –ed in English. Moreover, corroborating findings of an earlier study by Carroll (2001), R. Ellis et al. (2006) found that only feedback containing metalinguistic information helped learners generalize the form to new contexts. On the other hand, other classroom studies point to no advantages for more explicit types of feedback. For instance, for the acquisition of English questions, Loewen and Nabei (2007) reported no significant differences between no feedback, recasts, prompts, and metalinguistic feedback on an untimed grammaticality judgment test and an oral production test, and they found that all feedback groups were equally effective and outperformed the no feedback group on a timed grammaticality judgment test.

Research on the effects of different types of computer-delivered feedback, as in the present study, has also led to conflicting findings. For example, in comparing two different levels of computer feedback, Nagata (1993; Nagata & Swisher, 1995) found that “intelligent feedback” (detailed metalinguistic explanations based on errors) was more effective than less explicit “traditional feedback” (information about what was missing or not expected in participants’ production) for the acquisition of Japanese passive structures. Similarly, Heift (2010) found that detailed metalinguistic explanations regarding participants’ errors with German grammar resulted in more uptake (i.e., learner responses to feedback) than metalinguistic clues that provided a hint as to the nature of the highlighted error. While studies such as these suggest that
L2 learners benefit from more explicit automated feedback that specifically indicates the location and nature of the error (Ammar & Spada, 2006; Heift, 2010; Rosa & Leow, 2004; Sauro, 2009), other studies involving computer-delivered feedback have found no advantages for more explicit types of feedback. For instance, Sanz (2004; Sanz & Morgan-Short, 2004) found that both the group receiving implicit feedback (indication of correct/incorrect response) and the group receiving metalinguistic feedback during processing tasks related to Spanish word order improved on interpretation and production tasks, with no significant differences between the groups.

R. Ellis et al. (2006) and Li (2010) suggest that the inconsistent findings are due to methodological issues such as differences in feedback delivery, the nature of the target form/structure, the treatment activities, treatment duration, assessment measures, and the operationalization of explicit versus implicit feedback. In spite of substantial differences in research purposes and designs that make it difficult to draw generalizations, the overall results of studies comparing explicit types of feedback to more implicit types in production-based treatments (several of which were reviewed above) suggest an advantage for explicit over implicit corrective feedback (R. Ellis et al., 2006; Li, 2010). However, it is important to keep in mind that there is also a large body of research demonstrating benefits of recasts (the most frequent form of negative feedback used in response to erroneous production (Long, 2007)), but, due to different objectives, many of these studies did not compare the efficacy of recasts to the most explicit forms of feedback (explicit correction, metalinguistic feedback, explicit elicitations/prompts) (e.g., Sachs & Suh, 2007; Sagarra, in press; See Long, 2007, for a comprehensive review of earlier studies). Relevant to the present study is the fact that the vast majority of research has examined differential effects of feedback in some form of production-
based treatment, and thus any conclusions or implications made are only directly applicable to the provision of feedback on learners’ production errors, not on their input processing problems.

Only Rosa (1999; Rosa and Leow, 2004) and Sanz (2004; Sanz & Morgan-Short, 2004) have specifically examined the role of explicit versus implicit feedback in combination with practice processing input in activities involving task essentialness (i.e., attention to the grammatical form/structure is essential to the task) like the training employed in the current dissertation. While Rosa (1999; Rosa and Leow, 2004) found that the provision of immediate metalinguistic feedback during input processing was beneficial in enhancing learners’ ability to acquire generalizable knowledge regarding the past conditional in Spanish that they could apply in both comprehension and production, as mentioned above, Sanz (2004; Sanz & Morgan-Short, 2004) found no advantages for explicit over implicit feedback following structured input activities on Object(clitic pronoun)+Verb+Subject constructions in Spanish. This led these authors to conclude that explicit feedback was not necessary and positive evidence alone in the form of task-essential practice (specifically involving structured input activities in these studies) is sufficient to promote acquisition (Sanz, 2004, p. 252; Sanz & Morgan-Short, 2004, p. 69, 72).

Yet, as Doughty (2004) and Sanz and Morgan-Short (2004, p. 72) themselves recognize, without a control group receiving task-essential practice without any feedback, their research cannot establish whether it was exposure to structured input practice alone, or this practice in combination with their form of implicit feedback (correct vs. incorrect), that resulted in similar gains by both groups. This form of implicit feedback, which some researchers argue should be more accurately classified as “semiexplicit” (R. Ellis et al., 2006, p. 348; Henshaw, 2011) or “explicit inductive” (DeKeyser, 2003, p. 325), may in fact be necessary because it enables learners to adjust whenever they discover that they are not processing appropriately and to
ultimately figure out the system (Doughty, 2004; DeKeyser, 2003). The same limitation applies to Rosa (1999; Rosa and Leow, 2004) because their no-feedback control group also lacked exposure to task-essential practice (i.e., they simply read experimental sentences for meaning). Thus, while metalinguistic information indicating the nature of learners’ errors clearly explains the superiority of groups receiving explicit feedback over those receiving implicit feedback in Rosa (1999; Rosa and Leow, 2004), it is impossible to determine what accounts for the considerably greater learning gains evidenced by the implicit feedback group as compared to the control group (the task-essential practice, the “semiexplicit” feedback, or a combination of both). Moreover, in that the few studies examining the effects of feedback during task-essential input practice fail to isolate “semiexplicit” feedback as a variable and point to contradictory findings with respect to the role of metalinguistic feedback, more research is clearly necessary to determine whether some sort of explicit feedback can facilitate the processing and subsequent acquisition of L2 grammatical forms and structures. Moreover, potential effects of the different feedback types on processing could have been masked in these prior studies due to the short treatment durations (only one session). This dissertation addresses these methodological concerns through a longitudinal design that should inform the debate as to whether the provision of any feedback, and specifically explicit metalinguistic information, during task-essential processing-based training can facilitate the processing and learning of L2 grammar, specifically verbal inflections.

2.5 Working Memory (WM) in L2 Processing and Learning

In addition to external instructional variables (see Section 2.4), internal cognitive processes can also play an important role in L2 processing and learning. Working memory (WM) is one such process that has received increasing attention over the past twenty years in L2
research. Broadly defined, WM refers to the system responsible for the temporary storage and active maintenance of information during complex cognitive actions (Baddeley, 2007), which is operative in everyday tasks such as remembering a phone number, mental arithmetic, problem solving, and comprehending language. In addition to requiring short-term storage of information, such tasks demand a great amount of control of the information in order to maintain access to it while performing further processing, and to inhibit interference from task-irrelevant information (e.g., Kane et al., 2007). Thus, there are two fundamental components of WM: short-term memory and the Central Executive (Baddeley, 2007) or executive attention (Kane et al., 2007), which is responsible for managing the information.

Not surprisingly, individuals vary considerably in their working memory capacity (WMC), and individual differences in WMC have been associated with performance in various aspects of L2 learning and use, giving rise to the claim that WM may be an important component of language aptitude and predictor of L2 learning success (e.g., Dörnyei, 2005; Linck & Weiss, 2011; Miyake & Friedman, 1998; Williams, 2012). In particular, WM has been found to affect L2 reading comprehension (e.g., Harrington & Sawyer, 1992), writing (e.g., Adams & Guillot), speaking (e.g., O’Brien et al., 2007), sentence processing (Miyake & Friedman, 1998), vocabulary development (e.g., Speciale, N. Ellis, Bywater, 2004), and grammar learning (e.g., Williams & Lovatt, 2003). The goal of this section is to provide a general review of the literature on WM (esp., research examining its role for language) so as to better frame a discussion on the relationship between this cognitive individual difference and L2 processing and grammar learning.

This section is organized as follows. In Section 2.5.1, I describe and compare some of the most prominent models of WM, including domain-specific models as well as domain-general
models that emphasize the role of attention. I will highlight the similarities and differences between the models and focus on how they account for individual differences in WM and the connection to language abilities. In Section 2.5.2, I present the tests widely used to measure WM. Then in Section 2.5.3 I review the research examining the role of WM in a) L2 processing of (morpho)syntax and b) L2 attentional control with regard to relevant aspects of grammar learning (explicit instruction, intentional induction, feedback on error).

2.5.1 Models of WM

2.5.1.1 The Original Model of WM and other Domain-Specific Models

Baddeley and Hitch (1974) introduced the first model of WM, and it developed into what remains the most well-known and cited model of WM today. The early model emerged out of a need to modify prevailing theories of short-term memory (STM). Particularly, Baddeley and Hitch responded to the modal model of STM presented by Atkinson and Shiffrin (1968), itself a descendent of Broadbent (1958), which viewed STM as a single and unitary storage system that served as a necessary antechamber to long-term memory (LTM). By the early 1970s, evidence materialized that undermined several of the basic assumptions of Atkinson and Shiffrin’s model, such as the notion that long-term learning was dependent upon the short-term store and that the limited-capacity short-term store plays a critical role in general cognition (e.g., Craik & Watkins, 1973; Shallice & Warrington, 1970). In short, the STM model could not account for how temporary memory and processing are involved in the performance of complex cognitive tasks such as comprehension, retrieval, reasoning, and learning (Baddeley & Hitch, 1974). In light of this crucial shortcoming, Baddeley and Hitch introduced their seminal model of WM. It differed from Atkinson and Shiffrin’s STM model in two major ways: 1) it abandoned the notion of a single unitary storage system in favor of a multi-component system, and 2) it moved away from
focusing on the memory system itself, and rather emphasized the function of the system in complex cognition.

Baddeley and Hitch’s (1974) three-component model, represented in Figure 1, comprises a central executive that controls attention and two domain-specific subsidiary ‘slave’ systems, known as the articulatory loop (subsequently phonological loop) and the visuospatial scratchpad (subsequently sketchpad), which specialize in the temporary maintenance of information within the relevant domain. More specifically, the phonological loop comprises both a passive, phonological store that holds verbally coded information susceptible to decay within a few seconds, and an active articulatory rehearsal system where subvocal rehearsal can serve to refresh decaying representations in the phonological store. Similarly, the visuospatial sketchpad is responsible for the processing and temporary maintenance of visual, spatial, and possibly kinesthetic information (Baddeley & Hitch, 1974; Baddeley, 1986).

Figure 1. Baddeley & Hitch’s (1974) Three-Component WM Model

While the functions of the slave systems were rather extensively explored and illustrated in early outlines of Baddeley and Hitch’s model, Baddeley (1986, 1996, 1998) admits that their initial specification of the central executive was vague. In fact, he claimed it was the “area of residual ignorance” (1986, p. 225) serving as little more than a conceptual ragbag to account for any complex process or control strategy falling outside of the functions of the slave systems. Baddeley (1996) draws on empirical research to specify the responsibilities of this crucial component of WM. The central executive is proposed to serve as the supervisory attentional system that regulates control processes in working memory, including the coordination of the slave systems, the regulation of encoding and retrieval strategies, the switching of attention, the
manipulation of information stored in the slave systems (Baddeley, 1986, 1996). The central executive has limited processing resources, and the efficiency with which it can fulfill a role depends upon other concurrent demands placed upon it. In that the central executive itself is not equipped with storage capacity, it is supplemented by the limited-capacity slave systems. Nevertheless, the central executive is deemed the most important component of working memory, particularly with respect to general cognition (Baddeley, 1996). Baddeley (2003) claimed that executive processes are likely the most important factors determining individual differences in working memory capacity (WMC).

Baddeley (2000) made a structural change to the original model by adding an additional component, the episodic buffer, as shown in Figure 2. Baddeley proposed this new slave system to handle some problems with the three-component model, such as how verbal information is maintained under articulatory suppression. Evidence from patients with short-term memory deficits, particularly phonological loop deficits, indicate that their language performance (e.g., recall) is not impaired to the degree predicted by the original model, which attributes all storage and manipulation of verbal information to the phonological loop. Thus, Baddeley posits the episodic buffer as an additional limited-capacity back-up store, which is basically an immediate-memory version of episodic memory where multidimensional representations of objects and events are integrated and maintained. To date, though, there has been little research on the role of the episodic buffer in specific cognitive tasks such as language learning, and certain scholars claim that this new component offers little to research on individual differences in WM (Kane et al., 2007).
Following the proposal of Baddeley and Hitch’s (1974) model, and particularly Baddeley’s (1986) detailed elaboration of it, there was a surge of research exploring the concept of working memory, resulting in the development of measures to test the model as well as alternative theoretical proposals. Daneman and Carpenter (1980) was one of the first studies to test the model’s predictions through a reading span test they devised to tap simultaneous processing and storage (i.e., processing sentences and storing sentence-final target words). Unlike traditional measures of short-term memory (e.g., digit span and word span), Daneman and Carpenter demonstrated that the reading span test, which will be described in detail in Section 2.5.2, correlated strongly with three diverse measures of reading comprehension: fact questions, pronoun reference questions, and the Verbal Scholastic Aptitude Test. Thus, the researchers hypothesized that reading or processing efficiency, rather than static storage capacity, is the critical source of individual differences in both their span test and the reading comprehension.
tasks. Accordingly, they claimed that the reading span test reflects WMC, and this capacity is a critical source of individual differences in language comprehension. More specifically, good readers, who have fast and efficient reading processes, require less WMC to comprehend sentences as compared to poor readers. As a result, following Baddeley and Hitch’s limited-capacity view of WM, good readers then have functionally more capacity for storing and maintaining information in WM than poor readers, who need to devote so many resources to processing that they have less capacity left for storage. Yet, good and poor readers may very well be equally adept at other, non-reading processing tasks. Thus, strong language skills were assumed to lead to a larger functional verbal WMC, rather than a larger WMC leading to stronger language comprehension skills. This view taken by Daneman and Carpenter (1980, 1983), which garnered more support in subsequent studies by Daneman and colleagues (e.g., Daneman & Green, 1986), has been called the task-specific processing hypothesis.

The early experimental studies by Daneman and Carpenter proved to be highly influential. Not only did their research provide substantial validation for WM theory, it initiated an important individual-differences approach to WM research, stimulating particular interest in the potentially critical role of WM in language comprehension (Baddeley, 2007; Kane et al., 2007). In this vein, Just and Carpenter (1992) proposed a capacity theory focusing specifically on the role of working memory in language comprehension. The theory posits that WM consists of a single pool of limited resources that corresponds approximately to the part of the central executive in Baddeley’s (1986) theory dealing with language comprehension. Their conception of WM does not include modality-specific storage components, such as the phonological loop, that are structurally separate from the parsing process. Rather, Just and Carpenter claimed that the trading relation between processing and storage suggests that the two functions draw on a
common pool of resources within a single system. Thus, in this view the two functions of WM compete with each other for the limited cognitive resources when there is high demand for these resources. Moreover, this theory holds that both processing and storage are fueled by activation, where, in line with connectionist models, language comprehension can be considered to involve manipulations of activation. Just and Carpenter then define capacity as the maximum amount of activation available in WM to support either processing or storage. Information becomes activated during language comprehension by being encoded from written or spoken language, triggered by a computation, or retrieved from long-term memory. An element is deemed part of working memory as long its’ level of activation is above some minimum threshold. Individual differences in WMC, then, are considered to derive from the amount of activation available for processing and storage during language comprehension, which varies across individuals. If the amount of activation required to perform a comprehension task exceeds the amount of activation the individual has available, some of the activation preserving older elements will be deallocated to scale back total activation within the maximum bound. According to the theory, this process of scaling back activation results in a type of forgetting, where representations generated early in a sentence, for instance, may be forgotten before arriving to the end of the sentence when they are needed. The researchers submit that this notion of variability in the amount of activation available in WM accounts for differences among individuals in the speed, accuracy, and efficiency with which they comprehend language. Just and colleagues (Just & Carpenter, 1992; Miyake, Carpenter, & Just, 1994) draw on experimental evidence from high- and low-span readers (classified based on Daneman and Carpenter’s (1980) reading span test) as well as computer simulations of WMC to lend support to their conceptualization of how verbal working
memory capacity constrains language comprehension in both normal and aphasic adult populations.

Just and Carpenter’s (1992) capacity theory undoubtedly made a significant contribution to the field, particularly with respect to the role of WM in language comprehension and the need for models that elucidate this relationship. Nevertheless, their theory was soon critiqued by Waters and Caplan (1996a), who called into question their view that humans have a single pool of verbal processing resources that supports all verbal tasks. Waters and Caplan challenged the capacity theory’s claim that sentence processing uses the same resource pool that is measured by Daneman and Carpenter’s reading span test. Waters and Caplan (1996a, 1996b) contended that unlike much of natural language processing, the reading span task requires controlled processing and conscious retrieval of items, and it imposes a memory load that is unrelated to the ongoing computations required by the task. The researchers undermined the empirical support for and explanatory power of Just and Carpenter’s model, and they reviewed neuropsychological research showing that the verbal working memory system tapped by Daneman and Carpenter’s (1980) reading span task is not implicated in many aspects of language comprehension, namely online syntactic processing. Thus, Waters and Caplan (1996a) proposed an alternative to the single-resource capacity model, according to which there are specializations within the verbal-processing resource system for different verbally mediated tasks. They posited two resource pools, one for online psycholinguistic processing (e.g., syntactic parsing), and one that supports conscious, controlled, and verbally mediated processes (e.g., explicit reasoning). The researchers outlined three related predictions of their separate language-processing resources theory to illustrate how it differs from the single-resource capacity theory. First, in contrast to Just and Carpenter’s model, Waters and Caplan’s model holds that performance on general verbal
working-memory tasks like the reading span task will not predict online language-processing efficiency due to the separate resource pools. Second, their theory claims that there will not be interference between comprehension-external factors (memory loads) and comprehension-internal factors since they draw on different resource pools. Thus, their separate-resources account also predicts that low-span learners will perform less well under conditions of increased external memory load, but that this effect should not be greater for syntactically complex sentences since online syntactic processing does not compete for resources from the same pool.

Subsequent research by Caplan and Waters (1999; Caplan, Waters, & Dede, 2007) directly tested the aforementioned predictions and further developed their concept of verbal working memory, particularly the specialized component devoted to online syntactic processing. Citing data from normal young adults, elderly participants, and patients with various neurological diseases (involving impaired WM and executive control functions), the researchers provided evidence that participants whose verbal working memory capacity is reduced on standard span tasks (i.e., they are classified as low-span participants) do not show reduced syntactic processing abilities. In line with their theory, Caplan and colleagues claim that evidence points to no relation between variability in utilization (or efficiency) of WM in online first-pass syntactic processing and overall WM capacity as measured by standard span tests. Yet, replicating other studies, they did find correlations between other aspects of language processing (e.g., text memory and comprehension) and performance on standard span tests of WM. Based on these findings, the researchers proposed a more refined and precise conceptualization of the divisions of the verbal working memory system. They assert that their model operates within a Baddeley and Hitch-type framework that has a separate capacity-limited verbal WM system capable of processing and maintaining representations (Caplan et al., 2007). The focus is the
central executive component of WM, which the researchers hypothesize is fractioned into at least two components: 1) a specialized verbal WM system (svWM) that supports interpretive processing, or the initial, online, unconscious processes involved in assigning syntactic structure and using the structure in conjunction with lexical meaning to construe aspects of the propositional meaning of the utterance, and 2) a more general verbal WM system (gvWM) devoted to post-interpretive processing, or the use of the meaning construed to accomplish other verbally mediated tasks. Only these more conscious, controlled post-interpretive aspects of the comprehension process (e.g., plausibility judgments, recalling specific information), and not interpretive processing, are related to WMC as measured by standard tests. Thus, with respect to their general view of WM, Caplan et al. (2007) argue that the central executive is fractionated, and that variability in a given subpart, involving particular domain-specific knowledge and operations, may not overlap with that of other parts. In general, Caplan and colleagues conceptualize the central executive as divided not only according to types of representation (verbal, visual), but also computationally according to the specific types of operations required to perform a task.

While there is a stark contrast between the views of Just and Carpenter and Caplan and colleagues with respect to the structure of verbal working memory, the two theories have a similar overarching understanding of WM. They both posit a critical division between language processes and other aspects of WM, adopting a more domain-specific view of WM. This is consistent with the earlier proposed task-specific hypothesis of Daneman and colleagues (Daneman & Carpenter, 1980, 1983; Daneman & Green, 1986) as well as with related proposals such as those of Daneman and Tardif (1987) and Shah and Miyake (1996). Differences among these theories stem primarily from their conceptualization of working memory resources. On the
one hand is Just and Carpenter’s model that posits a single pool of resources designated for language within the central executive, without any separate modality-specific buffers or subsystems like those of Baddeley’s influential multicomponent model. Yet, while the domain-specific subsystems in Baddeley’s model act predominately as temporary storage systems with limited processing capacities, serving as “slave” systems for the central executive containing general-purpose resources, Daneman and Tardif (1987) and Shah and Miyake (1996) emphasized a less unitary view of this same model comprising at least two entirely separate pools of domain-specific resources that not only support maintenance, but also the processing and control of domain-specific information. Although Caplan and colleagues claim that the separate verbal WM system corresponds to a part of the central executive, their position arguably goes the furthest in the non-unitary direction, fractioning verbal WM into separable subsystems for different language-comprehension processes. However, there is another camp of researchers that disagree with the general idea of a structural approach to WM, rejecting the very notion of separate domain-specific components or buffers. I now turn to outline the most prominent theories within this divergent approach to WM.

2.5.1.2 Domain-General Models of WM

Turner and Engle (1986, 1989) challenged Daneman and colleagues’ (Daneman & Carpenter, 1980; Daneman & Green, 1986) claim that WM capacity is specific to a given task such as reading, where good readers have more available WM capacity while reading because their efficient reading skills demand less from the overall WM resources for processing and thus leave more residual resources for storage. Turner and Engle investigated an alternative hypothesis that WM capacity is independent of the task being performed, and consequently the reason people are good readers is because they have a large, general WM capacity that is
independent of specific task strategies. To this end, the researchers tested participants on four complex WM span measures (a sentence-word span task most similar to the reading span test, a sentence-digit task, a math operations-word task, and a math-operations digit task) to determine whether the processing task in a memory span test needs to involve reading, as Daneman and Carpenter suggested, in order to predict reading comprehension. In support of their hypothesis, the researchers found that the operations span tasks correlated just as well with reading comprehension as the reading-based span tasks. Good readers were able to recall more words and digits than poor readers, regardless of whether the processing task required reading or arithmetic skills. Thus, the Turner and Engle (1989) concluded that it is not more efficient reading skills that resulted in a larger span score on reading span tasks, but rather a larger WM capacity, which is likely a unitary individual characteristic, independent of the type of task in which the individual utilizes it. The open question posed by the authors is how these findings jibe with the prevailing multi-component WM model of Baddeley and Hitch (1974). Upon accepting Turner and Engle’s evidence that complex span tasks reflect a general WM system, the challenge was then to either determine which structural component(s) the complex span tasks measured in Baddeley and Hitch’s model (the central executive, the phonological loop, or some interaction of both) or to propose an altered model that re-conceptualizes WM as a more unitary system supporting the wide array of control functions as well as the temporary maintenance of representations.

One of the earlier attempts to explain the WM system in a more unitary, interconnected manner was Cowan’s (1988) prominent embedded-processes theory, which was further developed in Cowan (1995, 1999, 2005). In place of the multiple, separate storage systems and central executive of Baddeley (1986), Cowan adopted a functional approach and proposed a
structure that emphasized the intricate relation between attention and memory. Attention was considered to be enhanced processing of some information to the exclusion of other, concurrently available information. Figure 3 illustrates the three embedded levels of activation that comprise the WM system in Cowan’s model: long-term memory units that are currently inactive, the subset of long-term memory that is currently activated, and the subset of activated memory that is the current focus of attention and awareness.

Figure 3. Cowan’s (1988) Embedded-Processes Model

Similar to Just and Carpenter (1992), Cowan assumed that WM entails a subset of long-term memory, which is stimulated or activated above a critical threshold. Yet, unlike Just and Carpenter, Cowan distinguished between activation and attention, describing WM as involving not only memory in the focus of attention, but also memory out of the focus of attention (but nonetheless temporarily activated) as well as inactive elements of memory with sufficiently pertinent retrieval cues. As seen in Figure 3, Cowan postulated a central executive mechanism that controls the focus of attention for voluntary, effort-demanding processes; however, habitual or involuntary processing can also regulate the focus of attention. Cowan (1995) posited two sources of WM limitations: a time limit of memory activation and a capacity limit of the focus of
attention. Cowan (2001, 2005) particularly emphasized the limited-capacity attentional focus, which he argued could hold only a small subset of currently relevant information. He specifically posited that the capacity for young adults is approximately three or four chunks, but this amount can vary from about two to six chunks, and it also varies among children and the elderly. Thus, Cowan suggested that individual differences in WMC reflect the size, or capacity, of attentional focus.

Schneider and Detweiler (1987) developed a model that Cowan (1995) considered to be one possible instantiation of many of the ideas of his 1988 embedded-processes model. Within a connectionist framework, Schneider and Detweiler proposed a detailed conceptualization of WM that aimed to simulate much of the automatic and controlled (attention-directed) processing behavior of humans. The model comprised eight regions of modules (visual, auditory, speech, lexical, semantic, motor, mood, context) that differ in their inputs and outputs, but have a similar structure. Schneider and Detweiler represented their conceptualization of WM at three levels of detail: a) the microlevel, which depicts a neural-like network capable of producing associative processing and attentional phenomena, b) the macrolevel, which illustrates the attentional control and message transmissions, and c) the system level, which shows the interactions among regions. The system level is presented in Figure 4.
As seen in Figure 4, Schneider and Detweiler’s model assumes memory and processing take place in a multilayered hierarchy of the modular processors that are interconnected on a circuit referred to as the innerloop, which allows cross-area transmissions. The flow of information among modules is regulated by a hierarchical attentional control system with both regional and central control. Memory is stored in the connection weights as well as in activation patterns across cells. Thus, WM, which consists of activation within modules and the central executive as well as fast connection weights, is always influenced by long-term memory. Schneider and Detweiler emphasize the importance of their context module, which they describe as a critical context-based storage mechanism that serves to recover relevant task information after attention is shifted away from the current task due to an interruption or distraction. The researchers claimed that the inclusion of such a mechanism, which is not provided in previous models such as Baddeley (1986), is necessary to account for the ability to cope with interference, an essential characteristic of working memory. According to Schneider and Detweiler’s model, WM
limitations are the result of communication cross-talk interference, storage limitations (decay rate for connection weights), and executive control limitations (e.g., attention allocation is limited). Schneider and Detweiler’s model is considered less unitary than that of Cowan (1988) in that it incorporates from Baddeley and Hitch’s model the concept of separate buffers or slave processors, referred to as modules. However, in contrast to the domain-specific approaches outlined in the previous section, Schneider and Detweiler emphasize how the different modules share the same structure, are governed by the same processing principles, and interact to account for a wide range of cognitive performance. In fact, Schneider (1999) generalized this earlier WM model so that it is operative for all skilled processing and learning, rendering WM an integral aspect of this more general model.

Similar to Schneider and Detweiler (1988) and Cowan (1988, 1995, 1999), Engle and colleagues developed Baddeley and Hitch’s notion of WM to make it more relevant to higher level cognition in general. Based on evidence from Turner and Engle (1986, 1989) described above, Engle, Cantor, and Carullo (1992) proposed the general capacity hypothesis, a conceptualization of WM that they claimed has much in common with the theories of Schneider and Detweiler (1987) and Cowan (1988) outlined above. The general capacity theory also holds that activation is what drives the system, in which WM consists of the temporary or permanent knowledge units in long-term memory that are currently active. The knowledge units are assumed to vary in their ambient level of activation, and the total amount of activation within the system is limited. On this theory, individuals differ in the total level of activation available to their system, which is a relatively stable characteristic that accounts for individual differences in WMC. Accordingly, Engle et al. (1992) argue that complex WM span tasks provide a measure of the amount of activation or storage capacity, general to a wide variety of tasks, available for
maintaining temporary representations while concurrently switching attention back and forth to the processing task. By using variations of WM span tasks that allowed for the independent measurement of processing speed, processing accuracy, and storage, Engle et al. (1992) provided evidence in support of their general capacity view, and against competing hypotheses that assume individual differences in WMC are the result of task-specific processing skills, general processing efficiency, or strategy use. Specifically, they found that the correlation between the WM span score and verbal ability, based on the verbal Scholastic Aptitude Test (VSAT), was just as strong for the operation-span task requiring the solution of mathematical problems as the reading-span task, which replicated Turner (1986, 1989). More importantly, partialing out processing skills and strategy use did not weaken the correlation between the WM span score and verbal ability, thus corroborating the general capacity theory.

A study by Conway and Engle (1994) led to a qualification of the general capacity model. The researchers examined individual differences in WMC, and how those differences affect performance on two types of memory search tasks under either high- or low-interference conditions. They found that WMC was important to retrieval of well-learned information from active memory only when the task involved conflict or interference. Conway and Engle argued that overcoming interference involves inhibiting or preventing the activation of irrelevant information, which is resource demanding and would only occur to the extent that an individual had the available resources to do so. The researchers interpreted these findings as indicating that high-span participants outperformed low-span participants due to more available attentional resources. Moreover, the results suggest that WMC is only taxed by tasks that require controlled effortful processing, not by automatic processing. Thus, the researchers concluded that it is insufficient to simply claim that individuals differ in the total amount of activation they have
available. Rather, the qualification they posited is that individuals differ in the amount of general, controlled, effortful, attentional resources they have available, which is the capacity measured by complex WM span tasks that require controlled processing to continually switch attention from the storage aspect of the task to the processing aspect.

Conway and Engle (1996) is one of various studies that provided support for this qualified version of the general capacity model. Similar to Engle et al. (1992), Conway and Engle tested this theory against competing hypotheses for the relationship between WM and reading comprehension. Rather than statistically controlling for processing skill, though, Conway and Engle equated, across subjects, the processing demands of the span task by tailoring the difficulty of the task to the mathematical ability of each participant. The results showed that a strong correlation between operation span scores and verbal ability remained even when the difficulty of the processing component is equated. Moreover, although the participants spent more time processing the operation-word pair in the more difficult version of the span task as compared to the easy one, there were no statistical differences between the number of words recalled in each task. Thus, the researchers claimed that it is not the demand of the processing component or the trade-off in resources that is critical, because if it were, the level of difficulty should have had an effect on the number of words recalled, which was not the case. Rather, Conway and Engle argued that these findings support their view that the critical component of the span tasks is the attention switch itself, and, thus, individuals who perform well on the task (classified as high-span subjects) are considered to have more limited-capacity attentional resources to engage in controlled processing in situations involving interference or distraction.

Having designated domain-general attention capability (as opposed to processing skill, strategy use, or short-term storage) as the likely cause for individual differences in WMC, Engle,
Tuholski, Laughlin, and Conway (1999) set out to shed light on the nature of the relationship between WMC and general fluid intelligence (Gf). They investigated whether short-term memory (STM) and WM were dissociable constructs and, if so, whether WMC better predicted Gf. Engle and colleagues tested participants on various WMC tasks (operation span, reading span, counting span), STM tasks (forward and backward word span), and tests of Gf involving novel figural and spatial reasoning. Structural and equation modeling revealed that STM and WM are highly related but separable constructs. After the shared variance between the two constructs was removed, which they assumed to reflect primarily storage, there was still a strong correlation between the unique residual variance in WMC and Gf, but no such correlation was found between STM and Gf. The researchers concluded that the WM-Gf correlation is driven by whatever complex WM span tasks demand beyond simple storage, which they submit is the strong executive attention demands elicited by these dual-tasks. The strong correlations between WMC and Gf suggest WMC is an important mechanism of general cognitive ability, especially in light of the fact that WMC was derived from verbal, symbolic span tasks, and Gf was tested with nonverbal, figural reasoning tasks. Yet, as seen in the previous section, other studies (e.g., Shah & Miyake, 1996) indicated domain specificity in WM span due to low correlations between verbal and visuospatial span tasks as well as between ability and cross-domain WM span. Suspecting that these domain-specific findings were due to measurement error with few tasks and/or a restricted range of high general intellectual ability in the university-student sample, Kane, Hambrick, Tuholski, Wilhelm, Payne, and Engle (2004) tested a large, diverse sample of participants on multiple tests of verbal and spatial WM and STM. The findings revealed strong correlations between verbal and spatial WM span tasks (with shared variance between 70-85%), yet much weaker correlations between verbal and spatial STM measures (only 40% shared
variance). Interestingly, by then dividing the sample into a high-Gf group and a low Gf-group, they found that WMC and STM were significantly more domain specific in the high-Gf group as compared to the low-Gf group and the entire sample. Like Engle et al. (1999), Kane and colleagues also found that WM span tasks primarily tap general executive attention processes while STM tasks tap primarily domain-specific storage and rehearsal processes. Moreover, once again executive attention was shown to strongly predict Gf, leading the authors to conclude that WMC, which reflects predominately general executive attention, is one important source of variance in general fluid intelligence.

In addition to these large-scale studies implementing structural modeling, another line of investigation conducted by many of the same researchers has provided direct evidence for the relationship between WMC and the constructs of attention and executive control. Studies such as Conway and Engle (1994), discussed above, offer evidence that WMC plays an important role in memory retrieval in the face of distraction or interference. Kane and Engle (2000) is another such study that found low-span subjects to be much more vulnerable than high-span individuals to the effects of interference. Participants were presented with three 10-word lists, one word at a time, belonging to the same semantic category (e.g., occupations). After the presentation of each list, they took part in a short rehearsal-prevention task prior to a free-recall task. High- and low-span participants showed the same recall performance for List 1, where there was no interference. However, when there was proactive interference from previous lists on the subsequent recall tests, low-span participants’ recall performance dropped more drastically than that of high-span participants, showing a steeper interference effect for low- as compared to high-spans. Other participants performed the same procedure but under dual-task conditions requiring them to simultaneously tap their finger in a complex sequence during either the presentation or recall-
phase for each list. The secondary tapping task had no effect on low-span participants, yet it increased the high-span participants’ vulnerability to interference, resulting in equal recall performance by both groups in this load condition. The authors claimed that the fact that high-span participants were more vulnerable to proactive interference when their attention was divided confirms that they do normally use attentional control processes (e.g., inhibition, monitoring) to cope with interference, as evidenced in the no-load condition. However, in the load condition high-spans were not able to use attentional control near as effectively. On the other hand, since low-span participants normally do not engage in any more controlled processing under interference conditions compared to noninterference conditions, it makes sense that they showed equivalent vulnerability to interference regardless of whether or not they were performing a concurrent task.

