COGNITIVE AND PSYCHOSOCIAL FACTORS IN
THE LONG-TERM DEVELOPMENT OF
IMPLICIT AND EXPLICIT SECOND LANGUAGE KNOWLEDGE IN
ADULT LEARNERS OF SPANISH AT INCREASING PROFICIENCY

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Spanish

By

Ellen Johnson Serafini, M.S.

Washington, D.C.
June 13, 2013
This study examined the second language (L2) development of adult learners of Spanish at three levels of proficiency during and after a semester of instruction. A fundamental goal was to identify cognitive and psychosocial individual differences (IDs) that can explain between-learner variation over time in order to expand our understanding of the role learner IDs play in L2 learning and retention. This study also advances a line of research aiming to improve the validity and reliability of tests designed to measure knowledge of (implicit) and about (explicit) L2 grammar.

Eighty-seven learners in their 2nd, 4th, and 6th semester of university Spanish instruction completed four tests of cognitive ability measuring L2 aptitude, working memory capacity (WMC), phonological short-term memory (PSTM) and processing speed as well as two Likert-type questionnaires measuring different facets of L2 motivation. Implicit and explicit knowledge of ten linguistic structures in Spanish were measured through an elicited oral imitation task and untimed grammaticality judgment task used in previous research with several methodological improvements to strengthen their construct validity.

A multi-faceted statistical approach was taken to answer research questions employing both traditional and innovative hierarchical linear modeling techniques. Results suggest that 2.5 months of classroom instruction significantly enhanced performance on both measures of L2 knowledge for all proficiency groups with continued gains one month later. Higher proficiency
groups significantly outperformed lower proficiency groups at each point in time and there was evidence of faster and more efficient use of implicit linguistic knowledge with increasing proficiency. However, exploratory and confirmatory factor analyses revealed that the validity of a two-factor model varied according to learners' proficiency level, suggesting that implicit and explicit knowledge in instructed learners is best viewed as a continuum.

For lower proficiency groups, several strong relationships were found between cognitive capacity, L2 motivation, and performance on measures of implicit and explicit knowledge; however, learner IDs did not explain performance variance for the advanced group. Relationships among ID constructs also highlight their likely interaction and dynamic influence on L2 development over time. Findings are discussed in terms of their implications for second language acquisition (SLA) theory, assessment, and pedagogy.
ACKNOWLEDGEMENTS

Writing a dissertation has been truly humbling and I feel tremendous gratitude for the support and lessons I have received along the way.

To Cristina Sanz, my mentor and friend: I admire you in so many ways, but especially as a scholar, teacher, and superwoman in general. Thank you for believing in my ideas and in my ability to pursue them from the beginning, and for the laughter and many entertaining metaphors in your office that involved onion layers and forests. I am beyond grateful for your clarity, scholarly guidance, excellent feedback, and for the many times you have renewed my confidence in this process during the last five years. To Ron Leow, my 'boss' and friend: Thank you for the many opportunities you have given me, for encouraging me to think deeper and question assumptions in our field, and for your many famous reminders, my favorite of which is to have fun and enjoy the journey! You have undoubtedly pushed me to grow as a scholar, teacher, and administrator. To Nuria Sagarra, who has been an invaluable committee member and friend: I remember meeting in a café in Barcelona to "chat" about working memory, automatization, and L2 proficiency and feeling grateful, though maybe surprised, that you were on board with this kitchen sink project. Thank you for your continuous flexibility and sharp feedback, which has truly improved my dissertation and future research. I also owe a great deal to our resident statistician, Rusan Chen, for his ever cheerful guidance in carrying out statistical analyses, even when I asked questions I should have already known the answer to, and for the refreshing view of the Potomac river from his office window. It brightened my day many times.

To the beginning, intermediate, and advanced Spanish instructors at Georgetown who let me invade their classrooms and to their students who volunteered to participate in this research: You deserve many years of good karma. Thank you for your curiosity and commitment to taking
part in this study, despite the frequent classroom interruptions, annoying reminder emails, and endless hours spent with me in a language lab. Your smiles and positive attitudes were contagious. To my own students who I have had the privilege of teaching over the last five years: Thank you for being an outlet of creativity in the classroom, for your good humor and hard work, and for reminding me on a daily basis of why I am lucky to teach and research language.

I am very thankful for my Hoya family of friends and colleagues who have shared in the highs and lows of this experience with me. To Sarah Grey, Julio Torres, Germán Zárate-Sández, Jason Robinson, Jessie Cox, Annie Calderón, Kaitlyn Tagarelli, Julie Lake, Marisa Filgueras, Charlie Nagle, Sergio Adrada, Ana María Ferreira, Eli Kissling, Aubrey Logan-Terry, Rachael Albritten, Luis Cerezo, Missy Baralt, and Laura Gurzynski-Weiss: thank you for meeting me at some point along this seemingly endless, but always entertaining road of being a Ph.D. student. Your friendship and work ethic have been inspiring and provided much needed laughter many times along the way.

This dissertation would not have been possible without my parents, who have loved and supported me unconditionally throughout every chapter in my life, and to whom I dedicate this dissertation. To Mom and Dad: Thank you for being so very proud of me and for believing in my ability to achieve my goals, even when that meant moving across the country and world. Dad, you are a constant example of hard work and dedication and I will always carry your lessons and advice with me, though I am still refining the art of 'working smarter, not harder'. Mom, your excitement about each small step forward and comfort during each minor setback always renewed my drive to keep going and I have felt stronger after every phone call. To Mom and Ed: Thank you for your boundless pride and confidence in my decisions, for celebrating each milestone, and for continuously emphasizing the value of investing in myself. Your generous
support and reminders to take it one day at a time have helped me keep the bigger picture and the 
end goal in mind. To my 'baby' sisters, Masa and Gigi: Thank you for ensuring I kept a sense of 
humor and didn't take this thing too seriously, for always listening to the good and the bad, and 
of course, for advising me on what to wear during this process. To my suegros, Phil and Mary 
Alice: Thank you for reminding me to be kind to myself and to reboot and realign the plan when 
needed. Your genuine interest and pride in my work has truly been a source of motivation. To 
my Grandma and Grandpa Duffy, Grammy, and Papa Phil: Thank you for always asking 
questions about my students and research, and sincerely wanting to hear the answers. I feel 
fortunate to have you in my life and to have shared this experience with you. To my Grandma 
and Grandpa Johnson: You have been present during this journey even though you are no longer 
here; thank you for your never-ending pride in me. To my aunts, uncles, cousins, nieces, brother- 
in-law, and sister-in-law who live near and too far away: Thank you for being supportive and 
curious about what was happening in my world at every family gathering. You are a constant 
reminder of what matters most.

Finally, to my husband and best friend, Ted Serafini: Thank you for climbing this 
mountain with me. I would not have wanted to do it without you or your jokes, even the ones 
you tend to repeat. Thank you for always making me laugh, for making me take a break from the 
computer to walk or run in the sunshine and feel human again, for picking me up at campus 
during data collection marathon weekends, for bringing me plants, flowers, coffee, wine, 
chocolate, etc., for going salsa dancing when you needed to study, for always supporting and 
listening to me no matter what you had on your plate, and above all, for continually teaching me 
about persistence and determination. You are an unfailing source of strength and love.
TABLE OF CONTENTS

CHAPTER ONE: Introduction......................................................... 1
Statement of the Problem.......................................................... 1
Goals and Context of the Current Study........................................ 8
Potential Contribution to SLA Theory and Pedagogy....................... 10
Organization of the Study.......................................................... 11
Definition of Terms......................................................................... 12
  Second language (L2) knowledge............................................... 12
  Implicit L2 knowledge.............................................................. 12
  Explicit L2 knowledge.............................................................. 13
Development over time............................................................... 14
  Retention of L2 knowledge....................................................... 14
Learner individual differences (IDs)............................................. 15
  L2 aptitude................................................................................. 15
  Working memory capacity (WMC)............................................. 16
  Phonological short-term memory (PSTM)................................. 16
  L2 motivation............................................................................ 17
L2 proficiency.............................................................................. 17

CHAPTER TWO: Review of the Literature........................................ 19
Implicit and Explicit L2 Knowledge............................................... 19
Overview..................................................................................... 19
  Defining implicit and explicit L2 knowledge............................... 20
  Measuring what L2 learners know............................................ 21
  Empirical research aiming to improve the measurement of L2
    knowledge............................................................................. 22
  Further refining how we measure L2 knowledge....................... 25
  Implicit and explicit L2 knowledge: Some issues..................... 29
Summary..................................................................................... 32

Learner Individual Differences.................................................... 33
Overview..................................................................................... 33
  Historical perspective of learner ID research.............................. 35
  Cognitive dimension of L2 learning.......................................... 37
    Defining and measuring L2 aptitude....................................... 41
    Empirical research on L2 aptitude......................................... 43
  Summary................................................................................... 48
  Defining working memory capacity (WMC): Models in
    cognitive psychology.............................................................. 49
Overview
Description of Participants
Participant biodata questionnaire
Operationalizations
Implicit and explicit L2 knowledge
Elicited oral imitation task (OIT)
Untimed grammaticality judgment task (UGJT)
Cognitive IDs
L2 aptitude: Modern Language Aptitude Test (MLAT)
Working memory capacity: Operation span (OSpan) task
Phonological short-term memory: Digit span task
Processing speed: Digit symbol coding task (DSCT)
Psychosocial IDs
L2 motivation
Attitudes and Motivational Test Battery (AMTB)
Possible selves questionnaire
L2 proficiency level
Other Materials
Classroom teaching style questionnaire
Outside exposure/use questionnaire
Research Design
Procedure
Counterbalanced design
Scoring and coding procedures
OIT
UGJT
MLAT (III, IV, V)
OSpan task
Digit span task
DSCT
AMTB and Selves questionnaires
Statistical Procedures
Overview
Power analysis
Tests of reliability
Research question 1: Development and retention of implicit/explicit knowledge over time by proficiency level
Research questions 2 & 3: Relationship between learner IDs and development/retention of implicit (RQ2) and explicit (RQ3) knowledge by proficiency level and predictive ability over time
Research question 4: Relationships between cognitive and
Test version comparability.............................................................. 198
Test validity.................................................................................. 199
Correlation between placement exam and performance.............. 199
Factor analyses............................................................................ 199

Research question 2: Relationship between Learner IDs and Development/Retention of Implicit Knowledge by Proficiency Level (2a) and Predictive Ability over Time (2b).................................................. 205
Correlations: IDs and implicit knowledge (OIT)......................... 205
Correlations: Cognitive IDs and latency (OIT)..................... 212
Multiple regressions: IDs and implicit knowledge (OIT)...... 214
Level-2 linear growth curve analysis (OIT).............................. 224

Research question 3: Relationship between Learner IDs and Development/Retention of Explicit Knowledge by Proficiency Level (3a) and Predictive Ability over Time (3b) .................................................. 226
Correlations: IDs and explicit knowledge (UGJT).............. 227
Multiple regressions: IDs and explicit knowledge (UGJT).... 232
Level-2 linear growth curve analysis (UGJT)...................... 239

Research question 4: Relationships between Cognitive and Psychosocial IDs 242
Correlations: Cognitive ID components................................. 242
Correlations: Psychosocial ID components......................... 243
Correlations: Cognitive & psychosocial ID components........ 245

Summary of the Results.......................................................... 249
Research question 1................................................................. 249
Research questions 2 and 3..................................................... 252
Research question 4................................................................. 256

CHAPTER FIVE: Discussion and Conclusions........................................ 259
Overview.................................................................................... 259
Research question 1: Development and retention of implicit/explicit L2 knowledge over time by proficiency level.................. 259
Test reliability and validity.................................................... 264
Research questions 2 and 3: Relationships between learner IDs and development and retention of implicit (RQ2) and explicit (RQ3) L2 knowledge.............................................. 273
Cognitive ID components...................................................... 273
L2 aptitude.............................................................. 274
Working memory capacity and phonological short-term memory............................................................................. 276
Processing speed...................................................... 283
Cognitive IDs and oral latency.............................................. 284
Cognitive IDs and item grammaticality.................................. 285
Psychosocial ID components............................................................... 286
L2 motivation as measured in the AMTB........................................ 286
L2 Motivation as measured on adapted Selves questionnaire....................... 294
Psychosocial IDs and item grammaticality...................................... 299
Relative predictive ability of cognitive and psychosocial IDs............. 300
Research question 4: Relationships among ID components within the
cognitive and psychosocial dimensions of L2 learning...................... 302

Summary Conclusions of the Study....................................................... 307

Implications................................................................................................. 310
Theoretical implications.............................................................................. 310
1. Implicit and explicit L2 knowledge is a continuum................. 310
2. The role of IDs diminishes at increasing proficiency
levels........................................................................................................... 311
3. Working memory capacity defines foreign language
aptitude........................................................................................................ 313
4. Learner IDs mediate the use of both implicit and explicit
knowledge.................................................................................................. 314
5. The utility of learner internal IDs varies over time................. 315
6. The cognitive and psychosocial dimensions of L2 learning
interact...................................................................................................... 316
Practical implications for assessment of L2 performance and learner
IDs............................................................................................................... 317
1. Verifying how we know what our learners know................. 317
2. Measuring learner differences: Aptitude, working memory
and motivation.......................................................................................... 319
Pedagogical implications for the classroom........................................ 321
1. Know how your students differ from one another and apply
what you know in the classroom............................................................. 321
2. Use L2 Motivation questionnaires to encourage
self-awareness and use of motivational strategies among your
students....................................................................................................... 322

Limitations of the Study and Directions for Future Research.................. 323

APPENDIX A. Target Structures............................................................... 327
APPENDIX B. Participant Biodata Questionnaire...................................... 328
APPENDIX C. Classroom Teaching Style Questionnaire........................... 332
APPENDIX D. Outside Exposure/Use Questionnaire................................. 334
APPENDIX E. Elicited Oral Imitation (OI) Task Script............................... 337
LIST OF FIGURES

Figure 1. A model of language acquisition and use................................................. 38
Figure 2. Multi-component model of working memory (Baddeley, 2000).................. 52
Figure 3. Materials and procedure........................................................................ 127
Figure 4. Gains in meaning comprehension (OIT) over time by group................. 173
Figure 5. Gains in sentence repetition accuracy (OIT) over time by group.......... 174
Figure 6. Gains in sentence repetition accuracy (OIT) over time by group with error bars.......................................................... 174
Figure 7. Individual growth curve trajectories: Implicit knowledge.................... 178
Figure 8. Mean RTs over time from the beginning to the end of the semester (2.5 mos.) by group.................................................. 184
Figure 9. Mean CVs from the beginning to the end of the semester (2.5 mos.) by group........................................................................ 185
Figure 10. Gains in grammaticality judgment accuracy over time (3.5 mos.) by group........................................................................ 191
Figure 11. Individual growth curve trajectories: Explicit knowledge.................... 193
Figure 12. Judgment accuracy by source attributions (%).................................... 196
Figure 13. Relationship between performance on the OIT and phonemic coding ability (T1)........................................................................ 215
Figure 14. Relationship between performance on the UGJT and working memory capacity (T1)........................................................................ 232
Figure 15. Relationship between performance on the UGJT and working memory capacity (T3)................................................................. 233
Figure 16. RQ1 Results: Summary of significant one-way and Repeated Measures ANOVAs for performance on tests of implicit and explicit knowledge........................................................................ 251
Figure 17. RQ1 Results: Test validity based on Principal Component Confirmatory Factor Analysis across groups................................. 251
Figure 18. RQ1 Results: Test validity based on Confirmatory Factor Analysis across groups and Exploratory Factor Analyses by group............ 252

Figure 19. RQ2 Results: Summary of Pearson Product Moment correlations between cognitive and psychosocial ID variables and performance on test of implicit knowledge................................................................. 255

Figure 20. RQ3 Results: Summary of Pearson Product Moment correlations between cognitive and psychosocial ID variables and performance on test of explicit knowledge................................................................. 255

Figure 21. RQ2 & RQ3 Results: Summary of Multiple Regressions with significant cognitive and psychosocial ID predictors of performance on tests of implicit and explicit knowledge by testing time.......................... 256

Figure 22. RQ4 Results: Summary of Pearson Product Moment correlations between cognitive and psychosocial ID variables........................................... 258
LIST OF TABLES