Additional evidence that WMC reflects an attentional control construct comes from studies revealing effects of WMC differences even on tasks that do not involve memory retrieval. For example, in Conway, Cowan, and Bunting (2001), participants took part in a selective listening experiment in which they were instructed to listen to and repeat the words presented to their right ear while doing their best to ignore the distractor words presented to their left ear. The participant’s name was presented among the distractor words a few minutes into the task. While 65% of the low-span participants reported hearing their name, only 20% of high-span participants did, suggesting that they are less vulnerable to salient distractors because they are better able to block, or inhibit, task-irrelevant information. Kane and Engle (2003) provided further support for an association between WMC and the attentional control functions of goal maintenance and inhibition of interference. Participants were tested on several versions of the Stroop color-word paradigm, a task commonly used to investigate action control, selective
attention, and inhibitory control. The variations of the task differed in the percentage of congruent trials (e.g., the word “blue” appearing in the color blue) versus incongruent trials (e.g., the word “blue” appearing in the color orange). When the majority of the trials were congruent (75-80%), low-span participants showed significantly greater interference, measured by errors, than high-span participants on the few incongruent trials where accurate response required having immediate access to the task goal, which required responding based on the color and ignoring the word. However, when 80-100% of the trials were incongruent, which reinforced the goal of the task, there were no differences between the groups on errors, but there were modest differences in reaction times. In another study conducted around the same time, Long and Pratt (2002) similarly found that WMC is differentially related to the Stroop interference effect depending on the proportion of incongruent trials. Kane and Engle thus concluded that WMC predicts attentional control over goal-directed behavior, where high-span participants are much better able to keep goals of novel tasks accessible, which is necessary for proactive inhibition of competition in cognitive tasks like the Stroop test.

Converging evidence of a strong relationship between WMC and the domain-general ability to control attention, from studies such as those reported above, greatly influenced Engle and colleagues’ conception of WM. Over the years, their emphasis gradually shifted toward WM as an executive attention and control system, as opposed to a “memory” system (e.g., Engle, 2002; Kane et al., 2007). Engle, Kane, and Tuholski (1999) replaced the general-capacity model with a measurement model that emphasized the fundamental role of controlled attention in achieving activation of long-term traces through controlled retrieval, maintaining activation, and dampening activation through inhibition. Moreover, they abandoned the notion of “working memory capacity” as reflecting storage or even memory per se, and posited that it reflects the
capacity, or capabilities, of only one element of the system: controlled attention. WMC thus referred to the “capacity for controlled, sustained attention, particularly in the face of interference or distraction” (Engle, Kane, et al., 1999, p. 104). Yet, this theory was predominately inferential at the time it was first posed, as there was still little direct evidence to support it (Engle and Kane, 2004). However, since its initial proposal, there has been substantial evidence, some of which was reviewed above, that lends support to their theory and has contributed to its development. The most recent version of the measurement model (Kane et al., 2007) is shown in Figure 5, which is a further elaboration of the second version proposed by Engle and Kane (2004).

![Diagram of Working Memory System](image-url)

**Figure 5. Kane et al.’s (2007) Measurement Model of the WM system (adapted from Engle, Kane et al., (1999) [version 1] and Engle & Kane (2004) [version 1.2])**

Although the earlier versions of the model recognized the involvement of the central executive in inhibiting interference, this role of controlled attention became more central in the subsequent versions. In fact, the latest version, outlined in Figure 5, posits that the extent to which executive attention is active during a task – for maintenance, retrieval, or inhibition – is critically
determined based on the degree of conflict or interference present (Kane et al., 2007). It is under
conditions of interference that we most need working memory, as Engle (2007) emphasized: “the
primary purpose of the working memory system is to allow us to function in the face of proactive
interference and highly prepotent behaviors that are counter to the task facing us at the moment”
(p. 161). The mounting research demonstrating an association between WMC and executive
attention in resolving response competition also led to a more precise conception of “capacity”
that takes into account high-span individuals’ better ability to actively maintain task-relevant
information outside of consciousness and re-activate it despite interference. This view of
capacity stands in contrast to that of Cowan (2001, 2005) – despite substantial similarites in
their overall views of WM (Kane et al., 2007) – who views WMC as a true “capacity”, reflecting
the size or number of representations that can be maintained within the attentional focus.
Contrastingly, Engle, Kane, and colleagues’ executive attention theory of working memory
capacity proposes that high WMC does not necessarily entail the ability to hold more
representations in consciousness, but high WMC does facilitate the ability to control goal-
directed behavior by maintaining and recovering relevant information outside of consciousness
in the face of interference. In sum, according to this theory, there are two principle functions of
the central executive which are measured by complex WM span tasks: 1) the ability to maintain a
task-relevant goal in an active state during performance, and 2) the ability to inhibit interference
or prepotent responses that conflict with the task goals (Engle, 2007; Engle & Kane, 2004; Kane
et al., 2007).

The two functions proposed to account for individual differences in WMC, or attentional
control, have also been posited as crucial to executive functioning or frontal lobe functioning in
behavioral and neuroscience research (e.g., Braver & Cohen, 2000; Duncan, 1995; Duncan et al.,
1996), which Engle, Kane, and colleagues recognize as having strongly influenced their own ideas and research with respect to WM (Engle & Kane, 2004; Kane & Engle, 2003). A recent large, cross-sectional study examining the relationship between these constructs using a factor analytic approach provides additional support for the domain-general executive attention theory of WMC. McCabe, Roediger, McDaniel, Balota, and Hambrick (2010) administered four tests of WMC (reading span task, computation span task, letter rotation span task, match span task) and four tests of executive function (Wisconsin Sorting Test, verbal fluency test, mental control test, mental arithmetic test) to over 200 subjects ranging between 18 and 90 years of age, along with tests of episodic memory and processing speed. The battery of tasks used to measure executive function were selected based on their extensive use in previous research and a general consensus that they provide a valid assessment of multiple executive functions including flexibility of thinking, inhibition, problem solving, impulse control, concept formation, and abstract thinking. In contrast to Miyake and colleagues (e.g., Miyake et al., 2000), who have used multiple tests of each of three distinct executive functions (shifting, updating, and inhibition), McCabe and colleagues follow other research groups in attempting to measure more than just these few functions using tests that have been shown to each require multiple functions. The results revealed very strong correlations between executive functioning and WMC ($r = .97$) and weaker correlations between WMC or executive functions and general abilities such as processing speed ($r \approx .79$). Moreover, WMC and EF accounted for similar proportions of variance in episodic memory performance. Based on these findings, the researchers concluded that WM span measures and tasks of executive function share a common underlying executive attention construct that is highly predictive of complex cognition. This finding lends strong support for the view that complex WM span tasks predominately measure the functioning of a domain-general
central executive, which is directly in line with Engle, Kane, and colleagues’ executive attention theory.

In sum, this review has illustrated that there are clearly differing notions of WM in the literature. Yet, as Williams (2012) noted, the differences are primarily in focus rather than in overall conception of WM. For the most part, researchers agree that the WM system entails both short-term storage and an executive component, with some emphasizing the domain-specific storage components or resources (e.g., Baddeley and colleagues, Caplan and colleagues, Just and Carpenter) and others focusing on the domain-general central executive and its role in controlling attention (e.g., Cowan, Engle and colleagues). Nonetheless, there remains a poignant debate as to whether individual differences in WM stem from differences in a true capacity of some sort, processing efficiency, or the ability to control attention in the face of distraction.

In this dissertation, I will work within a domain-general framework of WM and the terminology proposed by Engle and colleagues with respect to the executive attention component of WM because of the substantial recent research lending support for this view (see above) as well as the fact that it accords best with the experimental research findings reported in Section 2.5.3, which highlight attentional control abilities as the reason for the positive relationship between WM and certain aspects of L2 learning. In any case, it is important to emphasize that, regardless of the model one ascribes to, there is generally a consensus that WM is an important source of individual difference in at least some language abilities (e.g., Gathercole, 2009; Williams, 2012). Before discussing the research examining the relationship between WM capacity (or attentional control ability) and the specific L2 abilities relevant to the current dissertation in Section 2.5.3, I describe in the following section (2.5.2) the various measures used to assess WM.
2.5.2 WM Measures

In that the WM system is also generally agreed to be responsible for both the active maintenance and manipulation of information, these aspects or components of WM have been measured separately and in combination. The key tests used to measure both verbal short-term storage capacity and the WM system as a whole will be discussed in this section, focusing in particular on the tasks that have been used in language research.

2.5.2.1 Short-Term Memory Tests

Verbal short-term memory (STM) has traditionally been measured with tasks that entail the recall of sequences of unrelated digits, words, or nonwords presented either visually or aurally. Participants are presented with increasingly larger sets of items, and the maximum number of items that can be accurately recalled back in the correct order is typically considered their ‘digit span’ or ‘word span’. In line with Baddeley and Hitch’s multi-component model, performance on these STM span tasks is typically assumed to reflect the capacity of the phonological loop, or its ability to actively maintain representations through subvocal rehearsal (Shah & Miyake, 1996). The implication of the rehearsal process in these tasks has been considered problematic, though, in that it renders interpretation of the results ambiguous (e.g., Baddeley, Gathercole, & Papagno, 1998; Daneman & Carpenter, 1980; Cantor, Engle, & Hamilton, 1991). It is difficult to determine to what degree performance is based on rehearsal strategies employed as opposed to passive storage capacity or decay rate. Moreover, the word span test poses problems, particularly for language learning research, in that prior knowledge of the word, or lack of knowledge, can affect performance (Gathercole, 1995).

Non-word repetition tasks have been employed to attenuate the involvement of language knowledge and rehearsal strategies based on evidence that it is phonological storage, not
rehearsal, that is the critical factor related to vocabulary learning, particularly in L1 acquisition (Baddeley et al., 1998). The task involves the presentation and immediate repetition of nonsense words that vary in length such as *prindle*, *frescovent*, and *woogalamic* (Baddeley et al., 1998). The test measures the temporal capacity of verbal STM. The larger the individual’s capacity, the longer the non-word that he or she can successfully recall. Contrary to original claims, the non-word repetition task does not always tap phonological storage in a completely knowledge-free manner, though, as repetition accuracy is boosted when non-words resemble actual words or conform to phonotactic rules of the native language (Baddeley, 2006). Thus, in addition to storage capacity, nonword repetition taps certain perceptual abilities as well as motor processes to reproduce the phonological forms. In that all of these abilities are also implicated in vocabulary acquisition, it makes sense that performance on non-word repetition tasks has been linked not only to vocabulary learning in the L1 (e.g., Baddeley et al., 1998), but also in the L2 (e.g., Cheung, 1996; Speciale et al., 2004). More surprising is that the digit span task has also been shown to correlate with word learning and the nonword repetition task (Gathercole, Willis, Emslie, & Baddeley, 1992; Gupta, 2003). This suggests that immediate serial recall tasks like digit span and nonword repetition tasks both tap a common storage capacity, or perhaps a capability for maintaining serial sequences, that is implicated in tasks such as vocabulary acquisition (Gupta, 2003).

While verbal STM appears to play an important role in vocabulary learning, STM tasks often fail to correlate, or only correlate weakly, with higher order cognitive tasks including more complex language processes such as reading and listening comprehension (Cantor et al., 1991; Daneman & Carpenter, 1980, 1983; Daneman & Merikle, 1996; Miyake & Friedman, 1998; Shah & Miyake, 1996). As Daneman and Carpenter (1980) first revealed in their seminal study,
there is considerable evidence that inclusion of a concurrent processing component is critical for the predictive power of WM tasks in this domain. In that sentence processing typically requires greater control of attention and thought than word learning, it follows that general WM span tasks, which tap primarily executive attention processes would be better predictors of processing performance than STM tasks, which tap primarily domain-specific storage and rehearsal processes (e.g., Kane et al., 2004). This does not, however, rule out a potential role for simple STM in language acquisition beyond vocabulary learning (see Section 4). Nevertheless, the majority of the evidence supporting a role for WM in language processing and acquisition comes from studies using complex span tasks, the term often used to differentiate between simple STM span tasks that primarily tap storage and dual tasks that also incorporate a processing task following Baddeley & Hitch’s (1974) conception of WM. These complex span tasks will be the focus of the subsequent discussion.

2.5.2.2 Complex WM Tests

The first complex span task developed with the goal of jointly tapping the storage and processing components of WM was Daneman and Carpenter’s (1980) reading span test. The original version of the task involved reading aloud unrelated sentences while simultaneously attempting to remember the last word of each sentence. The sentences were presented in sets of increasing size, from two to six sentences, and there were three sets at each level. Participants read the sentences at their own pace, and then after each set they recalled the sentence-final words in the order in which they occurred. Testing was terminated when the participant failed all three sets at a particular level. Span size was then defined as the level at which the participant correctly recalled two out of three sets. In a subsequent experiment, Daneman and Carpenter implemented three modified reading span tasks in which participants had to judge whether
sentences were true or false while reading the sentences out loud, reading them silently, or
listening to them. They referred to this latter condition as a listening span task. The true-false
component was incorporated to ensure that participants processed the whole sentence for
meaning and could not just focus on the final word. However, participants’ accuracy in their
true-false responses was ignored in scoring, which was conducted in the same manner as the
original reading span task, with the exception that the sentence-final words could be recalled in
any order. As discussed above in Section 2.5.2.1, Daneman and Carpenter found that scores from
these reading span tasks, but not from the STM word span task, strongly correlated with
performance on various measures of reading comprehension.

Various modifications of the original reading span tasks were subsequently developed by
other research groups, but the fundamental goal of the task remained the same. For instance,
Turner and Engle (1989) developed a version that asked participants to determine whether the
sentences were semantically and syntactically plausible, and they used overall accuracy scores
on this component to exclude participants who scored below 80%, an indication of not having
paid sufficient attention to the processing aspect of the task. Another widely used variant of the
reading span task is that of Waters and Caplan (1996b), which they proposed after providing
evidence that the test-retest reliability of Daneman and Carpenter’s task was low, as was the
classification of participants into groups based on their span score. One possible explanation for
this is variation across and within participants with respect to the resources they allocated to the
sentence-reading component of the task. Thus, Waters and Caplan devised a more reliable and
stable reading span task to take into account participants’ performance on both the processing
and storage component of the task. The task involves the computerized presentation of sentences
to individual participants, who read them silently and then make a judgment about their
acceptability by pressing a key. Span scores are calculated by taking into account speed and accuracy on the sentence task as well as accuracy on the recall task. In other slightly modified versions of the reading span task, the to-be-remembered word is an isolated, unrelated word (Engle et al., 1999) or a single letter (Kane et al., 2004) following each sentence, both of which are variations designed to prevent the use of reading-comprehension ability in recalling the words.

As discussed in Section 2.5.1.2, Turner and Engle (1989) devised an operation span task to test the domain-specific or task-specific view of Daneman and Carpenter (1980). They illustrated that a WM span task does not need to involve reading in order to predict reading ability. In their operation span task, participants solved mathematical equations (e.g., \(9/3 - 2 = 1\)) while simultaneously trying to remember the isolated words that followed each equation. The original version of this task simulated their reading span task in that participants read and determined the accuracy of the operations out loud. Subsequent modifications of the operation span task simulate those made for the reading span task. The counting span task (Case, Kurland, & Goldberg, 1982) is another language independent task that is ideal for a wide variety of populations, including children and patients, given the simplicity of the processing aspect, counting shapes (Conway et al., 2005). In Engle et al.’s (1999) version of the counting span task targeted for adults, participants count the number of target items (e.g., dark blue circles) in visual displays containing distractors (e.g., dark blue squares and light blue circles), and then after each set they recall the total number of target items from each display in serial order. Another non-linguistic task of this sort is the spatial or symmetry span (Shah & Miyake, 1996), in which participants make symmetry judgments and then have to remember spatial locations. They are presented with a series of letters, numbers, or shapes, one at a time, appearing in different spatial
locations or orientations. The original version of the task, for example, involved remembering the original orientation of each letter while simultaneously making decisions as to whether the letter was normal or mirror-imaged. At the end of each set, the participant recalled the orientation of each of the letters on a grid.

In addition to complex span tasks, other tasks used to tap WMC involve alterations of traditional STM tasks to require both storage and processing. The task of this sort that is probably most widely used is the backward digit span task (Botwinick & Storandt, 1974), which serves as the main test of general WM in the two latest versions of the Wechsler Adult Intelligence Scale (WAIS) (Gathercole & Alloway, 2008). The task requires participants to repeat back a sequence of random aurally presented digits in the reverse order, which demands both storage and manipulation of the digits. The main difference between this type of task and the complex span tasks is that the processing component here – in this case manipulating the order of the digits – involves the same material that is recalled. In contrast, the information recalled in the complex span tasks (e.g., single words or letters) is different from the information used in the processing task (e.g., sentences, mathematical equations). Another task similar to the backward digit span is letter-number sequencing, which was first introduced as a supplemental WM task in the third version of the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 1997) to enhance the assessment of WM and attention. The task involves listening to a series of alternating letters and numbers and then recalling them, placing the numbers in ascending numerical order followed by the letters in alphabetical order. Studies have indicated that letter-number sequencing is a valid measure of overall WMC that is not only verbal in nature, but also appears to activate visuospatial memory, perhaps because participants use visuospatial memory to recognize the letters and visualize them as well as the numbers during the dual-part
reorganization process (placing numbers in numerical order and alphabetizing) (Crowe, 2000; Haut, Kuwabara, Leach, Arias, 2000). In addition to backward digit span and letter-number sequencing, there are a number of other variations that similarly require processing on the target recall items themselves, such as the alphabet span (Craik, 1986) and the subtract 2 span described in Waters and Caplan (2003).

A potential advantage for using tasks such as letter-number sequencing, backward digit span, and operation span in L2 research is that they implicate substantially less language use and knowledge as compared to the reading span task (Juffs & Harrington, 2011). The inclusion of a strong language component renders it difficult to determine the degree to which individual differences in language ability or proficiency, particularly if the task is conducted in the L2, lead to differences in performance (Conway et al., 2005). This potential confound can be lessened, though, by administering the reading span task in the native language since evidence suggests that WMC is language independent (Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993). Although the use of less language-dependent WM tasks is increasing in L2 research, the vast majority of studies published to date have employed a reading span task, as will be seen in the review of literature below in Section 2.5.3.

For the current study, I assessed the WM capacity of my participants using three different measures described above (reading span, operation span, and letter-number sequencing) pace the methodological prescription of several recent studies concerned with the reliability of WM measures (e.g., Caplan et al., 2007; Conway et al., 2005; McCabe et al., 2010). As Conway et al. (2005) states, “WM span tasks are not perfect or process pure. Given such imperfection, an optimal research strategy is to administer multiple WM span tasks and then use the average (or weighted average) of scores on all the tasks as the measure of WMC (p. 780).” Moreover, as
McCabe et al. (2010) argue, in that each WM task measures task-specific skills in addition to the common executive attention component, the problem with using one task is that it remains unclear the degree to which task-specific abilities are influencing correlations as opposed to executive attention. Another reason Conway and colleagues mention for using multiple measures is that outliers can usually be detected and the impact of their data can be controlled. For instance, a participant who is overly anxious about math could perform poorly on the operation span even though she has high WMC. Results revealing a high score on reading span and letter-number sequencing but a low score on operation span would draw attention to this participant as an outlier given that the three measures should correlate with each other.

Although WM was measured with all three tests, for this dissertation I only analyzed and report the results from one task (as most studies reported below have done), the letter-number sequencing task. I opted to use this task because unlike the reading span task and even the operation span task (which incorporates individual words), it is virtually non-linguistic in that it merely requires the manipulation of letters and numbers. Thus, this task is least likely to reflect language-specific abilities, and as McCabe and colleagues (2010) argue, using a task without a language-related processing component avoids the potential confound of language ability affecting scores. Thus, of the three tasks used, the letter-number sequencing task could arguably be the best measurement of the executive attention component of WM that is central to domain-free frameworks of WM (e.g., Kane et al., 2007) and appears to play an important role in language processing and particularly language learning, as will be discussed in Section 2.5.3. Further support for this claim comes from the fact that, as mentioned above, this task was designed and tested specifically to enhance the assessment of the attentional component of WM in the revised version of the WAIS (Weschsler, 1997). The letter-number sequencing task is also
ideal in that its relative language independence lends itself best to future replication and
crosslinguistic studies carried out with learners of other L1s for comparison with results from the
L1 English participants used in the current study. While beyond the scope of this dissertation, I
will in future work examine the results from the other two WM tests in order to compare the
predictive power and reliability of the three measures, and then perhaps use a composite score
comprised of the average score across the three measures, in line with the methodological
suggestion above.

2.5.3 WM in L2 Processing and Learning

There has been substantial research throughout the past two decades exploring potential
links between WMC and L2 learning and use in hopes of shedding light on the nature of
individual differences in L2 acquisition. Many of the tasks described in Section 2.5.2 have been
used to measure WMC in L2 studies, although nonword repetition tasks and reading span have
been the preferred measure of phonological STM and overall WM respectively. This research
has resulted in increasing evidence for the role of WM in L2 learning and use, which has in turn
implicated these variables as important components of language aptitude and as potential
predictors of L2 learning success (e.g., Dörnyei, 2005; Linck & Weiss, 2011; Miyake &
Friedman, 1998). In particular, individual differences in WM have been associated with L2
reading (e.g., Harrington & Sawyer, 1992), writing (e.g., Adams & Guillot), speaking (e.g.,
O’Brien, Segalowitz, Freed, & Collentine, 2007), and vocabulary development (e.g., Speciale, N.
Ellis, Bywater, 2004). There is also evidence that WM influences L2 (morpho)syntactic
processing (e.g., Miyake & Friedman, 1998) and grammar learning (e.g., Brooks et al., 2006).
The purpose of this section is to review some of the main studies that have investigated the role
of WM in these latter two domains, since they are the focus of this dissertation. As mentioned in
Section 2.5.2, a considerable amount of the L2 research on WM has focused solely on the role of phonological STM. Yet, given the presumed role of executive attention in (morpho)syntactic processing and explicit grammar learning, the focus of most of the literature on these topics, as well as my own focus, is to examine the effects of WM in general.

2.5.3.1 Studies on WM and L2 Processing of (Morpho)syntax

As discussed in relation to WM models in Section 2.5.1, there is robust evidence in monolingual research for the involvement of WM in reading comprehension (e.g., Daneman & Carpenter, 1980; 1983; Daneman & Green, 1986; Engle, Cantor, & Carullo, 1992; see Daneman and Merikle, 1996, for a meta-analysis of early studies). Although the L2 research on the topic is considerably less, there is good evidence that WMC also influences L2 reading comprehension (e.g., Harrington, 1992; Harrington and Sawyer, 1992; Leeser, 2007; Walter, 2004). However, the evidence is not as robust and definitive with respect to the role of WM in (morpho)syntactic processing. In L1 research, there is a good amount of evidence suggesting that effects of WMC are at least evident in high load processing conditions that require maintenance and integration of elements across substantial distance to resolve syntactic ambiguities (e.g., Felser & Roberts, 2007; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; Roberts, 2007; Vos, Gunter, Schriefers, & Friederici, 2001). In light of this finding and the fact that processing the L2 as a late learner is generally much more effortful than L1 processing, it seems reasonable to predict that WMC may have a stronger influence on L2 as compared to L1 real time (morpho)syntactic processing. Yet, as the following review will show, there is conflicting evidence among the body of research that has addressed this issue.

Miyake and Friedman (1998) was among the first studies to explore how WM affects L2 syntactic processing and whether cue preference might mediate this relation. They tested
Japanese-English learners (with over six years of English instruction) on digit span tests in the L1 and L2, listening span tests in the L1 and L2, an agent identification task to assess cue preferences, and a syntactic comprehension task where learners listened to complex sentence structures in English and drew diagrams to depict the thematic roles of the nouns in each sentence. Previous studies had shown that Japanese learners of English struggle to rely effectively on word order to process English syntax (e.g., Kilborn & Ito, 1989), a finding that Miyake and Friedman interpreted within the framework of the competition model (MacWhinney & Bates, 1989). Native speakers of Japanese rely heavily on case markings, less so on animacy, and very little on word order to interpret Japanese sentences. Since English lacks their preferred cue (particle-based case markings), Japanese learners often rely on their second preferred cue, animacy, when processing English sentences. Thus, they have to adjust their cue preferences and learn to rely on word order, the most crucial cue in English. Consistent with findings in L1 research mentioned in Section 2.5.2, Miyake and Friedman found no correlations between the STM measures and cue preference or syntactic comprehension. Both L1 and L2 WM tests did, however, correlate with the two measures of syntactic processing. The model produced by their path analysis of the correlations revealed a direct connection between L1 WM and L2 WM, corroborating Osaka and colleagues’ (1992, 1993) findings that L1 WM and L2 WM share resources. More interestingly, L2 WM was found to directly influence both cue preferences (the degree to which they were native-like) and syntactic comprehension, and cue preferences in turn also affected syntactic comprehension. High-span learners showed more nativelike sensitivity to word order in processing English sentences as compared to mid-span learners, who in turn were more sensitive than low-span learners. The researchers claimed that learners with higher WMC are better able to rely on word order because this global cue places more demands on WM as
compared to other cues that are locally processed, such as their preferred L1 case-marking cue. Miyake and Friedman interpreted their results as evidence of an important role for WMC in the L2 acquisition and processing of appropriate linguistic cues, and thus more generally, in the comprehension of complex sentence structures in the L2. The researchers took their findings as support for their claim that WM for language may be the most important component of language aptitude.

The seminal study by Miyake and Friedman provided some of the first evidence that the effects of WMC may extend beyond global L2 reading comprehension to also constrain more specific L2 processing abilities such as syntactic parsing. Another offline study that lends some support for this finding is Leeser (2007), which revealed an indirect role for WM in the processing of L2 verbal morphology. The study examined how learners’ familiarity with text topics and individual differences in WM affect learners’ overall comprehension as well as their processing of Spanish future tense morphology, a novel grammatical form. In addition to completing a computerized version of Waters and Caplan’s (1996b) reading span test in their L1 English, the participants carried out recall protocols to measure their comprehension of familiar and unfamiliar passages. They were also administered form recognition and tense identification tasks to assess their processing of temporal morphology. In line with previous research, results from the recall task revealed that WM does play a role in comprehension. However, higher WMC was only beneficial if the topic of the passage was familiar to the participants. With respect to the processing of future tense morphology, on the other hand, WMC enhanced morphological processing only for passages with unfamiliar topics. In these conditions, higher WMC appeared to compensate for a lack of prior knowledge about the topic, resulting in better performance on the form recognition tasks. Thus, although WM influenced overall
comprehension as well as the processing of novel verbal morphology, topic familiarity was the strongest predictor of performance. The limited role of WM in morphological processing here may reflect more specifically the limited effects of WM on incidental processing or learning, a topic we will return to in Section 2.5.3.2.

Over the past decade there has been a surge of interest in exploring the role of WM in the online (or real time) processing of L2 (morpho)syntax, but the findings are somewhat mixed. Some of the first of these studies were conducted by Juffs (2004, 2005), who examined whether WMC could account for individual differences in the processing of Garden Path sentences, which are known to impose a very high processing load. The participants included L1 Chinese, Japanese, and Spanish learners of English living in the United States, who varied in proficiency, as well as native speakers of English. They completed a Daneman and Carpenter (1980) type reading span test in L1 and L2, a word-span test in L1 and L2, and a self-paced reading task involving grammaticality judgments. Corroborating previous findings, Juffs found strong correlations between L1 and L2 reading-span scores. However, neither correlational analyses nor analyses performed by grouping participants according to span scores revealed any effects of WM on mean reading time scores at the point in the Garden Path sentences where the processing load is greatest (i.e., the critical disambiguating verb in sentences such as After the children cleaned the house looked very neat and tidy). This contradicts findings from L1 online studies that revealed strong effects of WM in similar contexts where processing load is highest (e.g., Just et al., 1996). Based on STM word-span scores, Juffs did, however, find that low-span learners are slower overall in processing as compared to high-span learners, which he suggested may be due to chunking ability, a skill that may be more associated with word span measures given its relation to vocabulary acquisition. Juffs concluded that the lack of effect for WM in his data is
consistent with Waters and Caplan’s (1996a) view of the general verbal WM system – which is tapped by complex WM span measures – as being independent of the separate verbal WM system required for online sentence processing.

In a subsequent paper, Juffs (2005) reported the results of other experimental conditions examined at the same time as the Garden Path condition in Juffs (2004). Thus, the participants and the procedure were identical, but the target structures involved ungrammatical wh-extractions violating island constraints (e.g., *Who did Tom hear the woman who saw __ on television?) and grammatical long distance wh-movement with subject and object extractions from finite clauses (e.g., Who does the nurse know __ saw the patient at the hospital? and Who does the nurse know the doctor saw __ in his office?) and nonfinite clauses (e.g., Who does the boss expect __ to meet the customers next Monday? and Who does the boss expect to meet ___ next Monday?). Unlike the previous study, Juffs considered the role of the L1, predicting that the participants may process the L2 structures differently based on the canonical word order in their L1s (Chinese and Spanish are predominately SVO whereas Japanese is SOV). Consistent with Juffs (2004), the analyses revealed no effects of WM in the L2 processing of these complex structures. They did, however, indicate effects of the L1. Japanese speakers experienced the greatest difficulty, followed by the Chinese speakers, which Juffs attributed to the fact that neither of these languages have wh-movement, and in addition, Japanese has verb-final word order. Juffs acknowledged, though, that because the three groups were not matched in proficiency, these results must be interpreted cautiously. Juffs restricted this caution to the role of the L1, yet collapsing the data across such varying levels of proficiency (with Japanese learners at a particularly low level based on proficiency measures and reading time) also leaves open the possibility that differences in proficiency obscured any potential effects of WM. The different
language backgrounds may have also been a potential confound when considering the role of WM. Moreover, the fact that Juffs failed to even find effects in the English-speaking control group belies a rather well-documented finding in L1 processing research (see for e.g., Novick, Trueswell, and Thompson-Schill, 2005). Juffs and Harrington (2011) acknowledged that the lack of effects of WM in Juffs (2004, 2005) studies could be due to the manual administration of the reading span task and/or the use of the absolute span method in scoring (scores based on level reached and thus restricted in range from two to six, as in Daneman and Carpenter’s original version). Moreover, scoring was based only on words recalled, thus reflecting only the storage function of WM. Given that Juffs was specifically examining L2 processing, of sentences that incur a great processing load nonetheless, it may be critical that the measurement of WM also take into account reaction times and accuracy for the processing component to even potentially discover reliable relationships between WMC and processing performance.

In another study investigating the processing of wh-dependencies in L2 English, Felser and Roberts (2007) also found no effects for WM. Using a cross-modal picture priming task (involving auditory presentation of sentences and pictures), they examined the processing of complex sentences containing indirect-object relatives (e.g., John saw the peacock to which the small penguin gave the nice birthday present in the garden last week) by advanced Greek-speaking learners of English and native English speakers. The participants also performed an adapted version of Daneman and Carpenter’s (1980) reading span in their L2 English. Corroborating findings from previous studies, the results indicated that WM influenced antecedent priming in both native-speaking adults and children. However, WM had no influence on L2 processing, which also differed notably from the L1 processing of both high- and low-span natives. The WM data are subject to question, though, for two important reasons. First, the L2
learners completed the reading span test in their L2, which generates a potential confound with participants’ English proficiency. Moreover, as was the case in Juffs (2004, 2005), the reading span test was manually administered and scoring was based only on words recalled, and thus scores did not reflect performance on the crucial processing component.

In a study that reexamined the processing of long-distance *wh*-questions using stimuli similar to that of Juffs (2005), Dussias and Piñar (2010) found that WM does in fact constrain L2 syntactic processing. Like Juffs, Dussias and Piñar also used a self-paced reading task to test L2 learners of English with L1 Chinese, which has no overt *wh*-movement, as well as English monolinguals. However, in contrast to Juffs (2004, 2005) and Felser and Roberts (2007), the researchers manipulated plausibility to investigate how L2 learners use this information to recover from an initial misparse. The stimuli consisted of four conditions controlling for extraction site and whether the *wh*-word could be a plausible object filler for the main clause verb: subject extraction-implausible (*Who did the police declare __ killed the pedestrian?*), object extraction-implausible (*Who did the police declare the pedestrian killed __?*), subject extraction-plausible (*Who did the police know __ killed the pedestrian?*), and object extraction-plausible (*Who did the police know the pedestrian killed __?*). The participants also completed the Waters and Caplan (1996b) reading span test in English, which includes measurement of reaction times and accuracy in plausibility judgments as well as correctly recalled words.

Consistent with previous studies, the results indicated that all participants experienced more difficulties with subject-extraction sentences as compared to object-extraction sentences in both plausibility conditions. However, only the high-span L2 learners’ processing of the more difficult subject-extraction sentences resembled that of native speakers. The high-span L2 group and the native speakers displayed longer reading times when the *wh*-word was a plausible object of the
main verb, indicating that they were exploiting plausibility information to recover from their initial misanalysis. The low-span L2 learners, on the other hand, did not show this pattern. The researchers interpreted the findings as evidence that relatively high WMC is necessary for L2 learners to access and integrate syntactic and semantic information during L2 processing in a nativelike way. Although the reading span measure was in the L2, creating a potential confound with L2 proficiency, the effects of WM were rather robust, and given the high-level of L2 proficiency of the participants, this factor may have been inconsequential.

Havik, Roberts, van Hout, Schreuder, and Haverkort (2009) provided additional evidence for the role of WM in the L2 online processing of complex sentences containing ambiguities. Using self-paced reading, Havik et al. examined the processing of subject-object relative clause ambiguities by intermediate L2 learners of Dutch with L1 German. Their objective was to determine whether the L2 learners could exploit morphosyntactic information like native Dutch speakers to disambiguate between a subject-relative and an object-relative reading in sentences. To assess WM, both the native-speaking control group and the L2 learners completed a computerized version of the Daneman and Carpenter (1980) Reading Span test in Dutch, and the L2 learners completed an additional test in their L1 German. WM results revealed similar mean scores for the native Dutch speakers and the L2 learners when the task was completed in their respective L1s. The L2 speakers’ scores were significantly lower when they completed the task in L2 Dutch. The offline verification responses following each sentence indicated that all participants (native speakers and L2 learners alike) had more difficulty with object- as opposed to subject-relative clauses. However, only the high-span learners patterned like the native Dutch speakers in showing an online processing advantage for subject relatives in the short condition. The processing behavior of the high-span learners specifically resembled that of the low-span
native speakers for the long conditions. Both of these groups were considerably less accurate than high-span natives for object relatives in the long condition, and neither high-span learners nor low-span natives extended their processing preference for subject relatives to the long condition as high-span natives did. However, in the second experiment where participants’ attention was no longer focused on the experimental manipulation through constant verification questions, none of the L2 learners patterned like the native speakers, and the effects of WMC were only observed in native speakers. Thus, Havik et al. postulated that WM might only affect processing decisions when task demands are high (in which case high enough WMC can even result in native-like processing), but not when just processing for comprehension without structural information to guide parsing decisions. While this may be the case, such a conclusion would be presumptuous given the level of proficiency and limited immersion experience of the participants, who had only received four and a half weeks of intensive Dutch instruction since their recent move to the Netherlands. As the researchers recognize, a more highly proficient L2 group could quite possibly exhibit native-like processing in both contexts with high enough WMC.

As discussed briefly in Section 2.2.2, some recent studies have also investigated the influence of WMC in the online processing of nominal and verbal agreement in the L2. For convenience, these studies will be re-explained in this section, focusing on the WM results. For instance, Sagarra (2007b) examined the role of WM in the online processing of gender agreement by low proficient English-Spanish learners. The experimental tasks included a Waters and Caplan (1996b) version of reading span in L1 English as well as a self-paced moving window test in which participants read sentences containing gender agreement and gender violation within a determiner phrase (e.g., *La mujer lava la blusa blanca/*blanco en la cocina
‘The woman washes the[Fem] blouse[Fem] white[Fem]/*white[Masc] in the kitchen’) as well as across clauses (e.g., *La mujer lava la blusa que era blanca/*blanco en la cocina ‘The woman washes the[Fem] blouse[Fem] that was white[Fem]/*white[Masc] in the kitchen’) and then answered a yes/no comprehension question after every sentence. The results revealed that high-span learners were more sensitive than low-span learners to gender agreement (as evidenced by an increase in mean reading times for adjectives that disagree in gender with the noun) and significantly more accurate in their responses to the comprehension questions. In a subsequent study following a similar procedure, Sagarra and Herschensohn (2010) examined the processing of gender as well as number agreement in both beginning and intermediate English-Spanish learners as well as Spanish monolinguals. In addition to the reading span test and the self-paced reading task, the participants also completed an offline grammaticality judgment task. Results indicated that beginning learners were not sensitive to gender or number agreement errors, whereas intermediate learners showed sensitivity to both types of errors during self-paced reading (i.e., longer reading times for disagreement as compared to agreement). No effects were found for WM among native speakers or beginning learners, which the authors suggested is likely due to their performance at ceiling (natives) and floor (beginners). However, WM was found to influence intermediate L2 learners’ processing. While there were no differences between high- and low-span learners for number agreement, high WMC within the intermediate group was associated with greater online sensitivity to gender agreement, which is more cognitively taxing than number agreement. Moreover, intermediate learners with higher WMC were also more accurate on offline grammaticality judgments for gender agreement. The authors concluded that late learners can exhibit nativelike processing of gender and number agreement...
once they attain a certain level of proficiency, but that individual differences in working memory modulate processing ability.

Sagarra (2007c) also investigated whether WM modulates L2 processing of redundant verbal morphology. Specifically, using a self-paced reading task with comprehension questions based on meaning, the researcher examined third-semester English-Spanish learners’ processing of adverb-verb tense agreement and adverb-verb tense disagreement (e.g., Ayer el estudiante miró/*mira una película de terror en el cine. ‘Yesterday the student watched /*watches a horror movie at the theatre’). The overall results revealed no significant differences between mean reading times on verbs in the tense agreement condition and verbs in the tense disagreement condition, indicating that the L2 learners were generally not sensitive to the tense violations. However, working memory results revealed that the non-significant differences in processing agreement versus disagreement were in fact due to the insensitivity of low-span learners. High-span learners, on the other hand, were able to process tense disagreement significantly more than low span learners. Thus, Sagarra concluded that L2 learners with basic L2 knowledge do not process redundant verbal morphology when focused on comprehending meaning unless they have high WMC.

LaBrozzi (2009, 2011) used an eye-tracking experiment to further explore the L2 processing of redundant verbal morphology for temporal reference. He investigated whether immersion experience and cognitive abilities (overall working memory and inhibitory control) might influence processing of morphological and lexical cues in L2 Spanish. The sample pool consisted of two groups of intermediate-level English-Spanish learners, matched in proficiency: classroom learners with and without study-abroad experience. The experimental sentences consisted of the two conditions used in Sagarra (2007b) described above, but LaBrozzi also
manipulated the position of the adverb so it was in sentence-initial position in half of the stimuli and in post-verbal position in the other half. After reading each sentence, the participants answered comprehension questions based on meaning. The findings revealed that classroom learners relied predominately on lexical cues (temporal adverbs) to resolve a tense conflict, whereas learners with immersion experience exhibited less dependence on lexical cues and greater reliance on verbal morphology. Moreover, WMC, assessed by a Waters and Caplan (1996b) version of the task, was found to modulate the processing of verbal morphology among study abroad learners, but not classroom learners. LaBrozzi interpreted these findings as suggesting that learners with higher WMC are better able to attend to verbal morphology in the immersion setting where they are exposed to exponentially more input than in the classroom setting. Results from the Simon task revealed no effects for inhibitory control, independent of WM, in the processing of lexical or verbal cues. The overall findings suggest that immersion experience, particularly in conjunction with high WMC, can facilitate L2 processing of verbal morphology. While clearly important, the extent to which these factors are beneficial cannot be determined since the researcher used a between-subjects design with separate groups of learners, rather than examining the same group of participants before and after immersion experience.

In sum, this growing body of research exploring potential links between WM and L2 processing of (morpho)syntax has led to largely mixed results. Some support for Caplan and colleagues’ (1996a, 2007) independent WM system for online syntactic processing comes from Juffs (2004, 2005) and Felser and Roberts (2007), who found no effect of WMC on L2 syntactic processing in high load conditions. Felser and Roberts (2007) did, however, find effects of WM on L1 processing under such conditions, as have others mentioned above, which is not consistent with Caplan and colleagues’ position and has led to proposals for shallow online parsing in the
L2 independent of WMC (Clahsen & Felser, 2006; Felser & Roberts, 2007). However, Dussias and Piñar (2010) and Havik et al. (2009) indicated an important role for WM in complex syntactic processing in the L2, where learners with high WMC, but not low WMC, managed to resemble natives in certain aspects of processing. Moreover, there is good evidence, albeit limited, that WM plays at least some role in the L2 processing of nominal and verbal morphology (Sagarra, 2007b, 2007c; Sagarra & Herschensohn, 2010; LaBrozzi, 2009). Yet, taken together, the results from this body of research reviewed do not allow us to definitively discern a role for WM in L2 (morpho)syntactic processing for two reasons. First, when there are effects of WM in L2 processing, they often seem to be mediated by other factors including proficiency level, learning context, and prior knowledge (including L1 influence). For instance, Juffs and Harrington (2011) postulate that L1 influence overpowers the within-group differences based on WMC in L2 processing. Second, although all of the processing studies reported used a reading span task to measure WM, there were drastic differences across the versions, administration procedures, and scoring methods, as should be clear based on the discussion above. This not only renders it difficult to compare results, but it also leaves open the possibility that in some cases reported findings for WM were highly influenced by methodological factors. For instance, it is possible that the failure to find reliable effects of WM in some studies was due to administering the task in the L2 (Felser & Roberts, 2007) or neglecting the processing component of the task and only considering storage when measuring WMC (Felser & Roberts, 2007; Juffs, 2004, 2005). Studies that found an effect of WM despite one of these methodological problems (e.g., Dussias & Piñar, 2010; Havik et al., 2009; Miyake & Friedman, 1998) could potentially have found even stronger relationships between WM and L2 processing.
had they been avoided. Thus, given the disparity of findings, more research is clearly needed in order to better understand how WM may be implicated in this important domain of SLA.

2.5.3.2 WM and Aspects of L2 Learning

Compared to the L2 processing research outlined above in Section 2.5.3.1, research exploring the role of WMC in language learning has rather consistently indicated an important relationship between WMC and more explicit aspects of the L2 learning process. Although this body of research is still limited and also suffers from methodological problems outlined in the previous section, it reveals a clear role for WM when learning involves the conscious, attentional control processes regulated by the central executive construct of WM (e.g., Kane et al., 2007).

2.5.3.2.1 WM and Explicit Instruction

Several recent studies indicate the importance of WM in learning from explicit instruction. French and O’Brien (2008), for instance, provided evidence that phonological memory, or verbal STM, is involved in explicit-grammar learning in a classroom situation. In light of the strong evidence for a relationship between verbal STM and vocabulary learning, as discussed previously, one of the primary aims of French and O’Brien’s study was to determine whether there is also a relationship between verbal STM and grammar learning that is not mediated by lexical knowledge. The participants were 11-year-old French speakers in an intensive 5-month English language program. The tasks included an English and Arabic nonword repetition task, a measure of nonverbal intelligence, receptive and productive vocabulary measures, and a discrete-point test that assessed the morphosyntactic structures that had been explicitly taught (tense, aspect, morphological inflections, negation, questions, adjective order, question form, and possessive determiners). Hierarchical regression analyses revealed that both nonword repetition tasks significantly predicted gains in performance on the
grammar test over the course of the program, and this relationship was not mediated by nonverbal intelligence, L2 contact, or lexical knowledge. Following emergentist and connectionist approaches to language learning (N. Ellis, 1996, 1998), French and O’Brien propose that the independent contribution of verbal STM to grammar learning makes sense when considering the similarity between learning new morphosyntactic structures and learning new vocabulary. The researchers concluded that the role of verbal STM in grammar learning may involve establishing long-term representations of novel grammatical morphemes, but it more likely aids in storing in long-term memory, or learning, accurate *sequencing* of these morphemes. Their interpretation of the role of verbal STM remains speculative, though, especially considering that gains in morphosyntactic knowledge were determined based on improvement on a multiple-choice test.

Early evidence for a link between overall WM and language learning under explicit conditions comes from Ando, Fukunaga, Kurahachi, Suto, Nakano, and Kage (1992, described in Mackey et al. 2002). This study examined the role of WM in L1 Japanese fifth graders’ ability to learn from explicit form-focused aural instruction in L2 English. WM was measured using both reading and listening span tasks. Ando and colleagues found that WM capacity predicted the participants’ relative success with L2 learning after nine hours of the form-focused instruction, although this relationship was only evident on the two-month delayed posttest. Nonetheless, these findings suggest that learners with higher WM capacity were better able to attend to and retain in memory forms and explicit, metalinguistic information while simultaneously comprehending language.

Kormos and Safár (2008) provided additional evidence for a relationship between overall WM capacity and language learning under explicit conditions. This study examined the language
learning achievements of Hungarian learners of English after one year of intensive instruction, which the researchers described as involving primarily explicit explanation of grammatical constructions and vocabulary followed by practice. The vast majority of the participants were classified as beginners at the start of the program, and given the small sample size and less reliable results for the intermediate learners, only the results for the large beginner group will be reported here. They were tested on a general proficiency test, the Cambridge First Certificate Examination, which included assessments in the five major skill areas (linguistic knowledge, reading, writing, speaking, and listening). The participants performed a non-word repetition task to measure verbal STM memory and many of them also completed a backward digit span test to assess general WMC. Results revealed no significant correlation between these two measures, which the authors interpreted as further support for the separability of verbal STM and general WMC. Moreover, verbal STM was not related to learning in any of the skill areas or overall proficiency. However, scores on the backward digit span task correlated strongly with overall language proficiency test scores as well as with gains in reading, listening, speaking, and use of English (evidence of both vocabulary and grammar learning). Kormos and Safár interpreted this finding as support for the important role overall WM plays in explicit learning processes (involving memorization and application of metalinguistic rules) that place great demands on learners’ attentional control.

In another recent study employing a longitudinal design, Linck and Weiss (2011) provided additional evidence for the role of overall WM in explicit L2 acquisition. The participants included English-speaking university students enrolled in either a first-semester German course or a third-semester Spanish course. They all completed measures of L2 proficiency, motivation to learn the L2, WM (measured with a modified version of Turner and
Engle’s (1989) operation span task), and inhibitory control (measured with the Simon task). The proficiency tests were given at the beginning and end of the semester, and they assessed explicit knowledge of both L2 vocabulary and grammar. Results revealed that L2 motivation, SAT scores, grade-point average, and WM all predicted performance on the initial proficiency tests, and WM remained a significant predictor even when the influence of the other factors were controlled for in the regression model. Crucially, though, the relationship between WM and proficiency was almost twice as strong at the end of the semester when participants were retested, which the researchers suggested may reflect a more robust effect of WM at higher levels of proficiency. WM was in fact the only reliable predictor of learning across the testing sessions; the higher the WMC, the greater the increase in L2 proficiency. The authors interpreted the results as direct evidence supporting the claim that the executive control function of WM is an important predictor of explicit L2 learning and thus an important aspect of L2 aptitude.

2.5.3.2.2 WM and Intentional Induction

Further evidence for a relationship between WM and conscious, explicit learning processes involving attentional control is provided by experimental studies that have examined the learning and generalization of artificial or novel grammar. Robinson (1997) is among the first of the studies that directly tested either phonological STM or general WM, rather than indirectly inferring a relationship based on a task that seemed to implicate WM (e.g., Williams, 1999). In Robinson’s study, Japanese, Korean, and Chinese learners of English (intermediate level) were exposed to English sentences based on two unfamiliar syntactic rules under one of four training conditions: implicit (memory task), incidental (comprehension task), rule-search (task promoting rule induction), instructed (rule explanation plus metalinguistic task). Rule learning was assessed via a grammaticality judgment task. Phonological STM was assessed by the paired-associates
memory subtest on the Modern Language Aptitude Test (Caroll & Sapon, 1959), in which participants viewed a list of paired associates (English and Kurdish words) for three minutes and then completed a multiple-choice task in which they selected the correct Kurdish translation. Results revealed that verbal STM was strongly related to rule learning, but only among participants in the rule-search and instructed groups. Provided that the paired-associates task does in fact reflect verbal STM and scores are not overly influenced by prior language knowledge or rehearsal strategies, this study offers evidence of a connection between verbal STM and explicit grammar learning that entails consciously applying or intentionally inducing new metalinguistic rules.

Further evidence for this relationship comes from an experiment conducted by Williams and Lovatt (2003). English-speaking participants were tested on their ability to induce determiner-noun agreement rules after exposure to noun phrases in an artificial language with little resemblance to languages participants knew. Phonological STM was measured with a nonword span task that involved immediate recall, in the order of presentation, of three nonwords at a time from the target artificial language. Participants then learned the meaning and form for all of the articles and nouns that would be incorporated in the training. During the training, grammatical combinations of articles and nouns were presented in sets to the participants, and they were instructed to try and remember them for the recall phases. A generalization task involving production of novel combinations of the articles and nouns was used to assess learning of the underlying determiner-noun agreement rules. Results showed that both verbal STM and previous experience with gendered languages (e.g., Spanish, German) independently influenced rule learning. Debriefings with participants indicated that only those who were most successful at inducing the rules on the generalization test were able to explicitly
describe the determiner-noun system and rules, confirming the researchers’ assumption that the learning processes involved in the training were largely explicit. Moreover, the researchers claimed that subsequent experiments in their lab indicated that the same artificial language was not learnable under implicit or incidental training conditions. Thus, phonological STM clearly had an influence on grammar rule learning under explicit conditions, but the extent to which phonological STM may have been implicated in the process of intentional induction beyond storing the new forms is only speculative, especially given the similarity between the verbal STM task and the training task.

Given the demands that conscious, inductive learning of rules during exposure to input would presumably place on attention, it certainly makes sense that the executive control function of WM would be implicated in such learning processes, and likely even play a more critical role than phonological STM. A series of training experiments by Brooks, Kempe, and colleagues (Brooks, Kempe, & Sionov, 2006; Kempe & Brooks, 2008) provide direct evidence that this in fact the case. The researchers examined the learning of a subset of the multidimensional Russian inflectional system, which consisted of noun morphology for gender (masculine and feminine) and certain cases. In Brooks et al. (2006), the English-speaking participants completed six training sessions, which consisted of listening to the Russian phrases and completing comprehension and production tasks that incorporated feedback. Following the training, the participants were tested on their ability to inflect familiar nouns from training as well as new nouns. Participants also completed a separate vocabulary test, a test of non-verbal fluid intelligence, a nonword span task, and a Daneman and Carpenter (1980) version of the reading span task (with the exception that scoring was based on the total number of correctly recalled words). The results revealed that phonological STM had no effect on learning success, yet both
WM (reading span) and IQ, independently, had a strong influence on learning the nouns during training. Only IQ, though, had a significant independent effect on generalization of the rules to new nouns. Consistent with previous research (Engle et al., 1999; Kane et al., 2004), the researchers claimed that the contribution of WM to success in generalizing the inflectional system reflects the executive attention component, which is tapped by both reading span and the IQ test. Thus, they interpreted the data as evidence that only participants with sufficient attentional resources were able to extract complex morphosyntactic regularities and generate rules while processing and interacting with the input.

Kempe and Brooks (2008) replicated the same training procedure as the earlier study, but they also manipulated the transparency of the inflectional system. While Brooks et al. (2006) only used nouns that were transparently gender-marked in the nominative case (feminine nouns ended in –a and masculine nouns ended in consonants), Kempe and Brooks’ (2008) used non-transparent nouns in one experiment to determine how transparency affects learning of the system, and the influence of WM and IQ. In this study, the researchers also examined the effect of prior experience with languages that have a gender system and they tracked the trajectory of learning during training. Results revealed that learning was more successful when transparent cues to the underlying gender categories were available. Moreover, for the transparent system, IQ, WM, and knowledge of languages with transparently marked gender all had an independent facilitative effect on performance during training and on testing of familiar nouns, yet only IQ made an independent contribution to generalization performance, which replicates the findings of Brooks et al. (2006). Overall performance was much poorer with the non-transparent system, and in light of the difficulty learners faced, they focused more on trying to memorize individual noun-suffix associations rather than attending to distributional patterns. Interestingly, WM was
the strongest predictor in learners’ success in acquiring the nontransparent system, which Kempe and Brooks propose is due to the crucial role of the storage component in memorizing suffixes. Knowledge of gendered languages and IQ were again also related to performance, although IQ was mainly linked to performance later in training when learners had presumably begun to notice distributional patterns of the suffixes. Based on the findings from both experiments, the researchers concluded that the learning of complex inflectional systems appears to involve both the capacity for memorizing unfamiliar items and effective allocation of attention to the distributional characteristics of the input, and their relative impact depends on the degree of transparency of the system. If verbal WM storage capacity is crucial, as the Kempe and Brooks suggest, it is surprising that they failed to find any significant relations between phonological STM and learning in Brooks et al. (2006), which is also why they did not test phonological memory in the subsequent study. In that simple phonological STM has been directly related to pure storage functions of WM, further investigation of STM in explicit learning of inflectional systems seems to preclude any conclusions regarding a critical role for verbal storage in this learning process. Nevertheless, these findings provide good evidence for the role of general WM in learning processes involving intentional induction.

2.5.3.2.3 WM and Feedback on Error

Additional evidence for a facilitative role for WM in learning processes requiring attentional control comes from research examining learners’ ability to notice and make use of corrective feedback. Mackey, Philp, Egi, Fujii, and Tatsumi (2002) were among the first to directly explore this relationship in a classroom-based study involving intermediate Japanese-English learners exposed to corrective feedback during communicative interactions with native speakers. The participants completed a nonword recall task to assess phonological STM as well
as a listening span task in L1 and L2, adapted from Daneman & Carpenter (1980), to assess WM. The pretests, posttests, and treatment consisted of task-based communicative tasks designed to elicit the use of target question structures. During the treatment, the native speaker provided the participant with feedback in the form of recasts. A subset of the participants took part in stimulated recall protocols and questionnaires to assess learners’ noticing of the interactional feedback. The immediate posttests as well as a delayed posttest, completed by those who did not take part in the stimulated recall protocols, measured learners’ development in terms of the question forms. The results revealed that WM was related to learners’ ability to notice the feedback; learners with high WMC reported noticing more recasts during interaction than those with low WMC. With respect to L2 development, the low WMC learners surprisingly appeared to benefit more from the interactional feedback on the immediate posttests. However, the low WMC learners who took the delayed posttests two weeks later did not sustain this development. Conversely, while the high WMC learners did not demonstrate as much progress initially, those who took the delayed posttest all showed significant development at that time. Overall, learners with high WMC were more likely to benefit from the feedback, although this was after a time lapse. Mackey and colleagues suggested that high WMC learners may have processed the feedback more efficiently, comparing it to their own production and existing knowledge, but this process of consolidating the feedback and encoding changes into long-term memory took time. While suggestive, their results are based on a very small number of participants (only 7 performed the delayed posttest). Also, WM results were based on a combined score from the STM task and the two listening span tasks (L1 and L2), rendering it impossible to discern the relative contribution of the phonological store, general verbal WM, and even perhaps L2 proficiency given the use of the L2 WM measure.
In a recent larger-scale study, Mackey and colleagues (Mackey, Adams, Stafford, and Winke, 2010) further explored the relationship between WMC and the production of modified output during task-based communicative interactions. The participants were fourth-semester students of Spanish whose native language is English. They each carried out four interactive tasks with a native speaker who provided feedback based on errors related to a wide range of morphosyntactic forms and vocabulary. The interactions were transcribed and coded for instances of feedback that offered opportunity for learner response as well actual learner responses. WM was assessed with a L1 listening span task that simulated Waters and Caplan’s (1996) reading span task. Regression analysis revealed a significant relationship between WMC and production of modified output; high WMC learners produced more modified output than low WMC learners. A second regression model examining the separate effects of the processing and recall scores of the listening span task indicated that only the recall component was a significant predictor of modified output. Overall WM accounted for a little under 20% of the variation in the production of modified output, implying that other factors are also important for benefiting from feedback. Yet, post hoc analysis revealed differential effects of the WM components on the production of distinct types of modified output. While only recall scores significantly predicted the tendency to repeat recasts, only processing scores significantly predicted the tendency to alter original responses following feedback. This suggests that the influence of WM on modified output may vary for distinct types of modifications and feedback, which calls for future research to better control feedback types and explore the predictive power of phonological STM, WM, and other cognitive tasks using various measures that are both language dependent and independent. Nevertheless, Mackey and colleagues’ study provides good evidence that WMC enhances the ability to make use of feedback to subsequently modify output. They attribute this
relationship to the important role of WM in both storage and attentional control, which are required (although to varying degrees depending on the context) to quickly shift attention away from the meaning-based interaction to focus on form in the feedback, identify the mismatch between the incorrectly produced utterance and the correct structure, and then retrieve and produce the correct structure.

Two studies conducted by Sagarra (2007a; Sagarra & Abbuhl, in press) provide strong evidence that WM also constrains noticing and benefiting from feedback during computer-based tasks that lack meaning-focused interaction. In Sagarra (2007a), first-semester English learners of Spanish completed exercises focused on noun-adjective gender and number agreement, a novel structure for them. The treatment consisted of filling in basic sentences, presented one at a time, with the appropriate adjective. The control group received no feedback, and the experimental group received computer-delivered oral recasts on their errors. All learners completed three written posttests (immediate, delayed one week, delayed one month) that followed the same fill-in-the-blank format as the treatment and included familiar adjectives and new ones. A subset of the participants also carried out oral face-to-face posttests that examined accuracy in production of the target structure and immediate repair in response to feedback. WM was assessed with an adapted version of Waters and Caplan’s (1996) reading span test. Results showed that WM greatly facilitated the development of linguistic accuracy for the group receiving feedback. WM was positively related with target-like production of the target structure on all written posttests as well as in the face-to-face interaction tasks. Moreover, WM predicted the amount of targetlike modified output produced following the interactional feedback. Sagarra and Abbuhl (in press) expanded this line of research using the same basic procedure, but they also investigated feedback type (no feedback, utterance rejection, recasts, enhanced recasts) and
the mode of delivery (written or oral). Results revealed that recasts were the most effective type of feedback overall, and WM and mode influenced their efficacy. Specifically, WM enhanced the performance of the oral recast groups (orally enhanced and unenhanced) on both the written and interactional posttests, but WM was not related to the performance of the written recast group. Among the learners that received oral recasts, high WMC learners were both more accurate with agreement on posttests, and they produced more correctly modified output during face-to-face interaction. Sagarra and Abbuhl suggested that the lack of role for WM among the written recast group might be due to the lower processing demands associated with processing written input as compared to oral input. This proposal certainly follows previous research reviewed that suggests a more important role for WM in more difficult cognitive tasks that place a higher load on attentional resources. As a whole, this research in conjunction with that of Mackey and colleagues provides initial support for the role of WM in noticing, retaining, and benefiting from the rule-based information contained in corrective feedback, at least in the form of recasts. WM appears to constrain learners’ ability to focus on form in a meaningful context, make comparisons between their own interlanguage and the information in corrective feedback, and then subsequently improve their linguistic accuracy. Clearly more research is needed in this area, in particular studies examining the role of WM in learning from other types of corrective feedback frequently provided in classroom and computer-based settings. The current dissertation will examine this issue.

In sum, the research reviewed on WM and L2 learning provides compelling evidence that WM plays an important role in learning from explicit instruction, inductive learning of underlying grammatical rules and categories, and learning from corrective feedback. In all of these cases, the learning processes are rather explicit and require selective attention to retain and
make use of metalinguistic, rule-based information while concurrently comprehending and producing language. These findings corroborate Roehr’s (2008) theory that explicit metalinguistic language processes requiring conscious attention and effort should draw heavily on WM resources while non-conscious or implicit operations may remain largely unaffected by WM limitations.

2.6 The Present Study

The research presented in this chapter has illustrated and raised questions related to late L2 learners’ difficulty processing and acquiring L2 grammatical morphemes. The present section will explain the purpose of my dissertation in light of such. Broadly, the dissertation will examine how certain factors (internal and external) modulate the processing and learning of notoriously difficult L2 morphological cues. More specifically, the experimental design is set up so as to forward three aims:

(1) To shed further light on how learners coming from a morphologically weak L1 process verbal inflections in a morphologically rich L2

(2) To evaluate the efficacy of form-focused training for overriding L1 cue biases and for promoting the processing and learning of L2 inflectional morphology, as well as evaluate the role of feedback on such

(3) To elucidate the role of working memory in processing and learning L2 inflectional morphology

Below I will address these broad issues and the more focused aims in light of the literature review presented above.

While there is disagreement amongst the research on whether L2 learners can achieve native-like L2 morphological processing, the research is unanimous in one respect: there are significant obstacles that modulate learners’ processing and learning of L2 morphological cues. VanPatten’s Input Processing Model (1996, 2004, 2006) and associated offline research
highlight late L2 learners’ tendency to process meaningful lexical cues (e.g., adverbs, explicit subjects) over redundant inflectional morphology that encodes the same meaning, a processing strategy that is disadvantageous for the acquisition of L2 grammatical morphemes. Consistent with this, online studies investigating L2 morphosyntactic processing of agreement indicate that even proficient L2 speakers who demonstrate native-like knowledge of nominal or verbal agreement on offline tasks have failed to show native-like sensitivity to agreement violations in online measures including non-cumulative self-paced reading (e.g., Jiang, 2004), eye-tracking (e.g., Keating, 2009), and ERP (e.g., Chen et al., 2007). This dissertation explores this general problem area by examining early English-Spanish learners’ processing and learning of Spanish verbal morphology for tense and person.

What factors, then, modulate the processing of inflectional morphology? Above I discussed studies that attribute possible contributing factors to: (a) L1-L2 morphological similarity (e.g., Jiang et al., 2011; Tokowicz & MacWhinney, 2005), (b) proficiency (e.g., Osterhout et al., 2008; Wen et al., 2010), (c) linguistic characteristics of the input (e.g., Frenck-Mestre et al., 2009; Keating, 2009; Osterhout et al., 2008), (d) and working memory (e.g., Sagarra & Herschensohn, 2010; Sagarra, 2007). The Associative-Cognitive model provides a framework that unifies these various factors and takes into account the empirical evidence presented in the literature review above by drawing on cognitive abilities, linguistic characteristics of the input, and language experience to explain late learners’ difficulty acquiring a L2. Since I adopt this theoretical approach in this dissertation, I will briefly review this framework in relation to the target structure examined in the current study (Spanish verbal inflectional morphology) and the target population (early L1 English learners of Spanish). Because processing limitations often inhibit attention to multiple cues in the L2 input, learners
have to choose which aspects of the input they will process. The low salience, low-reliability, and frequent redundancy of morphological cues partially explains why they are less likely to be processed and learned than salient lexical cues (adverbs, explicit subjects) encoding the same information. Yet, since these linguistic characteristics are operative in both L1 and L2 acquisition, they themselves cannot be the critical factor for L2 learners’ difficulty. Rather, according to Associative-Cognitive model, the burden rests on prior language experience. As a result of L1 experience, regardless of the language, learners know that there are reliable and salient lexical cues to express time (adverbs) and person (explicit subjects); so if frequently present in the L2 input, these known cues might block the processing and acquisition of inflections. Moreover, the degree to which the learners’ native language makes extensive use of inflections influences sensitivity to these cues in the second language. Learners coming from a morphologically weak language, such as English, are less able to process and learn the inflectional cues than a learner from a morphologically rich language, such as Italian. In short, following this model, it is the conjunction of unfavorable linguistic factors with strong attentional biases based on a life-time of prior L1 usage that render verbal inflections especially difficult for English-Spanish learners.

The process of overcoming L1 cue biases is, in principle, difficult, but things can be compounded by the learning context. First, for many L2 learners, their primary exposure to the L2, at least initially, occurs in the classroom, but research has shown that modified and simplified input often results in the overrepresentation, underrepresentation, and altered pattern of use of L2 cues in the classroom (e.g., Dracos, 2010; Goodall, 2008; Santilli, 1996; Sanz, 1999). Particularly problematic for the processing of L2 verbal morphology is the overuse of lexical cues in ‘teacher talk’ as well as in other learners’ speech. LaBrozzi (2009) found that
exposure to intensive amounts of morphological cues during study abroad helped some intermediate English-Spanish classroom learners, particularly those with higher working memory, rely more on verbal inflections as native speakers do. Yet, study abroad is not always an option, especially in the early stages of L2 learning. Moreover, the fact that studies have shown that many who are otherwise proficient in the L2 and have spent significant time immersed in the target language still fail to show sensitivity to morphosyntactic violations suggests that even exposure to intensive amounts of input might not be enough to override learned attentional biases. Therefore, and in line with all of the above, N. Ellis (2008b, p. 373) claims that “form-focused instruction is a necessary component of SLA.”

With this background in place, the general inertia of this dissertation can now be better explained in relation to three relevant sub-questions. First, how do learners coming from a morphologically weak L1 process verbal morphology in a morphologically rich L2? As explained above, the Associative-Cognitive Model predicts that English learners of Spanish, particularly at lower levels of proficiency, will process lexical cues and ignore inflections when both cues are present. Thus, it would also be predicted that such low-level learners would be insensitive to adverb-verb tense violations as well as subject-verb agreement violations. Some recent experiments reviewed above by N. Ellis and Sagarra have provided direct support for these predictions, corroborating earlier research by VanPatten and colleagues using less direct and less informative offline measures (e.g., think-aloud protocols, written recalls in the L1, multiple-choice recognition tasks, form production tasks). In this dissertation, I employ a research design that will shed further light on how learners coming from a morphologically weak L1 process redundant L2 verbal inflections in a morphologically rich L2.
Next, and more important for this dissertation, what is the efficacy of form-focused training for overriding L1 cue biases and for promoting the processing and learning of L2 inflectional morphology, and what role does corrective feedback have on such? As N. Ellis (2006b, 2008b) stresses, although exposure to intensive amounts of input is crucial at the earliest stages of SLA, it may not be sufficient to overcome learned attentional biases and it may have to be combined with some sort of pedagogical intervention. N. Ellis and Sagarra (2010b) in fact found that preexposure to verbal morphology and textual enhancement of inflections increased L1 English and L1 Chinese (morphologically weak/absent languages) participants’ reliance on and learning of verbal morphology in a short, one-session laboratory study involving a small subset of Latin (3 temporal adverbs and 1 verb in the present, past, and future tense of the first-, second-, and third-person singular). While suggestive, it is unknown whether refocusing learners’ attention through instructional practices can retune the effects of long-term learned attention with real classroom learners that are actually attempting to acquire a complex language. Nor is it known whether such instruction has any durative effects. This dissertation will address these theoretical questions through longitudinal form-focused training, which will allow for the relevant hypotheses testing and be additionally informative for the development of efficacious methods of L2 instruction.

The form-focused training in this study exposes learners to intensive amounts of morphological cues and makes use of two instructional techniques: 1) input activities that lack lexical cues and require processing of morphology for successful interpretation of meaning (similar to structured input in VanPatten’s Processing Instruction), and 2) feedback on error. With respect to the first instructional technique, while there is robust evidence that such activities improve learners’ ability to produce the target form and process it in salient, nonredundant
contexts (e.g., VanPatten, 2002, 2004; Benati & Lee, 2010), research to date has not examined whether this instructional practice is effective in altering learners’ attentional biases and promoting attention to the form in subsequent processing of input with redundant morphology accompanied by lexical cues, as is the case when processing morphosyntactic (dis)agreement. The current dissertation will test this pedagogical practice for this purpose. With respect to the second instructional practice, research suggests that corrective feedback can facilitate the noticing and learning of targeted grammatical forms by helping learners validate or reject interlanguage hypotheses and make correct form-meaning connections (e.g., Doughty, 2001; Gass, 1997). Yet, as the studies reviewed in this chapter illustrate, there is a debate concerning the relative merits of explicit feedback as opposed to more implicit types (recasts, correct/incorrect responses). While the research as a whole leans toward more advantages for explicit metalinguistic feedback (e.g., R. Ellis et al., 2006), the vast majority of the studies focus on feedback during production and the meager amount of research on feedback during processing is completely inconclusive (Rosa, 1999; Rosa & Leow, 2004; Sanz, 2004; Sanz & Morgan-Short, 2004). Moreover, the processing research did not have control groups receiving no feedback to test whether task-essential processing alone may have been sufficient and, given that these studies involved only one session, potential effects of the different feedback types could have been masked due to short treatment durations. The current dissertation addresses these methodological concerns through a longitudinal design that should inform the debate as to whether the provision of any feedback, and specifically explicit metalinguistic information, facilitates the processing and learning of L2 grammar.

Lastly, what is the role of working memory in processing and learning L2 inflectional morphology? Working memory has been shown to play a role in various aspects of L2 learning
and use (Juffs & Harrington, 2011; Williams, 2012). Particularly relevant for the current dissertation is the research investigating whether individual differences in working memory capacity might account for variation in sentence processing as well as explicit grammar learning. With respect to the former, some studies (e.g., Havik et al., 2009; Sagarra and Herschensohn, 2010) have found significant correlations between working memory and language processing ability, whereas other studies (e.g., Juffs, 2004, 2005) show no effects. Notable methodological differences and distinct target structures likely account for the conflicting findings, and they underscore the need for more research to discern the relationship between working memory and processing in a L2. With regard to grammar learning, there is increasingly strong evidence that working memory affects explicit L2 learning processes involving conscious, attentional control, presumably because such processes draw heavily on the executive attention component of working memory. Specifically, research indicates that working memory is involved in learning from explicit instruction (e.g., Kormos & Safár, 2008; Linck & Weiss, 2011), inductive learning of underlying grammatical rules and categories (e.g., Brooks et al., 2006; Kempe & Brooks, 2008), and learning from corrective feedback (e.g., Mackey et al., 2010; Sagarra & Abbuhl, in press). This dissertation extends these lines of research by exploring the possible influence of individual differences in working memory capacity on learners’ cue biases, their online processing of verbal agreement violations, and their ability to learn from form-focused instructional strategies in training.