Table 1. Five tests measuring implicit/explicit L2 knowledge...................... 22
Table 2. Historical overview of individual differences (IDs) in L2 research 35
Table 3. Participant background............................................................... 114
Table 4. Counterbalanced design in the administration of tests of implicit and explicit L2 knowledge................................................................. 129
Table 5. Sample size determined by a priori power analyses (Spring 2012 pilot study).................................................................................. 135
Table 6. Pilot study: Descriptive statistics on the oral imitation task........... 146
Table 7. Pilot study: Oral imitation test reliability and correlation between scoring procedures................................................................................ 148
Table 8. Pilot study: Descriptive statistics on the untimed grammaticality judgment task.................................................................................. 148
Table 9. Pilot study: Untimed grammaticality judgment test reliability........ 149
Table 10. Descriptive statistics (%) on the Departmental Spanish Placement Exam (36 items) by group.............................................................. 155
Table 11. Summary item statistics (%) on measures of cognitive capacity for all participants (N = 87)........................................................................ 159
Table 12. Summary item statistics (%) on OSPAN measure by group for all participants (N = 87)........................................................................ 161
Table 13. Reliability estimates and summary item statistics on L2 motivation measures (avg. Likert response 1-7) from all participants at Time 1 (N = 86) and Time 2 (N = 82-83)................................................................................ 162
Table 14. Reliability estimates for all scale categories on AMTB measure for all participants................................................................. 164
Table 15. Overall reliability on the oral imitation test (40 items) and correlation between scoring methods on each test version (V1, V2, V3)....................... 166
Table 16. Overall reliability on the untimed grammaticality judgment test (N = 60 items) and by item grammaticality (N = 30 GR/30 UGR) on each test version (V1, V2, V3)....................................................... 169
Table 17. Descriptive statistics for meaning processing score on the oral imitation task by group and testing time................................. 170
Table 18. Descriptive statistics for sentence repetition accuracy on the oral imitation task by group and testing time.................................................... 170
Table 19. Results of one-way ANOVAs with test scores (OIT) by group and testing time.................................................................................. 171
Table 20. Unstandardized coefficients for growth curve parameters and variances: OIT.......................................................................................... 176
Table 21. Descriptive statistics for latency (ms) on the oral imitation task by group and testing time.................................................................. 180
Table 22. Results of one-way ANOVAs with latency scores (OIT) by group and testing time........................................................................ 182
Table 23. Descriptive statistics for difference (milliseconds) between L2 mean RT (Spanish) and L1 mean RT (English) by group and testing time... 186
Table 24. Descriptive statistics for judgment accuracy on the untimed grammaticality judgment task (UGJT) by group and testing time............... 188
Table 25. Results of one-way ANOVAs with test scores (UGJT) by group and testing time................................................................. 190
Table 26. Unstandardized coefficients and variances for growth curve parameters: UGJT......................................................................................... 192
Table 27. Grammaticality judgment source attributions (UGJT) (%).............. 194
Table 28. Judgment accuracy by source attributions (% (SD))...................... 195
Table 29. Pearson Product Moment correlational matrix for tests of implicit (OIT) and explicit (UGJT) L2 knowledge (Time 1)................................. 200
Table 30. Principal component factor analysis across groups (Time 1)............ 201
Table 31. Loadings for principal component factor analysis across groups (Time 1).................................................................................. 201
Table 32. Exploratory factor analysis by group (Time 1)............................... 203
Table 33. Loadings for rotated factor matrix by group (Time 1)...................... 203
Table 34. Correlation matrix: Cognitive IDs and performance on the OIT at Time 1 (N = 87), Time 2 (N = 83) and Time 3 (N = 33)............................ 206
Table 35. Correlation matrix: Psychosocial IDs and performance on the OIT at Time 1 (N = 86), Time 2 (N = 83) and Time 3 (N = 33)............................ 209
Table 36. Correlation matrix: Cognitive IDs, latency (RT) and CV (OIT) at Time 1 (N = 76), Time 2 (N = 59) and Time 3 (N = 21)............................ 213
Table 37. Cognitive ID explanatory model: Implicit knowledge (T1)........... 216
Table 38. Cognitive ID explanatory model: Implicit knowledge (T2)........... 217
Table 39. Cognitive ID explanatory model: Implicit knowledge (T3)........... 218
Table 40. Psychosocial ID explanatory model: Implicit knowledge (T1)..... 219
Table 41. Psychosocial ID explanatory model: Implicit knowledge (T2).... 220
Table 42. Psychosocial ID explanatory model: Implicit knowledge (T3).... 222
Table 43. Cognitive and psychosocial ID combined explanatory model: Implicit knowledge (T1).............................................................. 223
Table 44. Cognitive and psychosocial ID combined explanatory model: Implicit knowledge (T2).............................................................. 223
Table 45. Cognitive and psychosocial ID combined explanatory model: Implicit knowledge (T3).............................................................. 223
Table 46. Final model fitted for implicit knowledge (OIT) with cognitive & psychosocial ID predictors.............................................................. 225
Table 47. Correlation matrix: Cognitive IDs and performance on the UGJT at Time 1 (N = 86), Time 2 (N = 83) and Time 3 (N = 33).............. 227
Table 48. Correlation matrix: Psychosocial IDs and performance on the UGJT at Time 1 (N = 86), Time 2 (N = 83) and Time 3 (N = 33).............. 229
Table 49. Cognitive ID explanatory model: Explicit knowledge (T1)........ 234
Table 50. Cognitive ID explanatory model: Explicit knowledge (T2, T3).... 235
Table 51. Psychosocial ID explanatory model: Explicit knowledge (T1).... 236
Table 52. Psychosocial ID explanatory model: Explicit knowledge (T2)................................................................. 237

Table 53. Cognitive and psychosocial ID combined explanatory model: Explicit knowledge (T1)........................................................................................................................................................................ 238

Table 54. Cognitive and psychosocial ID combined explanatory model: Explicit knowledge (T2) ................................................................................................................................................................. 238

Table 55. Cognitive and psychosocial ID combined explanatory model: Explicit knowledge (T3) ................................................................................................................................................................. 238

Table 56. Final model fitted for explicit knowledge (UGJT) with cognitive and psychosocial ID predictors.................................................................................................................................................................................. 240

Table 57. Correlation matrix: Relationship between cognitive ID components (N = 87)........................................................................................................................................................................ 242

Table 58. Correlation matrix: Relationship between psychosocial ID components at Time 1 (N = 86) and Time 2 (N = 82)................................................................................................................. 244

Table 59. Correlation matrix: Relationship between components of cognitive and psychosocial ID variables at Time 1 (N = 86-87) & Time 2 (N = 82-83) .................................................................................................................. 246
CHAPTER ONE: Introduction

Statement of the Problem

While there are few undisputed areas of theoretical and empirical inquiry in the field of second language acquisition (SLA), one universal observation is that adults learning a second (L2) or foreign language (FL) vary considerably in their attainment of that language. Even if two adult learners receive a similar quantity and quality of input in the target language in a comparable setting, they will be differentially successful in their ability to read, listen, write, and speak the L2. Furthermore, there is no guarantee they will develop proficiency at the same speed or retain their skills over time. This stands in clear contrast to child learners acquiring their first language (L1), or additional languages, as they uniformly reach native or native-like competence with regularity and ease (e.g., Birdsong, 2005; DeKeyser, 2000; Johnson & Newport, 1989; Long, 1982, 2007).

As a result, SLA theory has the 'problem' of explaining why such varying learning outcomes exist for adults, or those who begin learning an L2 after puberty. Individual differences (IDs) have been proposed as one way to explain differential learning success for adults, particularly within cognitive or cognitive-interactionist approaches whose common goal is to assess the impact of and interaction between internal and external factors that constrain the rate and outcome of adult SLA (Long, 2007; Norris & Ortega, 2012; Sanz, 2005). While the last forty years of research into the role of learner IDs have been fruitful in identifying positive relationships between learner-internal factors and measures of L2 achievement (see Altman, 1980; Skehan, 1989, 1991; Larsen-Freeman & Long, 1991; Oxford & Ehrman, 1993; Sagarra, 2000; Robinson, 2002; Bowden, Sanz & Stafford, 2005; Dörnyei, 2005; R. Ellis, 2008a), there is
still little research that considers the potential joint impact of the cognitive and affective underpinnings of learning other languages as an adult (Segalowitz & Trofimovich, 2012).

While cognitive perspectives of SLA equate L2 learning in adults with other complex skills and emphasize the cognitive mechanisms constraining its development (DeKeyser, 2007; McLaughlin, 1987; McLaughlin, Rossman & MacLeod, 1983), psychosocial perspectives emphasize the uniqueness of L2 development as a process that is fundamentally constrained by the psychological, social and cultural context in which it takes place (Dörnyei, 2003, 2005, 2009; Gardner, 1985, 2001, 2010; Ushioda & Dörnyei, 2012). In related disciplines like cognitive psychology, it has long been observed that while “isolated traits often have a substantial impact on learning outcomes, it may be that combinations of traits have more predictive power than traits in isolation” (Ackerman, 2003, p. 92; see also Ackerman, Sternberg, & Glaser, 1989) and that cognition includes not only cognitive abilities but affect and motivation as well (Snow, 1989, 1994). Similarly, some researchers in SLA (e.g., Hulstijn, 2002; Larsen-Freeman & Cameron, 2008; Segalowitz & Trofimovich, 2012) have begun to recognize that the various dimensions of L2 learning could interact in complex ways with a number of potential “self-reinforcing loops between the cognitive, motivational, social, and experiential aspects of L2 use” (Segalowitz & Trofimovich, p. 187). Thus, investigating the impact of learner IDs in isolation seems to be a considerable limitation when attempting to understand the complexity of L2 performance variation.

Furthermore, the majority of research investigating relationships between learner IDs and L2 development has been cross-sectional in nature, focused on learner performance at only one point in time (see Ortega & Iberri-Shea, 2005 for a discussion), despite the gradual nature of gaining proficiency in another language. Given that learners' interlanguage systems are
permeable and dynamic, or in a constant state of flux (Selinker, 1972, 1992; Tarone, 1983), our understanding of the differential role IDs play in facilitating or constraining how form-function relationships in interlanguage evolve in long-term development remains quite limited.

The empirical gaps pertaining to the relationship between learner IDs and SLA are important to address in light of two fundamental goals driving SLA theory and research: 1) to define and describe L2 linguistic knowledge and 2) to explain how this knowledge develops over time by specifying the external and internal variables involved (R. Ellis, 2005; Norris & Ortega, 2012). Selinker (1992) has argued that only a longitudinal description of the adult's learner language will provide convincing evidence required in order to "make certain inferences about the learning process by describing successive states of learner language" (p. 151). Information about the nature of interlanguage competence can be indirectly derived by examining how learners use or access their knowledge in performance, which requires reliable and replicable empirical indices of learner knowledge (Norris & Ortega, 2012; Dörnyei & Ushioda, 2009).

To effectively discern what L2 learners know, a common division in SLA is made between knowledge of (implicit) and about (explicit) L2 grammar. This distinction stems from those made in cognitive psychology between controlled and automatic processing at the intake stage of learning (Shiffrin & Schneider, 1977) as well as declarative (knowing that) and procedural (knowing how) use of knowledge at the output stage (Anderson, 1983). These distinctions were discussed early on in terms of learned versus acquired knowledge (Krashen, 1981) and analyzed versus unanalyzed knowledge (Bialystok, 1994). More recent operational definitions describe explicit knowledge as consciously monitored during controlled processing whereas access to implicit knowledge is thought to proceed automatically during fluent, spontaneous language use (N. Ellis, 1994, 2005; R. Ellis, 1993, 1994, 2004, 2009; Hulstijn, 2002,
Furthermore, implicit knowledge is characterized as unverbalizable whereas explicit knowledge can, “in principle, be talked about” (Hulstijn, 2002, p. 205); however the criterion of (non)verbalizability is nonetheless problematic given that a learner may have explicit knowledge of some aspect of the target language but not be able to verbalize it. Thus, there may be levels of explicit knowledge with metalinguistic knowledge being a higher level that includes aspects or rules of the target language that learners are able to describe, though this need not entail the use of metalanguage (i.e., technical terminology, R. Ellis, 2004, p. 239).

There are several methodological challenges in assessing interlanguage knowledge given the difficulty in establishing clear separability of knowledge constructs, especially since learners likely draw on both types of knowledge in performance. Nonetheless, developing theory-based, replicable assessments of learner knowledge is a valuable endeavor (Norris & Ortega, 2012) and can improve how we know what our learners know based on behavioral evidence (Lakshmanan & Selinker, 2001). A recent line of research has set out to improve and refine a test battery based on operational definitions, or behavioral criteria, in order to examine L2 knowledge in learners of English (e.g., R. Ellis, 2004, 2005, 2009b; Erlam, 2006; Elder, 2009; Han & Ellis, 1998; Philp, 2009) and heritage and L2 learners of Spanish (Bowles, 2011; Gutiérrez, 2012).

These studies report tests to be internally consistent, and exploratory and confirmatory factor analyses have consistently produced a two-factor solution in which tests of implicit knowledge loaded on one factor and tests of explicit knowledge loaded on the other, indicating relatively separable knowledge constructs. However, the validity of these tests merit further investigation and replication with learners at different proficiency levels given that the memory systems underlying use of knowledge may potentially change at increasing levels of proficiency (e.g., Ullman, 2001, 2005). Moreover, some researchers (e.g., Shiu & Spada, 2012; Spada, 2013)
have recently questioned the validity of one of the tests designed to tap implicit knowledge, the elicited oral imitation task, positing that it is actually a measure of *automatized* explicit knowledge when used with instructed learners, who have likely learned their L2 explicitly.

In relation to the second goal in SLA theory and research concerned with variables that differentially impact L2 development over time, second language aptitude and motivation are the two most powerful learner-internal factors found to have positive, significant relationships with L2 achievement and a level of predictive success exceeding any other proposed causal variable in terms of magnitude of correlations (Skehan, 1998, p. 192). L2 aptitude, a cognitive, stable ability, is regarded by some researchers to be “the single best predictor of subsequent language learning achievement” (Sawyer & Ranta, 2001, p. 320) with medium to high effect sizes (.4-.6, Carroll, 1981). However, many researchers taking a componential view of aptitude have influentially proposed that working memory capacity (WMC) is actually the key construct underlying the relationship between aptitude and L2 achievement (e.g., Miyake & Friedman, 1998; Robinson, 2001a, 2002, 2005a; Sanz, 2005; Sawyer & Ranta, 2001; Skehan, 1998, 2002, 2012; Wen, 2012; Wen & Skehan, 2011).

Indeed, people vary in their ability to code, store and process information and these abilities are regulated by WMC, conceived primarily as a mental workspace where information is simultaneously stored and processed (Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Baddeley, 1986, 2000, 2003, 2007, 2010). WMC has been demonstrated to impact a range of L2 phenomena, particularly those occurring at the input processing phases of learning regulated by attention and noticing (Schmidt, 1990, 1994, 2001), reflecting an "encoding-dominated perspective [in aptitude research]" (Skehan, 1998, p. 202). In general, people with more efficient WM resources are better language learners in relation to processing L2 agreement morphology
(Coughlin & Tremblay, 2013; Sagarra, 2007b, 2008; Sagarra & Hershensohn, 2010; though see Foote, 2011), under both implicit and explicit instructional conditions (Erlam, 2005; Robinson, 2002, 2005; Sanz, under review), and in terms of their ability to benefit from corrective feedback (e.g., Goo, 2011; Lado, 2008; Mackey, 2012 (see Chapter 6); Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Sagarra, 2007a; Sagarra & Abbuhl, forthcoming).

Despite the fact that “a major role for working memory is retrieval of stored long-term knowledge relevant to the tasks in hand” (Baddeley & Logie, 1999, p. 31), far less empirical work has been conducted in relation to its involvement in the latter stages of learners’ developing system (Internal system > Output). Empirical evidence pertaining to the role of WMC in L2 grammatical performance that does exist has been mixed (Erçetin & Alptekin, 2013; Kormos & Sáfár, 2008; Linck & Weiss, 2011; Roehr & Gánem-Gutiérrez, 2009; Sagarra, 2000). Some have supported a more is better view of working memory capacity (Miyake & Friedman, 1998) (Erçetin & Alptekin; Kormos & Sáfár; Linck & Weiss) while others have found nonsignificant relationships (Roehr & Gánem-Gutiérrez) or found support for a less is more view (Newport, 1990) (Sagarra). However, results from the latter two studies may be explained in terms of methodological limitations in tapping the construct of working memory capacity.

Furthermore, the majority of research in this realm has not probed IDs in WMC at a range of proficiency levels within the same study, often limited to initial stages of development (e.g., Kormos & Sáfár, 2008; Linck & Weiss, 2011; Sagarra, 2000; Sagarra & Hershensohn, 2010) or an upper proficiency level only (Coughlin & Tremblay, 2013; Erçetin & Alptekin, 2013; Foote, 2011). On the other hand, studies investigating the phonological loop component of WMC in L2 learning (e.g., Kormos & Sáfár, 2008; O'Brien, Segalowitz, Freed & Collentine, 2007; O'Brien, Segalowitz, Collentine & Freed, 2006; Winke, 2005) have found a changing role
for phonological working memory at different stages of L2 proficiency. This provides support for the view espoused by Robinson and others that "learning a language involves different abilities at different stages of development" (2013, p. 2), but more empirical evidence based on between-learner comparisons are needed to clarify the differential impact cognitive resources may exert on use of linguistic knowledge in learners across the L2 proficiency spectrum.

However, the cognitive mechanisms underlying L2 learning are not the only piece to the puzzle. As previously argued, a broader view of cognition includes the psychosocial dimensions of L2 learning as well (Larsen-Freeman & Cameron, 2008; Segalowitz & Trofimovich, 2012). As noted by Juffs and Harrington (2011), "although WM may indeed be a factor in explaining SOME variability among learners, other factors such as L1 and motivation may turn out in the end to be much more powerful explanatory variables in L2 learning" (p. 146). L2 motivation, as originally conceived within Gardner’s socioeducational model of SLA (1985, 2000, 2001, 2006, 2010), has indeed stood the test of time as a robust individual difference variable capable of explaining a large amount of variance in L2 achievement with a “more than medium effect size” (.29-.39) (Masgoret & Gardner, 2003, p. 151). The standardized measure used in this model, the Attitudes and Motivation Test Battery (AMTB), has proven to be an effective measure of L2 motivation in learners studying different target languages in varying contexts, but only a couple empirical studies have investigated relationships between L2 motivation and long-term L2 development over a year (Gardner, Masgoret, Tennant, & Mihic, 2004) and a semester (Yanguas, 2007) of instruction for university French L2 learners and university Spanish heritage language learners (HLLs), respectively.