Thus, to restate the above succinctly, the overarching goal of the current dissertation is to elucidate how both external (form-focused instructional strategies) and internal (working memory) factors modulate learners’ processing and learning of L2 morphological cues. The specific aims of this study, as discussed, are to 1) shed further light on how learners of a
morphologically weak L1 process verbal morphology in a morphologically rich L2, 2) evaluate the efficacy of form-focused training for overriding L1 cue biases and for promoting the processing and learning of L2 inflectional morphology, as well as the role of feedback on such, and 3) elucidate the role of working memory in the processing and learning L2 inflectional morphology. To examine this, late English learners of Spanish received longitudinal processing-based training on both familiar and novel L2 verbal inflections, which are non-salient, often redundant, and considerably less present and important in their native language. In order to assess processing behavior and learning outcomes, the participants carried out four assessments before and after training (pretest, posttest, and delayed posttest design) examining their aural processing of inflections (aural processing task), their production of inflections (written production task), their cue biases (cue reliance task), and their online processing of agreement violations (non-cumulative self-paced reading task). In addition, learners’ working memory was assessed using the letter-number sequencing task. In the following chapter, I outline my method and procedure in detail as well as present the specific research questions and results for each task.
CHAPTER 3: Experimental Study

My objectives in this chapter are to describe the methods and procedures used in the present study, provide descriptive statistics for the L2 learner variables and training accuracy, and present the results for each of the assessment tasks. First, in Section 3.1, I provide an overview of the method and materials, which includes a description of the participant pool, the target structure, the general experimental design, and the screening and proficiency measures, the working memory (WM) task, and the training. Descriptive statistics for these measures will also be presented in this section. Then, in Sections 3.2 – 3.5, I describe the specific method and materials as well as present the results of the analyses conducted for each of the four assessment measures administered before and after the training: the aural processing task (Section 3.2), written production task (Section 3.3), cue reliance task (Section 3.4), and self-paced reading task (Section 3.5). I made the decision to outline the methodology for each assessment immediately prior to the corresponding results to make it easier for the reader to keep in mind the specific procedures, participant exclusion criteria, and scoring methods for the given assessment measure during the presentation of the statistical analyses and results.

3.1 Overview of Method and Materials

In this section, I present the general method and materials. In Section 3.1.1, I describe the participants and the basic inclusion criteria. In Section 3.1.2, I provide an overview of the target structure: verbal inflectional morphology in Spanish. Next, in Section 3.1.3, I present the experimental design of the study, which includes a detailed description of the screening measures, the working memory (WM) task, and the training. Then, the descriptive statistics for these same measures are presented in Section 3.1.4.
3.1.1 Participants

The participant pool comprised 310 novice Spanish students at the Pennsylvania State University who volunteered to participate in the study in exchange for extra credit. At the start of the study, all of the learners were in their seventh or eighth week of a second-semester, computer-enhanced Spanish course with two hours of contact class time per week. Because the course is standardized across individual class sections, all the participants were exposed to the same textbook and curriculum.

For the present dissertation, data was only analyzed from participants who were native English speakers that began learning Spanish post-puberty, had never spent time abroad in a Spanish-speaking country, and had little or no exposure to another language (two years or less of classroom instruction). Additional inclusion criteria required that the participants scored at least 80% on the pre-training vocabulary test, indicating that they had sufficient knowledge of the target vocabulary, and that they exhibited a lack of knowledge of the future tense (the new tense) on the pretest production task. These two tasks and these criteria will be further discussed in Section 3.1.3. There were 264 participants who met these general inclusion criteria, and thus analyses were only conducted on the data from these participants. Moreover, additional exclusions were made for analyses of some of the assessments, and these will be outlined in the section devoted to the specific assessment measure.

Of the 264 participants who met the inclusion criteria, 148 were involved in the control group research with no exposure to training, and 116 participants were involved in the experimental training. The participants who underwent training were randomly placed in groups that differed only in the feedback they received during the training (training materials and procedure are described in Section 3.1.3.3). The breakdown into training groups was as follows:
no feedback (n = 38), yes-no feedback (n = 38), and metalinguistic feedback (n = 40). For the statistical analyses (presented in Sections 3.2 – 3.5) that involved comparison of the three experimental groups to the control group, a randomly selected subset of the control group participants was used so the groups were comparable in size. Specific details regarding the participant pool used for analyses of each assessment are outlined in the corresponding section below.

3.1.2 Target structure: Spanish verbal inflectional morphology

The target structure consisted of Spanish regular verbal morphology in the present, past (preterite), and future tenses. As discussed earlier in Chapter 2, Spanish is a morphologically rich language in which verbs are inflected for person-number, tense, mood, and aspect (in the past tense). Of concern here are inflections for tense and person-number. Spanish verbs have distinct morphological forms for person-number in the present, preterite, and future tenses. This is the case for all three verb classes in Spanish (which in the infinitive end in -ar, -er, or -ir), as exemplified in Table 1 below.

<table>
<thead>
<tr>
<th>Verb Class</th>
<th>Present</th>
<th>Past (preterite)</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td>-AR</td>
<td>escucho</td>
<td>escuché</td>
<td>escucharé</td>
</tr>
<tr>
<td></td>
<td>‘I listen’</td>
<td>‘I listened’</td>
<td>‘I will listen’</td>
</tr>
<tr>
<td></td>
<td>escuchas</td>
<td>escuchaste</td>
<td>escucharás</td>
</tr>
<tr>
<td></td>
<td>‘you–SG-INF listen’</td>
<td>‘you–SG,INF listened’</td>
<td>‘you–SG,INF will listen’</td>
</tr>
<tr>
<td></td>
<td>escucha</td>
<td>escuchó</td>
<td>escuchará</td>
</tr>
<tr>
<td></td>
<td>‘he/she/you–SG,F listen(s)’</td>
<td>‘he/she/you–SG,F listened’</td>
<td>‘he/she/you–SG,F will listen’</td>
</tr>
<tr>
<td></td>
<td>escuchamos</td>
<td>escuchamos</td>
<td>escuchamos</td>
</tr>
<tr>
<td></td>
<td>‘we listen’</td>
<td>‘we listened’</td>
<td>‘we will listen’</td>
</tr>
<tr>
<td></td>
<td>escuchás</td>
<td>escuchasteis</td>
<td>escucharéis</td>
</tr>
<tr>
<td></td>
<td>‘you–PL-INF listen’</td>
<td>‘you–PL,INF listened’</td>
<td>‘you–PL,INF will listen’</td>
</tr>
<tr>
<td></td>
<td>escucharan</td>
<td>escucharan</td>
<td>escucharan</td>
</tr>
<tr>
<td></td>
<td>‘they/you–PL,F listen’</td>
<td>‘they/you–PL,F listened’</td>
<td>‘they/you–PL,F will listen’</td>
</tr>
<tr>
<td>-ER</td>
<td>aprendo</td>
<td>aprendi</td>
<td>aprenderé</td>
</tr>
<tr>
<td></td>
<td>‘I learn’</td>
<td>‘I learned’</td>
<td>‘I will learn’</td>
</tr>
<tr>
<td></td>
<td>aprendes</td>
<td>aprendiste</td>
<td>aprenderís</td>
</tr>
<tr>
<td></td>
<td>‘you–SG,INF learn’</td>
<td>‘you–SG,INF learned’</td>
<td>‘you–SG,INF will learn’</td>
</tr>
<tr>
<td></td>
<td>aprenda</td>
<td>aprendió</td>
<td>aprenderá</td>
</tr>
<tr>
<td></td>
<td>‘he/she/you–SG,F learn(s)’</td>
<td>‘he/she/you–SG,F learned’</td>
<td>‘he/she/you–SG,F will learn’</td>
</tr>
<tr>
<td></td>
<td>aprendemos</td>
<td>aprendimos</td>
<td>aprendemos</td>
</tr>
<tr>
<td></td>
<td>‘we learn’</td>
<td>‘we learned’</td>
<td>‘we will learn’</td>
</tr>
<tr>
<td></td>
<td>aprendéis</td>
<td>aprendistes</td>
<td>aprendéis</td>
</tr>
<tr>
<td></td>
<td>‘you–PL,INF learn’</td>
<td>‘you–PL,INF learned’</td>
<td>‘you–PL,INF will learn’</td>
</tr>
<tr>
<td></td>
<td>aprendan</td>
<td>aprendieron</td>
<td>aprenderan</td>
</tr>
<tr>
<td></td>
<td>‘they/you–PL,F learn’</td>
<td>‘they/you–PL,F learned’</td>
<td>‘they/you–PL,F will learn’</td>
</tr>
</tbody>
</table>
In contrast to Spanish, English is morphologically weak in that there is little inflectional richness in its verbs, as is evident in the English glosses of the Spanish verb forms in Table 1. There is only one person-number contrast for regular verbs, which is the 3rd singular form in the present tense: he/she writes but I/you/we/they/all write. English also makes less use of tense morphology than Spanish. In the regular past tense, verbs are marked with the same –ed [-t]/[-d] inflection regardless of person-number. The future tense in English does not carry any bound inflection, but rather, the auxiliary verb will is placed prior to the verb to denote future. Given that English is morphologically weak, native English-speaking learners of Spanish do not begin the L2 acquisition of Spanish with either a grammar that expects rich verbal morphology or with a parsing system that expects rich surface agreement.

In the current dissertation, 18 target verbs (9 regular -ar verbs and 9 regular -er/-ir verbs) were used to examine English-Spanish learners’ knowledge and processing of verbal morphology. The target verbs were used in all three tenses, and in four of the six person/number combinations: first singular (1SG), first plural (1PL), third singular (3SG), and third plural (3PL). The verbal paradigm was simplified in this manner to make the learning task more manageable, and to avoid the unwieldy challenges of incorporating 2nd person forms in computer-administered input. The target verbs were selected from chapters that had been covered in the course textbook to ensure that all of the students had been exposed to them. Although the participants had different instructors, all of the course sections used the same materials and
completed the same assignments. By the start of the study, the participants had learned and had been tested on both present and past (preterite) tenses. The future tense, though, was an entirely new tense that participants had not yet learned. A new tense was incorporated in order to compare the effects of training on familiar verbal inflections as opposed to new inflections.

3.1.3 Experimental Design

The study was administered over a period of nine weeks in eight sessions (training groups) or three sessions (control group) held in computer laboratory classrooms. Figure 6 outlines the experimental design, which included screening and proficiency measures, four assessment measures (self-paced reading task, cue reliance task, aural processing task, and production task), three working memory (WM) tasks, one inhibitory control task, and five training sessions (experimental groups only). In that the assessment measures were designed to measure the effects of training on learners’ processing and acquisition of verbal inflections, a pretest, immediate posttest, and delayed posttest procedure was adopted, with the delayed posttests administered three weeks after the training was completed. Three versions (A, B, C) of each assessment were created to avoid practice effects from exposure to identical stimuli. While the target verbs, conditions, and overall procedure were identical in all versions of a given assessment, the sentence stimuli were unique. The three versions of the assessments allowed for six possible orderings (ABC, BCA, etc.), and each participant was randomly assigned to one of these test-administration orders.
The specific materials and procedure for each task are described in the corresponding sections below that also include the results, but the general procedure will be outlined here. At the start of the first experimental session, all participants signed a consent form and agreed to the release of their exam grades and final course grade for inclusion in the study. Next, language background and proficiency in Spanish was assessed via (a) a language background questionnaire, (b) a vocabulary test, and (c) a grammar proficiency test. The participants then
completed the pretest assessments as well as two WM tasks (the letter-number sequencing task and operation span), either on the same day (control group) or the following week (experimental groups). The assessment measures were always administered in the following order (shown in Figure 6): self-paced reading task, cue reliance task, aural processing task, and production task. The self-paced reading assessment was always administered first to minimize biasing learners’ attention to form in that the objective of the task was to assess learners’ online processing of agreement violations while they are focused on comprehending overall meaning. Subsequently, the learners completed the cue reliance assessment, which like the self-paced reading task involved experimental stimuli with adverb-verb tense violations and subject-verb agreement violations, but the cue reliance task was designed to draw learners’ attention to form in order to assess their reliance on lexical cues (adverbs / explicit subject pronouns) as opposed to morphological cues (verbal inflections). Next, the learners completed the aural processing assessment, which tested the learners’ accuracy in aurally processing the target verbal morphology in a sentence context void of lexical cues. The final assessment was a short production task, which assessed the learners’ ability to produce the target verbal inflections. While the assessment measures were consistently presented in this same order at all three test times, they will be discussed below in Sections 3.2 – 3.5 in an alternate order that makes more sense given the results and discussion.

The experimental groups underwent their first training session approximately two weeks after they completed the pretest assessments. The remaining four training sessions took place within the following two weeks, with a minimum of 48 hours and a maximum of 72 hours between sessions. Each session lasted between 15-18 minutes, and the objective of the training was to improve visual and aural processing of verbal inflections in sentential context. The
experimental groups then completed the immediate posttest assessments following the final training session. The control group completed the immediate posttests the same week, which was four weeks after the pretests. Following the immediate posttests, the procedure was identical for the experimental and control groups. There was no training or exposure to the experimental materials in the three weeks between the immediate posttests and the delayed posttests. On the day of the delayed posttests, the participants also completed the final WM test (reading span) and the Flanker task, which assessed inhibitory control. In the corresponding sections below, I will describe the materials and discuss results for all of the tasks mentioned in this section (and shown in Figure 6) with the exception of the Flanker task and two of the three WM tasks (Operation span and Reading Span). The data from these three tasks were not used in the current dissertation since individual differences in inhibitory control were not examined here, and the participants’ WM capacity was determined based on their performance on the Letter-Number Sequencing task (described in Section 3.1.3.2). Future work will examine results from the excluded tasks to investigate the role of inhibitory control and to further explore effects of WM as well as the reliability of the WM measures.

3.1.3.1 Screening and Proficiency Tasks

As previously mentioned, the participants completed three screening tasks designed to assess their language background, knowledge of target vocabulary, and their general grammar proficiency in Spanish. Moreover, the pretest of the production assessment was used to test participants’ knowledge of the target grammar (verbal inflections). Each of these tasks will be discussed in turn below, and results will be presented in Section 3.1.4.

The language background questionnaire (see Appendix A) consisted of multiple-choice and open-ended questions that thoroughly addressed the participants’ history with Spanish,
English (if not native), and any other languages they might know. They provided specific information regarding previous formal instruction in Spanish, other language-learning experiences, and any exposure they may have had to Spanish or another non-native languages outside of the classroom. The information from the questionnaire was used to make the exclusions described above in Section 3.1.1 and thus ensure that all the learners included in the analyses were truly native English speakers, that they had no other significant experience with other languages, and that they had received a quantitatively and qualitatively similar amount of L2 instruction in Spanish post-puberty.

The vocabulary test (see Appendix B) was completed online in ANGEL, the university’s course management system, in the computer laboratory setting. The participants were monitored closely, and they were not allowed to use any resources to help them with the task. The vocabulary test required learners to match the 18 target Spanish verbs and 9 adverbial expressions with the English equivalents. This test served to ensure that learners knew the meaning of the target vocabulary used in the experimental tasks so that the results could not be attributed to unfamiliarity with the stimuli. All of the target vocabulary had already been covered in either the participants’ current Spanish course or in the previous Spanish course at the Pennsylvania State University.

The grammar proficiency test (see Appendix C) was also completed online in ANGEL in the computer laboratory, and it was used to measure homogeneity within and between the learner groups to ensure that general grammatical knowledge was comparable among the groups. Pilot testing revealed that the lowest level DELE test (Diploma de Español como Lengua Extranjera) – a widely used measure of Spanish proficiency in language studies – was too complex for the low-level learners in this study to demonstrate any of their grammatical knowledge (some of the
pilot participants failed to get any questions correct). Thus, to avoid floor effects and allow for some range in scores, a more appropriate proficiency measure was developed by adapting the proficiency tests offered by Transparent Language and Spanish Steps, two reputable companies that provide language learning software and resources to educational institutions, language schools, and L2 learners. The test consisted of 24 multiple-choice questions covering a wide array of grammatical forms and structures of varying levels of difficulty.

The pretest written production assessment (see Appendix D) served to assess learners’ proficiency specifically with the target grammatical structure (verbal inflections in the present, preterite, and future tenses). Moreover, it served to exclude the few participants who demonstrated having some knowledge of the new future tense by scoring above 38% accuracy in the written production of this tense. Like the previous screening measures, the production task was completed online in ANGEL in the computer laboratory, and it required the participants’ to produce the target verb forms in sentence context. The simple task, approximately 7 minutes in length, consisted of 12 sentences with blanks where verbs should be. The participants were instructed to type the correct past (preterite), present, or future form of the infinitive verb provided according to the subject preceding the blank. Following previous studies (e.g., VanPatten & Cadierno, 1993), a liberal scoring procedure was followed in order to reveal any partial effects of the training on the immediate and delayed posttest results. Responses were awarded 2 points if fully correct, and 1.5 points if the only error was a missing or wrongly placed accent. Responses were given 1 point if they were incorrectly spelled but the ending was correct, or if they were a correct form in the given tense but the wrong person. All other responses received a score of zero.
3.1.3.2 The Letter-Number Sequencing Working Memory Task

All participants completed the Letter-Number Sequencing (LNS) (see Appendix E) test after completing the four pretest assessment measures. This task was adapted from the working memory test used in the two latest versions of the Wechsler intelligence test (1997, 2008). The task was administered in E-Prime and involved aural presentation of strings of alternating letters and numbers at a rate of approximately one item per second. It took approximately twelve minutes to complete. After listening to each string, the participants were instructed to recall it by placing the numbers in ascending numerical order followed by the letters in alphabetical order. Participants could take as much time as they needed to recall and type the sequences they had heard. The task began with a series of two-item strings (e.g., C-6) and continued up until eight items (e.g., 5-X-9-N-3-R-6-C). Participants were presented with three trials at each series length. Participants therefore heard a total of 21 letter-number strings, which is the maximum possible raw score for the task. In addition to calculating the participants’ overall accuracy (total number of correctly recalled sequences out of 21), each participant’s memory length was calculated based on the maximum string length (2-8) that they could recall correctly for all three trials. However, following previous research (Crowe, 2000; Hanson, 2012) only the overall accuracy scores were used for analyses in the current dissertation in that they provide a more fine-grained distinction among the participants’ WM ability. It is important to note, though, that correlation analyses showed that the two scoring measures (memory length and overall accuracy) were strongly correlated (Pearson correlation of .860, with \( p < .001 \)) in the present study.

3.1.3.3 Training

The training, which spanned across 2.5 weeks (see Figure 6), was designed to improve visual and aural processing of verbal inflectional morphology in a sentential context. The five
training programs (see Appendix F) were administered with E-Prime, and each one lasted between 15-18 minutes. The participants wore headphones and worked individually through each program. All of the trainings consisted of exposure to visual and aural sentences in which lexical cues (adverbial expressions and explicit subject (pro)nouns) were omitted in order to push learners to process the inflections for temporal and subject reference. The sentences all comprised a prepositional phrase (either in sentence-initial or sentence-final position), a verb, and a direct object. The following are two examples of the training sentences:

(a) *Escucho las instrucciones de la profesora.*
    ‘[I] listen to the instructions from the professor.’

(b) *Antes del partido comerán espaguetis.*
    ‘Before the game [they] will eat spaghetti.’

During each training session, participants were exposed to 103 sentences, which included 6 practice trials and 96 experimental sentences, half presented visually and half presented aurally. Visual stimuli always preceded aural stimuli so that learners had an opportunity to first visualize the forms. All of the sentences were read by a female native Spanish speaker and recorded, and then the audio files were edited for sound quality.

The participants were required to respond to the stimuli in a way that encouraged them to attend to the inflections and make form-meaning connections. As seen in the example training trial in Figure 7, after each sentence was visually or aurally presented, the participants were asked to determine either the tense or the subject using designated keys that were labeled. The visual sentences remained on the screen for 4 s before the participants were presented with the tense or subject question. This duration period is consistent with previous studies involving block presentation of a single sentence of similar length (e.g., Bialystok & Miller, 1999), and it was also judged by pilot testing to be a sufficient amount of time for beginning learners to read each
sentence one time, but not multiple times. Since the sentences were randomly presented, the participants did not know which question they would be asked. In addition, following the tenets of structured input activities (VanPatten, 1996, 2002), so as not to overtax their processing resources, the learners were progressively exposed to additional forms of the verbal paradigm throughout the trainings. For training 1, learners were only exposed to the 3SG form of the present, past (preterite), and future tenses. In trainings 2 and 3, the input contained both the 1SG and 3SG forms of all three tenses. The two plural forms (1PL and 3PL) were then incorporated in trainings 4 and 5. The sentences were always balanced for tense, person, and number (only tense in training 1).

---

**Figure 7. Example of a Training Visual Trial with a Tense Question in Feedback Conditions**
As previously mentioned, there were three training conditions based on feedback provided: 1) none, 2) yes-no (saw correct or incorrect after each response), or 3) metalinguistic feedback (saw correct or incorrect with a metalinguistic explanation of the form incorrectly processed). Figures 8 and 9 provide examples of the metalinguistic feedback. While feedback differed between the groups, it is important to emphasize that exposure to the input practice was kept constant across all participants.

Figure 8. Example of Metalinguistic Feedback Following a Tense Question During Aural Training Practice
Figure 9. *Example of Metalinguistic Feedback Following a Subject Question During Visual Training Practice*

Accuracy and response times were recorded for each of the five training programs. In addition to the posttest assessments that measured processing and learning outcomes, data from the training is useful in evaluating learning trajectories and the potential influence of feedback and/or working memory on these trajectories. However, results from the assessments conducted before and after training are the focus of this dissertation, and thus I will only present and briefly discuss general accuracy results for the training (see Section 3.1.4.2).

3.1.4 Descriptive Statistics for Screening, Proficiency, WM, and Training Measures

Below in Section 3.1.4.1 I report the descriptive statistics by group for the screening and proficiency measures as well as the WM test. I also report the results of one-way ANOVAs run on these measures since it is important to compare the groups in order to determine if there were any differences between them prior to the study. Then, in Section 3.1.4.2 I provide the descriptive statistics for the experimental groups’ overall accuracy on the training as well as report results from one-way ANOVA analyses conducted on these scores. These basic training results will reveal whether the feedback conditions had differential effects on the experimental groups’ accuracy on the training itself prior to more thorough analyses of the effects of training on performance on the four assessment measures (reported in Sections 3.2-3.5).

3.1.4.1 Means for Screening, Proficiency, and WM Measures

Table 2 presents the means and standard deviations (SDs) for the age, years of Spanish instruction, and the first exam grade of the learner groups as well as their performance on the three screening tests (vocabulary test, general grammar proficiency test, and written verb production task) and the working memory (WM) task.
### Table 2. Means and Standard Deviations for Screening, Proficiency, and WM Measures by Group

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>Experimental</td>
<td>116</td>
<td>19.43</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>19.29</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>19.21</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>19.78</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>19.07</td>
<td>1.42</td>
</tr>
<tr>
<td><strong>Spanish Instruction in Years</strong></td>
<td>Experimental</td>
<td>116</td>
<td>2.23</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>2.30</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>2.26</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>2.13</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>2.24</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>First Course Exam Grade</strong></td>
<td>Experimental</td>
<td>116</td>
<td>82.15</td>
<td>10.60</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>81.89</td>
<td>10.63</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>81.97</td>
<td>10.82</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>82.57</td>
<td>10.63</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>84.89</td>
<td>9.25</td>
</tr>
<tr>
<td><strong>Grammar Proficiency</strong></td>
<td>Experimental</td>
<td>116</td>
<td>50.47</td>
<td>12.28</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>49.12</td>
<td>12.30</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>51.54</td>
<td>11.96</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>50.73</td>
<td>12.76</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>52.98</td>
<td>11.15</td>
</tr>
<tr>
<td><strong>Vocabulary Test</strong></td>
<td>Experimental</td>
<td>116</td>
<td>96.71</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>95.98</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>97.23</td>
<td>4.85</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>96.91</td>
<td>4.46</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>96.80</td>
<td>5.00</td>
</tr>
<tr>
<td><strong>Production Pretest (Present tense)</strong></td>
<td>Experimental</td>
<td>116</td>
<td>82.65</td>
<td>18.71</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>82.24</td>
<td>21.08</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>82.24</td>
<td>17.83</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>83.44</td>
<td>17.54</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>81.84</td>
<td>19.97</td>
</tr>
<tr>
<td><strong>Production Pretest (Preterite tense)</strong></td>
<td>Experimental</td>
<td>116</td>
<td>58.08</td>
<td>23.40</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>56.89</td>
<td>21.65</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>58.88</td>
<td>27.48</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>58.44</td>
<td>21.26</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>56.25</td>
<td>25.17</td>
</tr>
<tr>
<td><strong>Production Pretest (Future tense)</strong></td>
<td>Experimental</td>
<td>116</td>
<td>5.50</td>
<td>11.60</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>4.44</td>
<td>10.37</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>7.24</td>
<td>14.13</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>4.84</td>
<td>10.06</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>6.00</td>
<td>9.80</td>
</tr>
<tr>
<td><strong>Working Memory (LNS)</strong></td>
<td>Experimental</td>
<td>116</td>
<td>12.44</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td>- No feedback</td>
<td>38</td>
<td>12.32</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>- Yes-No feedback</td>
<td>38</td>
<td>12.05</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>- Metalinguistic feedback</td>
<td>40</td>
<td>12.93</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>148</td>
<td>12.74</td>
<td>2.63</td>
</tr>
</tbody>
</table>

* 1 year of high-school instruction = 1 semester of college-level instruction
A series of one-way ANOVAs were conducted on the screening measures in Table 2 to determine if there were any differences among the groups prior to the training. The analyses revealed no significant differences between the four groups on any of the measures (age, $F(3,260) = 2.377, p = .070$, years of Spanish instruction, $F(3,260) = .336, p = .799$, first exam grade, $F(3,260) = 1.700, p = .167$, grammar proficiency, $F(3,260) = 1.286, p = .280$, vocabulary test, $F(3,260) = .449, p = .718$, production pretest for present tense, $F(3,260) = .071, p = .976$, production pretest for past (preterite) tense, $F(3,260) = .167, p = .919$, production pretest for future tense, $F(3,260) = .562, p = .641$, letter-number sequencing WM task, $F(3,260) = 1.001, p = .393$). This suggests that the sample pool was homogeneous and that any between-group differences observed in the training results (between the three experimental groups) or on the posttest results (between all four groups) can likely be attributed to the training received. One-way ANOVA analyses will also be reported for the pretest scores of each assessment measure in their respective sections (3.2-3.5) to rule out any between-group differences on these tasks prior to the training.

### 3.1.4.2 Training Accuracy Means

The overall mean accuracy across all of the training sessions was 72.72 out of 100 (SD of 16.61) for the 116 learner participants in the experimental groups. Overall accuracy for determining tense was 73.40 (SD of 17.12), and overall accuracy for determining the subject was 70.19 (SD of 16.10). The means and SDs for each training and condition by group are presented in Tables 3 and 4. As a reminder, the participants were not asked to determine the subject in the first training (they only made decisions with respect to tense as all verbs were in the 3sg form). Thus, for the subject condition, results are only reported for trainings 2-5.
Table 3. Means for Training Accuracy in the Tense Condition by Group (SD in parentheses)

<table>
<thead>
<tr>
<th>Group</th>
<th>Training 1</th>
<th>Training 2</th>
<th>Training 3</th>
<th>Training 4</th>
<th>Training 5</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback (n = 38)</td>
<td>71.33 (25.80)</td>
<td>61.95 (27.36)</td>
<td>66.45 (26.48)</td>
<td>55.48 (22.30)</td>
<td>69.52 (24.81)</td>
<td>62.61 (15.81)</td>
</tr>
<tr>
<td>Yes-No Feedback (n = 38)</td>
<td>89.25 (17.83)</td>
<td>78.18 (18.01)</td>
<td>75.42 (21.88)</td>
<td>61.18 (18.31)</td>
<td>79.71 (18.91)</td>
<td>65.68 (15.54)</td>
</tr>
<tr>
<td>Metalinguistic Feedback (n = 40)</td>
<td>94.11 (10.31)</td>
<td>81.09 (15.28)</td>
<td>87.29 (13.41)</td>
<td>66.15 (19.73)</td>
<td>86.25 (17.86)</td>
<td>71.35 (11.05)</td>
</tr>
</tbody>
</table>

Based on the means presented in Tables 3 and 4, it is evident that, as predicted, all the learner groups were more accurate on the visual training practice as compared to the aural practice.

While it will be interesting in future work to examine visual and aural learning trajectories separately, analyses of training results are not a focus of this dissertation and thus here I will only report the results of one-way ANOVAs run on overall tense and subject accuracy (visual and aural practice combined). These basic results are only intended to provide a general picture of how the different feedback conditions of the training affected the experimental groups’ accuracy on the training itself prior to examining the effects of training on learners’ performance on the four assessment measures (Sections 3.2 – 3.5).
With respect to overall accuracy for the tense condition across the trainings, a one-way ANOVA revealed that there were significant between-group differences, $F(2,113) = 7.593, p = .001$. Bonferroni post hoc analyses indicated that these differences were due to the metalinguistic feedback group (80.18% accuracy with a SD of 12.81) significantly outperforming the no feedback group (65.87% accuracy with a SD of 19.85). There were no significant differences for overall tense accuracy between the metalinguistic group and the yes-no feedback group or between the yes-no feedback group and the no feedback group. Tense accuracy results for each training by group are shown in Figure 10, and significant between-group differences are indicated with asterisks.

![Tense Accuracy Overall for Trainings 1-5](image)

* Figure 10. *Training Accuracy for the Tense Condition by Group*

* Between-group difference is significant at $p < .05$

For the subject condition, a one-way ANOVA revealed that there were also significant between-group differences in accuracy, $F(2,113) = 4.712, p = .011$. Bonferroni post hoc analyses indicated that, as we saw with the tense condition, the metalinguistic feedback group (75.82%
accuracy with a SD of 14.45) significantly outperformed the no feedback group (65.06% accuracy with a SD of 16.26) in determining the subject. There were no significant differences between the metalinguistic feedback group and the yes-no feedback group or between the yes-no feedback group and the no feedback group. Subject accuracy results for each training by group are shown in Figure 11, and significant between-group differences are indicated with asterisks.

![Subject Accuracy Overall for Trainings 2-5](image)

Figure 11. Training Accuracy for the Subject Condition by Group
* Between-group difference is significant at $p < .05$

These basic results indicate that metalinguistic feedback in training consistently led to better accuracy in determining the tense and subject of the sentence as compared to receiving no feedback during training. Future work will further investigate the effects of the training conditions as well as cognitive individual differences on learning trajectories during the training by examining both accuracy and reaction times for the different trained tenses and persons. In the present dissertation, though, I focus on whether exposure to the training had effects on the
learners’ processing and production of verbal inflectional morphology, as measured by the four assessment measures discussed in the following sections (3.2 – 3.5).

3.2 Assessment 1: Aural Processing Task

The aural processing assessment measure was a brief task designed to assess the participants’ aural processing of the target verbal inflections in sentential context. The materials and procedure were nearly identical to that of the aural training, with the only differences being that every experimental sentence was followed by both a subject and tense question (instead of seeing either the subject or tense question as in the training) and none of the participants received any feedback. Thus, the aural processing task served as an independent assessment of aural processing, and the results will reveal any effects that the training may have had on the learners’ processing of verbal inflections. The research questions to be answered with the pretest, posttest, and delayed posttest aural processing task data presented here are:

(a) Does processing-based training that promotes attention to verbal morphology in the input lead to improved accuracy in aural processing of inflections?

(b) Is working memory (WM) capacity related to learners’ ability to accurately process inflections and/or learn from training to improve over time?

3.2.1 Participants

Two hundred and twenty-five of the participants (109 in experimental groups; 116 in the control group) who met the general inclusion criteria outlined in Section 3.1.1 successfully completed the pretest, immediate posttest, and delayed posttest versions of the cue reliance task. Exclusions were due to technological error that resulted in the loss of some pretest data. The breakdown into groups was as follows for the 109 participants who completed the training: no feedback group (n = 37), yes-no feedback group (n = 37), and metalinguistic feedback group (n = 35). Of the 116 participants in the control group who met the criteria, a randomly selected subset
of the participants (n=40) was used for the analyses. Table 5 presents the descriptive statistics for the screening tests, proficiency measures, and WM task corresponding to the subset of the participant pool used for the aural processing task analyses.

Table 5. Descriptive Statistics for the Subset of the Participant Pool Used in the Aural Processing Task Analyses on Screening, Proficiency, and WM Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Proficiency</td>
<td>No feedback</td>
<td>37</td>
<td>49.66</td>
<td>12.00</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>51.58</td>
<td>12.12</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>52.38</td>
<td>12.22</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>54.06</td>
<td>11.50</td>
</tr>
<tr>
<td>Vocabulary Test</td>
<td>No feedback</td>
<td>37</td>
<td>96.02</td>
<td>5.10</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>97.16</td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>96.92</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>97.11</td>
<td>5.29</td>
</tr>
<tr>
<td>Written Production Pretest (Present tense)</td>
<td>No feedback</td>
<td>37</td>
<td>82.77</td>
<td>21.11</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>82.43</td>
<td>18.03</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>82.86</td>
<td>18.21</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>83.59</td>
<td>21.54</td>
</tr>
<tr>
<td>Written Production Pretest (Preterite tense)</td>
<td>No feedback</td>
<td>37</td>
<td>58.09</td>
<td>20.63</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>58.11</td>
<td>27.44</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>58.39</td>
<td>20.45</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>56.88</td>
<td>22.01</td>
</tr>
<tr>
<td>Written Production Pretest (Future tense)</td>
<td>No feedback</td>
<td>37</td>
<td>4.56</td>
<td>10.48</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>7.43</td>
<td>14.27</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>5.18</td>
<td>10.56</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>7.81</td>
<td>11.39</td>
</tr>
<tr>
<td>Exam 1 Grade</td>
<td>No feedback</td>
<td>37</td>
<td>81.87</td>
<td>10.78</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>81.89</td>
<td>10.96</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>83.45</td>
<td>9.02</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>83.76</td>
<td>9.28</td>
</tr>
<tr>
<td>Working Memory (LNS)</td>
<td>No feedback</td>
<td>37</td>
<td>12.30</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>37</td>
<td>12.14</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>35</td>
<td>12.66</td>
<td>3.26</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>12.83</td>
<td>2.85</td>
</tr>
</tbody>
</table>

One-way ANOVAs performed on the preliminary measures presented in Table 5 revealed no significant differences between the groups on the grammar proficiency test, $F(3,145) = .896, p = .445$, the vocabulary test, $F(3,145) = .422, p = .738$, the production pretest for present tense, $F(3,145) = .024, p = .995$, the production pretest for preterite tense, $F(3,145) = .034, p = .992$, and the content.
the production pretest for future tense, $F(3,145) = .708, p = .549$, the first exam grade (taken just prior to the start of the study), $F(3,145) = .378, p = .769$, or the letter-number sequencing WM task, $F(3,145) = .486, p = .693$.

### 3.2.2 Materials and Procedure

Like the aural training, the aural processing task (administered using E-Prime) involved processing sentences that were void of lexical cues to the subject and temporal reference of the sentence. Thus, the participants were forced to rely on the target verbal morphology in order to determine the tense and subject of each sentence. This brief task, which took approximately 7 minutes to complete, was designed similarly to the paper-and-pencil interpretation tasks used in Processing Instruction research (see, for example, studies in VanPatten, 2004; see Chapter 2 for relevant discussion). The participants were exposed to 3 practice items followed by 12 experimental sentences, 4 in each tense (past, present, future) and three for each subject (1SG, 1PL, 3SG, 3SG). That is, each of the four subjects was represented once in each tense. As in the training, the sentences all comprised a prepositional phrase (either in sentence-initial or sentence-final position), a verb, and a direct object, as exemplified in Section 3.1.3.3. While the sentence structure of the experimental stimuli was identical to that of the training sentences, it is important to note that the participants were not exposed to any of the same sentences in the aural processing assessment that they saw in training. Appendix G presents the experimental sentences used in the three versions of the aural processing assessment task. All of the sentences were read by a female native Spanish speaker and recorded in a sound booth, and then the audio files were edited for sound quality. As illustrated below in Figure 12, after listening to each sentence (upon the offset of the sound file), participants were asked to select the correct tense and subject of the sentence at their own pace using the designated keys that were labeled with the numbers.
corresponding to the given tense or person. Participants then moved on to the next trial; no feedback was provided. The maximum possible score for the task was 24 points, with 12 points maximum for each condition (tense and subject). Although both accuracy and response times were recorded, only accuracy results will be reported in the current dissertation.

3.2.3 Results

In the current section, I present the results of analyses on the aural processing task data with the aim of shedding light on whether the training conditions and working memory (WM) had any effects on the learners’ ability to accurately process the target inflections for temporal and subject reference. The analyses involved a mixed design examining whether the within-
subject variable of Time (pretest, posttest, delayed posttest) and the two between-subjects variables of Group (no feedback training, yes-no feedback training, metalinguistic feedback training, control group) and WM capacity affected the participants’ accuracy in aurally processing the target verb forms (the dependent variable). The statistical analyses involved two repeated-measures ANOVAs, one for the tense condition and one for the subject condition. Before conducting these analyses to address the research questions, one-way ANOVAs were conducted on the pretest scores to rule out differences among the groups prior to the training. The results revealed no significant differences between the groups on the pretest (tense accuracy, $F(3,145) = .806, p = .492$; subject accuracy, $F(3,145) = .613, p = .608$). This suggests that any differences found between the groups on the immediate and delayed posttests can be attributed to the training. Table 6 presents the descriptive statistics for the aural processing task, which entails group mean percentages for accuracy in aurally processing the target verbal inflections at each test time.

<table>
<thead>
<tr>
<th>Group</th>
<th>Condition</th>
<th>Tense</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback, $n = 37$</td>
<td>Tense</td>
<td>44.37 (16.56)</td>
<td>57.88 (25.23)</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>65.32 (20.74)</td>
<td>77.25 (16.51)</td>
</tr>
<tr>
<td>Yes-No Feedback, $n = 37$</td>
<td>Tense</td>
<td>46.17 (16.74)</td>
<td>65.54 (22.24)</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>65.09 (17.11)</td>
<td>75.90 (22.97)</td>
</tr>
<tr>
<td>Metalinguistic Feedback, $n = 35$</td>
<td>Tense</td>
<td>41.67 (15.79)</td>
<td>79.05 (16.47)</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>69.76 (15.40)</td>
<td>88.10 (12.34)</td>
</tr>
<tr>
<td>Control, $n = 40$</td>
<td>Tense</td>
<td>46.67 (11.45)</td>
<td>49.79 (17.24)</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>67.92 (14.81)</td>
<td>72.71 (20.76)</td>
</tr>
</tbody>
</table>

The group mean percentages for tense accuracy (presented visually below in Figure 13) reveal that all three groups that underwent training showed improvement in accurately processing verbal inflections to determine the tense of the sentence. The control group, on the
other hand, showed little improvement. The means also suggest that there were differences across the three training groups at the immediate and delayed posttests with respect to their accuracy in processing inflections for temporal reference. The metalinguistic group showed the greatest improvement from the pretest to the immediate posttest (37% improvement), followed by the yes-no feedback group (19% improvement), and then the no feedback group (14% improvement). Although accuracy decreased from the immediate posttest to the delayed posttest for all training groups, they maintained most of the gain and the same apparent differences between the groups at the immediate posttest were evident at the delayed posttest. A 3 (Time) x 3 (Group) repeated-measures ANCOVA for tense accuracy, with WM as a covariate, confirmed that there was a significant main effect for Group, $F(3,144) = 7.837, p < .001$, as well as a significant interaction between Time and Group, $F(6,288) = 14.141, p < .001$. Pairwise comparisons revealed the following between-group differences as a result of training: (a) both the metalinguistic feedback and the yes-no feedback training groups were more accurate than the control group in processing inflections for temporal reference (all $p < .05$), and (b) amongst the training groups, the metalinguistic group was more accurate than the no feedback group ($p < .05$).

In brief, while exposure to training without feedback was no more effective than no training at all, exposure to training with feedback, regardless of whether it was basic yes-no feedback or feedback that also entailed a metalinguistic explanation, resulted in improved accuracy in processing inflections for temporal information. The significant effects found for WM in processing inflections for tense will be discussed below in conjunction with WM results for the subject condition.
Figure 13. Accuracy in Aural Processing of Verbal Inflections for Tense

Note: Error bars are +/-2 standard error.

Regarding the subject condition, the group mean percentages for subject accuracy (presented visually below in Figure 14) indicate that while all four groups improved, even if only minimally, in their processing of verbal inflections for subject reference, there were clear differences between the groups. The metalinguistic feedback group showed the greatest improvement at the immediate posttest (18%), followed by the no feedback group and the yes-no feedback group (12% and 11% improvement, respectively), and then the control group (5% improvement). At the delayed posttest, the metalinguistic feedback group maintained the greatest improvement gain from the pretest (15%), followed by the yes-no feedback group (10%), with the no feedback group and control group maintaining little overall gain from the pretest to the delayed posttest (5% and 3% respectively). A 3 (Time) x 3 (Group) repeated-measures ANCOVA for subject accuracy, with WM as a covariate, confirmed that there was a significant main effect for Group, $F(3,144) = 4.445, p = .005$, as well as a significant interaction between Time and Group, $F(6,288) = 2.789, p = .012$. Pairwise comparisons indicated that between-group
differences were due to the metalinguistic feedback group significantly outperforming both the control group and the no feedback group (all \( p < .05 \)). In short, training with metalinguistic feedback was more effective than no training and training without feedback for improving aural processing of verbal inflections for subject reference.

![Figure 14. Accuracy in Aural Processing of Verbal Inflections for Subject](image)

*Figure 14. Accuracy in Aural Processing of Verbal Inflections for Subject*

*Note: Error bars are +/-2 standard error.*

With respect to the role of WM in learning to process verbal inflections for temporal and subject reference, the RM ANCOVA analyses revealed a main effect for WM for both conditions (tense, \( F(1,144) = 22.042, p < .001 \); subject, \( F(1,144) = 39.047, p < .001 \)). In addition, there was a significant interaction between WM and Time for the tense condition, \( F(2,288) = 5.018, p = .007 \), and this interaction approached significance for the subject condition, \( F(2,288) = 2.497, p = .084 \). These results indicate that WM capacity was overall positively related to accuracy in processing verbal inflections for both temporal reference and subject reference. In addition, with exposure to training over time, WM facilitated the learning and in turn accurate processing of verbal morphology, particularly in the processing of inflections for temporal information. The
greater facilitative effects of WM for learning to process inflections for tense may be related to the inclusion of the new future tense. However, further speculation will not be offered here given that separate analyses were not conducted for each tense and the interaction between WM and Time for the subject condition did approach significance. The differential effects of WM capacity on the three tenses will be addressed when examining results from the written production data in Section 3.3.3.

3.2.4 Summary of Results

The aural processing task was used to assess the effects of training as well as WM on the learners’ aural processing of verbal inflections in sentential context for temporal and subject reference. In order to summarize the findings, the specific research questions relevant to the aural processing assessment are restated and addressed briefly in turn below.

(a) Does processing-based training that promotes attention to verbal morphology in the input lead to improved accuracy in aural processing of inflections?

While training without feedback was no more effective than no training, exposure to training with feedback led to greater accuracy in processing inflections. Training with both yes-no feedback and metalinguistic feedback facilitated accurate processing of inflections for temporal reference. However, only training with metalinguistic feedback had significant benefits for the accurate processing of inflections for subject reference. Thus, exposure to training with metalinguistic feedback was most effective overall in promoting accurate aural processing of the target verbal inflections.

(b) Is working memory (WM) capacity related to learners’ ability to accurately process inflections and/or learn from training to improve over time?
Yes. WM capacity was positively associated with greater accuracy overall in aural processing of verbal inflections, both when processing them for temporal reference and subject reference. Higher WM capacity was also positively related to learners’ ability to learn from the training and improve their processing of inflections over time (on the posttests), particularly when processing inflections for temporal reference.

3.3 Assessment 2: Written Production Task

The written production measure was a simple task designed to assess the participants’ ability to produce the target verbal inflections. The results from the pretest of this task were reported in Section 3.1.4.1 since they served to evaluate learners’ proficiency prior to the training with respect to verbal inflections in the present, preterite, and future tenses as well as ensure comparability across the groups. Moreover, as discussed in Section 3.1.3.1, the pretest for the future tense served to eliminate the participants who demonstrated prior knowledge of this new tense. Although the method and materials for the written production task were presented in Section 3.1.3.1, they will be repeated in this section for convenience. The research questions to be answered with the pretest, posttest, and delayed posttest production data presented here are:

(a) Does the processing-based training have any effects on accuracy in written production (for familiar or new verbal inflections)?

(b) Is working memory (WM) capacity related to learners’ accuracy in written production of inflections (familiar or new) and their ability to learn from training and/or improve over time?

3.3.1 Participants

Two hundred and forty-five of the participants (106 in experimental groups; 139 in the control group) who met the general inclusion criteria outlined in Section 3.1.1 successfully completed the pretest, immediate posttest, and delayed posttest versions of the written production
task. Exclusions were all due to failure to follow instructions on the immediate or delayed posttest (e.g., completing the wrong version of the task, putting all present tense instead of preterite). The breakdown into groups was as follows for the 106 participants who completed the training: no feedback group (n = 34), yes-no feedback group (n = 36), and metalinguistic feedback group (n = 36). Of the 139 participants in the control group who met the criteria, a randomly selected subset of the participants (n=40) was used for the analyses. Table 7 presents the descriptive statistics for the screening tests, proficiency measures, and WM task corresponding to the subset of the participant pool used for the production task analyses.

Table 7. Descriptive Statistics for the Subset of the Participant Pool Used in the Production Task Analyses on Screening, Proficiency, and WM Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Proficiency</td>
<td>No feedback</td>
<td>34</td>
<td>49.88</td>
<td>12.33</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>36</td>
<td>51.97</td>
<td>12.14</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>51.74</td>
<td>12.65</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>53.65</td>
<td>11.98</td>
</tr>
<tr>
<td>Vocabulary Test</td>
<td>No feedback</td>
<td>34</td>
<td>96.28</td>
<td>5.19</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>36</td>
<td>97.73</td>
<td>4.18</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>97.22</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>96.78</td>
<td>5.48</td>
</tr>
<tr>
<td>Written Production Pretest</td>
<td>No feedback</td>
<td>34</td>
<td>84.56</td>
<td>20.88</td>
</tr>
<tr>
<td>(Present tense)</td>
<td>Yes-No feedback</td>
<td>36</td>
<td>82.29</td>
<td>17.77</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>82.99</td>
<td>18.21</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>84.06</td>
<td>18.56</td>
</tr>
<tr>
<td>Written Production Pretest</td>
<td>No feedback</td>
<td>34</td>
<td>59.91</td>
<td>20.32</td>
</tr>
<tr>
<td>(Preterite tense)</td>
<td>Yes-No feedback</td>
<td>36</td>
<td>58.85</td>
<td>28.16</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>59.20</td>
<td>21.36</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>60.31</td>
<td>24.50</td>
</tr>
<tr>
<td>Written Production Pretest</td>
<td>No feedback</td>
<td>34</td>
<td>4.96</td>
<td>10.86</td>
</tr>
<tr>
<td>(Future tense)</td>
<td>Yes-No feedback</td>
<td>36</td>
<td>7.64</td>
<td>14.42</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>5.38</td>
<td>10.47</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>7.03</td>
<td>11.16</td>
</tr>
<tr>
<td>Aural Processing Pretest</td>
<td>No feedback</td>
<td>34</td>
<td>56.74</td>
<td>14.76</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>36</td>
<td>56.37</td>
<td>14.65</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>57.06</td>
<td>11.65</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>55.18</td>
<td>17.47</td>
</tr>
<tr>
<td>Exam 1 Grade</td>
<td>No feedback</td>
<td>34</td>
<td>82.41</td>
<td>11.03</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>36</td>
<td>82.22</td>
<td>10.78</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>36</td>
<td>84.05</td>
<td>10.03</td>
</tr>
<tr>
<td>Working Memory (LNS)</td>
<td>Control</td>
<td>Yes-No feedback</td>
<td>Metalinguistic feedback</td>
<td>Control</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------</td>
<td>-----------------</td>
<td>-------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>No feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.44</td>
<td>12.19</td>
<td>12.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.41</td>
<td>2.70</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>Yes-Yes feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metalinguistic feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One-way ANOVAs performed on the preliminary measures presented in Table 7 revealed no significant differences between the groups on the grammar proficiency test, $F(3,142) = .580, p = .629$, the vocabulary test, $F(3,142) = .573, p = .634$, the production pretest for present tense, $F(3,142) = .106, p = .957$, the production pretest for preterite tense, $F(3,142) = .029, p = .993$, the production pretest for future tense, $F(3,142) = .422, p = .738$, the aural processing pretest, $F(3,142) = .112, p = .953$, the first exam grade (taken just prior to the start of the study), $F(3,142) = .525, p = .666$, and the letter-number sequencing WM task, $F(3,142) = .488, p = .692$.

The lack of between-group differences on the pretest production task suggests that any differences found between the groups on the immediate and delayed production posttests can be attributed to the training.

### 3.3.2 Materials and Procedure

The written production assessment was completed online in ANGEL, the university’s course management system. The simple task, which took approximately seven minutes to complete, consisted of 12 sentences (4 for each tense) with blanks where verbs should be. The participants were instructed to type the correct present, preterite, or future form of the infinitive verb provided according to the subject preceding the blank. Each of the four target person-number combinations (1SG, 1PL, 3SG, 3SG) was used once in each tense. The following are three examples, one from each tense, of the sentences used in the production task. The remaining stimuli are provided in Appendix D.

(a) *Generalmente los estudiantes _______ (comer) en la cafetería de la universidad.*

‘Generally the students _______ (to eat) in the cafeteria of the university.’
La semana pasada yo _______ (estudiar) para el examen final.

‘The week last I _______ (to study) for the exam final.’

Mañana el hombre _______ (discutir) el problema con su esposa.

‘Tomorrow the man _______ (to discuss) the problem with his wife.’

Following previous studies (e.g., VanPatten & Cadierno, 1993), a liberal scoring procedure was followed in order to reveal any partial effects of the training on the immediate and delayed production posttest results. Responses were awarded 2 points if fully correct, and 1.5 points if the only error was a missing or wrongly placed accent. Responses were given 1 point if they were incorrectly spelled but the ending was correct, or if they were a correct form in the given tense but the wrong person. All other responses received a score of zero.

3.3.3 Results

In the current section, I present the results of analyses on the written production data with the aim of shedding light on whether the form-focused training conditions and working memory (WM) had any effects on the learners’ ability to accurately produce the target inflections, which included both familiar verbal inflections (present and preterite tense) and new inflectional forms (future tense). The analyses involved a mixed design examining whether the within-subject variable of Time (pretest, posttest, delayed posttest) and the two between-subjects variables of Group (no feedback training, correct-incorrect feedback training, metalinguistic feedback training, control group) and WM capacity affected the participants’ accuracy in producing the target verb forms (the dependent variable). The statistical analyses involved three repeated-measures ANOVAs, one for each of the three tenses. As reported above in Section 3.3.1, there were no significant differences between the groups with respect to their accuracy in producing the target present, preterite, or future inflections prior to the training. This means that any differences found between groups on the immediate and delayed posttest can be attributed to the
training. Table 8 presents the descriptive statistics for the written production task, which entails group mean percentages for accuracy in producing the target verbal inflections at each test time.

Table 8. *Descriptive Statistics for the Written Production Assessment (SD in parentheses)*

<table>
<thead>
<tr>
<th>Group</th>
<th>Tense</th>
<th>Pretest Accuracy (%)</th>
<th>Posttest Accuracy (%)</th>
<th>Delayed Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback n = 34</td>
<td>Present</td>
<td>84.56 (20.88)</td>
<td>87.13 (18.07)</td>
<td>86.40 (21.40)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>59.91 (20.32)</td>
<td>70.96 (26.11)</td>
<td>62.50 (27.57)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>4.96 (10.86)</td>
<td>41.54 (35.59)</td>
<td>36.95 (35.20)</td>
</tr>
<tr>
<td>Yes-No Feedback n = 36</td>
<td>Present</td>
<td>82.29 (17.77)</td>
<td>84.20 (22.93)</td>
<td>89.24 (14.99)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>58.85 (28.16)</td>
<td>72.23 (27.20)</td>
<td>64.06 (26.24)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>7.64 (14.42)</td>
<td>44.79 (34.80)</td>
<td>43.23 (35.37)</td>
</tr>
<tr>
<td>Metalinguistic Feedback n = 36</td>
<td>Present</td>
<td>82.99 (18.21)</td>
<td>93.92 (12.27)</td>
<td>91.49 (12.24)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>59.20 (21.36)</td>
<td>76.22 (19.47)</td>
<td>72.22 (20.29)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>5.38 (10.47)</td>
<td>72.05 (28.83)</td>
<td>60.24 (29.43)</td>
</tr>
<tr>
<td>Control n = 40</td>
<td>Present</td>
<td>84.06 (18.56)</td>
<td>79.06 (23.06)</td>
<td>85.31 (13.83)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>60.31 (24.50)</td>
<td>70.16 (24.66)</td>
<td>67.81 (22.76)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>7.03 (11.16)</td>
<td>15.94 (22.11)</td>
<td>18.44 (27.88)</td>
</tr>
</tbody>
</table>

The group means for the present tense (presented visually below in Figure 15) reveal that there was little change in accuracy across the three test times, although there was some improvement, mainly for the metalinguistic feedback group (exhibiting an improvement of approximately 11% from the pretest to immediate posttest, and maintaining an improvement of 9% at the delayed posttest). A 3 (Time) x 3 (Group) repeated-measure ANCOVA, with WM as a covariate, indicated that there were no significant main effects for Time, $F(1.91,269.34) = .444, p = .633$, or for Group, $F(3,141) = 1.532, p = .209$. While the interaction between Time and Group was also not significant, it did approach significance, $F(5.73,269.34) = 1.981, p = .072$, likely due to the greater improvement in accuracy over time for the metalinguistic feedback group as compared to the no feedback and control groups. With respect to WM, the analyses revealed a significant main effect for this factor, $F(1,141) = 6.598, p = .011$, but the interaction between Time and WM was non-significant, $F(1.91,269.34) = .377, p = .677$. These WM
findings indicate that WM capacity was overall positively related to higher accuracy in producing the present tense.

Figure 15. Accuracy in Written Production of Present Tense Verbal Inflections. 
Note: Error bars are +/-2 standard error.

The group means for the preterite tense (presented visually below in Figure 16) indicate that there was some improvement on the part of all groups at the immediate posttest (between 10-17% across the groups, with the metalinguistic group showing the greatest improvement). At the time of the delayed posttest, the metalinguistic feedback group and the control group exhibited the least drop in accuracy producing the preterite tense from the immediate posttest, maintaining a 13% and 8% improvement in accuracy, respectively, from the pretest. Nonetheless, a 3 (Time) x 3 (Group) repeated-measure ANCOVA, with WM as a covariate, indicated that there were no significant main effects for either Time, $F(2,282) = .631, p = .533$, or for Group, $F(3,141) = .329, p = .804$. There were also no significant effects of WM. Yet, while the interaction between Time and WM was highly non-significant, $F(2,282) = .020, p = .981$, the main effect for WM did approach significance, $F(1,141) = 3.328, p = .070$. This suggests that
WM capacity was trending toward having overall effects on accuracy in producing the preterite tense, which is what was shown be the case for the present tense.

![Figure 16. Accuracy in Written Production of Preterite Tense Verbal Inflections.](image)

In contrast to the production results for the known present and preterite tenses, the group means for the new future tense (presented visually below in Figure 17) reveal robust group differences in production accuracy at the time of the immediate and delayed posttests. All three of the training groups exhibited substantial improvement at the immediate posttest, although the improvement was greatest for the metalinguistic feedback group, who showed an improvement of 67% as compared to the 37% improvement exhibited by both the no feedback group and the yes-no feedback group. The control group showed minimal improvement of only 9%. The control group as well as the no feedback and yes-no feedback groups exhibited little change in the accuracy with which they produced the future tense at the delayed posttest. While the metalinguistic feedback group’s accuracy decreased by 12% from the immediate to the delayed posttest, this group maintained most of the knowledge of the future tense that they gained and
still showed the greatest overall improvement in production accuracy from pretest to delayed posttest. A 3 (Time) x 3 (Group) repeated-measure ANCOVA, with WM as a covariate, confirmed that there was a significant main effect for Group, $F(3,141) = 15.635, p < .001$, as well as a significant interaction between Time and Group, $F(5.85,274.82) = 16.605, p < .001$. Pairwise comparisons revealed the following between-group differences as a result of training: (a) all experimental training groups were more accurate in producing the future tense than the control group (all $p < .01$), and (b) the metalinguistic feedback group was significantly more accurate in producing future verb forms as compared to both the no feedback group and the yes-no feedback group (all $p < .05$). In short, exposure to training, regardless of feedback, resulted in greater accuracy in producing the new future tense. Nonetheless, training with metalinguistic feedback was overall more effective than training without feedback or even with yes-no feedback for learning to produce future tense verbal inflections.

![Figure 17. Accuracy in Written Production of Future Tense Verbal Inflections.](image)

*Note: Error bars are +/-2 standard error.*
With respect to the role of WM in learning to produce future tense inflections, the RM ANCOVA revealed both a main effect for WM, $F(1,141) = 17.055, p < .001$, as well as a significant interaction between WM and Time, $F(1.95,274.82) = 6.645, p = .002$. This indicates that WM was positively related to accuracy in producing future inflections both in general as well as over time (which meant exposure to the future tense in training for the experimental groups, and exposure to the future tense in the processing-based assessment measures for all groups). These findings suggest that while WM had limited effects on the learners’ ability to accurately produce the familiar present and preterite tenses, WM capacity did facilitate the learning and in turn accurate production of the new future tense. The reason why individual differences in WM capacity had robust effects on learning to produce the future tense over time, but not the present and preterite tenses, may be related to the fact that the future inflections were new to the learners and thus required ‘true learning’ of grammar, which has been shown to be affected by WM capacity (e.g., Brooks et al., 2006; Kempe & Brooks, 2008; Kormos & Sáfar, 2008; Martin & N. Ellis, 2012). Conversely, the above-chance group means and scores at ceiling on the pretest for the present and preterite tenses indicate that there was less learning taking place during the training for these familiar tenses as compared to the new future tense. Thus, the limited role of WM in learners’ ability to produce these familiar tenses may have been due to the related background factors of (a) previous knowledge, which resulted in ceiling effects, and (b) motivation. With respect to motivation, it is possible for example that some learners with lower WM capacity put forth considerably more effort than some high-span learners into memorizing the forms of these familiar tenses for the course exam prior to the study, and consequently WM capacity may have had minimal effects on accuracy in producing the present and preterite inflections. These same factors may also at least partially account for why training had no effects
on the production of these familiar tenses, yet it greatly facilitated the learning of the future tense regardless of whether feedback was provided.

3.3.4 Summary of Results

The written production task was used to assess the effects of training as well as WM on the learners’ ability to produce the target verbal inflections. In order to summarize the findings, the specific research questions relevant to the written production assessment are restated and addressed briefly in turn below.

(a) Does the processing-based training have any effects on accuracy in written production (for familiar or new verbal inflections)?

While the training had no effects on the L2 learners’ ability to produce the familiar present and past (preterite) tenses, exposure to training did facilitate the learning of the new future tense and thus resulted in greater accuracy in the written production of these future inflections on the posttests. While all the L2 learners who received training, regardless of feedback, significantly outperformed the control group for the future tense, the provision of metalinguistic feedback significantly enhanced the effectiveness of the training for learning to produce the new future tense. This suggests that processing-based training that pushes learners to process inflections can even benefit the written production of new verbal morphology, especially when learners receive metalinguistic feedback during training.

(b) Is working memory (WM) capacity related to learners’ accuracy in written production of inflections (familiar or new) and their ability to learn from training and/or improve over time?
Although WM capacity was associated with overall accuracy in producing the present tense and this relationship approached significance for the preterite tense, WM was not related to the learners’ ability to improve in their production of these tenses after exposure to the training. However, for the future tense, WM capacity was positively associated with learners’ ability to learn to produce the new future inflections after their extensive exposure to these forms. That is, learners with higher WM capacity benefited more from training and exhibited greater improvement in their written production of the future tense from the pretests to the posttest.

3.4 Assessment 3: Cue Reliance Task

The cue reliance task was designed to assess learners’ reliance on lexical as opposed to verbal inflectional cues to determine temporal and subject reference, before and after training.

The research questions to be answered with the data presented here are:

(a) Do English-Spanish beginning learners rely predominately on lexical cues rather than verbal inflections for temporal and subject reference as previous research suggests?

(b) Does training that promotes attention to morphological cues in the input help English-Spanish learners override their L1-based processing strategy (relying predominately on lexical cues) and depend more on verbal inflections for temporal and subject reference?

(c) Is working memory (WM) capacity related to learners’ cue biases and processing of verbal inflections?

3.4.1 Participants

Two hundred and thirty-nine of the participants (104 in experimental groups; 135 in the control group) who met the general inclusion criteria outlined in Section 3.1.1 successfully completed the pretest, immediate posttest, and delayed posttest versions of the cue reliance task. Additional inclusion criteria for the cue reliance task required that the participants scored at least 70% in the grammatical condition for familiar tenses (further explained in Section 3.4.2), which
reduced the number of participants in each group. There were 84 participants in the experimental groups who met the criteria, and the breakdown was as follows: no feedback group (n = 26), yes-no feedback group (n = 26), and metalinguistic feedback group (n = 32). Of the 121 participants in the control group who met the criteria, a randomly selected subset of the participants (n=40) was used for the analyses. Table 9 presents the descriptive statistics for the screening tests, pretest measures, and WM task corresponding to the subset of the participant pool used for the cue reliance task analyses.

Table 9. Descriptive Statistics for the Subset of the Participant Pool Used in the Cue Reliance Task Analyses on Screening, Proficiency, and WM Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Proficiency</td>
<td>No feedback</td>
<td>26</td>
<td>50.32</td>
<td>11.96</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>52.40</td>
<td>10.49</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>53.13</td>
<td>12.48</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>54.17</td>
<td>11.04</td>
</tr>
<tr>
<td>Vocabulary Test</td>
<td>No feedback</td>
<td>26</td>
<td>96.56</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>98.28</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>97.29</td>
<td>4.20</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>97.83</td>
<td>4.29</td>
</tr>
<tr>
<td>Written Production Pretest (Present tense)</td>
<td>No feedback</td>
<td>26</td>
<td>83.17</td>
<td>22.90</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>79.81</td>
<td>18.73</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>88.67</td>
<td>14.67</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>84.38</td>
<td>19.76</td>
</tr>
<tr>
<td>Written Production Pretest (Preterite tense)</td>
<td>No feedback</td>
<td>26</td>
<td>58.39</td>
<td>20.93</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>57.93</td>
<td>25.34</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>61.72</td>
<td>20.06</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>57.66</td>
<td>24.33</td>
</tr>
<tr>
<td>Written Production Pretest (Future tense)</td>
<td>No feedback</td>
<td>26</td>
<td>2.40</td>
<td>7.09</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>4.81</td>
<td>11.22</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>3.52</td>
<td>6.53</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>2.50</td>
<td>5.06</td>
</tr>
<tr>
<td>Aural Processing Pretest</td>
<td>No feedback</td>
<td>26</td>
<td>56.73</td>
<td>13.34</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>55.61</td>
<td>13.54</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>57.16</td>
<td>11.75</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>56.67</td>
<td>12.04</td>
</tr>
<tr>
<td>Exam 1 Grade</td>
<td>No feedback</td>
<td>26</td>
<td>83.09</td>
<td>10.83</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>82.40</td>
<td>10.37</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>32</td>
<td>84.82</td>
<td>9.81</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>40</td>
<td>84.65</td>
<td>9.29</td>
</tr>
<tr>
<td>Working Memory (LNS)</td>
<td>No feedback</td>
<td>26</td>
<td>12.77</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>Metalinguistic feedback</td>
<td>Control</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>32</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.38</td>
<td>12.63</td>
<td>12.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.73</td>
<td>3.10</td>
<td>2.81</td>
<td></td>
</tr>
</tbody>
</table>

One-way ANOVAs performed on the preliminary measures presented in Table 9 revealed no significant differences between the groups on the grammar proficiency test, $F(3,120) = .607, p = .612$, the vocabulary test, $F(3,142) = .573, p = .634$, the production pretest for present tense, $F(3,120) = 1.068, p = .365$, the production pretest for preterite tense, $F(3,120) = .222, p = .881$, the production pretest for future tense, $F(3,120) = .632, p = .595$, the aural processing pretest, $F(3,120) = .076, p = .973$, the first exam grade (taken just prior to the start of the study), $F(3,120) = .419, p = .740$, or the letter-number sequencing WM task, $F(3,120) = .092, p = .964$.

### 3.4.2 Materials and Procedure

Following the lead of Competition Model studies on cue use (e.g., MacWhinney, 1987) as well as recent research examining the effects of learned attention on cue preferences (N. Ellis & Sagarra, 2010a, 2010b, 2011), the cue reliance task involved placing the target linguistic cues in conflict to determine learners’ cue biases when asked to judge the temporal and subject reference of each sentence. The logic of the experiment is simple; every experimental sentence contains two temporal references – an adverbial cue and a verbal inflectional cue – and two references to the subject – an explicit subject pronoun and the verbal inflection. If participants pay equal attention to the lexical and morphological cues, then their responses to the sentences with agreement violations should be equally affected by both cues. If, however, they rely more on the lexical cues than the morphological ones, or vice versa, then their responses will be swayed toward the corresponding cues.

In this task, which was administered with E-Prime, participants were visually presented with a series of short Spanish sentences containing two cues to temporal reference (adverbial...
expressions and inflections) as well as two cues to the subject/agent (explicit subject pronouns and inflections) of the sentence. Following each sentence, the participants were asked to determine the tense and subject they thought best applied to each given sentence. The task took approximately 12 minutes to complete.

The experimental stimuli comprised three 58-sentence sets, one of which was randomly presented to each participant. Each set contained 4 practice trials and 54 experimental sentences with 18 items in each of the three conditions exemplified in Table 10. Appendix H includes the experimental sentences used in the three versions of the cue reliance task.

Table 10. Examples of Experimental Conditions for the Cue Reliance Task

<table>
<thead>
<tr>
<th>Condition type</th>
<th>Sample stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Agreement</td>
<td>*El mes pasado yo visit-é las escuelas.</td>
</tr>
<tr>
<td></td>
<td>‘Last month I visited the schools.’</td>
</tr>
<tr>
<td>(2) Adv-V tense agreement violation</td>
<td>*El próximo mes yo visit-é las escuelas.</td>
</tr>
<tr>
<td></td>
<td>‘Next month I visited the schools.’</td>
</tr>
<tr>
<td>(3) S-V agreement violation</td>
<td>*El mes pasado él visit-é las escuelas.</td>
</tr>
<tr>
<td></td>
<td>‘Last month he visited the schools.’</td>
</tr>
</tbody>
</table>

All experimental sentences comprised a three-word adverbial phrase, a subject pronoun, a verb, and a direct object. They were balanced for tense (preterite, present, future), person (1SG, 3SG, 1PL, 3PL), and position of the adverbial phrase (sentence-initial vs. sentence-final). Each of the 18 target verbs appeared once in each condition. Filler sentences were not included because the explicit focus-on-form nature of the task rendered them superfluous.

As seen in the example trial in Figure 18, each trial began with a 500-ms fixation sign (+) in the center of the computer screen before the sentence appeared. Then, each sentence remained on the screen for four seconds. This duration period is consistent with some previous studies involving block presentation of a single sentence of a similar length (e.g., Bialystok & Miller, 1999), and it was also judged by pilot testing to be a sufficient amount of time for beginning
learners to read each sentence one time, but not multiple times. After the sentence disappeared, the participants responded at their own pace to the questions with designated keys that were labeled with the numbers corresponding to the given tense or person, as shown in Figure 18.

Figure 18. Example of a Trial for the Cue Reliance Task

The participants’ responses to the tense and subject questions were recorded via E-Prime. Analyses then involved calculating participants’ percentage of reliance on the verbal inflection, the adverbial phrase, and neither for the tense violation condition and the subject-verb (S-V) agreement violation condition. Accuracy was also measured in the grammatical condition (tense and S-V agreement), and the data from this condition on the pretest was used as a baseline for performance that served to make exclusions due to low accuracy in determining the tense or subject even when both the lexical and morphological cue agreed and indicated the same
response. It was presupposed that if a participant could not accurately determine the tense or subject for the known tenses (present and preterite) in this grammatical condition the majority of the time, their data from the conditions of interest would not be reliable. Thus, as previously mentioned in Section 3.4.1, participants who scored lower than 70% on the pretest in the correct condition for known tenses were excluded.

3.4.3 Results

In the present section, I present the results of analyses on the cue reliance data with the aim of shedding light on whether form-focused training conditions and working memory (WM) capacity had any effects on learners’ cue biases; that is, their reliance on verbal inflections as opposed to lexical cues. The analyses involved a mixed design examining whether the within-subject variable of Time (pretest, posttest, delayed posttest) and the two between-subjects variables of Group (no feedback training, correct-incorrect feedback training, metalinguistic feedback training, control with no training) and WM capacity may potentially have effects on the participants’ percentage of reliance on verbal inflections (the dependent variable). The statistical tests run include a series of one-way and repeated-measures ANOVAs.

One-way ANOVAs were first conducted on pretest scores to determine if there were any between-group differences with respect to cue biases prior to the training. Results revealed that on the pretest the groups were not statistically different with respect to their percentage of reliance on verbal inflections in the tense condition, $F(3,120) = .521$, $p = .669$, or the Subject condition, $F(3,120) = 1.774$, $p = .156$. This means that any differences found between groups on the immediate and delayed posttest can be attributed to the training.