Later conceptions of motivation within the L2 Motivational Self-System framework (Dörnyei, 2005, 2009; Dörnyei & Ushioda, 2009) based on possible selves theory in social
psychology (e.g., Markus & Nurius, 1986; Markus & Ruvolo, 1989; Higgins, 1987) have also received accumulating empirical validation, accounting for up to 42% of variance in criterion measures (reported learning effort and achievement) on average (Dörnyei, 2005, p. 104). Overall, researchers have found that the core concepts in both Gardner and Dörnyei's frameworks, Integrativeness and the Ideal L2 Self, actually converge conceptually in terms of psychological and emotional identification (Dörnyei, 2003, p. 6) as well as empirically, with moderate-to-strong relationships (Csizér & Dörnyei, 2005; Dörnyei, 2005, 2009; MacIntyre, Mackinnon, & Clément, 2009). However, some studies (e.g., Kormos & Csizér, 2008; Ryan, 2009) have reported that the Ideal L2 Self is more closely related to criterion measures (i.e., intended learning effort) than integrativeness in university learners of English. Nonetheless, this relationship needs to be verified with objective performance measures, given that correlations with motivational variables tend to be higher with learner-reported effort to learn the target language compared to objective measures of L2 achievement (Masgoret & Gardner, 2003, p. 147). Furthermore, no study has investigated the relative capacity of each motivational construct to explain differential language learning success over time (MacIntyre, et al., 2009), which is a question that warrants further empirical investigation.

**Goals and Context of the Current Study**

In order to address the gaps outlined above, the main goal of the present study is to investigate the roles that cognitive (i.e., L2 aptitude, working memory capacity, phonological short-term memory, processing speed) and psychosocial (i.e., L2 motivation) learner-internal differences play in mediating the development of grammatical knowledge in instructed L2 adult learners of Spanish i) at three different points in the developmental continuum (i.e., proficiency level), ii) at three different times during and after a learning period (i.e., one semester (3 months)
and 1 month after instruction), and iii) in stages of the SLA process that have typically received less empirical attention (i.e., use of L2 knowledge: Developing system > Output). In pursuit of this goal, this study provides a macro as opposed to a micro view of interlanguage development over the course of a semester and across the L2 proficiency spectrum. By distinguishing between knowledge of and about L2 grammar, I also contribute to the growing line of empirical work aiming to improve and refine the reliability, validity, and replicability of measures designed to tap implicit and explicit linguistic knowledge.

This study is conducted within an instructed university context that is difficult to strictly define in terms of a FL versus a L2 environment. Oxford (1996) and Oxford and Shearin (1994) distinguish between these learning contexts in terms of the availability of the target language. According to these authors, L2 learning occurs in contexts in which the language is readily available in the community and that provides ample opportunities to experience it outside the classroom. On the other hand, FL learning occurs in contexts in which the language in not commonly used in the community and there is little opportunity to experience it. The learners in this study are native English-speaking adult learners of Spanish studying in a non-intensive university language program. However, both Spanish and English are spoken in the wider local community and there are arguably ample opportunities to 'experience' Spanish outside the classroom but this largely depends on the students' awareness of and willingness to take advantage of those opportunities, as pointed out by Oxford and Shearin (p. 13). In light of these factors, the current setting arguably shares elements defining both FL and L2 environments.

The classroom context and teaching approach of instructors in this study are also important to consider in light of possible relationships between implicit and explicit instruction (external manipulation of input), learning (internal processing of input), and knowledge (access
to developing internal system) (DeKeyser, 2003, R. Ellis, 2009). The majority of instructors (i.e.,
graduate teaching assistants (TAs)) in the non-intensive program have been educated from a
psycholinguistic perspective that promotes teachers’ awareness and understanding of the
cognitive aspects of L2 learning in the formal classroom (Leow, 1995). Teachers are encouraged
to implement a task-based curriculum and engage learners in communicative activities with a
focus on meaning in the classroom and promote in-depth study of the formal aspects of Spanish
at home (Leow, 1994). Learners are instructed in the target language for 150 minutes on a
weekly basis (50 minutes a day x 3 days a week) for approximately fourteen weeks a semester
and use their textbook and accompanying workbooks to complete grammar exercises at home¹.
In other words, teachers are encouraged to implement both implicit and explicit instructional
techniques with a dual focus on form and meaning. While type of instruction does not
necessarily align with internal learner processes or with the development of L2 knowledge
(Schmidt, 1994), it is acknowledged that learners may develop both implicit and explicit
knowledge of Spanish grammar over the course of the semester of instruction (see Definition of
Terms for operational definitions of L2 knowledge).

**Potential Contribution to SLA Theory and Pedagogy**

This study represents a comprehensive effort to identify what factors best account for
differential learner success over time, which is a fundamental question guiding research into
adult L2 learning. Results promise to contribute to SLA theory construction by investigating the
separate and joint impact of several individual differences on L2 performance variance and by
providing a finer-grained picture of the role of IDs in L2 learning at different points in time and
in learners at different levels of experience. Importantly, this study takes a componential
approach to measuring learner IDs and attempts to relate different components of both cognitive
and psychosocial factors to the development of implicit and explicit knowledge, which could refine and clarify our understanding of these long-standing constructs in SLA.

In addition, this line of investigation can be extended to address practical questions of interest to language practitioners involving learner-external factors, such as the effectiveness of instruction and feedback. To provide the most effective instruction possible, L2 teachers should be able to recognize significant differences in their students and understand how such differences may constrain the rate at which they develop L2 proficiency over time (Oxford & Ehrman, 1993). An explicit understanding of learner IDs can not only improve teachers’ preparation to make informed decisions while teaching (Gurzynski-Weiss, 2010; Leow, 1995) but also aid in identifying learners who may require specific types of intervention (Roberts & Meyers, 2012). The interaction between learner IDs and language instruction is of central importance in classroom research and curriculum development (Robinson, 2001a, 2001b, 2002) with the ultimate goal of matching learner aptitudes to optimal conditions of instructional exposure (Robinson, 2013).

**Organization of the Study**

In the following section, I first define the dependent variables (i.e., implicit/explicit L2 knowledge) and independent variables (i.e., L2 aptitude, working memory capacity, phonological short-term memory, processing speed, L2 motivation, proficiency level) under investigation as well as concepts relevant to the research design in this study (i.e., long-term development). In Chapter two, I review relevant theoretical frameworks and detail the methodology and findings of empirical research before formulating the research questions driving the rationale for this dissertation research. Chapter three describes the proposed design and methodology which includes in-depth detail of participants, materials, and research
procedure (see Figure 3), coding and scoring procedure, and the statistical analyses and results of a pilot study conducted to ensure the validity and reliability of all measures. In Chapter four I report the results of the analyses used to address my research questions. In the final chapter, I discuss the main findings in relation to previous theoretical and empirical literature, draw conclusions based on these findings, consider implications for theory, assessment, and pedagogy, and conclude with limitations and suggested directions for future research.

**Definition of Terms**

**Second language (L2) knowledge.**

L2 knowledge is a construct related to proficiency, or a state of interlanguage development (Selinker, 1972). Proficiency can be broadly construed as an index of the comprehension and production abilities that L2 learners develop across linguistic domains (e.g., lexical competence, grammatical competence, discourse competence) and modalities (spoken and written) to communicate in a target language (e.g., Bachman, 1990, 2004; Harley, Cummins, Swain, & Allen, 1990; Thomas, 1994; Tremblay, 2011). It has also been defined based on different types of knowledge that are drawn on in learner performance (Hulstijn, 2011), which has been the focus of much theoretical and empirical research. The current study focuses on knowledge of L2 grammar as assessed in learner performance.

*Implicit L2 knowledge.*

Early definitions of "knowing things in different ways” (Bialystok, 1994, p. 549) equated implicit L2 knowledge with fluent, automatic language use (Hulstijn & Hulstijn, 1984; Seliger, 1979; Sorace, 1985). Though no universally-agreed upon definitions exist, more recent operational definitions have proposed a cumulative list of criterial features to define this construct, which is characterized as intuitive, procedural, systematic, automatic, unverbalizable,
and available in fluent language use (N. Ellis, 1994, 2005; R. Ellis, 1993, 1994, 2002, 2005, 2009b). In line with these descriptors, implicit L2 knowledge is assumed to be associated with automatic processing (Segalowitz, 2003; Segalowitz & Hulstijn, 2005) and to most effectively be measured under time pressure with a focus on meaning to increase the likelihood that learners will draw on their implicit knowledge while performing a language task.

**Explicit L2 knowledge.**

On the other hand, the ability to articulate or verbalize rules underlying the correct use of the target language has traditionally served as an indicator of explicit knowledge (e.g., Bialystok, 1982; Hulstijn & Hulstijn, 1984). However, this early criterion has since been questioned as a reliable indicator of explicit rule-based knowledge given that "explicit knowledge exists independently of whether it can be verbalized" (R. Ellis, 2004, p. 239). Furthermore, verbalizability need not entail the use of technical metalanguage (e.g., verbal tense, aspect, mood), which is a more appropriate descriptor of a higher level of conscious knowledge, or metalinguistic knowledge (Hulstijn, 2002, p. 205). Taking these factors into account, explicit knowledge has more recently been construed as conscious, declarative, highly variable, potentially verbalizable, and accessible through controlled, effortful processing and in planned language use (e.g., N. Ellis, 1994, 2005; R. Ellis, 1993, 1994, 2004, 2005, 2009b; Hulstijn, 2002, 2005). The current study assumes that explicit knowledge involves some degree of attention to formal aspects of the language and is best measured without time constraints to encourage conscious reflection and increase the chance that learners will access their knowledge in a deliberate fashion.

As this study is concerned with access to the two knowledge systems during behavioral performance, it does not define the constructs in psychological (i.e., mental representation) or
neurological terms (i.e., brain location and mechanisms). However, in line with previous research, the present study recognizes the likely interaction between knowledge systems during performance (e.g., Bialystok, 1981, 1982; N. Ellis, 1994, 2005, 2008; R. Ellis, 1994, 2005, 2009a; Hulstijn, 2002, 2005; Paradis, 2009) and as a result, the methodological challenges in clearly measuring them as separate constructs.

**Development over time.**

The question of what constitutes longitudinal data in language learning hinges on how one defines "a period of time", which have previously been vague. R. Ellis (2008b) defines controlled longitudinal studies as those that “experimentally elicit data from a single learner over a period of time” and pseudo-longitudinal studies as those that compare “groups of learners at different stages of development” (p. 588). In SLA research on instructional effectiveness, a pre-post-delayed posttest design is typically used (Norris & Ortega, 2000; Ortega & Iberri-Shea, 2005) whereas recent longitudinal studies in SLA investigating some aspect of L2 learning by college-level populations typically span anywhere between three or four months and up to six years (Ortega & Iberri-Shea, 2005, p. 37). In an academic context, time can be measured in a variety of units—weeks, months, years, semesters, sessions, etc.—and Singer and Willett (2003) advocate choosing a time metric that makes most sense for one's research questions and implements a reasonable number of data collection waves (at least 3) sufficiently spaced apart (p. 11). As the current study is structured around three assessment waves to investigate within and between-learner change during and after a semester of instruction (3.5 months), it is considered to be ‘developmental’ in nature.

**Retention of L2 knowledge.**
Retention of L2 knowledge (third data collection wave) is defined in relation to L2 grammatical performance at the first and second testing times (i.e., beginning and end of the semester). It is defined as maintenance of L2 grammatical skills after a one-month period of disuse or reduced input (Bardovi-Harlig & Stringer, 2010, p. 34), also termed “incubation” period (Gardner, 1982).

**Learner individual differences (IDs).**

The broad definition of learner IDs provided by Dörnyei (2005) is adopted here, namely, “enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree” (p. 4). This study specifically focuses on relevant cognitive (i.e., L2 aptitude, WMC, processing efficiency) and psychosocial (i.e., L2 motivation) learner differences with proven relationships to L2 learning.

**L2 aptitude.**

Following Skehan (1989, 1998, 2002, 2012) and Robinson (2001b, 2005a, 2007, 2013) a componential approach is taken to define *second or foreign language learning aptitude*. Generally, it is considered a “talent for learning languages that is independent of intelligence” (Skehan, 1989, p. 276) and more specifically, it is conceived as a composite trait including the capacity to process input (phonetic coding ability), to analyze that input and make generalizations (grammatical sensitivity; inductive language learning ability), and the capacity to learn new sound-meaning associations and retain them (rote learning ability, or memory) (Carroll, 1981, p. 105). Grammatical sensitivity and inductive language learning ability are viewed jointly as language analytic ability (Skehan, 1989, 1998). Following Miyake and Friedman (1998), Robinson (2001a), Sanz (2005), Sawyer and Ranta (2001), Wen and Skehan
(2011) and others, working memory capacity is viewed as the key component defining the main functions of language aptitude.

**Working memory capacity (WMC).**

*Working memory* is conceived here to be a mental workspace where information is consciously and simultaneously stored and processed (Baddeley & Hitch, 1974; Baddeley, 1986, 2000, 2003, 2007, 2010) and is primarily used in "complex tasks such as reasoning, comprehension and learning" (Baddeley, 2010, p. R136). It is a complex, multi-component system comprised of a *central executive*, responsible for attentional control and the active maintenance of information, and two independent storage systems, the *phonological loop*, responsible for phonologically coding and storing sounds, and the *visuospatial sketchpad*, in charge of generating and retaining visual or spatial information. The *episodic buffer* (Baddeley, 2000), also controlled by the central executive, functions as a temporary interface between the systems and long-term memory (LTM). Working memory is seen as a cognitive mechanism not only underlying online processing, or encoding, but also “retrieval of stored long-term knowledge relevant to the tasks in hand” (Baddeley & Logie, 1999, p. 31). While retrieval of information maintained in the short-term store (STS) is assumed to be accessed relatively quickly and effortlessly, retrieval from the long-term store (LTS) is generally considered a slower and more effortful process (Shah & Miyake, 1999, p. 6).

**Phonological short-term memory (PSTM).**

In terms of the phonological loop subcomponent of WM, *phonological memory* is responsible for the storage of verbally coded information (Baddeley & Hitch, 1974). Its function is fractionated into a passive phonological store and an active rehearsal process. The
phonological store represents material in a phonological code, which is subject to temporal decay, whereas the rehearsal process refreshes the decaying representations (Baddeley, 1986).

**L2 motivation.**

*L2 motivation* is conceptualized in two ways in the present study. Generally, it is viewed here as a dynamic, complex construct that fluctuates according to different internal and external factors while learning a second language. More specifically, L2 motivation is defined in terms of three main components within Gardner’s Socioeducational Model of SLA (Gardner & Lambert, 1959; Gardner & Lambert, 1972; Gardner, 1985, 2000, 2001, 2006, 2010): 1) *Integrativeness* (integrative orientation, interest in foreign languages, and attitudes towards the L2 community); 2) *Attitudes toward the learning situation* (attitudes towards the language teacher and the course); and 3) *Motivation* (effort, desire, and attitude towards learning). At the core of this conceptualization are individuals’ attitudes toward the target language community and their desire to identify with speakers of the target language, or *integrativeness*. Levels of motivation are influenced and maintained by attitudes toward the learning situation and integrativeness.

The view of L2 motivation espoused here also aligns with current notions of future self-guides promoted within Dörnyei’s L2 Motivational Self System (2005, 2009; Dörnyei & Ushioda, 2009), which is rooted in mainstream theories of cognitive psychology (Higgins, 1987, 1996; Markus & Nurius, 1986). The basic tenet of Dörnyei’s theory is that if proficiency in the target language is integral to one’s *ideal or ought-to L2 self* (i.e., attributes one would *ideally* like to possess versus attributes one believes one *ought to* possess), this aspiration will serve as a powerful motivator to learn the language because of a psychological desire to reduce the discrepancy between current and future self states (Ushioda & Dörnyei, 2012, pp. 400-401).

**L2 proficiency.**
Proficiency level of L2 learners of Spanish in this study is primarily operationalized as institutional enrollment in beginning (2<sup>nd</sup> semester-SPAN 004), intermediate (4<sup>th</sup> semester-SPAN 022) and advanced (6<sup>th</sup> semester-SPAN 104) university Spanish courses in the non-intensive program in the Department of Spanish and Portuguese at Georgetown University. The majority of students enroll either based on results of a summer or fall placement exam that has been in place for seventeen years, results from performance on Spanish foreign language sections of the SAT and advanced placement (AP) exams, or previous course enrollment. To confirm proficiency level, learners also completed a grammar and vocabulary sections of the internal departmental placement exam at the beginning of the study.

Spanish classes meet three days a week (MWF) for fifty minutes each class period for approximately fourteen weeks a semester. Estimates of total previous classroom contact hours vary based on reported previous instruction at the secondary level, but are calculated here assuming at least one prior semester of high school instruction. In line with these criteria, beginners are estimated to have 70 hours previous instruction at the initial testing time (1 prior semester high school instruction & 1 prior semester university instruction: 2.5 hrs/wk x 14 wks x 2 semesters = 70 hrs); intermediate learners are estimated to have 140 hours of classroom instruction (2.5 hrs/wk x 14 wks x 4 prior semesters = 140 hrs); and advanced learners should have received 210 hours formal instruction more or less (2.5 hrs/wk x 14 wks x 6 prior semesters = 210 hrs).
CHAPTER TWO: Review of the Literature

Implicit and Explicit L2 Knowledge

Overview

Researchers, teachers and administrators generally rely on learner behavior, or performance, to make inferences about the knowledge of the target language they have developed up to that point. Whether the goal is to determine their proficiency level in order to appropriately place them into a curricular sequence, or to evaluate their performance for a course grade, the availability of reliable and valid measures of language knowledge is paramount.

Norris and Ortega (2012) assert that the majority of SLA research concerns the assessment of learner knowledge in some way and L2 assessment tools should provide systematic and replicable techniques that allow researchers “to elicit, observe, and interpret indicators of L2 knowledge (however defined)” (p. 573). A long-standing distinction in SLA stemming from cognitive psychology is language knowledge that is implicit (i.e., knowledge of language) or explicit (i.e., knowledge about language). However, a key methodological challenge for researchers is to clearly demonstrate whether L2 assessment tools engage implicit or explicit knowledge (or both) (Norris & Ortega, p. 576). Though researchers in language testing have argued that SLA researchers tend to neglect demonstrating the validity and reliability of their testing instruments (e.g., Bachman, 1989; Bachman & Cohen, 1998; Douglas, 2001), basing tests on operational definitions of L2 knowledge is one way to move forward in this respect and in so doing, render "theories of implicit and explicit knowledge and learning testable" (Hulstijn, 2005, p. 137), including the long-standing question of how linguistic knowledge changes as proficiency increases.
In this section, I review theoretical definitions of implicit and explicit learning and knowledge and varying positions on the nature of their potential relationship. Then, I discuss criteria used to operationalize and distinguish between types of language knowledge used in performance and review findings from a strand of empirical research using these criteria to inform and refine a test battery designed to tap implicit and explicit linguistic knowledge.