Table 11 presents the descriptive statistics for the cue reliance task, including group mean percentages for reliance on verbal inflections and standard deviations at each test time.
Table 11. Descriptive Statistics for the Cue Reliance Assessment: Reliance on Verbal Morphology

<table>
<thead>
<tr>
<th>Groups</th>
<th>Statistic</th>
<th>Tense Condition</th>
<th>Subject Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td>No Feedback</td>
<td>Mean</td>
<td>12.61</td>
<td>51.50</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>17.89</td>
<td>35.68</td>
</tr>
<tr>
<td>Yes-No Feedback</td>
<td>Mean</td>
<td>13.46</td>
<td>60.90</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>14.42</td>
<td>32.45</td>
</tr>
<tr>
<td>Metalinguistic Feedback</td>
<td>Mean</td>
<td>17.71</td>
<td>83.16</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>19.47</td>
<td>22.47</td>
</tr>
<tr>
<td>Control</td>
<td>Mean</td>
<td>15.69</td>
<td>12.92</td>
</tr>
</tbody>
</table>

Note. The scores above are percentages, indicating the degree to which the groups rely on verbal morphology to determine the tense/subject at each testing time.

The group means at the time of the pretest reveal that, as expected, all groups relied little on verbal inflections prior to training given their overwhelming preference for the lexical cue (the adverb in the tense condition and the explicit subject in the subject condition). Group mean percentages on the posttests indicate that all three groups that underwent training showed an increase in reliance on verbal inflections while the control group, on the other hand, showed little change in cue reliance, continuing to rely predominately on lexical cues and little on inflections. Specifically, while the control group showed no increase in percentage of reliance on inflections from the pretest to the immediate posttest for either condition, the average increase in reliance on inflections from the pretest to the immediate posttest for all the participants who underwent training was 51% in both the tense and subject conditions. Moreover, the means suggest that the three training groups differed with respect to their increase in reliance on verbal inflections, with the metalinguistic feedback group showing the greatest increase, followed by the yes-no feedback group, and then the no feedback group. The groups’ mean percentage of reliance on verbal inflections at each test time is presented visually below in Figure 19 and Figure 20.
To determine whether the training conditions had significant effects on learners’ cue biases, two 3 (Time) x 3 (Group) repeated-measure ANCOVAs (one for the tense condition and one for the
subject condition), with WM as a covariate, were performed. The analyses revealed a significant main effect for Time in the tense condition, $F(1.76,209) = 7.19, p = .002$, although it was non-significant in the subject condition. Analyses also confirmed that there were significant main effects for Group in both conditions (tense condition, $F(3,119) = 29.44, p < .001$, subject condition, $F(3,119) = 23.96, p < .001$). Moreover, the analyses yielded significant Time x Group interactions (tense, $F(5.27,209) = 27.98, p < .001$, subject, $F(5.87,232.80) = 23.61, p < .001$). Pairwise comparisons revealed the following two consistent between-group differences as a result of training: (a) all experimental groups relied significantly more on verbal inflections than the control group for both tense and subject conditions (all $p < .001$), and (b) the metalinguistic feedback group relied significantly more on verbal inflections as compared to the no feedback group for both tense and subject conditions (all $p < .001$). Moreover, there were additional between-group differences for the training groups that applied to only one of the conditions. For the tense condition, the metalinguistic group not only relied significantly more on inflections as compared to the no-feedback group, but also as compared to the yes-no feedback group ($p = .009$). That is, the metalinguistic feedback group came to rely significantly more on inflections for tense than each of the three other groups. There were no significant differences between the yes-no feedback and no feedback groups in the tense condition. However, for the subject condition the yes-no feedback group did come to rely significantly more on inflections as compared to the no-feedback groups ($p = .028$). In sum, receiving training, regardless of feedback, led to a significantly greater reliance on inflections as compared to lexical cues. Nonetheless, training with metalinguistic feedback was overall more effective than training without feedback for altering cue biases, and it was also more effective in the tense condition.
than yes-no feedback, which in turn was only more effective than no feedback in the subject condition.

With respect to WM, there were no significant main effects, although for the subject condition the main effect approached significance, $F(1,119) = 3.111, p = .08$. There was, however, a significant interaction between Time and WM in the subject condition, $F(1.96,232.80) = 3.47, p = .034$, but not in the tense condition. This indicates that while WM was not associated with an increase in reliance on verbal inflections to determine tense over time, WM was positively related to an increase in reliance on verbal inflections (as opposed to explicit subject pronouns) to determine the subject of the sentence. There are two possible explanations, which are likely related, for this finding. First, it might be that WM ability was only helpful for learning to process the non-meaning-bearing person-number inflections that involve a purely grammatical dependency with the lexical cue (the subject). In that tense inflections, on the other hand, are a semantic marker that bears meaning – situating an event in time – WM ability may not facilitate the processing of these cues. Moreover, given that all the participants were L1 English speakers, and English does mark regular past tense with the meaning-bearing -ed inflection, the learners had some L1 experience with relying on inflections for temporal reference. Thus, WM ability may not have enhanced learners’ processing of tense inflections whereas greater cognitive resources did aid learners in processing the person-number inflections that are non-meaning bearing and also almost entirely absent in the L1 (the 3sg –s marker in regular present tense is one exception, but this marker is often ignored due to the anomaly of its presence and the obligatory use of explicit subjects in English).
3.4.4 Summary of Results

The cue reliance task was used to assess the effects of training as well as WM on learners’ cue biases, specifically their reliance on lexical as opposed to inflectional cues to determine tense and subject. In order to summarize the findings, the specific research questions relevant to the cue reliance task are restated and addressed briefly in turn below.

(a) Do beginning English-Spanish learners rely predominately on lexical cues rather than verbal inflections for temporal and subject reference as previous research suggests?

Yes. Consistent with previous research and unlike native speakers of a morphologically rich language like Spanish, prior to training the beginning English-Spanish learners exhibited overwhelming preference toward processing the lexical cue (the adverb or explicit subject) over the morphological one (verbal inflections) to determine the tense and subject of the sentence.

(b) Does training that promotes attention to morphological cues in the input help English-Spanish learners override their L1-based processing strategy (relying predominately on lexical cues) and depend more on verbal inflections for temporal and subject reference?

Yes. The training helped learners override their L1-experience-based strategy of attending predominately to lexical cues to determine temporal and subject reference. All the participants who received training, but not the control group, came to rely significantly more on L2 verbal inflections, the more valid cue in morphologically rich, null-subject languages like Spanish. This suggests that extensive exposure to input activities that lack lexical cues and require processing of inflections for accurate interpretation of the message can help learners continue to process morphological cues even when they are later encountered in redundant, less salient contexts. While training promoted increased reliance on inflections regardless of whether the learners received
corrective feedback, the provision of metalinguistic feedback significantly enhanced the effectiveness of the training overall, and yes-no feedback was more effective than no feedback in overcoming cue biases for the tense condition.

(c) Is working memory (WM) capacity related to the learners’ cue biases and processing of verbal inflections?

WM capacity had some limited effects on the learners’ cue biases. Higher WM capacity was positively related to learners’ ability over time (on the posttests) to overcome cue biases and rely more on verbal inflections for subject reference, but not for temporal reference.

3.5 Assessment 4: Self-paced Reading Task

A self-paced reading moving-window technique (Just et al., 1982) was used to assess beginning learners’ morphological sensitivity, or online processing of verbal agreement, before and after training. The research questions to be answered with the data presented here are:

(a) Are beginning L2 English-Spanish learners sensitive to tense and subject-verb (S-V) agreement violations during online processing for meaning?

(b) Does training that promotes attention to morphological cues in the input influence learners’ sensitivity to verbal agreement violations?

(c) How does WM capacity relate to L2 learners’ ability to process agreement online?

3.5.1 Participants

Two hundred and twenty-one of the participants (105 in experimental groups; 116 in the control group) who met the general inclusion criteria outlined in Section 3.1.1 successfully completed the pretest, immediate posttest, and delayed posttest versions of the self-paced reading task. These participants’ accuracy on the comprehension measure in the self-paced reading task (described below in Section 3.5.3.1) was also taken into account as a criterion for inclusion in the
final analyses, with a requirement of 72% correct on all of the experimental items. This comprehension accuracy threshold (justified in Section 3.5.3.1) decreased the number of participants in each group. There were 83 participants in the experimental groups who met the criteria, and the breakdown was as follows: no feedback group (n = 28), correct-incorrect feedback group (n = 26), and metalinguistic feedback group (n = 29). There were 101 participants in the control group who met the criteria, of which a subset of approximately 30% (n = 31) was randomly selected by SPSS for the analyses involving between-group comparisons. Table 12 presents the descriptive statistics for the screening tests, pretest measures, and WM task corresponding to the subset of the participant pool used for the self-paced reading task analyses.

Table 12. Descriptive Statistics for the Subset of the Participant Pool Used in the Self-Paced Reading Analyses on Screening, Proficiency, and WM Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar Proficiency</td>
<td>No feedback</td>
<td>28</td>
<td>47.62</td>
<td>14.72</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>53.69</td>
<td>10.30</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>29</td>
<td>52.59</td>
<td>10.59</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31</td>
<td>52.96</td>
<td>10.61</td>
</tr>
<tr>
<td>Vocab Test</td>
<td>No feedback</td>
<td>28</td>
<td>93.80</td>
<td>6.88</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>98.38</td>
<td>4.15</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>29</td>
<td>97.28</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31</td>
<td>97.71</td>
<td>4.54</td>
</tr>
<tr>
<td>Written Production Pretest (Present tense)</td>
<td>No feedback</td>
<td>28</td>
<td>83.04</td>
<td>21.57</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>83.65</td>
<td>18.29</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>29</td>
<td>87.07</td>
<td>13.15</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31</td>
<td>86.09</td>
<td>23.15</td>
</tr>
<tr>
<td>Written Production Pretest (Preterite tense)</td>
<td>No feedback</td>
<td>28</td>
<td>54.89</td>
<td>18.20</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>68.51</td>
<td>23.22</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>29</td>
<td>61.85</td>
<td>19.36</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31</td>
<td>56.85</td>
<td>26.53</td>
</tr>
<tr>
<td>Written Production Pretest (Future tense)</td>
<td>No feedback</td>
<td>28</td>
<td>4.02</td>
<td>9.65</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>7.21</td>
<td>13.77</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>29</td>
<td>4.96</td>
<td>9.36</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31</td>
<td>7.86</td>
<td>11.18</td>
</tr>
<tr>
<td>Aural Processing Pretest</td>
<td>No feedback</td>
<td>28</td>
<td>52.68</td>
<td>15.43</td>
</tr>
<tr>
<td></td>
<td>Yes-No feedback</td>
<td>26</td>
<td>58.97</td>
<td>14.89</td>
</tr>
<tr>
<td></td>
<td>Metalinguistic feedback</td>
<td>29</td>
<td>57.61</td>
<td>12.80</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>31</td>
<td>57.12</td>
<td>12.37</td>
</tr>
<tr>
<td>Exam 1 Grade</td>
<td>No feedback</td>
<td>28</td>
<td>80.85</td>
<td>11.15</td>
</tr>
</tbody>
</table>
One-way ANOVAs performed on the preliminary measures presented in Table 12 revealed no significant differences between the groups on the grammar proficiency test, $F(3,110) = 1.565, p = .202$, the vocabulary test, $F(3,142) = .573, p = .634$, the production pretest for present tense, $F(3,110) = .277, p = .842$, the production pretest for preterite tense, $F(3,110) = 2.042, p = .112$, the production pretest for future tense, $F(3,110) = .783, p = .506$, the aural processing pretest, $F(3,110) = 1.067, p = .366$, the first exam grade (taken just prior to the start of the study), $F(3,110) = 1.222, p = .305$, or the letter-number sequencing WM task, $F(3,110) = 1.524, p = .212$. There were, however, significant differences between the groups on the vocabulary test, $F(3,110) = 4.593, p = .005$. Bonferroni post hoc tests showed that between-group differences were due to the lower accuracy of the no feedback group on the vocabulary test as compared to the correct/incorrect feedback group ($p = .007$) and the control group ($p = .021$). In that the mean accuracy for the no feedback group was still very high (93.8%), it is evident that the group had a solid grasp of the relevant experimental vocabulary and their slightly lower accuracy compared to the other groups was not considered justification for additional exclusions, or reason to believe that the no feedback group was of a lower proficiency than the other groups.

### 3.5.2 Materials and Procedure

In the self-paced reading task, which was implemented with E-Prime and lasted approximately 15 minutes, participants read Spanish sentences word by word in a non-cumulative manner and then chose which of two pictures accurately depicted the meaning of the
sentence. The stimuli consisted of three 74-item sets, one of which was randomly presented to each participant. Each set included 4 practice trials, 36 experimental sentences, and 36 filler sentences. Half of all the sentences were grammatical, and the other half were ungrammatical.

The conditions for the experimental sentences are illustrated in Table 13. Appendix I presents the experimental sentences used in the three versions of the task.

Table 13. Examples of Experimental Conditions for the Self-Paced Reading Task

<table>
<thead>
<tr>
<th>Condition type</th>
<th>Sample stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Agreement</td>
<td>*El año pasado ella recibió el diploma en la ceremonia.</td>
</tr>
<tr>
<td></td>
<td>*The year last she-3SG received-3SG the diploma at the ceremony.</td>
</tr>
<tr>
<td></td>
<td>‘Last year she received the diploma at the ceremony.’</td>
</tr>
<tr>
<td>(2) Adv-V tense agreement violation</td>
<td>*El próximo año ella recibió el diploma en la ceremonia.</td>
</tr>
<tr>
<td></td>
<td>*The next year she-3SG received-3SG the diploma at the ceremony.</td>
</tr>
<tr>
<td></td>
<td>**‘Next year she received the diploma at the ceremony.’</td>
</tr>
<tr>
<td>(3) S-V agreement violation</td>
<td>*El año pasado yo recibió el diploma en la ceremonia.</td>
</tr>
<tr>
<td></td>
<td>*The year last I-1SG received-3SG the diploma at the ceremony.</td>
</tr>
<tr>
<td></td>
<td>**‘Last year I received the diploma at the ceremony.’</td>
</tr>
</tbody>
</table>

All of the experimental sentences followed the same syntactic structure and contained exactly 10 words. Verb tense (preterite vs. future) was matched across conditions, and each of the 18 target verbs was used once in each tense. The filler sentences were also controlled for vocabulary (incorporating familiar, non-target vocabulary from the textbook) and were of a similar length (7 to 11 words) to the experimental sentences, but they tested other aspects of processing (subject and object relative clauses, long-distance gender agreement, and clitic pronouns). All of the sentences were presented in a random order to avoid consecutive presentation of experimental sentences.

As illustrated in Figure 21, each trial began with a 500-ms fixation sign (+) in the center of the computer screen. Following this, the press of the space bar revealed the first word of the sentence, left justified on the screen, along with place-holding dashes marking the position of all the letters and words in the sentence. Each dash represented a letter, the words were separated with spaces, and a period marked the end of each sentence. This helped make reading more
natural by visibly displaying the characters normally available during reading. The participants were instructed to read each word silently and to use the space bar to make the visible word disappear and to display each consecutive word on the screen.

Figure 21. Example of a Trial for the Self-Paced Reading Task

Pressing the space bar after the final word of each trial sentence prompted the appearance of two pictures, and participants used one of two designated keys (one on the right side of the keyboard and one on the left side) to indicate which picture best represented the meaning of the sentence. This picture-based comprehension task was used for two main reasons. First, it pushed learners to focus on comprehending the meaning of the sentence as opposed to explicitly thinking about grammar. Thus, high response accuracy would indicate that evidence of longer reading times (RTs) on the critical region are due to sensitivity to agreement violations rather than to failure to
comprehend meaning. Second, pictures were used in lieu of comprehension questions in order to avoid biasing learners’ attention to temporal reference and S-V agreement. Pictures were professional clip-art drawings that were matched in size and style as well as pilot tested. To force participants to attend to all aspects of the sentences (experimental and fillers), the pictures referred to different aspects of the sentences. With respect to the experimental sentences, one-third of the pictures assessed comprehension of the prepositional phrase (location), one-third assessed comprehension of the verb (action), and the final third assessed comprehension of the direct object. For example, Figure 22 illustrates the choice between two pictures based on the example experimental sentence presented above in Table 13 and repeated here for convenience: *El año pasado ella recibió el diploma en la ceremonia.* ‘Last year she received the diploma at the ceremony.’ The sentence could have appeared in any of the three conditions (see Table 13) depending on the set of experimental sentences randomly assigned to the participant. A sample of the pictures used in the comprehension task, accompanied by the corresponding experimental sentence, can be found in Appendix I.

![Figure 22. Example of the Picture-Based Comprehension Measure in the Self-Paced Reading Task.](image)

The pictures depict two different objects that could be received at a ceremony, and based on the sentence, the correct response is the picture of the diploma on the left, and not the trophy.
Within each of the three picture conditions, the correct answer for half of the trials was the picture on the left, and for the other half, it was the picture on the right. The experiment began with a brief description of the moving-window procedure as well as instructions for choosing the correct picture in the comprehension task. Four practice items preceded the experimental stimuli in order to familiarize participants with the procedure.

Reading times (RTs) on each word and responses to the picture-based comprehension task were recorded via E-Prime. As previously mentioned, data was only analyzed from participants who scored 72% or higher on the picture comprehension measure at all three test times. In addition, following standard practice in this type of experiment, all RTs that were below 200ms or above 2,000ms, or that corresponded to sentences with incorrect responses to the comprehension measure, were discarded. In addition, RTs that exceeded +/- 2 standard deviations from a participant’s mean in each condition (agreement, Adv-V tense violation, S-V agreement violation) were eliminated. After performing this data trimming, analyses were conducted based on the mean RTs in milliseconds on the critical region (the verb), the word prior to the critical region, and the two words following the critical region. Details regarding the regions analyzed are discussed further in Results Section 3.5.3.2.

3.5.3 Results

In the present section, I provide the results of various analyses on the self-paced reading data with the aim of shedding light on how form-focused training conditions and individual cognitive differences may affect beginning L2 learners’ online processing of tense and S-V agreement violations. The analyses followed a mixed design examining the effects of two within-subjects variables, Agreement (tense/S-V agreement, tense violation, S-V violation) and Time (pretest, posttest, delayed posttest) as well as two between-subjects variables, namely Group (no
feedback training, correct-incorrect feedback training, metalinguistic feedback training, control group) and WM. The statistical analyses include a series of one-way and repeated-measures ANOVAs.

3.5.3.1 Accuracy on the Picture Comprehension Measure

As mentioned in 3.5.1, data was only analyzed from participants who scored 72% or above on the picture comprehension measure for the experimental trials in the self-paced reading task at all three test times. This inclusion criterion was implemented to ensure that the L2 learners read the sentences for comprehension and were generally able to understand their meaning. The specific cutoff of 72% was chosen for two reasons. First, it falls in the middle of the 60 – 80% accuracy cutoff range that has been employed for the comprehension measure in previous self-paced reading studies examining L2 morphosyntactic processing while reading for meaning (e.g., Jiang, 2004, 2007; Jiang et al., 2011; Sagarra & Herschensohn, 2010, 2011). Second, the 72% cutoff allowed for inclusion of the number of participants needed for statistical analyses; a higher cutoff would have eliminated too many participants for the present analyses. A higher cutoff will be used, though, in subsequent analyses reported in Section 3.5.3.4. The mean accuracy percentages for the four groups are displayed in Table 14.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest Accuracy (%)</th>
<th>Posttest Accuracy (%)</th>
<th>Delayed Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback (n=28)</td>
<td>85.46 (7.63)</td>
<td>86.46 (6.94)</td>
<td>86.86 (6.36)</td>
</tr>
<tr>
<td>Cor-Inc Feedback (n=26)</td>
<td>88.77 (5.78)</td>
<td>90.38 (6.70)</td>
<td>89.46 (5.84)</td>
</tr>
<tr>
<td>Metalinguistic Feedback (n=29)</td>
<td>87.14 (6.21)</td>
<td>90.24 (5.89)</td>
<td>90.48 (4.48)</td>
</tr>
<tr>
<td>Control Subset (n=31)</td>
<td>87.82 (6.28)</td>
<td>88.63 (6.17)</td>
<td>88.28 (5.73)</td>
</tr>
</tbody>
</table>

One-way ANOVAs conducted on these accuracy scores showed that there were no significant effects for group at the time of the pretest, $F(3,110) = 1.254, p = .294$, immediate
posttest, $F(3,110) = 2.247, p = .087$, or delayed posttest, $F(3,110) = 2.180, p = .094$). Thus, the sentences were not significantly harder or easier for a particular group to comprehend as compared to the others. However, as is evident by the means and $p$ values, there was a trend toward significant between-group differences in accuracy at the time of the immediate and delayed posttests. Comparison of the group means suggest that it is the lower accuracy of the no feedback group that drives the trend toward between-group differences on the posttests. Importantly, though, these differences were not significant, and the mean accuracy for all groups was high (above 85%). These comprehension rates suggest that all groups were reading the sentences for comprehension, as they were instructed, and understood the sentences well. Moreover, it is crucial to reiterate that the reading time results to be discussed in the subsequent sections correspond solely to items whose picture comprehension probe was answered correctly. This is a standard procedure followed to minimize the possibility that any latencies found in RTs could be due to a lack of comprehension (see e.g., Jiang, 2011; Sagarra & Herschensohn, 2010).

### 3.5.3.2 Reading Time Results Overall

As mentioned previously, the stimuli for the self-paced reading task were presented word-by-word in a non-cumulative manner. The relevant words or regions for the purposes of our analyses are words 4, 5, 6, and 7 as shown in Table 15.

<table>
<thead>
<tr>
<th>Table 15. Relevant Regions for Reading Time Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Agreement</td>
</tr>
<tr>
<td>Adv-V Tense Violation</td>
</tr>
<tr>
<td>S-V Violation</td>
</tr>
</tbody>
</table>

Reading times (RTs) at word 4 are pertinent to ensure that there are no differences in processing across the conditions prior to word 5, the critical verb, where latencies would be expected in the
ungrammatical conditions *if* the L2 learners are sensitive to Adv-V tense violations or S-V agreement violations. The RTs on words 6, and possibly 7, will reflect whether agreement violations result in spill-over effects, or longer RTs following the critical verb. Such effects have been found for L2 learners in some studies (see e.g., Jiang, 2011), which have been interpreted as evidence that it may take longer for non-native speakers’ sensitivity to grammatical errors to materialize for some structures.

Figures 23 and 24 show the overall processing behavior by all participants for agreement and agreement violations in the tense and S-V conditions, respectively, for the four words analyzed.

![Mean RTs for Tense Agreement and Violations Overall](image)

*Figure 23. Mean Reading Times for Tense Agreement and Violations Overall (All Groups Combined)*

* p < .05
As seen in Figures 23 and 24, mean RTs were increasingly faster at each test time. A series of repeated measures ANOVAs including Time as a within-subjects variable confirmed that there was a significant main effect for Time (all $p < .05$), indicating that the learners did in fact process the experimental stimuli faster overall at each test time regardless of Agreement, presumably as a result of task familiarity and practice. In light of this finding, separate repeated measures ANOVAs were run for each word at each time to determine if learners exhibited any sensitivity overall to agreement violations.

Results revealed no significant effects for Agreement for word 4 at any time for either the Tense or S-V conditions (all $p > .05$). This confirms that there were no differences in RTs based on congruency prior to the critical word 5 (the verb), which is to be expected since learners had not yet been exposed to a violation in the relevant conditions.
At the time of the pretest, there were no significant effects for Agreement at the critical word 5 for either the tense or S-V condition, nor at word 6 or word 7 (all \( p > .05 \)). This indicates that on the pretest, overall the learners did not exhibit sensitivity at the verb when there was an agreement violation with respect to the temporal adverb or the subject pronoun previously processed. Moreover, a lack of significant effects for Agreement at words 6 and 7 confirm that were also no spill-over effects, which indicates that learners did not exhibit latent sensitivity to agreement violations at the time of the pretest.

At the time of the posttest, as seen in Figures 23 and 24 mean RTs were longer on the critical word 5 (verb) when there was an agreement violation for tense or subject in relation to the previously encountered lexical cue (temporal adverb or subject pronoun, respectively). This pattern is what is expected if learners were sensitive to agreement violations. To determine whether Agreement was in fact a significant variable, repeated measures ANCOVAs were run on RTs for the immediate posttest, and WM was included as a covariate to also explore the role it potentially plays in processing. As indicated in Figure 23 with an asterisk, results for RTs at the verb confirmed a significant main effect for Agreement in the tense condition, \( F(1,112) = 12.51, p = .001 \), at the time of the immediate posttest. There was also a significant a significant main effect for WM, \( F(1,112) = 7.71, p = .006 \), and a significant interaction between Agreement and WM, \( F(1,112) = 16.25, p < .001 \). On the other hand, results for the S-V condition indicated a non-significant main effect for Agreement at the verb, \( F(1,112) = 2.96, p = .088 \), although there was a trend toward significance. There was, however, also a significant main effect for WM, \( F(1,112) = 5.15, p = .025 \), as well as a significant interaction between Agreement and WM, \( F(1,112) = 4.39, p = .039 \). Non-significant effects for Agreement, both for the tense and S-V
conditions, at words 6 and 7 (all \( p > .05 \)) confirm that the learners did not exhibit any latent sensitivities to agreement violations at the time of the posttest.

At the time of the delayed posttest, the repeated measures ANCOVA showed no significant effect for Agreement in the tense condition, \( F(1,112) = 1.013, p = .316 \), and no significant interaction between Agreement and WM, \( F(1,112) = 1.221, p = .271 \). There was, however, a significant main effect for WM, \( F(1,112) = 10.123, p = .002 \). With respect to the S-V condition at the time of the delayed posttest, results revealed a non-significant, yet trending, main effect for Agreement, \( F(1,112) = 3.075, p = .082 \), a significant interaction between Agreement and WM, \( F(1,112) = 4.351, p = .039 \), and a main effect for WM, \( F(1,112) = 10.867, p = .001 \). There were no significant differences in RTs based on Agreement for words 6 or 7 (all \( p > .05 \)), again indicating the absence of spillover effects.

In sum, with respect to the overall results for the tense and S-V agreement conditions, there were only significant main effects for Agreement at the immediate posttest and only for the tense condition. This indicates that the learner group as a whole only demonstrated sensitivity to adverb-verb tense violations, and this sensitivity was not maintained at the delayed posttest. However, the significant main effects for WM as well as the significant interactions between WM and Agreement suggest that WM may be playing an important role in the learners’ general processing of the experimental stimuli as well as their ability to process agreement violations, something we will explore in more depth in Section 3.5.3.4 where we compare the processing behavior of low-span learners to that of high-span learners. We now turn to more fine-grained analyses examining learners’ processing of agreement violations in the preterite tense as opposed to the future tense.
Figures 25 and 26 show the processing behavior by all participants for tense agreement and violations in the preterite and future tenses, respectively, for the four regions of interest. Consistent with the overall findings reported above, there were no significant effects of Agreement for any of the test times on the non-critical regions, namely word 4 prior to the verb, and words 6 and 7 following the verb (all $p > .05$). Thus, only results from analyses conducted for RTs at the verb (word 5) will be presented.

![RTs for Preterite Tense Agreement and Violations](image)

Figure 25. *Mean Reading Times for Preterite Tense Agreement and Violations (All Groups Combined)*
As expected based on the means displayed in Figures 25 and 26, at the time of the pretest there were no significant main effects for Agreement in the tense condition at word 5 for either the preterite tense, $F(1,112) = .240, p = .625$, or the future tense, $F(1,112) = .589, p = .445$, and there were also no significant interactions between tense Agreement and WM (preterite tense, $F(1,112) = .216, p = .643$; future tense, $F(1,112) = .611, p = .436$). There were also no main effects for WM on the pretest for the tense condition (preterite tense, $F(1,112) = 3.585, p = .061$; future tense, $F(1,112) = 2.463, p = .119$), although there was a trend toward significance in the preterite tense.

At the time of the immediate posttest, the repeated measures ANCOVA revealed no significant effect for Tense Agreement in the preterite tense, $F(1,112) = 2.040, p = .156$, and a trending but non-significant interaction between Tense Agreement in the preterite and WM, $F(1,112) = 2.784, p = .098$. There was, however, a significant main effect for WM, $F(1,112) = 300$.
At the time of the delayed posttest, there was also no significant main effect for Tense Agreement in the preterite, \(F(1,112) = 1.307, p = .255\), and no significant interaction between Agreement and WM, \(F(1,112) = 1.981, p = .162\). Yet there was a significant main effect for WM, \(F(1,112) = 13.121, p < .001\).

With respect to the future tense, results at the time of the immediate posttest revealed a significant main effect for Tense Agreement, \(F(1,112) = 12.046, p = .001\), a significant main effect for WM, \(F(1,112) = 8.882, p = .004\), and a significant interaction between Tense Agreement and WM, \(F(1,112) = 14.708, p < .001\). At the delayed posttest, there was no longer a significant main effect for Tense Agreement in the future tense, \(F(1,112) = .711, p = .401\), and no significant interaction between Tense Agreement and WM, \(F(1,112) = .552, p = .459\). Yet there remained a significant main effect for WM, \(F(1,112) = 7.676, p = .007\). These findings indicate that the learners developed sensitivity to tense violations for the unfamiliar future tense, evident at the immediate posttest, but they were not sensitive to tense violations for the known preterite tense. One possible explanation for learners exhibiting sensitivity to violations in the new future tense, but not the known preterite tense, is the salience of the future tense forms (all endings are stressed) and its consistency in formation across verb classes (the same set of endings added to the infinitive form of the verb) as compared to the shorter and more varied preterite forms (see Section 3.1.2 for a thorough description of the formation of these tenses in Spanish). Moreover, in that WM only had significant effects on processing agreement violations in the future tense, it may be that the learners with higher WM capacity are influencing these overall results based on a better ability to store and integrate the forms of the new future tense. As aforementioned, the effects of this cognitive individual difference will be further explored in Section 3.5.3.4.
Figures 27 and 28 show the learners’ processing behavior for S-V agreement and violations in the preterite and future tenses, respectively, for the four regions analyzed. In that there were again no significant effects of Agreement at the non-critical words 4, 6, and 7 (all $p > .05$), only results for RTs at the verb (word 5) will be presented.

![RTs for Preterite S-V Agreement and Violations](image)

**Figure 27. Mean Reading Times for Preterite S-V Agreement and Violations (All Groups Combined)**
Consistent with the results for the tense Agreement condition, on the pretest there were also no significant effects for S-V Agreement at word 5 for either the preterite tense, $F(1,112) = 1.120, p = .292$, or the future tense, $F(1,112) = .422, p = .517$. There were also no interactions between S-V Agreement and WM (preterite, $F(1,112) = 1.375, p = .243$; future, $F(1,112) = .206, p = .651$) at the time of the pretest. There was no main effect for WM at the time of the pretest for the preterite tense, $F(1,112) = 1.724, p = .192$, yet there was a significant main effect for WM in the future tense for the S-V condition, $F(1,112) = 4.635, p = .033$.

At the time of the immediate posttest, the repeated measures ANCOVA indicated no significant effect for S-V Agreement in the preterite tense, $F(1,112) = .402, p = .527$, as well as a non-significant interaction between S-V Agreement and WM, $F(1,112) = .731, p = .394$. Yet there was a significant main effect for WM, $F(1,112) = 4.484, p = .036$. For the delayed posttest, there was also no significant main effect for S-V Agreement in the preterite tense, $F(1,112)$
Turning to the future tense, results for the posttest revealed a main effect for S-V Agreement that approached significance, $F(1, 112) = 3.493, p = .064$, a significant interaction between S-V Agreement and WM, $F(1, 112) = 4.450, p = .037$, and a significant main effect for WM, $F(1, 112) = 5.520, p = .021$. At the time of the delayed posttest, however, there was a significant main effect for S-V Agreement in the future tense, $F(1, 112) = 6.726, p = .011$, a significant interaction between S-V Agreement and WM, $F(1, 112) = 7.949, p = .006$, and consistent with all previous analyses, a main effect for WM, $F(1, 112) = 11.043, p = .001$. Similar to the results for the tense condition, these findings for the S-V agreement condition demonstrate that the learners only exhibited sensitivity to agreement violations for the new future tense, but not for the known preterite tense. This further strengthens the possible interpretation I offered earlier regarding the nature of the future tense (it is new for the learners, its morphological form is more salient, and there is consistency in its formation across all verb classes). Moreover, significant interactions between S-V Agreement and WM at both the immediate and delayed posttests suggest that WM capacity had an influence on learners’ sensitivity to S-V agreement violations in the future tense, something that will be further explored in Section 3.5.3.4.

### 3.5.3.3 Reading Time Results for Training Groups over Time

In the previous section we examined the processing behavior of the participants as a whole. In this section we explore potential effects of training on learners’ processing of tense and S-V agreement. One-way ANOVAs with Group as a between-subjects variable were run on the difference scores (agreement condition subtracted from agreement violation condition) at each test time. Analyses are only reported for the critical word 5 (verb) as there were no effects of
Agreement at words 4, 6, or 7 for any group (all $p > .05$), which is consistent with the overall results reported above in Section 3.5.3.2. Table 16 presents the means and standard deviations for RTs at the verb based on group and condition.