**Defining implicit and explicit L2 knowledge.**

Early on in SLA implicit and explicit L2 knowledge was discussed in terms of unanalyzed versus analyzed knowledge (Bialystok, 1981) and acquired vs. learned knowledge (Krashen, 1981, 1985). Krashen claimed that acquired language knowledge, or competence, is primary and attributed a limited role to explicit knowledge, claiming that can only used as a monitor in L2 production. More recent definitions draw on dichotomies within an information-processing approach in cognitive psychology between automatic (without attention) and controlled (with attention) processing. (Shiffrin & Schneider, 1977) and a declarative/procedural distinction between controlled use of ‘chunks’ of declarative knowledge (Anderson, 1983, p. 23) and automatic, proceduralized use of knowledge, as proposed in Anderson’s Adaptive Control of Thought (ACT*).

Based on these distinctions, more recent characterizations of implicit and explicit L2 knowledge constructs highlight the respective contrast between intuitive, procedural, systematic, automatic, unverbalizable knowledge that is available in fluent, spontaneous language use and conscious, declarative, highly variable, potentially verbalizable knowledge that is accessible through controlled processing and in planned language use (e.g., N. Ellis, 1994, 2005; R. Ellis, 1993, 1994, 2002, 2005, 2009a, b; R. Ellis & Loewen, 2007; Hulstijn, 2002, 2005).
The following section reviews measurement of the constructs of L2 knowledge based on the above operational definitions.

**Measuring what learners know.**

As noted in Chapter One, the distinction between different ways of knowing language (Bialystok, 1994) has been pursued in psychological terms (i.e., mental representation), neurobiological terms (i.e., brain location and mechanisms), and behavioral terms (i.e., how knowledge is accessed) (Paradis, 2009), the latter of which is of most relevance to the present study. Early studies that examined learners’ implicit and explicit knowledge (e.g., Hulstijn & Hulstijn, 1984; Sorace, 1985) tended to determine the former in terms of the learners’ use of specific linguistic features in oral language and the latter as learners’ explanation of rules underlying the use of such features. Recent research has expanded on these indices by identifying seven criterial features underlying knowledge types (R. Ellis, 1994, 2004, 2005, 2009b; Han & R. Ellis, 1998): 1) **Degree of awareness** (i.e., the extent to which learners demonstrate the ability to analyze why structures are grammatical or ungrammatical); 2) **Time available** (i.e., whether learners are given a time limit to perform a task or not); 3) **Focus of attention** (i.e., whether the task prioritizes meaning over form and/or fluency over accuracy); 4) **Systematicity** (i.e., whether learners are consistent or more variable in their response to a task); 5) **Certainty** (i.e., how certain learners are that the linguistic forms they have produced conform to target language norms); 6) **Metalanguage** (i.e., learners’ knowledge of metalingual terms will be related to their explicit (analyzed) knowledge, but not to their implicit knowledge); and 7) **Learnability** (i.e., learners who acquired the L2 as a child are more likely to display high levels of implicit knowledge, while those who were first exposed as adolescents or adults, especially if
reliant on instruction, are more likely to display high levels of explicit knowledge) (R. Ellis, 2009b, pp. 38-9).

While these criteria are certainly not without controversy, they provide a step in the right direction as an example of what theory-based, replicable assessments might look like (Norris & Ortega, 2012, p. 583) and as a way to improve how we know what our learners know based on behavioral evidence (Lakshmanan & Selinker, 2001). Indeed, several studies have begun to evaluate the operationalization of implicit/explicit knowledge based on these criteria in a test battery developed for L2 learners of English (e.g., R. Ellis, 2004, 2005, 2009b; Elder & Ellis, 2009; Han & Ellis, 1998; Philp, 2009) and adapted for heritage and L2 learners of Spanish (Bowles, 2011; Gutiérrez, 2012).

**Empirical research aiming to improve the measurement of L2 knowledge.**

Han and Ellis (1998) originally developed a test battery that corresponded to a categorization of learner assessment tests based on controlled/free use of language measures and comprehension/production measures (Doughty, 2003) as shown in Table 1 below.

Table 1

*Five tests measuring implicit/explicit L2 knowledge*

<table>
<thead>
<tr>
<th>Test</th>
<th>Type of measure</th>
<th>Type of knowledge measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Elicited Oral Imitation Test (OIT)</td>
<td>Constrained, constructed response-production</td>
<td>Implicit knowledge</td>
</tr>
<tr>
<td>2. Oral Narrative (ON) Test</td>
<td>Free production</td>
<td>Implicit knowledge</td>
</tr>
<tr>
<td>3. Timed Grammaticality Judgment Test (TGJT)</td>
<td>Metalinguistic judgment</td>
<td>Implicit knowledge</td>
</tr>
<tr>
<td>4. Untimed Grammaticality</td>
<td>Metalinguistic judgment</td>
<td>Explicit knowledge</td>
</tr>
</tbody>
</table>
Forty-eight L2 English learners completed the test battery that targeted a range of seventeen grammatical structures in English as well as two general measures of English proficiency (the TOEFL and the Secondary Level English Proficiency test-SLEP). Han and Ellis reported significant relationships between measures of implicit and explicit L2 knowledge and both proficiency measures, with the exception of the metalinguistic knowledge test, which led them to conclude that the ability to use metalanguage is unrelated to L2 proficiency.

A subsequent series of studies assessed the construct and content validity of the above test battery with adult L2 English learners at varying proficiency levels (e.g., R. Ellis, 2004, 2005, 2009b; Elder & R. Ellis, 2009; Han & R. Ellis, 1998; Philp, 2009). This research has demonstrated high reliability of all tests for these learner populations, with Cronbach’s alpha coefficients and inter-rater agreement exceeding 0.80. Furthermore, it has consistently been found that the three tests measuring implicit knowledge (OIT, ON & TGJT) and the two measuring explicit knowledge (UGJT & MK test) load on separate factors in exploratory and confirmatory principal component factor analyses (see Isemonger, 2007 and R. Ellis & Loewen, 2007 for a comparison of these procedures). However, R. Ellis (2009b) reported that performance on the ON test and the Timed GJT (both implicit) had the lowest loadings on Factor 1 (.57 and .68 respectively) compared to the OIT which had the strongest loading (.87) (p. 51). In terms of tests of explicit knowledge, judgment scores of ungrammatical sentences on the untimed GJT loaded the strongest on Factor 2 (.91) compared to the MK test scores (.73). These results led R. Ellis (2005, 2009b) and Philp (2009) to conclude that the elicited OIT seems to afford the
“best” measure of implicit knowledge while performance on the ungrammatical sentences in the UGJT most effectively taps explicit knowledge.

Thus far, Bowles (2011) and Gutiérrez (2012) are the only studies to investigate the validity of R. Ellis and colleagues’ test battery for a target language other than L2 English. Following Ellis’s methodology, Bowles adapted the tests for L2 Spanish by selecting both early and late-acquired forms introduced at a range of instructional levels that are known to be problematic in L2 Spanish, including nominal, morphological, and verbal inflection as well as tense-aspect and mood morphology ($N = 17$ target structures). Bowles was interested in whether or not Spanish HL and L2 learners’ test scores differed according to their previous language experience and hypothesized that HL learners would score higher on implicit tests given their informal learning background and L2 learners would score higher on explicit tests given their formal classroom experience. Twenty HL and L2 Spanish learners at an intermediate proficiency level based on university course enrollment and ten Spanish native speakers completed the test battery. Classroom L2 Spanish learners were exposed to Spanish after puberty and had received an average 6 years of instruction whereas Spanish HL learners had learned Spanish naturalistically since birth with less than 2 years of formal instruction.

Reliability coefficients well exceeded 0.80 for each of the five tests (Cronbach’s alpha and interrater reliability estimates). One-way analyses of variance (ANOVAs) revealed that, as expected, HL learners significantly outperformed L2 learners on all tests designed to tap implicit knowledge and L2 learners significantly outperformed HL learners on the MK test. Whereas L2 learners scored highest on the MK and the untimed GJT, HL learners scored highest on the oral narration task. A confirmatory factor analysis replicated R. Ellis’s two-factor model, providing
support for the construct validity of the test battery and evidence that the tests adapted for this language and these target populations “indeed tap separate pools of knowledge” (p. 261).

Gutiérrez (2012) used three measures (timed and untimed GJT's and MKT) from the test battery adapted by Bowles with fifty-three L2 Spanish learners at two proficiency levels. The low proficiency group was in their third semester of university Spanish instruction and the high proficiency group was in their fifth semester. Gutiérrez further investigated test validity in terms of the grammaticality of the test items based on the hypothesis that learners are more likely to draw on their implicit knowledge for judging grammatical sentences and their explicit knowledge for judging ungrammatical sentences (R. Ellis, 2004, pp. 258-59). Based on factor loading in principal components and confirmatory factor analyses, he concluded that the grammatical sentences in the timed and untimed GJT's measured implicit knowledge and the ungrammatical sentences in both GJT's and the MKT measured explicit knowledge (p. 29).

Studies like those conducted by R. Ellis and colleagues for L2 English and by Bowles and Gutiérrez for L2 Spanish bring the field a step closer to having a set of reliable, valid measures to assess implicit and explicit learner knowledge. However, in R. Ellis and Gutiérrez's studies, factor analyses were conducted with performance scores across proficiency groups. Given that learners at different proficiency levels are at different stages of interlanguage development, the construct validity of the test battery should be further investigated to see if a two-factor model is replicated according to proficiency group. Furthermore, there is still room for methodological improvement on the 'best' measures of L2 knowledge in order to increase their internal validity, or the likelihood they are tapping the intended knowledge constructs.

Further refining how we measure L2 knowledge.
Given that learners' "interlanguage contains both types of knowledge" (Bowles, 2011, p. 248) and that learners may access both sources of knowledge during L2 performance, internal validity of the test battery is a key issue and challenge. For example, in a spontaneous oral production task like the oral narration task, learners who have not ‘automatized’ their skills are likely to use explicit knowledge as a tool to monitor and mediate their speech production. Likewise, any type of GJT, timed or not, could possibly invite conscious metalinguistic reflection given the nature of ‘judging’ sentences (Isemonger, 2007). The explicit test of metalinguistic knowledge can be critiqued as well based on its four option multiple-choice (MC) format asking learners to choose which rule best describes the ungrammaticality of a provided sentence. DeKeyser (1990) has argued that MC tests are not good measures of conscious knowledge based on the fact that they tend to either be ‘too easy’ (suggesting answers and eliminating choices the learner would have to make otherwise) or ‘too difficult’ (answer alternatives might cause unnecessary doubts and errors based on the learner’s developing knowledge) (pp. 149-150). Another issue to consider is what this type of test is actually measuring: knowledge of correct linguistic terminology or knowledge of underlying rules in the target language (Han & Ellis, 1998).

These critiques seem to be supported in the weaker loadings of performance on the ON and TJGT on Factor 1 (compared to the OIT) and the weaker loading of the MK test on Factor 2 (compared to the ungrammatical sentences on the UGJT). Thus, it does seem that the elicited OIT and the UGJT afford the most convincing measures of implicit and explicit knowledge, respectively. This argument is particularly effective in the case of the elicited OIT, which has received further empirical support as a valid test of implicit knowledge in other studies (e.g., Erlam, 2006; Philp, 2009; Tomita, Suzuki, & Jessop, 2009) in addition to being a reliable, valid
measure of L2 oral proficiency (Bowden, 2012; Ortega, 2000; Ortega, Iwashita, Norris & Rabie, 2002).

In the version of the OIT used by Ortega and colleagues, the learner’s task is to repeat grammatical sentences \( (n=30) \) increasing in syllable length (7-17 syllables) and scoring is more content-based (0-4 points possible). This version has been demonstrated to have high internal consistency \( (\alpha > 0.95) \) and been shown to strongly correlate with a standardized measure of oral proficiency (Simulated Oral Proficiency Interview—SOPI) \( (r = 0.87 \text{ to } 0.91) \) in L2 Spanish learners at a range of proficiency levels. Learners were at the beginning and end of their fifth semester of classroom instruction (Ortega, 2000) or in their fourth and sixth semester (Ortega et al.). Learners in Bowden's study (2012) were ‘low experience’ (2nd-4th semester), advanced (5th-7th semester), and very advanced learners with extensive immersion experience (> 19 months).

In contrast, in the version employed by R. Ellis, Bowles, Gutiérrez, and others, learners must make a meaning judgment about the content of sentence exemplars and then repeat sentences that are either grammatical or ungrammatical based on seventeen targeted grammatical structures \( (N = 34) \) in L2 Spanish and scoring is exclusively based on form (0-1 points possible) (see Appendix N for a detailed comparison of the methodological differences between both versions of the OI task). The current study is the first to implement both scoring procedures to assess their relative effectiveness (see the Scoring and Coding Procedure section in Chapter 3 for further detail) but follows R. Ellis and Bowles’ test format in eliciting a meaning judgment and including both grammatical and ungrammatical exemplars in order to focus on targeted linguistic knowledge as opposed to meaning comprehension of increasingly longer sentences.

While the OIT has been used in many studies, some researchers have questioned the test's validity for instructed L2 learners. In a validation study of the elicited imitation task, Shiu and
Spada (2012) administered a retrospective questionnaire to adult intermediate L2 English learners to ask them what they were thinking about during task completion. They found that the majority of students were exclusively thinking about form or form and meaning while completing the task. Thus, Spada (2013) claims that time pressure and a focus on meaning do not guarantee that learners will access implicit knowledge, and that the learners in her study appear to have been able to successfully complete the task by "rapidly accessing their explicit knowledge" (p. 9). One way to improve the interpretation that learners access their knowledge more or less automatically on the OIT would be to collect oral latency data given that time and automaticity are thought to play a role in distinguishing between use of implicit and explicit knowledge. However, no previous study employing the OIT has reported collecting such data.

The “best” measure of explicit knowledge, the untimed GJT, can also be improved. Many researchers have recommended using source attributions (feel, guess, rule) or certainty indices (e.g. R. Ellis, 2004; Gass, 1994; Mandell, 1999; Sorace, 1985, 1996) in order to increase confidence in the interpretation of judgments of non-native speakers given that their grammar is highly variable (Gass & Mackey, 2007). Isemonger (2007) has also pointed out that the criterion of verbalizability has not been sufficiently probed in R. Ellis and colleagues' studies, arguing that they "do not fully circumscribe the theoretical claim that explicit knowledge is potentially verbalizable" (p. 112). Thus, learners should also be asked to correct the ungrammatical element in the sentence or provide a rule explaining its use either in written format for those items for which learners report using a rule or verbally by eliciting online verbal reports while learners are engaged in making their grammaticality judgments (Bowles, 2010; Leow, Grey, Marijuan & Moorman, submitted).
All methodological modifications to the adapted versions of the OIT and the UGJT used here are further elaborated in the assessment materials section in Chapter Three.

**Implicit and explicit L2 knowledge: Some issues.**

The valid measurement of learner knowledge is important because it provides a way to clarify certain theoretical issues that have prompted much debate such as the relationship between implicit and explicit learning processes, knowledge, and instruction as well as how linguistic knowledge changes as proficiency increases (e.g., Bialystok, 1981, 1991, 1994; DeKeyser, 1998, 2003; N., Ellis, 1994, 2005, 2008; R. Ellis, 1993, 1994, 2005, 2009b; Hulstijn, 2002, 2005; Schmidt, 1990, 1994, 2001; Schwartz, 1993; Sharwood Smith, 1981). The latter inquiry is most relevant to the current study. SLA researchers generally take two stances regarding whether explicit knowledge can proceduralize, or become implicit, through certain learning mechanisms: 1) the noninterface position, rejecting the conversion of one type of knowledge into another through learning (Hulstijn, 2002, 2005; Krashen, 1981, 1982, 1985; Paradis, 2009; Schwartz, 1993), and 2) the interface position, with a strong version arguing that explicit knowledge can be automatized through practice (DeKeyser, 1997, 2003, 2007; Sharwood Smith, 1981) and a weak version acknowledging interaction between the two knowledge types with limitations on when or how explicit knowledge can become implicit through practice (N. Ellis, 1994, 2005, 2008; R. Ellis, 1993, 1994, 2005, 2009a).

Other positions in the interface debate do not posit a direct relationship but rather, as proficiency increases, use of implicit knowledge may substitute explicit knowledge over time. This is a central view espoused in Ullman’s (2001, 2005) declarative/procedural model that claims that at lower levels of L2 experience, declarative memory underlies the learning and use of lexical knowledge and complex linguistic representations whereas at higher levels of L2
experience, the procedural system is gradually involved (Ullman, 2005, p. 153). This view, which argues for a gradual substitution in memory systems, along with the strong version of the interface position, which argues for the possibility of automatization of explicit knowledge, are of most relevance to the current study, which is concerned with the use of implicit and explicit L2 knowledge at increasing levels of proficiency.

Both views reflect the development of automaticity and 'cognitive' fluency as conceived by Segalowitz and colleagues (e.g., Segalowitz & Hulstijn, 2005; Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998; Segalowitz, 2010). The study of automaticity is concerned with how efficiently learners access the knowledge they possess, which is a function of the degree to which the underlying processes are executed in an automatic as opposed to a controlled fashion (Segalowitz et al., 1998, p. 54). At higher levels of classroom experience in the L2, learners are presumed to have engaged in more classroom practice using the L2 and therefore, Segalowitz and colleagues postulate that they undergo a qualitative restructuring of the system underlying the relevant skill, rendering it more efficient (and not just faster) overall.

Segalowitz and Segalowitz (1993) proposed reaction time and reaction time variability as an effective means to indexing the development of automaticity in learners; specifically, they calculate a coefficient of variation (CV_{RT}) to distinguish between speed-up (improvement without increased automaticity) and restructuring (improvement with increased automaticity). However, as previously mentioned, this claim has mainly been investigated in the use of the lexicon as opposed to use of grammar. Studies investigating lexical processing and performance have generally provided support for this index of automaticity (e.g., Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz & Wood, 1998; though see Hulstijn et al. (2009) for an exception) while comparatively little research has looked at the role of automaticity in L2 morphosyntax.
(DeKeyser, 1997) despite the fact that non-target-like verbal inflection persists even at advanced stages of adult L2 acquisition (Montrul, 2004). Two studies by Hopp (2010) and Rodgers (2011) provide some recent evidence of automatization in this realm.

Hopp (2010) investigated L2 inflectional variability in highly proficient L2 learners. He conducted four experiments that tested knowledge and processing of morphosyntactic inflection (case marking in word order and subject-verb agreement) in fifty-nine highly advanced late L2 German learners with different native languages (English, Russian, Dutch). Based on their performance in off-line speeded and unspeeded grammaticality judgment tasks and an on-line self-paced reading task compared to performance by German native speaker control groups, Hopp reported that native-like attainment of L2 grammatical inflection is possible for advanced adult learners. Nonetheless, longer reading times in the online task and slower reaction times under timed judgment conditions led Hopp to conclude that the L2 system was computationally less efficient than the system underlying L1 processing and knowledge (p. 918) and that L1 influence can affect processing efficiency. However, claims related to automaticity in this study are based on response speed, rather than a decrease in variability (CV).

Rodgers (2011), on the other hand, used the CV formula proposed by Segalowitz and colleagues to investigate automatization of verbal morphology in a study with eighty-five instructed learners of Italian at increasing proficiency levels. Based on the results of a 21-item cloze test with high reliability, learners were divided into beginning (n = 30), intermediate (n = 34), and advanced (n = 21) proficiency groups; however, previous language learning experience was not experimentally controlled given that approximately 75% of the participant sample "knew a foreign language other than Italian (typically Spanish)" (p. 304). Results statistically supported increasing automatization with higher proficiency but this finding differed based on
comprehension and production demands as measured in a picture identification and description task; namely, while increasingly faster and more accurate responses were found in both tasks, a significant decrease in CV values and significant RT-CV correlations were found only for receptive skills at higher proficiency. Findings in Hopp's study are also limited to comprehension of L2 grammar. Overall, more research is needed to clarify the role of automaticity in the use of linguistic knowledge in L2 production and with learners who have limited exposure to other foreign language in the same family as the target language under investigation.

In sum, indices of automaticity such as that originally proposed by Segalowitz and Segalowitz (1993) may be one way to clarify the issue of what type of grammatical knowledge learners access in performance, especially on the oral imitation task as suggested previously. For example, it is possible that more advanced learners have automatized their explicit knowledge of the target language which creates the methodological challenge of distinguishing use of implicit knowledge from use of automatized explicit knowledge solely based on reaction time speed. RT variability is one possible solution in quantitatively distinguishing between speed up, or faster RTs, and a more efficiently organized system, as decreasing CVs would indicate. While faster RTs could be interpreted as use of automatized explicit knowledge, a reduced CV may be a more promising indicator of implicit knowledge, or restructuring of underlying skill components. Furthermore, at lower levels of proficiency learners' language skills are less automatized and as such, they are more likely to access explicit knowledge during performance under testing conditions designed to tap implicit knowledge (Paradis, 2009, p. 63). Therefore, collecting RT data could also be methodologically useful in establishing the degree to which low-proficiency learners access their explicit knowledge.

Summary
As many researchers have pointed out, measuring L2 knowledge is complex and methodologically challenging (Norris & Ortega, 2003; 2009). However, several theoretical debates hinge on its valid measurement such as whether classroom learners possess knowledge of and/or about their second language and how linguistic knowledge changes as proficiency in the language increases. Therefore, developing systematic and replicable assessment tools with proven validity and reliability with different learner populations and target languages is a valuable research endeavor. The implicit/explicit test battery developed by R. Ellis and colleagues for L2 English and adapted for L2 Spanish by Bowles is a useful step in the right direction though the interpretation of results can be strengthened by further replication and methodological improvement of assessment measures, the collection of oral latency data in particular.

The current study expands this line of investigation by not only measuring knowledge of L2 Spanish grammar in performance but also by seeking to explain variation in rate of development based on learner individual differences, a core goal driving SLA theory and research (N.C. Ellis, 2008; R. Ellis, 2005, 2009; Norris & Ortega, 2012).

Learner Individual Differences

Overview

Certain SLA theories take a modular view of language (Chomsky, 1976; Fodor, 1983) like generative linguistic approaches (e.g., Schwartz, 1993; White, 1989) which tend to assume similarities between individuals. Other approaches like the cognitive-interactionist approach to adult L2 learning focus on learner-internal differences (e.g., domain-general mechanisms like memory) and their interaction with external factors (e.g., type of instruction) as a way to explain variable L2 development (Norris & Ortega, 2012; Robinson, 2001a, b, 2007; Sanz, 2005). This
parallels an interactional perspective within cognitive psychology that views human behavior as a "continuous interaction between situation and person" (Snow, 1989, p. 45).

In light of the theoretical and practical need to explain the high variability associated with adult L2 learning outcomes, many scholars in SLA have stressed that no account of adult SLA can be complete without taking key learner IDs into account (e.g., Skehan, 1989, 1991, 2002; Roberts & Meyer, 2012). Roberts and Meyer (2012) have argued that a comprehensive psychological theory of a skill or type of behavior "should not only make predictions about average behavior but also about individual differences" (p. 2). Similarly, Skehan (1989) has argued that ID research is useful for establishing "how big the jigsaw puzzle is" as well as "how much it is likely to contain" (p. 145) in the pursuit of identifying and eventually solving 'problems' in SLA (Long, 2007).

Therefore, studying individual learner variation in L2 learning is a productive way to test hypotheses and refine SLA theory, which is rooted in a variety of disciplines such as linguistics, cognitive psychology, anthropology, sociology, and education. This interdisciplinary origin is reflected in the wide range of perspectives motivating L2 research (see SLA handbooks, e.g., Doughty & Long, 2003; Gass & Mackey, 2012). For example, research driven by a cognitive information processing perspective of L2 learning (e.g., McLaughlin, 1987) equates learning second languages as an adult with learning other complex cognitive skills (DeKeyser, 2007) and tends to emphasize the importance of domain-general cognitive mechanisms in explaining the development of L2 proficiency. On the other hand, researchers within a socio-psychological perspective emphasize that learning additional languages as an adult is different from learning other skills given that it is contextually constrained by certain social, psychological and cultural complexities (Dörnyei, 2003, 2005, 2009; Gardner, 1985, 2000, 2010). They stress the
fundamental importance of motivation or learners' reasons for learning the target language and the extent of their identification with the native-speaking community.

Overall, the explanatory power of SLA ultimately depends on embracing both views in order to gain a comprehensive understanding of what learners bring cognitively and psychologically to the task of learning another language as an adult. Below I provide a historical perspective of research on learner differences, with a focus on the cognitive and psychosocial (or socio-psychological) dimensions of L2 learning.

**Historical perspective of learner ID research.**

Age is included as a key variable that has been pervasive throughout SLA theory construction and empirical research since the introduction of the critical period hypothesis (CPH) (Penfield & Roberts, 1959; Lenneberg, 1967). However, while it consistently differentiates between the rate and ultimate attainment of L1 and L2 acquisition in children and adults, biological age is not a trait or ability, in contrast to other learner difference variables listed. Table 2 below provides a summary of the learner IDs that have most frequently been investigated over the last forty years of SLA research.

Table 2

*Historical overview of individual differences (IDs) in L2 research*

<table>
<thead>
<tr>
<th>ID</th>
<th>Pre-80’s</th>
<th>80’s</th>
<th>90’s</th>
<th>2000’s</th>
<th>ID Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Age</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1, 3, 4, 5, 6</td>
</tr>
<tr>
<td>2 Intelligence</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1, 2, 5, 8</td>
</tr>
<tr>
<td>3 Language aptitude</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1-5, 7, 8</td>
</tr>
<tr>
<td>4 Motivation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1-5, 7, 8</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>Learner strategies and/or beliefs</td>
<td>Personality</td>
<td>Cognitive/learning style</td>
<td>Working memory</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>----------------------------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>1-4, 7, 8</td>
<td>1-4, 7, 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4, 5, 6, 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7, 8</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


The study of L2 learner IDs originally gained momentum in the late 70’s and early 80’s with research aimed at constructing ‘good language learner’ profiles (e.g., Naiman, Fröhlick, Stern, & Todesco, 1978; Reiss, 1985; Rubin, 1975). Such profiles were based on identified strategies that led to successful learning outcomes. Good learners were generally those who i) made good guesses in the face of uncertainty; ii) made an effort to communicate and to learn through communication; iii) practiced the language whenever possible; iv) monitored their speech and that of others, and v) attended to form (i.e., grammar) and meaning (Rubin, 1975). However, this research strand declined in the 90’s based on unclear, or indirect, relationships between learner strategies and L2 proficiency as well as a limiting prescriptive approach to constructing a single profile of a “good language learner” (Griffiths, 2008). In the late 80’s and early 90’s the focus of empirical ID research shifted to the possibility of different routes to language learning success.
Two key ID variables that have received much theoretical and empirical attention are language *aptitude*, a cognitive, stable ability, and *motivation*, an affective, fluctuating construct. Both constructs have been the subject of five decades of investigation in L2 research since 1959 when Carroll and Sapon and Gardner and Lambert first published their seminal works. Since being reported to independently contribute to L2 achievement early on (Gardner & Lambert, 1972), they have consistently been identified as significant predictors of success in L2 learning. Below I elaborate on cognitive and psychosocial perspectives of L2 learning, the conceptual evolution of the constructs of aptitude and motivation within these frameworks, and subsequently detail empirical research investigating their role in L2 development.

**Cognitive dimension of L2 learning.**

A central question within a cognitive perspective of adult SLA concerns the underlying mechanisms that can explain the high variation in adult L2 learning outcomes. One reason posited for the highly variable nature of L2 developments in adults is that they are 'limited-capacity processors' (McLaughlin, 1987). That is, adults are cognitively mature language learners, in contrast to children, and the efficiency with which they acquire additional languages is likely constrained by domain-general cognitive mechanisms. To study these mechanisms more effectively, many theoretical perspectives of SLA, cognitive processing models in particular, emphasize three main stages in the learning process (Leow, forthcoming; McLaughlin, 1987; McLaughlin, Rossman & McLeod, 1983; Schmidt, 1990, 2001; Tomlin & Villa, 1994; VanPatten, 1996, 2004). These include i) *encoding*, or the input processing stage, ii) *storage*, or the incorporation of intake into the internal system, and iii) *output*, or retrieval from the developing system. The three basic phases of L2 learning are seen below.
Adult learners have limited attentional resources and this limited capacity constrains their ability to simultaneously attend to form (grammatical or linguistic form) and/or meaning (informational content of the input) in the input (Stage 1) (VanPatten, 1996, 2004). Noticing is a key construct posited to underlie processing aspects of form at this stage (Schmidt, 1990, 1994, 2001). Schmidt maintains that aspects of form must first be consciously attended to in the input before being processed as intake, or “that part of the input that the learner notices” (1990, p. 139) which is then available for restructuring and accommodation into the developing system. However, while language information that goes ‘in’ to the developing system is determined by noticing, it is not necessarily sufficient for further language development (p. 141).

Attention and noticing are regulated by working memory function, which is most commonly conceptualized as a cognitive workspace used for sequential processes involving the “temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities” (Baddeley, 2003, p. 189). As such, it has been posited to play a role in the comprehension and production of language, as these skills require the linear processing of sequences of symbols over time (Miyake & Friedman, 1998). While it is posited to underlie the cognitive processes of attention and noticing at play in the initial phases of language learning, it also plays a major role in access to the learner’s developing system (Baddeley & Logie, 1999; Skehan, 1998, 2002). Namely, Skehan posits an important role for WM not only in
the acquisition of new information (involving attention and noticing of input) but also in its retrieval during processing and output (1998, p. 203).

Another cognitive mechanism thought to play a role in the latter stages is monitoring. Monitoring involves controlled processing (i.e., attentional control) and is also postulated to be regulated by working memory capacity (Kormos, 1999). In Levelt’s L1 speech production model (1989, 1992; adapted for bilingual speakers in de Bot, 1992, 1996), noticing or detection of errors in the output relies on monitoring which is primarily based on successfully parsing one’s own inner and overt speech. Even if L2 speakers possess the underlying linguistic knowledge (e.g., know the particular linguistic rule to be applied under certain conditions), use of their interlanguage in oral or written production is variable (Tarone, 1983) and it is not guaranteed that they will detect, or notice, the errors that they make in the output. Therefore, working memory capacity could influence the extent to which learners not only successfully retrieve their knowledge, but also perceive gaps in their L2 knowledge.

These cognitive mechanisms are central to theories of adult L2 learning that equate learning language with learning other complex skills (i.e., skill acquisition theory-DeKeyser, 2007; McLaughlin, 1987). Under this view language learning involves the gradual automatization of subskills through practice and constant restructuring of internal representations with increasing degrees of proficiency. The process of automatization happens when controlled processes become available after information moves from short-term memory (STM) to long-term memory (LTM). Controlled processes, which are effortful, attention demanding and capacity-limited, are gradually replaced by automatic processes that do not depend on attentional or processing capacity (Shiffrin & Schneider, 1977, p. 155). Restructuring then proceeds by integrating hierarchical skills through practice. As such, retrieval of L2 knowledge becomes
faster (speed-up) at increasing levels of experience given that L2 learners at higher levels of proficiency have had more exposure to and practice using the language than learners at a novice skill level.

However, Segalowitz and colleagues (e.g., Segalowitz, 2003; 2010; Segalowitz & Hulstijn, 2005; Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998) have made a convincing case that fast processing is not necessarily the best indicator of the extent to which a learner's developing system has undergone automatization and restructuring. Rather, the efficiency with which a skill is carried out is likely a more useful signal of underlying processes that can be quantified in terms of a coefficient of variation (i.e., standard deviation/reaction time). In terms of the accumulation of L2 knowledge, "skill acquisition, manifested by increased processing efficiency, goes hand in hand with knowledge acquisition, manifested by increased response accuracy" (Hulstijn, Van Gelderen & Schoonen, 2009, p. 579). Thus, one can typically expect automatic behaviors to go hand in hand with accurate performance (DeKeyser, 2001) and the construct of automaticity in SLA is not only a reflection of how quickly but how efficiently and accurately L2 learners access their knowledge of the language.

In the preceding section, I briefly reviewed some key cognitive mechanisms posited to underlie different stages of L2 learning and proficiency development. In the following section, I elaborate on the traditional concept of L2 aptitude along with empirical research into its proposed auditory, linguistic, and memory subcomponents. Then, I review theoretical and empirical research investigating working memory capacity. In line with several SLA researchers (e.g., Miyake & Friedman, 1998; Robinson, 2001a, 2002; Sanz, 2005; Sawyer & Ranta, 2001; Skehan, 1998, 2002; Wen & Skehan, 2011), it is viewed here as the key component defining L2
aptitude, or the “arena in which the effects of other components of aptitude are integrated” (Sawyer & Ranta, 2001, p. 342).

**Defining and measuring L2 aptitude.**

Foreign language aptitude is popularly referred to as a ‘flair’ or ‘knack’ for learning languages (Dörnyei, 2005, p. 33). Within SLA, its conceptualization stems from attempts in the mid-20th century to relate the specific ability to learn a foreign language with that of general intelligence (Sternberg, 2002), commonly referred to as Spearman’s $g$ (1904), a single higher-order factor that describes a general ability underlying broad areas of performance. Intelligence tests, the earliest being the Binet-Simon Intelligence Scale (1905), were developed with the goal of identifying students with limited mental ability who would benefit from extra instruction outside the classroom.

Similarly, the earliest tests of second language aptitude in the 1920’s and 30’s stemmed from the need to address widespread classroom language learning failure (Spolsky, 1995) with the goal of predicting the likelihood of learners’ success in a language program. Later aptitude tests developed by John Carroll and Stanley Sapon, the *Modern Language Aptitude Test* (MLAT; Carroll & Sapon, 1959) and Paul Pimsleur, the *Pimsleur Language Aptitude Battery* (PLAB; Pimsleur, 1966), also intended to provide a prediction of the rate, or speed, of learning and improved the operationalization of foreign language aptitude by factoring out the components of the construct; nonetheless, second language aptitude testing in both the earlier and later periods were guided by assessment-driven approaches and lacked a strong theoretical foundation (Spolsky, 1995).

The original subcomponents of L2 aptitude were derived from a series of factor analyses conducted by Carroll and Sapon (1959) during the course of a five-year study conducted with
college students. A series of verbal tests were developed and administered to a large sample of approximately five thousand participants. The results produced four relatively uncorrelated subtests from which the components of L2 aptitude were derived, one dealing with the auditory aspects of language learning (*phonemic coding ability*), two with the processing of linguistic material (*grammatical sensitivity* and *inductive language-learning ability*), and one with memory (*associative memory*). Carroll (1981) offers a functional description of these subcomponents, with the outdated stimulus-response concept of memory revised as *rote learning ability* (p. 105):

1) *Phonemic coding ability*—an ability to identify distinct sounds, to form associations between those sounds and symbols representing them, and to retain those associations;

2) *Grammatical sensitivity*—the ability to recognize the grammatical functions of words (or other linguistic entities) in sentence structures;

3) *Rote learning ability for foreign language materials*—the ability to learn associations between sounds and meanings rapidly and efficiently, and to retain those associations; and

4) *Inductive learning ability*—the ability to infer or induce the rules governing a set of language materials, given samples of language materials that permit such inferences.