Table 16. Group Mean RTs on Critical Region (Verb) in the Tense and S-V Agreement Conditions (SD in parentheses)

<table>
<thead>
<tr>
<th>Group</th>
<th>Condition</th>
<th>Pretest RTs (ms)</th>
<th>Posttest RTs (ms)</th>
<th>Delayed RTs (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Feedback (n=28)</td>
<td>Overall</td>
<td>823.34 (360.55)</td>
<td>663.22 (286.36)</td>
<td>532.89 (229.65)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>793.02 (397.96)</td>
<td>634.78 (269.02)</td>
<td>478.99 (192.35)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>861.86 (383.12)</td>
<td>706.73 (345.69)</td>
<td>583.70 (278.35)</td>
</tr>
<tr>
<td></td>
<td>Tense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>824.30 (320.22)</td>
<td>710.69 (344.46)</td>
<td>563.24 (265.55)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>765.20 (328.82)</td>
<td>684.74 (346.59)</td>
<td>525.68 (256.34)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>872.91 (358.72)</td>
<td>724.25 (358.19)</td>
<td>590.66 (303.95)</td>
</tr>
<tr>
<td></td>
<td>S-V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>808.96 (275.60)</td>
<td>675.92 (304.29)</td>
<td>599.99 (304.03)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>722.63 (255.59)</td>
<td>627.16 (273.61)</td>
<td>570.68 (324.69)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>898.91 (357.01)</td>
<td>720.77 (352.78)</td>
<td>627.18 (338.01)</td>
</tr>
<tr>
<td>Yes-No Feedback (n=26)</td>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>788.56 (288.72)</td>
<td>559.00 (261.05)</td>
<td>503.81 (207.71)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>712.23 (251.86)</td>
<td>557.69 (279.03)</td>
<td>489.53 (205.51)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>870.09 (358.98)</td>
<td>564.56 (266.69)</td>
<td>520.84 (225.65)</td>
</tr>
<tr>
<td></td>
<td>Tense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>789.25 (301.63)</td>
<td>580.91 (303.29)</td>
<td>479.41 (192.91)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>724.58 (299.40)</td>
<td>548.31 (305.25)</td>
<td>482.36 (206.17)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>853.40 (375.05)</td>
<td>611.27 (309.83)</td>
<td>474.78 (197.58)</td>
</tr>
<tr>
<td></td>
<td>S-V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>766.22 (276.17)</td>
<td>573.69 (275.60)</td>
<td>480.80 (182.59)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>732.19 (285.10)</td>
<td>544.78 (254.22)</td>
<td>454.82 (164.14)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>859.62 (315.10)</td>
<td>602.37 (314.09)</td>
<td>507.90 (228.86)</td>
</tr>
<tr>
<td>Metalinguistic Feedback (n=29)</td>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>803.12 (280.24)</td>
<td>636.79 (266.73)</td>
<td>597.40 (265.23)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>750.22 (244.08)</td>
<td>579.63 (262.94)</td>
<td>583.00 (248.63)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>850.09 (368.29)</td>
<td>687.32 (332.48)</td>
<td>613.90 (301.26)</td>
</tr>
<tr>
<td></td>
<td>Tense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>814.55 (301.71)</td>
<td>672.25 (308.51)</td>
<td>621.68 (307.07)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>764.90 (313.91)</td>
<td>586.42 (220.57)</td>
<td>623.56 (312.87)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>867.48 (336.58)</td>
<td>760.99 (444.19)</td>
<td>624.60 (344.35)</td>
</tr>
<tr>
<td></td>
<td>S-V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>800.59 (279.75)</td>
<td>676.45 (283.71)</td>
<td>609.15 (285.92)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>777.18 (271.69)</td>
<td>644.63 (248.02)</td>
<td>610.39 (294.91)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>833.09 (328.35)</td>
<td>714.12 (367.59)</td>
<td>606.85 (305.09)</td>
</tr>
<tr>
<td>Control (n=31)</td>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>793.03 (338.96)</td>
<td>577.08 (267.88)</td>
<td>515.16 (254.10)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>770.51 (371.24)</td>
<td>560.75 (242.03)</td>
<td>496.03 (260.28)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>826.65 (405.35)</td>
<td>598.60 (323.31)</td>
<td>534.54 (260.28)</td>
</tr>
<tr>
<td></td>
<td>Tense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>784.60 (314.88)</td>
<td>608.47 (287.51)</td>
<td>506.70 (233.11)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>758.00 (309.47)</td>
<td>595.94 (294.77)</td>
<td>492.08 (232.51)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>813.43 (340.45)</td>
<td>617.05 (304.78)</td>
<td>516.92 (242.04)</td>
</tr>
<tr>
<td></td>
<td>S-V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>763.61 (294.09)</td>
<td>609.66 (280.71)</td>
<td>534.52 (286.41)</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>748.54 (336.48)</td>
<td>591.90 (312.27)</td>
<td>506.22 (276.91)</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>775.38 (285.61)</td>
<td>625.95 (295.31)</td>
<td>573.92 (350.33)</td>
</tr>
</tbody>
</table>
The one-way ANOVAs conducted on difference scores between agreement and violation conditions at the verb showed no main effects for Group (all $p > .05$). The results of the one-way ANOVAs are shown in Table 17.

Table 17. Between-Group Results of Statistical Analyses (One-way ANOVAs) on Difference Scores (Agreement Violation – Agreement RTs) for Critical Word 5 (Verb)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>Sig</td>
<td>$F$</td>
</tr>
<tr>
<td>Tense</td>
<td>Overall</td>
<td>.082</td>
<td>.970</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>.253</td>
<td>.859</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>.115</td>
<td>.951</td>
</tr>
<tr>
<td>S-V</td>
<td>Overall</td>
<td>.134</td>
<td>.939</td>
</tr>
<tr>
<td></td>
<td>Preterite</td>
<td>.995</td>
<td>.398</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>1.054</td>
<td>.372</td>
</tr>
</tbody>
</table>

The lack of significant differences between groups suggests that the form-focused training did not enhance sensitivity to tense or S-V agreement violations during online processing when focused on comprehension of meaning. The increase in sensitivity overall to agreement at the posttest, then, is likely influenced by familiarization with the task, and perhaps by further exposure to classroom input with tense and S-V agreement. In addition, WM was reported in Section 3.5.3.2 to have effects on learners’ processing, and thus we turn now to further explore the role of this cognitive individual difference in the online processing of tense and S-V agreement.

3.5.3.4 Reading Time Results for High vs. Low WM Groups

In Section 3.5.3.2, WM was included as a covariate (a continuous variable), and this cognitive individual difference was consistently shown to have a significant main effect on learners’ RTs on the critical verb. The fact that the covariate WM span score also interacted with Agreement in some conditions indicated that WM also had some effect on learners’ sensitivity to agreement violations. To further examine the extent to which the availability of more cognitive
resources modulates online processing of tense and S-V agreement, between-group analyses were run by using the median WM score (13) to divide the learners into a high and low WM group. In that no between-group differences were found based on training conditions (see Section 3.5.3.3), this factor was not taken into consideration, and the entire control group, as opposed to just a subset, was used to increase the participant pool to a total of 191 participants (84 participants received training and 107 were control) prior to exclusions based on comprehension accuracy. The cutoff for comprehension accuracy was also raised from 72% to 80%. This stricter cutoff was not an option in previous analyses as it would have eliminated too many participants to have enough statistical power for each group, but it was implemented here for three main reasons. First, a lower comprehension error rate ensures the participants had less difficulty comprehending the overall meaning of the sentences, and thus should show a clearer picture of the effects of WM on the processing of morphosyntactic agreement violations. Second, the above 80% cutoff generates high and low WM groups that are of more equal size. At the 72% cutoff there were almost twice as many low-span participants (n=116) as high-span participants (n=68), but with the stricter cutoff this difference is minimized since many more low-span participants are excluded, resulting in the following groups used for the analyses reported in this section: low-span (n=84) and high-span (n=58). It is also important to mention that although clearer effects of WM are seen at the above 80% cutoff, analyses run at the 72% revealed the same general pattern for significant effects of WM. All the results reported and displayed in this section are with the comprehension accuracy cutoff at above 80%. Table 18 shows the mean WM score for both groups and the mean RTs on the critical region (the verb) based on condition. The results of a one-way between-groups ANOVA confirmed that there was
a statistically significant difference in the two groups’ mean WM span scores, $F(1,140)=223.173$, $p < .001$.

Table 18. Mean WM Span Scores and Mean RTs (SDs) on the Verb in Tense and S-V Agreement Conditions for High and Low WM Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>WM</th>
<th>Condition</th>
<th>Pretest RTs</th>
<th>Posttest RTs</th>
<th>Delayed RTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low WM</td>
<td>11.55 (1.38)</td>
<td>Agreement Overall</td>
<td>830.36 (327.70)</td>
<td>593.19 (252.95)</td>
<td>521.51 (210.79)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterite</td>
<td>776.89 (338.89)</td>
<td>562.46 (238.50)</td>
<td>495.24 (204.24)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future</td>
<td>892.49 (392.79)</td>
<td>626.59 (314.33)</td>
<td>548.26 (239.28)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tense Violation Overall</td>
<td>836.03 (318.30)</td>
<td>620.82 (280.09)</td>
<td>513.47 (202.64)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterite</td>
<td>780.10 (326.67)</td>
<td>585.18 (253.59)</td>
<td>492.32 (185.52)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future</td>
<td>893.89 (369.89)</td>
<td>655.08 (334.10)</td>
<td>534.01 (253.23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-V Violation Overall</td>
<td>828.48 (301.96)</td>
<td>616.54 (263.27)</td>
<td>515.32 (214.43)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterite</td>
<td>783.72 (287.01)</td>
<td>584.53 (258.82)</td>
<td>490.57 (214.71)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future</td>
<td>872.83 (372.78)</td>
<td>646.19 (311.09)</td>
<td>541.20 (252.41)</td>
</tr>
<tr>
<td>High WM</td>
<td>15.48 (1.75)</td>
<td>Agreement Overall</td>
<td>920.80 (330.18)</td>
<td>709.02 (272.25)</td>
<td>629.26 (244.54)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterite</td>
<td>874.86 (339.47)</td>
<td>680.69 (276.02)</td>
<td>605.59 (233.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future</td>
<td>970.09 (361.57)</td>
<td>743.82 (308.97)</td>
<td>658.97 (291.14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tense Violation Overall</td>
<td>937.77 (309.78)</td>
<td>793.79 (322.62)</td>
<td>654.67 (284.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterite</td>
<td>871.11 (297.47)</td>
<td>762.14 (334.26)</td>
<td>642.98 (282.80)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future</td>
<td>1004.58 (378.09)</td>
<td>821.16 (367.82)</td>
<td>668.69 (323.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-V Violation Overall</td>
<td>921.39 (316.41)</td>
<td>767.73 (294.80)</td>
<td>690.09 (304.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Preterite</td>
<td>854.61 (325.49)</td>
<td>713.01 (252.01)</td>
<td>660.80 (301.15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Future</td>
<td>994.57 (364.12)</td>
<td>828.59 (372.79)</td>
<td>726.55 (383.48)</td>
</tr>
</tbody>
</table>

Interestingly, the mean RTs presented in Table 18 indicate that the high-WM group spent longer reading the verb, regardless of whether or not there was agreement, than the low-WM group at each test time. Results from one-way between-group ANOVAs performed on the mean RTs revealed that the high-WM group was in fact significantly slower than the low-WM group in all cases on the immediate and delayed posttests (all $p < .05$). Although between-group differences were non-significant on the pretest, the same trend of slower mean RTs for the high-WM group was evident and differences between the groups approached significance ($p < .10$) for all conditions except three (agreement overall, agreement future, and S-V agreement violation in the preterite tense). A side examination of RTs on the other three regions (word 4 prior to the verb, and words 6 and 7 following the verb) revealed the same pattern of slower reading on the part of high-WM learners, and one-way ANOVAs confirmed that between-group differences in
reading speed were significant ($p < .05$) or trending towards significance ($p < .1$) in almost all cases and conditions, irrespective of whether there was agreement, at the three non-critical regions on the pretest, posttest, and delayed posttest. That learners with higher WM span spent more time reading the stimuli than their low-WM counterparts is interesting and perhaps even initially counterintuitive given that higher WM capacity has been associated with greater storage and faster, more efficient processing (see Section 2.5 and the following reviews: Gathercole, 2009; Juffs & Harrington, 2011; Sagarra, in press). Nonetheless, this pattern was similar to that reported in Havik et al. (2009), in which L2 learners with higher WM also spent longer reading experimental stimuli involving subject-object ambiguities than learners with lower WM. In line with similar findings in two monolingual studies (MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995), Havik and colleagues conjectured that the slower reading on the part of learners with higher WM might be because they were able to maintain both interpretations of ambiguous sentences active in parallel while learners with low WM are unable to do this. This hypothesis does not, however, offer a convincing explanation for the findings in the present study given that the experimental items did not involve ambiguous sentences, but, rather, sentences with tense and S-V agreement and disagreement. Thus, it seems reasonable to posit that slower reading on the part of the high-WM group in the present study, and perhaps even in previous studies, may be related to a more general difference in online processing between high and low-WM learners. Further research is needed to shed light on this issue.

To investigate potential differences between high- and low-WM groups in their online processing of tense and S-V agreement violations, a separate repeated measures ANOVA was run on the critical verb per condition: tense overall, tense future, tense preterite, S-V overall, S-V future, and S-V preterite. In that I was interested in examining the effects of Agreement, WM,
Training (whether or not the participant underwent training), and namely the interactions between these factors, each repeated measures ANOVA included the within-subject factor Agreement (agreement/agreement violation) and the between-subject factors WM Group (high-WM learners/low-WM learners) and Training (yes training/no training). On the pretest, there were no significant effects of Agreement, WM, or Training, with lack of effects for Training expected given that the task was completed prior to the start of training. There were, however, main effects approaching significance ($p < .10$) for WM on the pretest for the tense condition overall and specifically for the tense condition in the familiar preterite tense. The lack of additional general effects for WM on the pretest are not surprising given the participants’ unfamiliarity with the task and the future tense. Consistent with previous analyses, on the immediate and delayed posttests WM was shown to have a significant main effect for all conditions ($all \ p < .05$), revealing again that high-span L2 learners processed the stimuli differently than the low-span learners, which could be driven mainly by the slower overall reading on the part of the high-span learners. Moreover, there were significant main effects for Agreement for every condition ($all \ p < .05$) on the immediate posttest. On the delayed posttest, there were significant main effects for Agreement in two conditions, specifically S-V agreement overall and S-V agreement in the future tense ($all \ p < .05$). There were, however, no significant effects of Training. The main effect findings for Agreement and WM are not particularly informative given that they do not shed new light on potential differences between high- and low-WM learners in their sensitivity to tense and S-V agreement violations, especially given that the overall slower processing on the part of the high-WM group confounds interpretation of these effects. Thus, of central interest here are any significant interactions between Agreement, WM, and Training.
As expected based on results from previous analyses, there were no significant interactions at the time of the pretest. At the time of the immediate posttest, there were significant interactions between Agreement and WM for the tense condition overall, $F(1,138) = 5.857, p = .017$, and the tense condition in the preterite, $F(1,138) = 3.981, p = .048$, indicating that the high-span learners were more sensitive than low-span learners to adverb-verb tense violations overall and specifically for the familiar preterite tense. On the delayed posttest, there were significant interactions between Agreement and WM for the S-V condition overall, $F(1,138) = 8.284, p = .005$, and the S-V condition in the future $F(1,138) = 5.711, p = .018$, but interactions between Agreement and WM also approached significance for the S-V condition in the preterite, as well as the tense condition overall and tense in the preterite (all $p < .10$). The fact that there were only significant interactions between Agreement and WM for the S-V conditions at the time of the delayed posttest suggests that it took longer for the L2 learners to exhibit sensitivity to S-V agreement violations as compared to tense violations, and then high WM again facilitated sensitivity to such violations.

With respect to the effects of exposure to training, analyses revealed only one significant interaction involving this variable. On the delayed posttest, there was a significant interaction between Agreement and Training for the tense condition in the preterite, $F(1,138) = 3.983, p = .048$. This indicates that exposure to training facilitated sensitivity to Adv-V tense violations in the preterite at the time of the delayed posttest. Overall, though, we can conclude that the training had very little effect on learners’ processing during online comprehension-based reading. This does not, however, eliminate the possibility that training could have a greater influence on this type of processing, as it did in the other assessments reported in Sections 3.2-3.4, with additional exposure. This question should be taken up in future research.
Although the significant interactions between Agreement and WM reported above show that high-span learners were more sensitive to agreement violations than low-span learners, the analyses conducted thus far do not reveal whether the low WM group exhibited any sensitivity to violations. To shed light on this issue, another series of repeated measures ANOVAs were run with split results for the high- and low-WM groups, each with Agreement as a within-subjects factor. Figures 29 – 32 show the mean RTs for both groups, with an asterisk denoting each case in which the mean group RT was significantly longer for a particular type of agreement violation as compared to when there was agreement, indicating sensitivity to grammatical violations.

Figure 29. Mean Reading Times on Critical Region for Low WM Group: Overall Tense and Subject-Verb Conditions

a = approaches significance at p = .078
**HIGH WM GROUP:**
Word 5 RTs for Conditions at each Test Time

- Agreement
- Tense Violation
- S-V Violation

Figure 30. *Mean Reading Times on Critical Region for High WM Group: Overall Tense and Subject-Verb Conditions*

* p < .05

**LOW WM GROUP:**
Word 5 RTs for Preterite & Future Conditions at each Test Time

- Agreement
- Tense Violation
- S-V Violation

Figure 31. *Mean Reading Times on Critical Region for Low WM Group: Preterite and Future Tense and Subject-Verb Conditions*
Figure 32. *Mean Reading Times on Critical Region for High WM Group: Preterite and Future Tense and Subject-Verb Conditions*

* p < .05; a = approaches significance at p = .061

As is evident by the lack of asterisks in Figures 29 and 31, results showed that there were in fact no significant effects of Agreement for the low-WM group in any condition on the pretest, posttest, or delayed posttest (although the main effect approached significance on the posttest for the overall tense condition, $F(1,83) = 3.177, p = .078$). That is, for the low-WM group there were no significant differences in mean RTs between sentences with tense/S-V agreement as compared to those with tense or S-V violations, which indicates that as a whole the low-WM group was insensitive to tense and S-V agreement violations. In contrast, results for the high-WM group revealed significant main effects for Agreement at the posttest and delayed posttest, as illustrated in Figures 30 and 32. Specifically, at the time of the immediate posttest mean RTs on the critical verb were significantly longer when there was an agreement violation for tense overall, $F(1,57) = 15.191, p < .001$, tense in the preterite, $F(1,57) = 7.673, p = .008$, tense in the
future, $F(1,57) = 5.181, p = .027$, S-V overall, $F(1,57) = 10.265, p = .002$, and S-V in the future, $F(1,57) = 6.976, p = .011$. At the time of the delayed posttest, main effects for Agreement were maintained for the S-V condition overall, $F(1,57) = 6.573, p = .013$, and the S-V condition in the future, $F(1,57) = 4.189, p = .045$, and there was a main effect approaching significance for Agreement in the S-V condition in the preterite, $F(1,57) = 3.648, p = .061$. These results indicate that the high-WM group showed sensitivity to both types of agreement violations, even if they only maintained sensitivity to S-V agreement violations at the delayed posttest. In line with the WM results reported in Section 3.4.3 for the cue reliance task, it appears that higher WM was particularly beneficial for facilitating and maintaining sensitivity to the non-meaning bearing person-number inflections and S-V grammatical dependency relations that are weak in L1 English. In short, the between-group results reported in this section illustrate that across the immediate and delayed posttests, the high-WM L2 learners exhibited sensitivity to both Adv-V tense violations and S-V agreement violations whereas the low-WM learners were entirely insensitive to these violations.

3.5.4 Summary of Results

The self-paced reading task was used to investigate whether early English-speaking L2 learners of Spanish would show sensitivity to Adv-V and S-V agreement violations, and whether the training or individual cognitive differences in WM would influence their processing of such agreement violations. In order to summarize the main findings, the specific research questions relevant to the self-paced reading task are restated and addressed briefly in turn below.

(a) Are beginning L2 English-Spanish learners sensitive to tense and subject-verb agreement violations during online processing for meaning?
As a whole, the learners were insensitive to agreement violations on the pretest, but then exhibited sensitivity to tense violations on the immediate posttest and to S-V agreement violations on the delayed posttest, in particular for the new future tense in both cases. Sensitivity to agreement violations in the future tense, but not the preterite tense, may be related to the newness of the future tense as well as the greater salience of its forms (they are all stressed) and the fact that its formation is the same for all verb classes, which is not the case for the preterite tense.

(b) Does training that promotes attention to morphological cues in the input influence learners’ sensitivity to verbal agreement violations?

The longitudinal form-focused training had no reliable effects on the L2 learners’ ability to process verbal agreement violations online.

(c) How does WM capacity relate to L2 learners’ ability to process agreement online?

Comparison of low-span to high-span learners confirmed that only learners with high WM, and not the low-WM group, were sensitive to tense and S-V agreement violations. This indicates that the high-WM learners are driving the overall finding reported in (a) indicating an apparent sensitivity on the part of the whole learner group to verbal agreement violations in certain conditions.
CHAPTER 4: Conclusions and Implications

Broadly this dissertation has explored how certain external (form-focused instructional strategies) and internal (working memory) factors modulate learners’ processing and learning of notoriously difficult L2 morphological cues. Now we can make use of the results to provide some answers to the general research questions outlined in Section 2.6.

(1) How do learners coming from a morphologically weak L1 process verbal morphology in a morphologically rich L2?

(2) What is the efficacy of form-focused training for overriding L1 cue biases and for promoting the processing and learning of L2 inflectional morphology, and what role does corrective feedback have on such?

(3) What is the role of working memory in processing and learning L2 inflectional morphology?

With respect to the first question, results from the cue reliance task indicated that, prior to training, the late English-Spanish learners exhibited overwhelming preference toward processing the lexical cue (adverb or explicit subject) over the morphological cue (verbal inflections). This finding was predicted based on previous research illustrating that native speakers of a morphologically weak L1 like English rely on meaningful lexical items over redundant inflections to derive meaning in a morphologically rich L2 such as Spanish, particularly at lower levels of proficiency (e.g., N. Ellis & Sagarra, 2010a, 2010b; LaBrozzi, 2009; Lee et al., 1997; Musumeci, 1989; Rossomondo, 2003; VanPatten, 1996, 2004, 2006). The additional support for this type of processing provided by the results of the cue reliance task is important given that most previous research either employed less direct and less informative offline measures to assess cue reliance (e.g., think-aloud protocols, written recalls in the L1, multiple-choice recognition tasks, form production tasks), or it involved a single laboratory-learning session of a
greatly simplified verbal paradigm in Latin, an unfamiliar L2 (but see N. Ellis & Sagarra, 2010b, classroom experiment; LaBrozzi, 2009). Thus, the pretest cue reliance assessment results are consistent with claims of both VanPatten’s Input Processing Model and N. Ellis’ Associative-Cognitive framework that difficulty with inflections stems from the tendency of many learners (particularly those whose L1 is morphologically weak) to ignore, at least initially, these grammatical markers in the input, consequently failing to make the appropriate form-meaning connections necessary for acquisition.

Moreover, results from the self-paced reading task demonstrated that the English-Spanish learners were completely insensitive to both tense violations and subject-verb agreement violations during online processing for meaning prior to the training. This directly corroborates the findings of N. Ellis and Sagarra (2010b, classroom experiment) and Sagarra (2007b) for adverb-verb tense agreement processing in beginning English-Spanish learners, and the finding is also consistent with other online studies of nominal and verbal agreement processing that report lack of sensitivity to violations, especially at low levels of proficiency (Keating, 2009; Sagarra & Herschensohn, 2010; Ojima et al., 2005; Wen et al., 2010). Thus, based on this previous research and the evidence that the learners relied predominately on lexical cues to resolve grammatical conflicts in the cue reliance task, it came as no surprise that the early learners would exhibit insensitivity to tense and subject-verb agreement violations during online processing for meaning prior to undergoing the form-focused training. These results are also in line with the Associative-Cognitive model, providing further support for long-term learned attention effects: first language experience indicating that lexical cues are simpler and more reliable in their interpretation than morphological cues blocks the acquisition of morphological cues in an L2 that uses these more extensively than lexical cues (N. Ellis, 2006b). Although the
findings of the present dissertation could be attributed to a general L2 reliance on lexical cues, the laboratory studies by N. Ellis and colleagues and the various studies indicating L1 influence on the processing of L2 inflectional morphology (e.g., Jiang et al., 2011; Frenck-Mestre et al., 2009; Tokowicz and MacWhinney, 2005) strongly suggest that cues learned in the L1 block the acquisition of later experienced L2 cues. However, future training studies comparing the processing and learning of inflections by learners of L2 Spanish whose L1 is morphologically weak (English) or rich (e.g., Italian) would provide stronger support for this argument as well as shed light on the extent to which the morphological richness of the L1 affects learning trajectories during and outcomes after form-focused intervention.

This dissertation is more innovative with respect to the second research question, which asks about the efficacy of form-focused training for overriding L1 cue biases and for promoting the processing and learning of L2 inflectional morphology, as well as the role of feedback on such. Results from the posttests indicated that extensive exposure in training to input activities that make morphological cues nonredundant was effective in promoting reliance on and accurate processing and production of verbal inflections. The cue reliance task data demonstrated that the training helps learners behave more like native Spanish speakers and increases their reliance on inflections, the more valid cue in morphologically rich, null-subject languages like Spanish, to resolve grammatical conflicts. Moreover, results from the aural-processing task illustrated the efficacy of the training for increasing accuracy in processing inflections in non-redundant contexts. The written production data revealed that the processing-based training can even benefit the written production of new verbal morphology, although it appears to have no effects for familiar inflections likely due to background knowledge that resulted in scores at ceiling on the production pretest. Taken together, these results suggest that the refocusing of attention to
morphological cues through training intervention can help real classroom learners of a complex language overcome learned attentional biases (based on L1 experience and L2 teachers’ overuse of lexical cues in the classroom) and facilitate the processing and learning of L2 inflections as assessed by form-focused tasks, even at beginning stages of L2 acquisition. These findings are in keeping with N. Ellis and Sagarra’s (2010b; study 3) one-session laboratory studies involving a small subset of Latin. Also, the results are in accordance with the general tenets of the Associative-Cognitive model, which holds that form-focused instruction that recruits explicit conscious processing can free learners from learned attentional biases that make certain L2 forms less learnable (N. Ellis, 2006b, 2008b).

However, results from the self-paced reading posttests revealed that the training had no effects on online processing of L2 morphology when the learners were focused on meaning as opposed to form, as was the case with the cue reliance, aural processing, and written production assessments. Thus, there is as of yet to my knowledge no direct evidence that form-focused instructional intervention involving extensive exposure to input activities that make morphological cues non-redundant and more salient (similar to VanPatten’s (1996, 2002, 2004) structured input) can facilitate sensitivity to L2 morphological cues during more implicit, real-time processing involved in reading for overall meaning. This could therefore appear to serve as support for the shallow structure hypothesis (Clahsen and Felser, 2006b) discussed in Section 2.2.2, which claims that due to deficiencies in the L2 grammar, L2 learners compute shallower, less detailed syntactic representations that prevent them from exhibiting nativelike processing of morphosyntactic features, such as detecting agreement errors. Since this theory posits underlying deficiencies in the competence of late L2 learners, it assumes that instructional intervention would not bring late learners to achieve nativelike processing. However, the lack of effects in the
The current study of the training on online morphological processing cannot be taken as support for the shallow structure hypothesis because the learners in this study were still at very early stages of L2 acquisition, and there is still a possibility that the training could have a greater influence on this type of processing with additional exposure and/or with learners at higher levels of proficiency. This question should be taken up in future research.

The latter part of the second research question addressed the role of feedback during training for overcoming learned attentional biases and promoting the processing and learning of L2 morphological cues. Although training had no effects, regardless of feedback condition, on online processing of agreement violations when focused on meaning, feedback did have effects on learners’ performance on the three form-focused assessments. In general, the findings indicated that the provision of feedback, particularly metalinguistic feedback, enhances the effectiveness of the training by facilitating learners’ ability to override cue biases and accurately process and produce verbal inflections. While metalinguistic feedback was shown to be an essential component of the training in order to promote the accurate processing of inflections in the aural processing task, the training alone, regardless of whether or not feedback is provided, is beneficial for overriding cue biases and the written production of verbal inflections. This suggests that extensive exposure to input activities that require the processing of the target grammatical form (i.e., task-essential practice), without any corrective feedback, is sufficient to have an affect on L2 development, at least with respect to verbal inflections. This lends support to the tentative conclusions drawn by Sanz (2004) and Sanz and Morgan-Short (2004), both of which suffered from the limitation of not including a control group that underwent treatment without any feedback.
 Nonetheless, given that training in conjunction with metalinguistic feedback consistently led to the most gains on all of the form-focused assessments as well as the highest accuracy on the training sessions themselves in the present dissertation, we can conclude that there are advantages for providing explicit metalinguistic evidence during exposure to input through task-essential practice, at least for the target structure (verbal inflections). These benefits found for metalinguistic feedback are in line with previous findings from various studies involving production-based treatments (e.g., Caroll & Swain, 1993; R. Ellis et al., 2006; Heift, 2010; Nagata & Swisher, 1995; Sauro, 2009). More importantly, the findings inform the very limited and conflicting research on feedback during processing. They corroborate the two studies that found advantages of processing input with explicit feedback involving a grammatical explanation (Rosa, 1999; Rosa and Leow, 2004), but they contradict the findings of the studies that found no such benefits (Sanz, 2004; Sanz & Morgan-Short, 2004). What could be the cause of this discrepancy? It could be due to differences in methodological approaches. Or, since the studies also focused on different grammatical forms, it may be that some target forms benefit more from metalinguistic feedback during processing than others. In any case, more research is needed. Lastly, the benefits found for explicit metalinguistic feedback in the present study lend support to a hypothesis present in the Associative-Cognitive model, which states that explicit feedback can help learners overcome learned attentional biases and promote L2 development by generating a “dialectic tension between the conflicting forces of the learner’s current stable states of interlanguage and the evidence of explicit form-focused feedback” (N. Ellis, 2008b, p. 373). However, since there were no effects of training (with or without feedback) on learners’ ability to detect agreement errors in the self-paced reading assessment, we do not have evidence here that explicit feedback can facilitate sensitivity to L2 verbal morphology during more implicit,
real-time processing when reading for meaning. Future research— including that involving other learner groups, target structures, and methodological designs— should further examine this issue in order to discern any impact that explicit feedback during instruction might have on subsequent online processing when attending to meaning, as is the case in natural language comprehension.

Finally, the third research question addressed the role of internal cognitive abilities, namely working memory, on the processing and learning of inflectional morphology. The results demonstrated significant effects of working memory in both of these domains. Participants with higher working memory were more accurate overall in aural processing of verbal inflections, and they were also able to learn more from the training and improve their accuracy in the processing and written production of inflections over time. However, with respect to written production, higher working memory capacity only benefitted learners’ ability to learn to produce the new future tense inflections, but not the familiar present and preterite tense inflections. This is consistent with previous research that has found a positive relationship between working memory and grammar learning because studies have primarily examined the learning of completely novel structures (e.g., Brooks et al., 2006; Kempe & Brooks, 2008; Kormos & Sáfar, 2008; Martin & N. Ellis, 2012). As noted in Chapter 3, the lack of effects for working memory (as well as training) on learners’ ability to produce the familiar tenses may have been due to variation in previous knowledge, which resulted in ceiling effects on the pretest. Thus, the present findings contribute to the still limited but growing body of evidence for a relationship between working memory and grammar learning, particularly when learning processes involve entirely new grammatical structures and fall toward the explicit end of the spectrum, presumably drawing on the executive attention component of working memory (Roehr, 2008; Williams, 2012).
Moreover, results from the cue reliance task indicated that working memory capacity was positively related to learners’ ability to overcome cue biases and rely more on verbal inflections for subject reference, but not for temporal reference. As discussed in Chapter 3, this finding suggests that higher working memory capacity might be particularly beneficial for facilitating processing of the non-meaning bearing person-number inflections that involve a purely grammatical dependency with the lexical cue (the subject) and are almost entirely absent in L1 English. On the other hand, because tense inflections are a semantic marker that bear meaning (situating an event in time) and L1 English learners have some experience relying on them (-ed inflection marking regular past tense), greater cognitive resources might not enhance the already robust effects of training for facilitating reliance on these cues. Nevertheless, the present study does suggest a role for working memory in learners’ ability to alter their cue preferences and come to rely on more valid L2 cues, a finding that lends support to the only two studies to my knowledge that examine this issue (LaBrozzi, 2009; Miyake & Friedman, 1998).

With respect to the role of working memory on online (morpho)syntactic processing, an area of contention in the literature, self-paced reading task results indicate that learners with higher working memory capacity can better process verbal agreement. In fact, comparison of low-span to high-span learners indicated that only learners with high working memory were sensitive to tense and subject-verb agreement violations. This finding is consistent with the few previous studies that have directly examined the relationship between working memory and L2 processing of verbal and nominal agreement in beginning and intermediate learners (LaBrozzi, 2009; Sagarra, 2008; Sagarra & Herschensohn, 2010). The present findings also inform the more general debate in the literature as to whether working memory capacity constrains online L2 sentence processing, suggesting that it can. Thus, although this dissertation did not set out to
determine whether nativelike morphological processing is attainable (highest-proficient learners would be necessary to address this), the results provide tentative support for perspectives that maintain that L2 learners’ inability to exhibit nativelike processing of morphosyntactic features (detecting agreement errors) are due to cognitive limitations in L2 processing (Hasegawa, Carpenter, & Just, 2002; Hopp, 2007, 2010; McDonald, 2006) rather than to underlying deficits in the L2 grammar (Clahsen & Felser, 2006). The fact that the L2 learners with high working memory, but not those with low working memory, showed sensitivity to agreement errors in the present study underscores the importance of cognitive resources in L2 morphological processing and suggests that with sufficient working memory, nativelike processing may in fact be attainable. Although more research specifically designed to address this question (including a native-speaking control group) would be necessary to make strong claims for this position, the findings here are suggestive. The results of the present study can, however, be taken as empirical support for assertions of both Van Patten’s Input Processing Model and N. Ellis’ Associative-Cognitive framework that L2 processing of redundant grammatical forms is constrained by working memory limitations. The benefits found for high working memory capacity accord with the model’s prediction that, for L2 learners to process redundant grammatical forms, the processing of overall sentential meaning must not tax their working memory and use up all of their available cognitive resources (2004, 2006). Nevertheless, the notable methodological differences across studies examining working memory in online processing (e.g., Dussias & Piñar, 2010; Felser & Roberts, 200; Havik et al., 2009; Juffs, 2004, 2005) – including differences in working memory tasks and scoring methods as well as in processing tasks and target structure – indicates the need for more replicable and comparable research on this topic. Also, pace the methodological prescription of several recent studies concerned with the reliability of working
memory measures (e.g., Caplan et al., 2007; Conway et al., 2005; McCabe et al., 2010), it would be valuable to employ multiple measures in studies in order to compare the predictive power and reliability of the measures, and then perhaps use a composite score. This is something I will do in future research using the data from the other two working memory measures (reading span and operation span) not analyzed for this dissertation.