The first three components are directly measured in five subtests composing the MLAT: I-Number Learning (NL); II-Phonetic Script (PS); III-Spelling Clues (SC); IV-Words in Sentences (WS); and V-Paired Associates (PA). However, no separate measure of *inductive language-learning ability* is included in the MLAT, as it is thought to be closely related to *grammatical sensitivity* (measured by the WS subtest). Skehan (1989, 1998) later proposed to collapse the two abilities into *language analytic ability*.

In a similar vein, Pimsleur (1966) conceptualized aptitude based on three factors:

5) *Verbal intelligence*—the knowledge of words and the ability to reason analytically in using verbal materials
6) **Motivation**—the degree of interest in studying a modern foreign language

7) **Auditory ability**—the ability to receive and process information through the ear.

The PLAB, developed for a younger, school-age learner population, operationalizes these three factors based on a six-part test: I-Grade Point Average (GPA), II-Interest in Foreign Language Learning, III-Vocabulary, IV-Language Analysis, V-Sound Discrimination, and VI-Sound-Symbol Association. One advantage of the PLAB is that it specifically targets ‘inductive learning ability’ (Carroll, 1965) in part IV, a component not specifically addressed in the MLAT.

Other aptitude batteries have been developed for different learner populations (e.g., Defense Language Aptitude Battery—Petersen & Al-Haik, 1976) but no test battery has demonstrated superiority over the MLAT in terms of predicting language-learning success (Sawyer & Ranta, 2001; Skehan, 2002, 2012; Sparks & Ganschow, 2001). Therefore, it is not surprising that the MLAT continues to be the dominant measurement of foreign language aptitude in current research. However, the conceptualization of the construct has been refined since Carroll & Sapon’s original conceptualization (Skehan, 1989, 1998, 2002, 2012; Sparks & Ganschow, 1991, 2001, 2002) with the memory component receiving “far and away the most research attention over the last thirty years” (Skehan, 2002, p. 75).

**Empirical research on L2 aptitude.**

Overall scores on the MLAT typically correlate highly with instructed language learning success in a range of institutional settings (Carroll, 1981; Dörnyei & Skehan, 2003; Sawyer & Ranta, 2001; Skehan, 1989, 1998, 2002), especially for learners in the early stages of development. Correlations are typically found with L2 achievement measures such as final course grades or ‘end-of-course’ performance (Carroll, 1981), instructor proficiency ratings of speaking and reading skills (Ehrman & Oxford, 1995), or objective language proficiency
measures (Harley & Hart, 1997). However, research into the role of foreign language aptitude has not been consistently productive over the years, in part due to an early association with formal learning contexts only during a period in SLA research when ‘acquisition’ was the ideal goal (Cook, 1986; Krashen, 1981). Nonetheless, researchers following a componential approach to defining L2 aptitude continue to investigate its core abilities and assume that an optimal aptitude test should not be limited to distinctions between competence and performance, implicit “acquisition” and explicit “learning,” or knowledge of language and ability for use, but rather, "should predict development of all of these" (Robinson, 2005, p. 58). In this section, I review several core issues discussed in the L2 aptitude literature including the nature of its relationship with intelligence and native language skills as well as its subcomponents and their role in different learning settings.

An early and somewhat persistent issue in L2 aptitude research is the extent of the relationship between intelligence and the ability to learn foreign languages. While the components underlying general intelligence and foreign language aptitude as measured by various intelligence batteries and the MLAT have been proven to overlap to some degree (Carroll & Sapon, 1959; Gardner, 1985; Gardner & Lambert, 1972; Skehan, 1982, 1986, 1989), they do not coincide completely (Robinson, 2002; Sasaki, 1996). Using factor analyses, Sasaki (1996) revealed that the two constructs are moderately correlated based on a second-order factor analysis but a first-order analysis revealed them to be separable constructs. In particular, language analytic ability (capacity to analyze input and make generalizations) was most closely related to intelligence, while phonemic coding ability (capacity to code and retain unfamiliar sounds) and rote learning ability (capacity to learn and retain associations between sounds and meanings rapidly and efficiently) still showed separation. Overall, researchers tend to agree that
aptitude is related to, yet distinguishable from, general cognitive constructs such as intelligence (Carroll, 1993).

Regarding the relationship between L1 and L2 aptitude, Sparks, Ganschow and colleagues have conducted much work in this realm (e.g., Sparks, 1995; Sparks & Ganschow, 1991, 2001; Sparks, Patton, Ganschow & Humbach, 2011). Based on work comparing high- and low-achieving and at-risk and not-at-risk L1 and L2 learners with learning disabilities, they focus on a central cognitive factor termed linguistic coding ability that serves as the core variable in their Linguistic Coding Differences Hypothesis (LCDH). This hypothesis proposes that the ability to learn an L2 is closely related to the individual’s L1 learning skills, based on L1 literacy such as phonological/orthographic processing and word recognition/decoding, and this proposal has accumulated empirical support. For example, Sparks et al. (2011) were interested in determining whether a factor analysis would reveal distinct factors composed of similar L1 and L2 skills including early L1 skills, L1 academic aptitude (IQ), L2 aptitude, and L2 affective measures to predict L2 proficiency in fifty-four high school students followed from Grades 1 through 10. They found that two of the factors that emerged—Language Analysis and Phonology/Orthography—were composed of both L1 skills and L2 aptitude subtests from the MLAT. They interpreted this finding to suggest that early L1 skills are linked to later L2 aptitude (p. 267).

While aptitude is considered to be a fairly stable, cognitive ability that is fixed and does not change, research investigating the subcomponents of aptitude and their relation to L2 proficiency (e.g., Harley & Hart, 1997; Skehan, 1986; Winke, 2005) has revealed different ‘routes’ to success, implicating a unique role for aptitudinal components according to age and proficiency level. Focusing on the language analytic and memory components of foreign
language aptitude, Skehan (1986) conducted a study with adults learning colloquial Arabic in an intensive 10-week British military course. Based on results from a cluster analysis, two learner ‘types’ emerged, one analytic and one memory-oriented, and both led to successful L2 learning outcomes. This finding was interpreted by Skehan (1989) as early evidence of different possible routes to success based on two different orientations to language development, one “which stressed the analysability of language while the other […] is more apt to rely on chunks of language and efficient memory” (p. 37). Similarly, Harley and Hart (1997) found that different components of the aptitudinal construct were more prevalent for Grade 7 and Grade 11 immersion school children; namely, memory components mattered more in determining performance in younger learners whereas language analytic subtests had the highest explanatory power for the older learners (later confirmed in Harley & Hart, 2002).

Regarding increasing proficiency levels, Winke (2005) found several relationships between phonological working memory, components of aptitude as measured on the MLAT and performance on a Chinese Proficiency test in one-hundred and thirty-four beginning and advanced adult learners of Chinese. She reported that memory (rote learning ability), phonetic coding ability and grammatical sensitivity measures of aptitude significantly related to learning for beginners but not for advanced learners. Phonological working memory, as measured in an aural version of a sentence span task (Waters & Caplan, 1996), correlated with writing, speaking, and reading abilities for beginners but only with listening ability for advanced Chinese learners. She interpreted these findings to support a changing role for aptitudinal constructs (WM included) as L2 proficiency increases.

Another influential approach in L2 aptitude research has been to investigate its role in informal, immersion and communicative instructional contexts of learning (e.g., Harley & Hart,
1997; Horwitz, 1987; Reves, 1983) as well as implicit and explicit instructional conditions within a laboratory setting (e.g., De Graff, 1997; Robinson, 1996, 1997). This approach has shed new light on the traditionally stigmatized notion promoted by Krashen (1981) and others (e.g., Cook, 1996) that aptitude is not relevant for communicative SLA where the focus is on meaning and only impacts L2 classroom learning where the focus is on form and rule learning. On the contrary, these studies suggest that aptitude may be relevant for language learning in both explicit and implicit learning settings. For example, in an artificial language learning experiment using a version of Esperanto ‘eXperanto’, De Graff (1997) investigated the interaction between aptitude (collapsed scores from the MLAT Words in Sentences and Paired Associates subtests) and type of instruction (implicit [-grammatical explanation] or explicit [+grammatical explanation]) in fifty-four university students. While participants in the explicit instructional condition outperformed those in the implicit learning condition on four tasks, results revealed that participants’ average aptitude score correlated with performance in all conditions, indicating that grammatical sensitivity and rote learning ability could matter for both explicit and implicit learning processes. However, the collapsed scores of both aptitude components obscures their relative ability to account for performance variance in each instructional condition.

Robinson (1996, 1997) also studied the effects of aptitude (separated subtest scores from Words in Sentences & Paired Associates) and instructional condition (implicit, incidental, rule-search, instructed) on the acquisition of two syntactic rules judged to be ‘easy’ or ‘hard’ by fifteen English second language (ESL) teachers. One-hundred four intermediate English L2 learners who were native speakers of Japanese, Korean and Mandarin Chinese, were randomly assigned to each condition. Results revealed that the grammatical sensitivity subcomponent of aptitude was significantly positively correlated with learning measured in a grammaticality
judgment task in all conditions, except the incidental condition. In the implicit condition, accuracy on the easy and hard rules produced the strongest correlations with grammatical sensitivity ($r = .69$ and $.75$, respectively). There was also a strong link between this aptitude component, learning and awareness (measured retrospectively) in the implicit condition. Namely, high aptitude learners in the implicit learning condition were found to be those who most likely reported searching for rules, and were able to verbalize those rules. Robinson interpreted this finding as the ability of grammatical sensitivity to positively predict awareness during implicit L2 exposure (despite measuring awareness after, as opposed to during, exposure to L2 input).

**Summary.**

In summary, the construct of L2 aptitude has undergone several refinements over the years since Carroll and Sapon’s (1959) original operationalization. L2 aptitude is still conceived as a relatively stable cognitive trait whose components overlap with, yet are distinguishable from intelligence and native language ability. Other research has examined the role of aptitudinal components in interaction with proficiency and different learning conditions. This research suggests an important role for phonetic coding ability, grammatical sensitivity and memory (rote learning ability) in explaining overall performance by beginning learners (Winke, 2005) as well as a facilitative role for grammatical sensitivity (Robinson) and grammatical sensitivity and rote learning ability (combined score) (DeGraff) under both explicit and implicit instructional conditions. Thus, it seems that for adults, aptitude matters regardless of type of L2 instruction, which may be because the same basic cognitive abilities such as noticing and rehearsal underlie learning across different conditions (Robinson, 2002; Schmidt, 2010). These findings not only undercut the claim that individual differences in implicit learning are minor compared to explicit learning ability (Reber, 1993; Stanovich, 2009) but also indicate that language analytic and
memory subcomponents of L2 aptitude could be implicated not only in learning but also in how learners use their knowledge in performance with the possibility of a diminishing role of aptitudinal components as learners develop in language skill.

The measurement of the construct of L2 aptitude and its components is fundamental in order to accurately interpret their relationship with L2 achievement. While the MLAT has proven durable as a measure of L2 aptitude throughout the years, the subtests still lack direct correspondence with certain components (i.e., inductive language learning ability). More importantly, there is still no standardized measure of aptitude that operationalizes a developmental perspective of L2 learning (Robinson, 2013), or that learning a language likely involves different abilities at different stages of development. Recent L2 aptitude research taking a componential approach propose that traditional measures of aptitude be supplemented, or even replaced, by measures of working memory capacity (e.g., McLaughlin, 1995; Miyake & Friedman, 1998; Robinson, 2005a; Sawyer & Ranta, 2001; Skehan, 1989, 1998, 2012; Wen, 2012; Wen & Skehan, 2011), following the view that they “could potentially be better predictors of L2 success than the MLAT” (Sanz, 2005, p. 15).

In the following section, theoretical models of working memory within cognitive psychology are first reviewed followed by measurement approaches and relevant empirical studies investigating its relationship to L1 and L2 learning, with a detailed focus on the latter.

**Defining working memory capacity (WMC): Models in cognitive psychology.**

The concept of working memory has been a central theoretical construct in cognitive psychology for several decades since its inception (Baddeley & Hitch, 1974). It refers to a specific set of cognitive processes comprising online processing, storage and retrieval and plays a key role in the complex, cognitive tasks we carry out every day such as reading a newspaper.
article or calculating the appropriate amount to tip in a restaurant (Miyake & Shah, 1999).
Simply, WMC is conceived as a mental workspace where information is simultaneously stored and processed (Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Baddeley, 1986, 2000, 2003, 2007, 2010) and a fundamental characteristic is that it has a limited capacity, which constrains cognitive performance.

Two main theoretical issues among different models of WMC involve the factors influencing individual differences in WM performance and the question of domain-generality versus domain-specificity (Miyake, 2001; Sagarra, 2013; Shah & Miyake, 1999). As such, its conceptualization has principally evolved within three models: i) the domain-specific, single-resource model (Just & Carpenter, 1992); ii) the domain-specific, multiple-resource model (Baddeley & colleagues); and iii) domain-free, attentional models (e.g., Engle, 2002; Engle, Kane, & Tuholski, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane, Conway, Hambrick & Engle, 2007). The construct of WMC is considered within each of these views below but for complete comprehensive overviews of all WM theories and models, including task-switching (e.g., Towse & Hitch, 2007; Towse, Hitch, & Hutton, 1998) and inhibition-based accounts (e.g., Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988), the reader should consult Baddeley (2007), Conway, Jarrold, Kane, Miyake, & Towse (2007), Goo (2010), Juffs and Harrington (2011), Miyake (2001), Miyake and Friedman (2012), Miyake and Shah (1999), Wen (2012, forthcoming) and Williams (2012).

Models of WMC view storage capacity, processing efficiency and executive ability as the main sources of variation between individuals (Sagarra, 2013). Within the domain-specific single-resource model, or Constrained Capacity model (Just & Carpenter, 1992), WM involves simultaneous storage and processing functions that compete for a shared pool of resources and
individual differences are revealed when attentional capacity limits reach their threshold, or ‘activation maximum’ (Just & Carpenter, p. 123). This results in a trade-off between storage and processing in cognitively demanding tasks that exceed an individual's WMC (i.e., storage decreases and processing slows down).

On the other hand, Baddeley’s domain-specific multiple-resource model (2000, 2003, 2007, 2010; Baddeley & Hitch, 1974) posits that processing and storage components operate relatively independently of one another. This model originally aimed to promote a ‘non-unitary’ view of WM with specialized components in an effort to refine the concept of short-term memory (STM) in information-processing models dominant at the time (e.g., Atkinson & Shiffrin, 1968; Broadbent, 1958). Under this view, WM is conceived as a multi-component system comprising a domain-general central executive, responsible for various attentional functions (e.g., inhibiting, switching, maintaining, integrating, coordinating & retrieving information), and two domain-specific short-term storage systems, the phonological loop, handling phonological and verbal information such as remembering a telephone number, and the visuospatial sketchpad, in charge of visual and spatial information such as visualizing chess moves. A third cross-modal store, the episodic buffer (Baddeley, 2000), communicates information between long-term memory (LTM) and the storage components. Thus, although processing (central executive) and storage components (phonological short-term memory (PSTM); visuospatial sketchpad) under this conceptualization operate independently to some extent, they communicate through the episodic buffer in encoding and retrieving information from LTM. In fact, one of the major roles for working memory is "retrieval of stored long-term knowledge relevant to the tasks in hand” (Baddeley & Logie, 1999, p. 31). Figure 2 below displays the multi-component WM model as conceived by Baddeley.
Domain-free attentional models of WM emphasize the inhibitory control function of the central executive in exerting control over the contents of STM (e.g., Engle, 2002; Engle, Kane, & Tuholski, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane, Conway, Hambrick & Engle, 2007). This theory draws on notions such as the supervisory attentional system (SAS) (Norman & Shallice, 1986) that emphasizes the executive functions of WM in managing attention (i.e., blocking interference, resolving conflict). In other words, WM is executive attention, or the general ability to maintain or suppress verbal or non-verbal information. As such, a controlled-attention view of WM does not make a distinction between processing and storage capacity and attributes individual variation in WMC to one's ability to maintain and recover “access to information that is relevant to ongoing tasks and in blocking access to task-irrelevant information” (Kane, et al., 2007, p. 22), processes subserved by prefrontal cortex functioning (Engle, Kane, & Tuholski, 1999). Furthermore, WM is seen as an activated part of LTM, rather than as an entryway to LTM.
While the above conceptualizations have expanded our understanding of the range of executive functions the central processor carries out including focusing and sustaining attention, task switching, updating, inhibition, encoding, and retrieval (Baddeley, 2007), some researchers have asserted the dominance of the multiple-component, resource-sharing view along with its associated dual-task methodology (Miyake, 2001, p. 164), given that it offers an overall coherent account of the role of WMC in different domains of complex cognition (Baddeley & Logie, 1999). Other researchers have argued that the varying accounts are in fact complementary to one another with each being more appropriate in investigating certain research questions (Logie, 2011, p. 242). To adequately address such inquiries, regardless of their focus, a key issue for all models of WMC is how to validly measure the limited capacity defining this construct.

**Measuring executive and phonological loop components of WMC.**

Working memory span tasks are among the most widely used measurement tools in cognitive psychology (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). Measures of WMC can be administered in the L1 and/or the L2 and are typically grouped along two broad dimensions: simple and complex span tasks and linguistic and non-linguistic tasks (also, verbal/non-verbal). The former dimension reflects a passive, storage-oriented versus active, dynamic concept of working memory whereas the latter dimension reflects domain-specific versus domain-general views of working memory.

Traditional linguistic simple span tasks such as word span require a storage component only (maintaining a target memory item like a word) and are believed to tap STM. A commonly used non-verbal simple span task is the digit span in which the target item for recall is a numerical digit. Complex span tasks such as reading, listening and operation span tasks align with current conceptualizations of WMC as they impose dual-task demands of both storage and
processing. The processing component in complex tasks typically consists of reading and listening to sentences (Reading & Listening span tasks (RSpan/LSpan) (Daneman & Carpenter, 1980; Waters & Caplan, 1996)) or calculating simple mathematical operations (Operation span (OSpan) Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005) while simultaneously maintaining target items (words or letters) in memory for later recall. In the RSpan task, the critical component of remembering words is embedded within a secondary comprehension task in which learners read sentences (Daneman & Carpenter—Experiment 1) or verify their truthfulness (Experiment 2). Waters and Caplan (1996) modified Daneman and Carpenter’s version by requiring plausibility judgments after silently reading each sentence to ensure semantic processing; they made the widely-cited claim that to ensure reliability, sentence span tasks should be measured compositely to reflect both sentence processing and recall components (p. 70). Kane, Hambrick and Conway (2004) changed the recall items to capitalized, isolated letters following each utterance, thus making the measure less language dependent. Nevertheless, the RSpan remains verbal in nature as it requires semantic processing of sentences.

The most widely used non-verbal complex span task is the OSpan (Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005) with Unsworth et al. modifying the to-be-recalled items (words) in Turner and Engle's version to letters. Others include the letter-number ordering task (LNOT) (WAIS-III Intelligence Scale—Weschler, 1997) involving reading mixed sequences of numbers and letters and then recalling and reordering them in numeric, alphabetical order and the digits back/forward tests involving listening to sets of random digits increasing in length and then recalling them in reverse or forward order.

An individual’s span score in these tasks is typically calculated according to the maximum number of words or letters accurately recalled in sets of increasing difficulty, usually
ranging from two to six sentences, and different scoring methods produce composite scores including processing and storage scores or separate scores for each. Overall, scoring procedure in dual tasks should reflect the “maximum possible score when stretching [learner] abilities to their limits” (Logie, 2011, p. 243).

Importantly, estimates of reliability based on internal response consistency (e.g., Cronbach’s alpha) in complex WM span tasks have been quite high, typically in the range of .70-.90 for span scores (Conway et al., 2005, p. 776). Furthermore, span scores are reported to be stable across time with test-retest correlations for adults ranging between .70-.85 for the reading and operation span tasks (e.g., Friedman & Miyake, 2004; Klein & Fiss, 1999; Unsworth et al., 2005) although test-retest reliability for composite scores is reportedly higher than sentence-final word recall scores only in the Waters and Caplan sentence span task and Daneman and Carpenter reading span task (Waters & Caplan, 1996, p. 66). Additionally, reliability of these measures has been derived from correlations between them, which are reportedly strong (r = .43 -.69) (Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004; Unsworth et al., 2005).

A crucial issue to keep in mind when administering WM tasks is that of language proficiency. Though some research views WMC as language-independent (e.g., Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993), there is growing evidence that WMC interacts with language proficiency (e.g., Gass & Lee, 2011; Miyake & Friedman, 1998; Sanchez, Wiley, Miura, Colflesh, Ricks, Jensen, & Conway, 2010; Service, Simola, Metsanheiro, & Maury, 2002; Van den Noort, Bosch & Hugdahl, 2006; Wen, 2012). In other words, low-proficiency L2 learners perform poorly in WM tasks conducted in the L2 due to their proficiency level and not due to their WM capacity. Therefore, to avoid a confound with L2 proficiency, many researchers argue that WMC should be measured in the learners’ L1 (e.g., Juffs & Harrington, 2011; Sagarra,
2013) and others emphasize the need to employ a domain-general span task like the OSpan to avoid a potential confound with verbal processing (e.g., Engle, 2002).

The most widely used measures of the phonological short-term storage component of the multiple-resource model include the digit-span task and non-word repetition and recognition tasks. The digit span is used in many standard measures of intelligence, such as the WAIS-III, and according to Baddelely, Gathercole, and Papagno (1998), it “provides a useful indication of the capacity of an individual’s phonological loop” (p. 159). In the digit span task, participants hear a set of numbers and repeat them aloud in the same order, usually in two sets each of two to nine digits. Participants are awarded one point for every set remembered correctly until they reach the last set or until they miss two sets in a row. In non-word repetition (NWR) tasks, learners repeat a series of non-word strings increasing in syllable length, frequently between two and nine syllables. Non-word repetition can be based on the L1 or L2 lexicon and performance can be measured either as the number of syllables in the longest correctly repeated non-word or as the number of non-word strings accurately repeated without deviating more than one syllable. Not surprisingly, repetition tends to be better for L1 non-words than for L2 non-words (Service, 1992; Thorn & Gathercole, 1999) based on language familiarity of L1 sound patterns (long-term memory effects). Serial NWR tasks require participants to determine whether the presentation order of a pair of identical non-word sequences (sequence lengths of 5 to 7 items) is the same or different. Thus, while the digit span task provides a relatively non-verbal measure of PSTM, serial non-word repetition and recognition tasks are more verbally based and depend to some degree on the L1 or L2 lexicon.

**Empirical research on IDs in WMC.**

*L1 learning.*
Individual differences in WMC as measured by verbal complex span tasks, the RS Span in particular, have been shown to strongly relate to higher-order abilities in the L1 like reading and sentence comprehension (e.g., Caplan & Waters, 1999; Daneman & Carpenter, 1980, 1983; Daneman & Hannon, 2001; Daneman & Merikle, 1996), verbal fluency (Daneman, 1991), the ability to learn novel-word referent connections (Daneman & Green, 1986), and the ability to handle syntactic ambiguity and process syntactically complex sentences (Just & Carpenter, 1992; King & Just, 1991). In support of a domain-general view of WM, non-verbal complex span tasks such as the OS Span task have also been shown to be a strong predictor of language comprehension (e.g., Daneman & Merikle, 1996; Turner & Engle, 1989), though Daneman and Hannon (2001) maintain that verbally based span tasks tend to be better predictors of language ability than numerically based span tasks, which is likely due to a confound between the verbal nature of the RS Span task and L1 ability.

Furthermore, the phonological loop component of the multi-component WM model, which is responsible for the processing and temporary retention of (un)known phonological information, has been proposed to play a fundamental role in child L1 development (Baddeley, 1986). In particular, its role has been demonstrated in children’s ability to acquire new words (e.g., Baddeley, Gathercole, & Papagno, 1998; Service & Kohonen, 1995) which has led Baddeley, Gathercole and Papagno (1998) to propose that phonological memory could be a “language learning device” (p. 887) for children. In addition to L1 vocabulary learning, PSTM capacity has also been linked to more general markers of oral fluency in terms of length of utterances and complexity of narratives (Juffs & Harrington, 2011, p. 139).

While the central executive and phonological storage components within Baddeley's multi-component model have been found to underlie child and adult L1 development, WMC is
hypothesized to occupy an even stronger role in L2 processing and performance for adults (Geva & Ryan, 1993; Miyake & Friedman, 1998, 2012) given that L2 learners are likely to find themselves in situations where their WM limitations "are pushed and exceeded" (Sawyer & Ranta, 2001, p. 341), depending on factors such as proficiency level and task demands. As such, L2 learning has been proposed to rely “to a greater extent than L1 acquisition on general learning mechanisms” (Miyake & Friedman, 1998, p. 340).

**L2 learning.**

The potential importance of the role of WMC in SLA stems from the need to attend to meaning and form, which consumes cognitive resources (Sagarra, 2013; VanPatten, 2004). Given that WMC requires attention in order to hold language information in a short-term memory store (PSTM), analyze it in reference to information in other memory systems and then move it to the LTM store, individual variation in WMC could determine the quality and quantity of L2 input that is successfully incorporated into the developing system (i.e., converted into intake), the first step in the learning process (Schmidt, 1990, 1994, 2001). As a result, some researchers have hypothesized that WM function is the only ID variable likely to have a direct impact on the rate of development of L2 proficiency (Miyake & Friedman, 1998; Sagarra, 2000) following the logic that "the more WM you have, presumably the more attentional resources you possess and the more you can potentially learn, because you have enough memory to process form together with meaning" (Juffs & Harrington, 2011, p. 147).

In line with this view, a wide range of research questions pertaining to WMC have been pursued in SLA including its role in L2 reading comprehension (e.g., Harrington & Sawyer, 1992), L2 pronunciation (Trude & Tokowic, 2011), possible effects of thinking aloud on L2 performance (Goo, 2010), the relative effectiveness of corrective feedback in L2 grammar...
learning (e.g., Goo, 2011, 2012; Lado, 2008; Mackey, 2012; Sagarra, 2007a; Sagarra & Abbuhl, forthcoming), spoken output production in classroom-based research (e.g., Mackey, Adams, Stafford, & Winke, 2010; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002), and lexical comprehension and production in study abroad contexts (Sunderman & Kroll, 2009). In general, these studies support the cognitive psychology and L1 learning literature as learners with higher WM spans outperform learners with lower WM spans in relation to the phenomenon under investigation, also known as the more is better view (Miyake & Friedman, 1998).

On the other hand, there is a lack of consensus pertaining to the role of WMC in processing L2 grammar, despite the prevalence of studies investigating this domain of inquiry for over fifteen years (Juffs & Harrington, 2011, p. 147). Moreover, different models of WMC make differential predictions regarding whether it constrains syntactic processing. Whereas low WMC individuals are expected to have reduced syntactic processing efficiency under a single-resource view (Just & Carpenter, 1992; Just, Carpenter & Keller, 1996), a multiple-resource model predicts that the capacity of the WM system will not determine the efficiency of online syntactic processing in language comprehension (Waters & Caplan, 1996, 2004, 2005). Several empirical self-paced reading studies have reported evidence that higher WM span L2 learners are more sensitive than lower span learners in processing L2 morphology and syntax (e.g., Dussias & Piñar, 2010; Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009; Sagarra, 2007b, 2008; Sagarra & Hershensohn, 2010) while other self-paced reading studies support the multiple-resource view (i.e., no WMC effects) (e.g., Foote, 2011; Juffs, 2004, 2005; Rodríguez, 2008). The discrepancy in findings in this strand of research may be due to the proficiency level of the learners investigated, with WMC effects most consistently being found for learners at low proficiency (Sagarra, 2013).
The phonological loop in Baddeley’s multi-component model of WMC is also thought to function as an important determinant of L2 learning (e.g., Baddeley, Gathercole & Papagno, 1988; N.C. Ellis, 1996; Williams & Lovatt, 2003). The role of PSTM capacity has mainly been studied in L2 vocabulary learning by children (e.g., Papagno, Valentine, & Baddeley, 1991; Service, 1992; Service & Kohonen, 1995) with a more prominent effect in the early stages of vocabulary learning and a diminishing role as learners develop L2 linguistic resources (e.g., Speciale, N.C., Ellis, & Bywater, 2004; Masoura & Gathercole, 1999). Nonetheless, PSTM has also been reported to be a robust, longitudinal predictor of vocabulary growth in both native English-speaking children and those who are learning English as a second language (Farnia & Geva, 2011).

For adults, individual differences in PSTM are proposed to contribute to sequencing, or chunking ability, which is related to variation in “learners’ facility to acquire second-language grammar” (N.C. Ellis, 1996, p. 92). Indeed, PSTM has been reported to play a role in grammatical rule learning (N.C. Ellis, 2012; Martin & N.C. Ellis, 2012; N.C. Ellis & Sinclair, 1996; Williams & Lovatt, 2003) as well as a range of L2 skills including reading, writing, speaking and listening abilities in L2 Chinese (Winke, 2005), oral fluency, lexical and grammatical skills in L2 Spanish (O’Brien, 2005; O’Brien, Segalowitz, Collentine & Freed, 2006; O’Brien, Segalowitz, Freed & Collentine, 2007) and achievement test performance measuring grammar, vocabulary and reading comprehension in L2 English (Hummel, 2009).

**L2 morphology and syntax.**

In this section, I elaborate further on studies investigating L2 grammatical processing that are of relevance to the current study. I first detail a classic study looking at WMC and comprehension of L2 syntax (Miyake & Friedman, 1998). Then, I review several studies that
investigate the relationship between IDs in WMC and L2 grammatical processing according to learner proficiency level (Coughlin & Tremblay, 2013; Foote, 2011; Sagarra, 2007b; Sagarra & Hershensohn, 2010) as well as the role of the phonological loop component of WMC by proficiency level (Kormos & Sáfár, 2008; O’Brien, Segalowitz, Collentine & Freed, 2006; Winke, 2005). Next, I review the interaction between IDs in WMC and type of instructional condition (Erlam, 2005; Robinson, 2002, 2005b; Sanz, under review) as well as type of L2 knowledge measured (Erçetin & Alptekin, 2013; Roehr & Gánem-Gutiérrez, 2009). Finally, I review the only two studies investigating the relationship between WMC and L2 development over time in a classroom setting (Linck & Weiss, 2011; Sagarra, 2000).

In a highly cited chapter reporting a study carried out by Miyake, Friedman, & Osaka (1998), Miyake and Friedman (1998) examined the relations among ID’s in L1 (Japanese) and L2 (English) WM, syntactic comprehension ability (i.e., the ability to decode “who is doing what to whom”), and the acquisition of appropriate linguistic cues in the L2 (i.e., assigning agency to nouns in English sentences based on word order versus the Japanese strategy of assignment based on animacy). Participants were fifty-nine native speakers of Japanese who had studied English as an L2 for at least six years. WM was measured in a listening span version of Daneman and Carpenter’s (1983) reading span task. Using path analysis, a form of structural equation modeling, results showed that L1 WM determines L2 WM which directly determines cue preference distance. In other words, high L2 WMC learners more accurately employed cue assignment strategies based on L2 English than low WMC learners who preferred the L1 Japanese based strategy. They concluded that L2 WM and cue preferences affect how well learners can interpret complex English sentences. However, these findings could also stem from
the relationship between L2 proficiency level and L2 WM function, as reported in other research (e.g., Van den Noort et al., 2006).

**Level of L2 proficiency.**

In this section, research exploring the relationship between IDs in WM and various aspects of L2 grammatical processing and performance is detailed in beginning and intermediate Spanish L2 learners (Sagarra & Hershensohn, 2010), advanced late English-Spanish bilinguals (Foote, 2011), and learners at 'mid and high' proficiencies in L2 French (Coughlin & Tremblay, 2013). A review of research investigating the role of the phonological storage component of WM is also provided for beginner and intermediate L2 English learners (Kormos & Sáfár, 2008) and low and high proficiency L2 learners of Spanish (O'Brien et al., 2006, 2007) and Chinese (Winke, 2005).

Building on previous findings of a link between WMC and L2 processing of noun-adjective gender agreement in beginning learners (Sagarra, 2007b, 2008), Sagarra and Herschensohn (2010) investigated L1 and L2 processing of gender and number agreement in learners at a beginning and intermediate proficiency level. Sixty-nine beginning (3rd semester) and sixty-four intermediate (7-8th semester) English-speaking Spanish L2 learners, and sixty-three Spanish monolinguals completed grammar portions of a Spanish proficiency test (*Diploma de Español como Lengua Extranjera* (DELE) for intermediate learners), an online self-paced reading task, and an offline GJT with sentences containing noun–adjective gender/number (dis)agreement. WMC was measured via the Waters and Caplan (1996) version of the reading span task in the participants’ L1 based on the possibility that deficits in L2 knowledge could affect the results of a test conducted in that language. Results revealed that only intermediate L2 learners with higher WMC were more sensitive to gender, but not number, disagreement on the
online task than those with lower capacity. Beginners did not show sensitivity to gender or number agreement violations, possibly due to an overall low WMC; this contrasts Sagarrà’s (2007b, 2008) studies reporting sensitivity to agreement violations for high WMC beginners.

Foote (2011) investigated the degree to which forty early and late English-Spanish bilinguals and twenty native Spanish speakers were sensitive to subject-verb number agreement errors and noun-adjective gender agreement errors in a word-by-word sentence reading task in Spanish. She specifically inquired whether late learners possess integrated (i.e., implicit) linguistic knowledge and can access such knowledge in an automatic, nativelike manner. She also asked whether the development of some structures in late L2 acquisition is 'selective' (Jiang, 2007), or (non)integratable, and whether processing depends on the distance between agreement source and target. To account for any distance effects found for late bilinguals, Foote administered a reading span task (Waters & Caplan, 1996) to measure WMC, though only scored participants based on words recalled rather than calculating a composite score. Results revealed that all three groups showed sensitivity to both subject–verb number agreement errors and noun–adjective gender agreement errors while reading for comprehension, leading Foote to conclude that late English-Spanish bilinguals do possess integrated knowledge of Spanish morphology (p. 214). However, sensitivity to agreement violations was greater when the agreement source and target were adjacent to one another and late bilingual working memory span scores did not correlate with a difference in sensitivity as a function of agreement distance for either structure. The lack of WMC effects, however, may be due to the advanced proficiency level of the learners or the scoring method used (i.e., words recalled).