In addition to the theoretical importance of the findings summarized above, this research clearly has important applied implications. In general, this research suggests that by exposing learners to input through controlled form-focused training intervention, it is possible to help them alter ineffective processing strategies from L1 experience and adopt L2 processing strategies, even at early stages of acquisition. In a similar vein, this research indicates that implementing a training paradigm (similar to that used in this dissertation) with metalinguistic feedback is a useful instructional method that can facilitate the learning of verbal morphology, a grammatical phenomenon that poses persistent difficulties for L2 learners, particularly those who come from a morphologically weak L1. Nevertheless, many open questions still remain. For example, how much exposure do learners need? Future work investigating learning trajectories across the training sessions, based on accuracy and reaction time data collected, will shed some light on this. Moreover, can this sort of form-focused training intervention influence online processing? This was not shown to be the case in this study, but as discussed above, future work involving more exposure to training and/or more proficient learners might reveal beneficial effects for such intervention. Likewise, does this sort of intervention offer benefits for other types of grammatical structures? Future research should carry out similar sorts of training studies with other phenomena such as the processing and learning of object-first clitic structures in Spanish, another structure known to cause difficulties for learners.
Practical considerations also need to be made in light of the clear effects found for working memory on the processing and learning of L2 morphological cues. This study provides evidence for the claim that the construct of working memory is a crucial component of language learning aptitude (e.g., Dörnyei, 2005; Williams, 2012), at least for the specific aspects of L2 learning and use examined in this dissertation. Some might take these findings to be a dour verdict for L2 learners with low working memory. But complete resignation does not follow. First of all, there is not near enough research to suggest that a certain level of working memory capacity is a necessary condition for L2 learning success. All that is being claimed is that working memory has an effect. Next, it is generally good pedagogy to conform instruction to learners’ abilities. And the case of working memory is no different. For example, instructors can help learners with lower working memory by using supplemental written and visual aids, by not asking for too much memorization, by increasing repetition, and by assisting learners with developing memory heuristics (Gathercole & Alloway, 2008; Hummel & French, 2010). And given the findings of this dissertation, maybe extra form-focused training would help. All of these issues should be examined further. More optimistically, there is now some promising evidence that working memory can be improved through its own style of training (Holmes et al., 2009; Klingberg, 2010). Future research should pursue this line vigorously because if working memory can in fact be improved through some type of cognitive training, then we might have recourse to help those individuals for whom low working memory capacity restrains their language learning abilities.
REFERENCES


Doughty, C. (2004). Commentary: When PI is Focus on Form it is Very, Very good, but When it is Focus on Forms... In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 257-270). Mahwah, NJ: Lawrence Erlbaum.


The Lexicon-Syntax Interface in Second Language Acquisition (pp. 21-44). Amsterdam: Benjamins.


APPENDIX A: Language Background Questionnaire

Please provide the following information. Write N/A if a question does not apply to you. All information you provide will remain confidential.

Full Name: ___________________________ Instructor: ___________________________

1. Year at PSU: Freshman Sophomore Junior Senior
2. Major: ___________________________ Minor (if applicable): ___________________________
3. Age: __________
4. Year of high school graduation: __________
5. What is your native language? (the first language you were exposed to and learned at home): __________________________________________________________

6. What language(s) do you speak with your parents? __________________________________________________________

7. What language(s) do you speak with your grandparents and other family members? __________________________________________________________

8. Have you ever lived outside of the United States? YES NO
   If YES, please provide details regarding where you lived, how old you were, and how long you lived there:
   ________________________________________________________________________________

9. Out of all of the languages that you know/speak (including English and your native language if it is not English), what do you consider your most dominant language now (i.e. the language you are most comfortable speaking, understanding, reading and writing)?
   ________________________________________________________________________________

10. Besides Spanish, English and your native language (if not English), have you ever studied or been extensively exposed to any other languages? YES NO
    If YES, what other language(s) have you studied/been exposed to? __________________________
Please provide details explaining when and where you studied or were exposed to the language(s) and for how many years:

________________________________________________________________________

**Experience with Spanish:**

11. Did you take Spanish 1 at Penn State University Park Campus? YES NO
12. Have you taken any Spanish courses at another PSU campus or a different college/university? YES NO
   If YES, please indicate when and where you took another college-level Spanish course:
   ______________________________________________________________________

13. Did you take Spanish in high school or before? YES NO
   If YES, please indicate how many years of Spanish you took *before* college:
   - Years of Spanish in high school? ________
   - Years of Spanish in junior high / middle school? ________
   - Years of Spanish in elementary school? ________

14. Is Spanish spoken by any of your family members or friends? YES NO
   If YES, please explain how often you had/have opportunities to speak or listen to Spanish with your family and/or friends:
   ______________________________________________________________________

15. Have you ever spent more than a week in a Spanish-speaking country? YES NO
   If YES, please provide details regarding where you went, the purpose of your stay there, how long you were there, and how old you were:
   ______________________________________________________________________

16. Based on your experience, what do you consider to be the most challenging aspect of learning Spanish for you? (Please be as specific as you can):
   ______________________________________________________________________

¡Muchas Gracias!
### APPENDIX B: Vocabulary Test

1. **Match the Spanish verb with the English meaning.**

<table>
<thead>
<tr>
<th>Spanish Verb</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>limpiar</td>
<td>A. to sing</td>
</tr>
<tr>
<td>discutir</td>
<td>B. to look/watch</td>
</tr>
<tr>
<td>beber</td>
<td>C. to drink</td>
</tr>
<tr>
<td>mirar</td>
<td>D. to argue/debate</td>
</tr>
<tr>
<td>vender</td>
<td>E. to clean</td>
</tr>
<tr>
<td>cantar</td>
<td>F. to sell</td>
</tr>
</tbody>
</table>

2. **Match the Spanish verb with the English meaning.**

<table>
<thead>
<tr>
<th>Spanish Verb</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>viajar</td>
<td>A. to take/to wear</td>
</tr>
<tr>
<td>llevar</td>
<td>B. to buy</td>
</tr>
<tr>
<td>correr</td>
<td>C. to live</td>
</tr>
<tr>
<td>buscar</td>
<td>D. to run</td>
</tr>
<tr>
<td>vivir</td>
<td>E. to travel</td>
</tr>
<tr>
<td>comprar</td>
<td>F. to look for</td>
</tr>
</tbody>
</table>

3. **Match the Spanish verb with the English meaning.**

<table>
<thead>
<tr>
<th>Spanish Verb</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>trabajar</td>
<td>A. to receive</td>
</tr>
<tr>
<td>recibir</td>
<td>B. to swim</td>
</tr>
<tr>
<td>abrir</td>
<td>C. to work</td>
</tr>
<tr>
<td>nadar</td>
<td>D. to practice</td>
</tr>
<tr>
<td>practicar</td>
<td>E. to study</td>
</tr>
<tr>
<td>estudiar</td>
<td>F. to open</td>
</tr>
</tbody>
</table>

4. **Match the Spanish verb with the English meaning.**

<table>
<thead>
<tr>
<th>Spanish Verb</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>descubrir</td>
<td>A. to paint</td>
</tr>
<tr>
<td>bailar</td>
<td>B. to cook</td>
</tr>
<tr>
<td>cocinar</td>
<td>C. to write</td>
</tr>
<tr>
<td>escribir</td>
<td>D. to learn</td>
</tr>
<tr>
<td>aprender</td>
<td>E. to dance</td>
</tr>
<tr>
<td>pintar</td>
<td>F. to discover</td>
</tr>
</tbody>
</table>

5. **Match the Spanish time expression with the English meaning.**

<table>
<thead>
<tr>
<th>Spanish Time Expression</th>
<th>English Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>el próximo año</td>
<td>A. next year</td>
</tr>
<tr>
<td>el mes pasado</td>
<td>B. in this instant</td>
</tr>
<tr>
<td>la próxima semana</td>
<td>C. last week</td>
</tr>
<tr>
<td>el año pasado</td>
<td>D. last year</td>
</tr>
<tr>
<td>el próximo mes</td>
<td>E. last month</td>
</tr>
<tr>
<td>en este momento</td>
<td>F. next week</td>
</tr>
<tr>
<td>la semana pasada</td>
<td>G. next month</td>
</tr>
<tr>
<td>en este instante</td>
<td>H. in this moment</td>
</tr>
</tbody>
</table>
APPENDIX C: Grammar Proficiency Test

This test is adapted from proficiency tests offered by:
• Transparent Language (http://www.transparent.com/learn-spanish/) and
• Spanish Steps (http://www.spanishsteps.co.uk)

Choose the most appropriate response.

1. Tomás dijo que _______ a la tienda pero no tiene tiempo.
   A) ir  
   B) va  
   C) iría  
   D) iba

2. Era probable que él lo _______.
   A) tenga  
   B) tuviera  
   C) tuvo  
   D) tenía

3. ¿ _______ museos hay en Nueva York?
   A) Cuántas  
   B) Cuántos  
   C) Muchos  
   D) Cuantas

4. Yo _______ el hermano de Pepe.
   A) soy  
   B) eres  
   C) es  
   D) estoy

5. Rosa y Miguel van _______ cine.
   A) de la  
   B) del  
   C) al  
   D) a la

6. _______ levanto a las siete.
   A) Yo  
   B) Mi  
   C) Conmigo
7. A _______ no le gusta levantarse temprano.
   A) la
   B) le
   C) ello
   D) ella

8. ¿Dónde vivían los aztecas a _______ venció Cortés?
   A) que
   B) los cuales
   C) quienes
   D) quien

9. Las niñas _______ jugando en la calle.
   A) está
   B) están
   C) somos
   D) son

10. ¡_______ Uds.!
    A) Levántense
    B) Levántanse
    C) Se levanten
    D) Levántese

11. La clase de español empieza a las _______.
    A) once y media
    B) mediodías
    C) una y cuarto
    D) una en punto

    A) era
    B) había
    C) estaba
    D) hacía

13. El domingo voy a comer con _______ abuelos.
    A) mis
    B) suyos
    C) su
    D) míos

    A) vi a
B) vieron  
C) veía  
D) viste  

15. Pepe toca ______ guitarra.  
   A) el  
   B) los  
   C) la  
   D) lo  

16. María siempre lava los platos y su esposo ______ seca.  
   A) les  
   B) las  
   C) los  
   D) lo  

17. En el museo ______ cuadros de Goya.  
   A) son los  
   B) está  
   C) hay muchos  
   D) muchos  

18. Hace un año que trabajo en ______ escuela.  
   A) esta  
   B) esto  
   C) este  
   D) ese  

19. Rosa me ______ ayer.  
   A) visité  
   B) visitasteis  
   C) visitó  
   D) visitaste  

20. ¿Qué ______ Uds.?  
   A) hago  
   B) hacéis  
   C) haces  
   D) hacen  

21. Todas mis amigas ______ a la fiesta por Marcos.  
   A) están invitado  
   B) han sido invitadas  
   C) son invitadas  
   D) han sido invitado
22. La película es _______ interesante.
   A) mucha
   B) muy
   C) mucho
   D) poco

23. Juan quiere que _______ temprano.
   A) llegaron
   B) lleguen
   C) llegan
   D) llegar

24. Mucha gente cree que dentro de veinte años todos ________ coches eléctricos.
   A) usarán
   B) están usando
   C) usaron
   D) usan
APPENDIX D: Written Production Experimental Stimuli

Version A:

Present Tense:

1. Todos los días yo __________(hablar) con mi mama por teléfono.
2. Normalmente Marta __________(compartir) su ropa con sus amigas.
3. Por las tardes los estudiantes __________(trabajar) en sus proyectos.
4. Generalmente mi amigo y yo __________(vender) las bebidas durante el partido.

Past (Preterite) Tense:

1. Ayer yo __________(discutir) el tema en el debate.
2. La semana pasada el chico __________(mirar) una película en el cine.
3. El año pasado nosotros __________(viajar) en avión a la isla.
4. El mes pasado Juan y Carlos __________(beber) unas cervezas en el bar.

Future Tense:

1. Mañana Alberto y Mario __________(limpiar) la casa para la fiesta.
2. El próximo mes la madre __________(comprar) un coche nuevo para su hijo.
3. La próxima semana yo __________(comer) pizza en mi cumpleaños.
4. El próximo año mi hermano y yo __________(vivir) con nuestros primos en la ciudad.

Version B:

Present Tense:

1. Todos los días yo __________(cocinar) la cena para mi familia.
2. Normalmente Elena y Rosa __________(escribir) una carta a su abuela.
3. Por las tardes el niño __________(visitar) los perritos en la tienda.
4. Generalmente nosotros __________(vivir) con nuestros amigos.

Past (Preterite) Tense:

1. Ayer ellos __________(comer) unas hamburguesas en el restaurante.
2. La semana pasada yo __________(hablar) con la profesora en su oficina.
3. El año pasado él __________(estudiar) la literatura española en la universidad.
4. El mes pasado mi hermana y yo __________(correr) en el parque.

Future Tense:

1. Mañana el hombre __________(discutir) el problema con su esposa.
2. La próxima semana nosotros __________(trabajar) en la oficina de mi papá.
3. El próximo año Juana y Javier __________(pintar) su casa con su familia.
4. El próximo mes yo __________(descubrir) la importancia de la tradición.
**Version C:**

Present Tense

1. Todos los días Fernando y yo _________ (escuchar) el programa en la radio.
2. Normalmente los dueños _________ (abrir) sus tiendas muy temprano.
3. Por las tardes el artista _________ (pintar) en el estudio de la universidad.
4. Generalmente yo _________ (comer) en la cafetería de la universidad.

Past (Preterite) Tense:

1. Ayer él _________ (vender) las revistas a sus vecinos.
2. La semana pasada yo _________ (estudiar) para el examen final.
3. El año pasado los empleados _________ (trabajar) en el edificio.
4. El mes pasado Gabriela y yo _________ (aprender) canciones en español.

Future Tense:

1. Mañana yo _________ (practicar) la flauta con mi amiga.
2. La próxima semana el nieto _________ (visitar) a su abuela en Florida.
3. El próximo mes Isabel y Esteban _________ (correr) juntos en la playa.
4. El próximo año nosotros _________ (recibir) el tratamiento en el hospital.
APPENDIX E: Letter-Number Sequencing Stimuli

1. 2-J
2. C-6
3. F-8
4. 4-G-1
5. J-7-O
6. 9-S-4
7. 5-W-3-L
8. A-8-C-1
9. 2-Y-8-E
10. P-4-H-6-K
11. 9-B-7-H-2
12. 3-I-1-Q-M
13. 4-N-9-B-5-R
14. T-8-V-6-C-1
15. 3-Y-2-D-7-K
16. F-8-L-2-O-5-V
17. 6-N-7-H-3-D-4
18. 9-R-4-Q-1-M-8
20. W-1-H-9-P-7-Q-3
21. 5-X-9-N-3-R-6-C
APPENDIX F: Training Stimuli

Below is a sample of the sentences used in each of the training sessions.

**Training 1** (*3SG form of the present, preterite, and future tenses*):

Discute la política con su padre. 
Después de la escuela practica deportes.

Cocinó desayuno para sus amigos. 
Durante la reunión compartió sus ideas.

Para el concierto aprenderá las canciones. 
Buscará un libro en la librería.

**Training 2** (*1SG and 3SG forms of the present, preterite, and future tenses*):

Como una ensalada con tomates. 
Durante el concierto escucho al cantante.

En la oficina visité a la profesora. 
Descubrió el virus de la computadora.

En la playa abriré un restaurante. 
Mirará los delfines en el acuario.

**Training 3** (*1SG and 3SG forms of the present, preterite, and future tenses*):

Para la prueba aprendo las fechas. 
Pinta las caras de las niñas.

Para el viaje llevé dos maletas. 
Estudió la cultura de los antepasados.

Compraré un condominio en la costa. 
En el capítulo estudiará las costumbres.

**Training 4** (*1SG, 1PL, 3SG, 3PL forms of the present, preterite, and future tenses*):

Busco una mochila en la tienda. 
Antes del almuerzo limpia el comedor.
Aprendemos las tradiciones de la gente. 
Con los niños visitan el zoológico.
Desde el avión miré las montañas.
Compró sillas para la peluquería.
Durante la vacación bebimos mojitos.
Escucharon la fiesta de los estudiantes.

Estudiaré la guerra entre los países.
Antes del viaje aprenderá el idioma.
Compartiremos el cuento con la niña.
Para el proyecto escribirán la propuesta.

**Training 5 (1sg, 1pl, 3sg, 3pl forms of the present, preterite, and future tenses):**

Escribo la introducción del artículo.
Para el carnaval compra una máscara.
Bebemos gaseosas en el cine.
Miran a los tenistas en el campeonato.

En la caminata llevé una mochila.
Abrió una carta de la universidad.
Para el viaje compraron unas sandalias.
Vendimos pinturas en la galería.

Recibiré los boletos por correo.
Con los turistas visitará los monumentos.
Escribiremos una petición contra la decisión.
En el seminario discutirán la teoría.
APPENDIX G: Aural Processing Experimental Stimuli

Version A:

1. Durante la clase trabajé en el proyecto.
2. Durante la cena compartió las noticias.
3. Después de la escuela corrimos con el equipo.
4. En el parque buscaron al perro.
5. Con el jefe discuto el sueldo.
6. En la televisión mira un programa.
7. En la exposición vendemos la colección.
8. Durante el viaje visitan la playa.
9. En la playa llevaré una gorra.
10. Para el aniversario compraré un anillo.
11. En Italia viviremos con el abuelo.
12. Durante el almuerzo comerán un sándwich.

Version B:

1. Durante la conversación discutí las posibilidades.
2. Para la fiesta cocinó la cena.
3. En la cafetería comimos papas fritas.
4. En la clase escucharon las reglas.
5. En el instituto estudio música.
6. En el centro pinta las tiendas.
7. Para el examen aprendemos las palabras.
8. Después de la película escriben un resumen.
9. En la celebración beberé champán.
10. En la carta descubrirá el secreto.
11. En el hospital visitaremos al abuelo.
12. Después de la escuela practicarán fútbol.

Version C:

1. Para la fiesta limpié la terraza.
2. Antes de la escuela practicó la natación.
3. En la feria vendimos caramelos.
4. Después de la fiesta abrieron las tarjetas.
5. En la cafetería cocino el almuerzo.
6. En el trabajo recibe un certificado.
7. Durante el desayuno comemos huevos.
8. Con el traje llevan una corbata.
9. Para el supermercado escribiré una lista.
10. En el mercado buscará frutas.
11. En los vasos pintaremos diseños.
12. En la clase aprenderán geografía.
APPENDIX H: Cue Reliance Experimental Item Sets

Below is the complete stimuli for one of the three versions of this task created. As explained in Section 3.1.3, although the sentences were unique in the other two versions, the target verbs, conditions, and overall procedure were identical. The three conditions are numbered as follows:

1 = Agreement
2 = Adverb-verb tense agreement violation
3 = Subject-verb agreement violation

**Version A:**

1. La semana pasada yo miré el documental.
2. En este momento yo miré el documental.
3. La semana pasada nosotros miré el documental.

1. En este momento yo miro la procesión.
2. La semana pasada yo miro la procesión.
3. En este momento nosotros miro la procesión.

1. La próxima semana yo miraré el maratón.
2. En este instante yo miraré el maratón.
3. La próxima semana nosotros miraré el maratón.

1. Ella limpió los apartamentos la semana pasada.
2. Ella limpió los apartamentos en este momento.
3. Yo limpió los apartamentos la semana pasada.

1. Ella limpia la cocina en este momento.
2. Él limpia la cocina la semana pasada.
3. Yo limpia la cocina en este momento.

1. Él limpiará el garaje la próxima semana.
2. Él limpiará el garaje en este instante.
3. Yo limpiará el garaje la próxima semana.

1. La semana pasada nosotros aprendimos las respuestas.
2. En este instante nosotros aprendimos las respuestas.
3. La semana pasada ellos aprendimos las respuestas.

1. En este momento nosotros aprendemos la información.
2. El mes pasado nosotros aprendemos la información.
3. En este momento ellas aprendemos la información.

1. La próxima semana nosotros aprenderemos los verbos.
2. En este momento nosotros aprenderemos los verbos.
3. La próxima semana ellos aprenderemos los verbos.
1. Ellos practicaron el fútbol el mes pasado.
2. Ellos practicaron el fútbol la próxima semana.
3. Nosotras practicaron el fútbol el mes pasado.

1. Ellos practican el instrumento en este instante.
2. Ellos practican el instrumento el próximo mes.
3. Nosotros practicarán el instrumento en este instante.

1. Ellas practicarán las danzas el próximo mes.
2. Ellos practicarán las danzas el mes pasado.
3. Nosotros practicarán las danzas el próximo mes.

1. Yo vendí la bicicleta el mes pasado.
2. Yo vendí la bicicleta la próxima semana.
3. Él vendió la bicicleta el mes pasado.

1. Yo vendo las verduras en este instante.
2. Yo vendo las verduras el próximo mes.
3. Él vendo las verduras en este instante.

1. Yo venderé la tienda el próximo mes.
2. Yo venderé la tienda el mes pasado.
3. Él venderé la tienda el próximo mes.

1. El mes pasado ellos escribieron la composición.
2. El próximo mes ellos escribieron la composición.
3. El mes pasado él escribieron la composición.

1. En este instante ellos escriben el informe.
2. La próxima semana ellas escriben el informe.
3. En este instante ella escriben el informe.

1. El próximo mes ellas escribirán la carta.
2. La semana pasada ellos escribirán la carta.
3. El próximo mes ella escribirán la carta.

1. En este momento yo descubrí el código.
2. La semana pasada nosotras descubrí el código.
3. La semana pasada yo descubrí el código.

1. La semana pasada yo descubro las ruinas.
2. En este momento nosotras descubro las ruinas.
3. En este momento yo descubro las ruinas.
1. En este instante yo descubriré la solución.
2. La próxima semana nosotros descubriré la solución.
3. La próxima semana yo descubriré la solución.

1. Ella estudió la lección la semana pasada.
2. Él estudió la lección en este momento.
3. Yo estudió la lección la semana pasada.

1. Él estudia los apuntes en este momento.
2. Él estudia los apuntes la semana pasada.
3. Yo estudia los apuntes en este momento.

1. Ella estudiará las fórmulas la próxima semana.
2. Él estudiará las fórmulas en este instante.
3. Yo estudiaré las fórmulas la próxima semana.

1. Nosotras bebimos las margaritas la semana pasada.
2. Nosotras bebimos las margaritas en este instante.
3. Ellos bebimos las margaritas la semana pasada.

1. En este momento nosotras bebemos el té.
2. El mes pasado nosotros bebemos el té.
3. En este momento ellos bebemos el té.

1. La próxima semana nosotros beberemos el chocolate.
2. En este momento nosotros beberemos el chocolate.
3. La próxima semana ellas beberemos el chocolate.

1. El mes pasado ellos escucharon los himnos.
2. La próxima semana ellos escucharon los himnos.
3. El mes pasado nosotros escucharon los himnos.

1. Ellos escuchan la conversación en este instante.
2. Ellas escuchan la conversación el próximo mes.
3. Nosotros escuchan la conversación en este instante.

1. Ellos escucharán la historia el próximo mes.
2. Ellos escucharán la historia el mes pasado.
3. Nosotros escucharán la historia el próximo mes.

1. El mes pasado él pintó el baño.
2. La próxima semana él pintó el baño.
3. El mes pasado ellos pintó el baño.

1. En este instante ellas pinta las uñas.
2. El próximo mes ella pinta las uñas.
3. En este instante ella pinta las uñas.

1. El próximo mes él pintará la pared.
2. El mes pasado él pintará la pared.
3. El próximo mes ellos pintarán la pared.

1. Yo compré los regalos la semana pasada.
2. Yo compré los regalos en este momento.
3. Él compré los regalos la semana pasada.

1. Yo compré la revista en este momento.
2. Yo compré la revista la semana pasada.
3. Él compré la revista en este momento.

1. Yo compraré los libros la próxima semana.
2. Yo compraré los libros en este instante.
3. Él compraré los libros la próxima semana.

1. La semana pasada ella recibió el cheque.
2. En este momento él recibió el cheque.
3. La semana pasada ellos recibió el cheque.

1. En este momento ella recibe la tarjeta.
2. La semana pasada ella recibe la tarjeta.
3. En este momento ellos reciben la tarjeta.

1. La próxima semana él recibirá los documentos.
2. En este instante él recibirá los documentos.
3. La próxima semana ellos recibirá los documentos.

1. Nosotras comimos los chocolates la semana pasada.
2. Nosotros comimos los chocolates en este instante.
3. Yo comimos los chocolates la semana pasada.

1. Nosotras comemos los sándwiches en este momento.
2. Nosotros comemos los sándwiches el mes pasado.
3. Yo comemos los sándwiches en este momento.

1. Nosotros comeremos el pavo la próxima semana.
2. Nosotros comeremos el pavo en este momento.
3. Yo comeremos el pavo la próxima semana.

1. El mes pasado ellas cocinaron el pan.
2. La próxima semana ellas cocinaron el pan.
3. El mes pasado nosotras cocinaron el pan.
1. En este instante ellos cocinan el jamón.
2. El próximo mes ellos cocinan el jamón.
3. En este instante nosotros cocinan el jamón.

1. El próximo mes ellos cocinarán la paella.
2. El mes pasado ellos cocinarán la paella.
3. El próximo mes nosotras cocinarán la paella.

1. El mes pasado yo visité las escuelas.
2. La próxima semana yo visité las escuelas.
3. El mes pasado nosotros visité las escuelas.

1. Yo visito la oficina en este instante.
2. Yo visito la oficina el próximo mes.
3. Nosotros visito la oficina en este instante.

1. El próximo mes yo visitaré el pueblo.
2. El mes pasado yo visitaré el pueblo.
3. El próximo mes nosotros visitaré el pueblo.

1. Él abrió el café el mes pasado.
2. Él abrirá el café el próximo mes.
3. Ellos abrió el café el mes pasado.

1. En este instante ella abre las cortinas.
2. La próxima semana él abre las cortinas.
3. En este instante ellos abre las cortinas.

1. Él abrirá el vino el próximo mes.
2. Ella abrirá el vino el mes pasado.
3. Ellos abrirá el vino el próximo mes.

1. Ellos compartieron las experiencias el mes pasado.
2. Ellas compartieron las experiencias el próximo mes.
3. Él compartieron las experiencias el mes pasado.

1. Ellos comparten las ideas en este instante.
2. Ellos comparten las ideas la próxima semana.
3. Él comparten las ideas en este instante.

1. Ellas compartirán el trabajo el próximo mes.
2. Ellos compartirán el trabajo la semana pasada.
3. Ella compartirán el trabajo el próximo mes.
APPENDIX I: Self-paced Reading Experimental Stimuli

Below is the complete experimental stimuli for one of the three versions of this task created. As explained in Section 3.1.3, although the sentences were unique in the other two versions, the target verbs, conditions, and overall procedure were identical. The three conditions are numbered as follows:

1 = Agreement
2 = Adverb-verb tense agreement violation
3 = Subject-verb agreement violation

Version A:

1. La semana pasada ella abrió el paquete en la sala.
2. La próxima semana ella abrió el paquete en la sala.
3. La semana pasada ellas abrió el paquete en la sala.

1. El próximo año él abrirá el restaurante en la plaza.
2. El año pasado él abrirá el restaurante en la plaza.
3. El próximo año ellos abrirá el restaurante en la plaza.

1. El mes pasado nosotras aprendimos el baile en el estudio.
2. El próximo mes nosotras aprendimos el baile en el estudio.
3. El mes pasado ellas aprendimos el baile en el estudio.

1. El próximo mes nosotros aprenderemos la fórmula en el laboratorio.
2. El mes pasado nosotros aprenderemos la fórmula en el laboratorio.
3. El próximo mes ellos aprenderemos la fórmula en el laboratorio.

1. El mes pasado nosotros bebimos la cerveza en el bar.
2. El próximo mes nosotros bebimos la cerveza en el bar.
3. El mes pasado ellos bebimos la cerveza en el bar.

1. El próximo mes nosotros beberemos el vino en el restaurante.
2. El mes pasado nosotros beberemos el vino en el restaurante.
3. El próximo mes ellos beberemos el vino en el restaurante.

1. El mes pasado ellas cocinaron el pollo en el horno.
2. El próximo mes ellas cocinaron el pollo en el horno.
3. El mes pasado nosotras cocinaron el pollo en el horno.

1. El próximo mes ellos cocinarán la pizza en el horno.
2. El mes pasado ellos cocinarán la pizza en el horno.
3. El próximo mes nosotros cocinarán la pizza en el horno.

1. El mes pasado nosotros comimos el pescado en el restaurante.
2. El próximo mes nosotros comimos el pescado en el restaurante.
3. El mes pasado yo comimos el pescado en el restaurante.

1. El próximo mes nosotros comemos el desayuno en el palacio.
2. El mes pasado nosotros comemos el desayuno en el palacio.
3. El próximo mes yo comemos el desayuno en el palacio.

1. La semana pasada ellas compartieron la responsabilidad en la oficina.
2. La próxima semana ellas compartieron la responsabilidad en la oficina.
3. La semana pasada ella compartieron la responsabilidad en la oficina.

1. El próximo año ellos compartirán la habitación en la residencia.
2. El año pasado ellos compartirán la habitación en la residencia.
3. El próximo año él compartirán la habitación en la residencia.

1. La semana pasada yo compré la comida en el centro.
2. La próxima semana yo compré la comida en el centro.
3. La semana pasada ella compré la comida en el centro.

1. La próxima semana yo compraré la computadora en el almacén.
2. La semana pasada yo compraré la computadora en el almacén.
3. La próxima semana él compraré la computadora en el almacén.

1. La semana pasada yo descubrí el anillo en la playa.
2. La próxima semana yo descubrí el anillo en la playa.
3. La semana pasada nosotros descubrí el anillo en la playa.

1. La próxima semana yo descubriré el oro en la isla.
2. La semana pasada yo descubriré el oro en la isla.
3. La próxima semana nosotros descubriré el oro en la isla.

1. El mes pasado ellas escribieron el poema en el cuaderno.
2. El próximo mes ellas escribieron el poema en el cuaderno.
3. El mes pasado ella escribieron el poema en el cuaderno.

1. La próxima semana ellas escribirán el ensayo en la computadora.
2. La semana pasada ellas escribirán el ensayo en la computadora.
3. La próxima semana ella escribirán el ensayo en la computadora.

1. El mes pasado ellas escucharon el tráfico en la calle.
2. El próximo mes ellas escucharon el tráfico en la calle.
3. El mes pasado nosotras escucharon el tráfico en la calle.

1. El próximo mes ellos escucharán la presentación en el auditorio.
2. El mes pasado ellos escucharán la presentación en el auditorio.
3. El próximo mes nosotros escucharán la presentación en el auditorio.
1. La semana pasada él estudió el mapa en la clase.
2. La próxima semana él estudió el mapa en la clase.
3. La semana pasada yo estudió el mapa en la clase.

1. La próxima semana ella estudiará la lección en la computadora.
2. La semana pasada ella estudiará la lección en la computadora.
3. La próxima semana yo estudiaré la lección en la computadora.

1. El año pasado ella limpió la piscina en el patio.
2. El próximo año ella limpió la piscina en el patio.
3. El año pasado yo limpió la piscina en el patio.

1. La próxima semana él limpiará el coche en el garaje.
2. La semana pasada él limpiará el coche en el garaje.
3. La próxima semana yo limpiará el coche en el garaje.

1. La semana pasada yo miré la televisión en la sala.
2. La próxima semana yo miré la televisión en la sala.
3. La semana pasada nosotros miré la televisión en la sala.

1. El próximo año yo miraré el camello en el zoológico.
2. El año pasado yo miraré el camello en el zoológico.
3. El próximo año nosotros miraré el camello en el zoológico.

1. El año pasado él pintó el cuadro en el estudio.
2. El próximo año él pintó el cuadro en el estudio.
3. El año pasado ellos pintó el cuadro en el estudio.

1. El próximo año él pintará la estatua en la plaza.
2. El año pasado él pintará la estatua en la plaza.
3. El próximo año ellos pintará la estatua en la plaza.

1. El año pasado ellos practicaron el esquí en la montaña.
2. El próximo año ellos practicaron el esquí en la montaña.
3. El año pasado nosotros practicaron el esquí en la montaña.

1. El próximo mes ellos practicarán el básquetbol en la cancha.
2. El mes pasado ellos practicarán el básquetbol en la cancha.
3. El próximo mes nosotros practicarán el básquetbol en la cancha.

1. El año pasado él recibió el diploma en la ceremonia.
2. El próximo año él recibió el diploma en la ceremonia.
3. El año pasado ellos recibió el diploma en la ceremonia.

1. La próxima semana ella recibirá la nota en la clase.
2. La semana pasada ella recibirá la nota en la clase.
3. La próxima semana ellas recibirá la nota en la clase.

1. La semana pasada yo vendí el periódico en la calle.
2. La próxima semana yo vendí el periódico en la calle.
3. La semana pasada él vendió el periódico en la calle.

1. El próximo año yo venderé la ropa en el mercado.
2. El año pasado yo venderé la ropa en el mercado.
3. El próximo año él venderé la ropa en el mercado.

1. El año pasado yo visité la catedral en el pueblo.
2. El próximo año yo visité la catedral en el pueblo.
3. El año pasado nosotros visité la catedral en el pueblo.

1. El próximo año yo visitaré la casa en la playa.
2. El año pasado yo visitaré la casa en la playa.
3. El próximo año nosotras visitaré la casa en la playa.
APPENDIX J: Examples of Self-Paced Reading Assessment Pictures

Three examples of pictures used to test comprehension of the verb:

1. El mes pasado nosotros comimos el pescado en el restaurante.

2. El próximo mes ellos escucharán la presentación en el auditorio.

3. El año pasado él pintó el cuadro en el estudio.
Three examples of pictures used to test comprehension of the direct object:

1. La semana pasada yo descubrí el anillo en la playa.

2. La próxima semana él limpiará el coche en el garaje.
3. El próximo año yo venderé la ropa en el mercado.

Three examples of pictures used to test comprehension of the prepositional phrase:

1. La próxima semana ellas escribirán el ensayo en la computadora.

2. El próximo mes nosotros aprenderemos la fórmula en el laboratorio.
3. El año pasado yo visité la catedral en el pueblo.
VITA

Melisa J. Dracos
mjdracos@gmail.com

Ph.D. in Spanish Linguistics and Language Science
Doctoral Option: Applied Linguistics 2013
The Pennsylvania State University, University Park, PA

M.A. in Spanish Linguistics 2009
The Pennsylvania State University, University Park, PA

M.A. in Hispanic Languages and Literature 2007
Stony Brook University, Stony Brook, NY

B.S.Ed. in Spanish and English (Secondary Education) 2002
Baylor University, Waco, TX

RESEARCH INTERESTS
Second language processing and acquisition
Cognitive Individual differences
Morphosyntax
Second language teaching methodologies

SELECTED TEACHING AND ADMINISTRATION EXPERIENCE
Courses taught at the Pennsylvania State University: 2007-2013
  SPAN 440 – Teaching Romance Languages: Theory and Practice
  SPAN 215 – Introduction to Spanish Linguistics
  SPAN 003 – Intermediate Spanish (full semester & intensive summer)
  SPAN 002 – Elementary Spanish II
Assistant Coordinator for Summer Study Abroad in Argentina, Baylor University 2010-2011
Supervisor for Spanish 3, Penn State University 2008-2009
Courses taught at Stony Brook University: 2005-2007
  SPAN 112 – Elementary Spanish II
  SPAN 111 – Elementary Spanish I
Lecturer of ESL, Universidad San Ignacio de Loyola, Lima, Peru 2004
Spanish Teacher, Flower Mound High School, Lewisville, TX 2003

SELECTED CONFERENCE PRESENTATIONS


SELECTED AWARDS
Teaching Excellence Award, Penn State University 2012
Enhanced Dissertation Award and Grant, Penn State University 2011-2012
Teaching Excellence Award, Penn State University 2009
Teaching with Technology Certificate, Penn State University 2008
Teaching Excellence Award, Stony Brook University 2007