Coughlin and Tremblay (2013) investigated the roles of WMC and learner proficiency level in processing number agreement dependencies in L2 French as a function of the length of
the agreement dependency. Fifty-two L1 English L2 French learners were evenly divided into 'mid' and 'high' proficiency groups based on a cloze test (Tremblay, 2011). They completed a self-paced reading task to measure sensitivity to violations, an acceptability judgment task that assessed whether they had explicit knowledge of number agreement between object-clitics and left-dislocated antecedents, and L1 (Eng) and L2 (French) RSpan tasks. They found that L2 learners with higher WMC (in French) were more sensitive to number agreement violations in French but WMC in English (L1) was not a significant predictor of learners' sensitivity. Furthermore, both mid- and high-level L2 learners were able to find and correct number agreement violations in object clitics in the offline acceptability judgment task, but only high-level proficiency learners showed sensitivity to violations in the online task (p. 23).

Overall, these findings indicate that WMC can be a good predictor of L2 learners' ability to process agreement morphology, but its explanatory power may depend on the proficiency level of the learner. Also, these studies highlight the need to measure this construct in the learners' L1 to avoid a confound with L2 proficiency (e.g., Van den Noort et al., 2006) and to use a composite score that reflects processing and storage components in order to accurately interpret the role of WMC in L2 grammatical processing.

The role of the phonological loop component of WMC has also been shown to interact with learner proficiency level. Kormos and Sáfár (2008) investigated the extent to which various L2 skills correlated with performance on the non-word repetition (NWR) task measuring PSTM, and the backward digit span task measuring general WMC in one-hundred twenty-one Hungarian adolescent learners of L2 English divided into beginner ($n = 100$) and ‘pre-intermediate’ ($n = 21$) groups. Beginners had either no previous English instruction or had been placed via a non-descriptive placement test. Pre-intermediate learners “had learned English before coming to school”
(p. 264). Only beginners completed the backward digit span task. Learner performance was based on the Cambridge First Certificate language exam measured at the end of a one-year intensive language-training program. Thus, by the end of the learning period, beginners were considered to be at a "near upper-intermediate level" of proficiency and intermediate learners were determined to have reached a "high upper-intermediate level" (p. 266).

Results revealed PSTM to be an important factor for the higher proficiency group only, with the average NWR scores correlating moderately with students’ overall performance and the writing, Use of English, oral fluency and vocabulary subcomponents. While PSTM was not significantly related to success at the end of the term for the students who were beginners at the start of the school year, WMC, as measured on the backward digit span task, was significantly related to overall proficiency in the lower proficiency group and to all sub-components of the end-of-year proficiency test, with the strongest relationship between WMC and the (explicit) Use of English test. Thus, these results seem to indicate a stronger role for WMC in mediating performance on tests that likely tap explicit knowledge for beginning learners and a stronger role for PSTM in the case of intermediate learners in an intensive language learning context.

Based on data gathered for her dissertation (O'Brien, 2005) and the finding that PSTM more strongly correlated with L2 oral fluency development in less proficient adult L2 learners (O'Brien et al., 2007), O'Brien et al. (2006) investigated the relationship between PSTM, as measured in a serial non-word recognition task, and L2 speech production in Spanish (i.e., productive vocabulary, narrative and grammatical abilities such as inflectional morphology and subordinate and coordinate clause use) measured at the beginning and end of a thirteen-week semester. Participants were forty-three native English-speaking adults learning Spanish who all had at least two prior semesters of formal instruction in Spanish and who were divided into low
ability \((n = 23)\) and high ability \((n = 20)\) groups based on their Spanish SAT II scores. At the beginning and end of the semester, each participant participated in a 20-30 minutes Oral Proficiency Interview (OPI) in Spanish administered by testers trained by the American Council of Teachers of Foreign Languages (ACTFL). The authors found that phonological memory was a significant predictor in the development of L2 narrative abilities for low proficiency participants, explaining 17.5% of the performance variance, whereas for the more proficient participants, PSTM was related to measures of grammatical competence (subordinate clause use) and accurate use of function words, predicting 15.7% of the variance.

The finding that phonological working memory may play a differential role according to the proficiency level of the learner and language skill measured was supported by Winke (2005), as previously detailed in the section on L2 aptitude. She found that phonological working memory, as measured in an aural version of the sentence span task (Waters & Caplan, 1996), correlated with writing, speaking, and reading abilities as measured in a standardized Chinese proficiency test for beginning Chinese L2 learners but only with listening ability for advanced learners. Taken together, this empirical evidence suggests that PSTM could be implicated more generally at early to intermediate stages of language learning (Kormos & Sáfár; O'Brien et al.; Winke) with a more specific role at increasing levels of proficiency in terms of listening comprehension (Winke) and accurate use of complex grammar (O'Brien et al.). Nonetheless, the task used to measure PSTM capacity should be considered as all three studies used different measures (serial non-word repetition vs. serial non-word recognition vs. aural sentence span), which could have affected the picture these results seem to portray.

*Different instructional conditions.*
Researchers in cognitive psychology have hypothesized that implicit and explicit learning processes may not be modulated by individual differences in the same way (Reber, 1993; Stanovich, 2009) and that differences in cognitive capacity only regulate explicit learning (Reber & Allen, 2000, p. 236). In SLA Krashen (1981, 1985) has also argued that unconscious L2 'acquisition' is unaffected by differences in aptitude whereas conscious L2 'learning' is affected by aptitude. However, some research has investigated the role of working memory capacity as a measure of aptitude that reflects the processing demands of simultaneous attention to form and meaning and reported that explicit instruction in fact may neutralize the role of cognitive capacity, which was more strongly related to learning under implicit instruction (Erlam, 2005; Robinson, 2002, 2005b; Sanz, under review), thus contradicting Reber and Krashen's claims. Nonetheless, while conditions designed to promote implicit or explicit learning are presumed to tap into separate processes, type of instruction never guarantees that implicit or explicit processes internal to the learner will actually take place (Schmidt, 1994).

In an effort to replicate a study of implicit artificial grammar (AG) learning (Reber, Walkenfeld, & Hernstadt, 1991) and generalize its findings to L2 learners in a natural language learning experiment (Samoan), Robinson (2002, 2005b) investigated the correlation between cognitive abilities and performance under implicit and explicit AG learning conditions and an incidental Samoan learning condition. As found by Reber et al. (1991), there was significantly greater variance in explicit compared to implicit learning outcomes on listening grammaticality judgment posttests. Results revealed statistically significant correlations between WMC (measured via a L1 reading span) and learner performance in the incidental group on immediate and delayed listening grammaticality judgment posttests ($r = .42$ and $.48$, respectively, $p < .01$ for both) as well as on two delayed production tests ($r = .33$ and $.44$, respectively, $p < .05$ for
both), providing some evidence that WMC may mediate incidental learning and retention of L2 rules of a natural language (Samoan). Interestingly, however, learner performance did not correlate with WMC in either the explicit or implicit AG learning conditions, which may be related to the differing demands imposed by natural versus artificial language learning.

In a classroom study, Erlam (2005) investigated whether more explicit (deductive; structured input) and less explicit (inductive) instructional methods varying in rule explanation, production practice and input-based activities differentially affected learning by sixty secondary-school learners of L2 French according to their L2 aptitude. L2 aptitude was operationalized as: 1) language analytic ability (Words and Sentences subtest of the MLAT), 2) phonemic coding ability (Sound discrimination test of the PLAB), and 3) working memory capacity (multisyllabic word test of the phonological loop testing storage only). Results revealed very few correlations with L2 aptitude for the group that received deductive instruction (2 out of 30 correlations conducted) whereas language analytic ability correlated positively with gain scores on the listening comprehension test (pre-test > immediate post-test) and the written production test (pre-test > delayed post-test) in the inductive group, and both language analytic ability and working memory correlated positively with immediate and delayed gain scores on the written production tests in the structured input group. Based on these findings, Erlam concluded that more explicit teaching methods involving rule explanation and production practice seemed to neutralize the beneficial effects of cognitive learner differences, supporting Skehan's (1989) claim that aptitude aids learners in dealing with 'less structured' material (p. 129, c.f. Erlam, p. 165).

Sanz (under review) extended this line of research by reporting on two experimental studies conducted in a laboratory setting investigating the role of WMC in the development of Latin morphosyntax. Forty-four novice learners were randomly assigned to two learning
conditions that varied exclusively in terms of the presence [+explicit] or absence [-explicit] of grammatical explanation prior to input-based practice and provision of feedback (right/wrong). Sanz reported that higher WMC, measured via a L1 sentence span test, was significantly associated with development under less explicit instruction but that this association depended on assessment task demands (aural vs. written), confirming that tests used to measure instructional outcomes make different demands on working memory (Robinson, 2005b; Sanz, 1997), an important methodological consideration when deciding how to measure language development.

Somewhat indirectly related to this line of research is a study that reported a meaningful association between basic cognitive ability (i.e., processing speed) and implicit (nonlinguistic) sequence learning (Kaufman, DeYoung, Gray, Jiménez, Brown, & Mackintosh, 2010). This study investigated the association between several cognitive and personality IDs and implicit (sequence) learning ability in a large sample of adolescent students ($N = 147$). Participants completed an implicit sequence learning serial reaction time (SRT) task, explicit associative learning task, tests of psychometric intelligence, a test of WMC (OSpan), three tests of processing speed, the Big Five personality questionnaire, and reported general academic achievement scores were also collected. The authors reported a nonsignificant relationship between implicit sequence learning and WMC (OSpan) but through structural equation modeling, provided support for the roles of verbal analogical reasoning and processing speed, two components of general intelligence. They concluded that meaningful IDs in implicit learning (of non-linguistic sequences) do exist, contrary to prior belief (Reber, 1993; Reber & Allen, 2000; Stanovich, 2009) (p. 334). However, WMC did not play a role in implicit learning, contrary to findings by previous SLA studies reported above. This difference may again be related to the need to process form and meaning in natural language learning, a key ability served by working
memory function, versus processing of nonlinguistic sequences that does not require joint processing of form and meaning.

**Implicit/explicit knowledge of L2 grammar.**

The empirical studies detailed above have investigated the role of cognitive capacity under conditions designed to promote implicit and explicit learning. However, researchers have also suggested that different components of working memory capacity may be differentially implicated in use of L2 knowledge (N.C. Ellis, 2005; Hulstijn, 2005). However, empirical research in this realm is lacking, especially behavioral evidence based on performance measured on tests designed to tap implicit and explicit linguistic knowledge, with a couple recent exceptions (Erçetin & Alptekin, 2013; Roehr & Gánem-Gutiérrez, 2009).

Roehr and Gánem-Gutiérrez (2009) specifically tested the hypothesis that metalinguistic knowledge (MLK) and verbal WM in adult L2 learners are associated and also explored whether length/type of prior language study and language aptitude predicted university learners’ level of MLK. Participants were L1 English, L2 German learners ($n = 19$) and L1 English, L2 Spanish learners ($n = 20$) who had studied their L2s for approximately four years each. Both German and Spanish learners had studied up to three other languages apart from the L2, including French, Italian, Portuguese, Latin and Dutch; thus previous language experience was operationalized generally without accounting for language type and potential effects of transfer. Learners’ mean age was 25.6 years (German learners) and 22 years (Spanish learners) though their ages ranged from 18-65 years, another potentially problematic variable when measuring cognitive capacities such as WMC. Participants completed five instruments consisting of a biodata questionnaire, German and Spanish MLK tests, the MLAT, a test of L1 reading span (Daneman & Carpenter), and a test of L2 reading span (Harrington & Sawyer, 1992). In the L1 span test, there was no
plausibility or grammaticality judgment requirement, but the L2 span test required learners to judge correctness of the sentences. Thus, it seems the L2 version tapped both storage and processing ability whereas the L1 test did not.

Results revealed that description/explanation ability and language analytic ability on the MLK tests moderately correlated with the grammatical sensitivity subcomponent of aptitude measured on the MLAT (Words in Sentences) ($r = 0.41$ to $0.45$). However, contrary to expectation, no significant relationships were obtained between L1 or L2 WMC and MLK components. The researchers attempted to explain this finding in terms of type of measurement employed and speculation that a higher-level mental faculty, such as analytic reasoning about language (i.e., MLK), might be more domain-specific than a lower-level, more generic mental faculty like online storage and processing of linguistic information (i.e., WMC) (pp. 175-6). On the other hand, four significant predictors were found to account for sixty percent of the variance in learners’ level of MLK, namely, cumulative years of study of other languages (26% of variance), years of formal L2 study (19%), and the MLAT IV (9%) and MLAT V (6%) sections. However, there are several reasons to warrant caution in interpreting the strong relationship between prior experience and MLK and the lack of an association between WMC and MLK.

First, both German learners and Spanish learners reported having previously learned other languages that belonged to the same Germanic (e.g., Dutch) or Romance (e.g., Italian) language families as their L2. Thus, the strong predictive capacity of this variable might be more related to transfer effects from previously studied languages and not to type/length of prior language study. Second, the age range of the participants is problematic. Scores of older adult participants (who were up to 65 years old) might have affected the WM performance mean as outliers. Also, the L1 WM span test arguably did not tap both processing and storage abilities, a
requirement of complex span tasks previously revealing strong relationships with L2 learning. Overall, the relationship between cognitive ID’s and type of L2 knowledge merits further research in studies that methodologically control for the above variables.

Erçetin and Alptekin (2013) reported on relationships between L2 explicit and implicit knowledge sources, L2 WMC and L2 reading comprehension in fifty-one Turkish university L2 English learners. Learners were determined to be at a relatively advanced proficiency level based on performance on a standardized test of English proficiency. Participants completed Daneman and Carpenter’s version of the RSpan in the L2 (English), one measure of explicit knowledge (UGJT) and two measures of implicit knowledge (TGJT, EI test) (following R. Ellis, 2009) as well as the Nelson-Denny, a standardized reading comprehension test of English. Pearson product-moment correlations among scores on all five tests indicated that both implicit and explicit knowledge correlated with WM capacity though the strength of the relationship was stronger in the case of the explicit measure ($r = .47^{**}$ (UGJT) versus $r = .35^*$ (TGJT) & $.31^*$ (EI)). L2 reading comprehension revealed statistically significant positive relationships with explicit knowledge and WMC, but not with performance on measures of implicit knowledge.

The novel and somewhat unexpected result was the meaningful association found between WM and implicit knowledge since theoretically, WM is linked to conscious manipulation of information regulated by controlled processing and attentional capacity. The authors interpret this finding as being “indicative of the coexistence of explicit and implicit knowledge sources and their underlying processes in WM” (p. 19). However, there are a few limitations of this study that warrant consideration. Given that the learners in this study were advanced, it is possible that the type of knowledge being assessed on the EI and TGJT could actually be *automatized* explicit knowledge (Spada, 2013), or knowledge that has been
automatized as a function of increasing exposure to and practice using the language; as a consequence it is rapidly and efficiently accessed in use (Segalowitz & Segalowitz, 1993; Segalowitz, 2010), and thus could be functionally equivalent to implicit knowledge (DeKeyser, 2003). Another possible explanation of this relationship could stem from a potential confound between L2 WM and L2 proficiency (Van den Noort et al.) as the authors used the target language (English) to administer the WMC task. Overall, Erçetin and Alptekin suggest that future research "take the proficiency factor into account as an important variable" (p. 747) and replicate their research using measures specifically designed to measure implicit and explicit linguistic knowledge with late adult L2 learners at different proficiency levels.

*Long-term L2 development.*

Finally, there is a gap in our understanding of the role WMC plays in predicting L2 development over time in classroom learners at different proficiency levels. Only three studies to date have investigated differences in WMC and development of L2 grammatical knowledge *over time* in early and late beginners (1.5 years—Sagarra, 2000), advanced learners (10 months—Wright, 2009) and learners at a “low” proficiency level (3 months—Linck & Weiss, 2011) and they have produced conflicting results. Wright’s (2009) study was conducted in a study-abroad setting with eleven advanced ESL learners and positive (though non-significant) correlations were found between linguistic and WM scores. However, given the small sample of learners and immersion learning context, this study is not further detailed.

In the long-term study conducted by Sagarra (2000), a large sample (*n = 375*) of native speakers of English learning L2 Spanish at the university level were tested at two points in the learning period, as ‘early beginners’ (2nd semester; 300 hours previous instruction) and ‘late beginners’ (4th semester; 420 hours). WMC was measured on an L1 version of the reading span
test (Daneman & Carpenter, 1980) but without a plausibility judgment component to tap simultaneous processing. Learners were divided into low span (score of 0-2.5), medium span (3-3.5) and high span (4-5) with relatively uniform participant numbers throughout the span groups, though with exceptionally fewer high span learners. L2 grammatical knowledge of Spanish was measured in grammar sections on the Mecartty test (1993) and the University of Iowa (UI) Spanish Placement test (1993). The Mecartty test included a multiple-choice sentence completion section and a GJT requiring judgment (‘Good’ or ‘Bad’) and error identification (circling). Time allotted for each test was 15 minutes. The grammar knowledge section on the UI Placement test was untimed and included modified cloze tests with short passages in Spanish missing key words or phrases. Given these test characteristics, Sagarra speculated that the two grammar tests might have been tapping different types of knowledge; namely that the Iowa cloze passage and sentence completion tests may have been testing learned grammatical knowledge whereas the GJT was evaluating acquired knowledge (p. 130).

WMC was analyzed as a categorical between-group variable in a Repeated Measures ANOVA and as a continuous variable in an ANCOVA. The initial hypothesis that WMC would be a significant factor in the L2 acquisition of grammatical knowledge over time was not supported as results showed a lack of significant differences among subjects with low, medium, and high WM span. In fact, learners who scored 2.5 on the reading span tended to obtain the highest score on the grammar tests, a surprising result that seems to support the less is more view of WMC in L2 learning (Newport, 1991, c.f., Sagarra, p. 143). However, the lack of WMC effects can likely be attributed to the use of Daneman and Carpenter's RSpan measure which did not require simultaneous processing of meaning, a key component of complex span tasks. Given this methodological limitation, Sagarra calls for future research to clarify "whether less or more
In a semester-long study assessing the predictive validity of executive function and learning outcomes, Linck and Weiss (2011) aimed to examine whether executive functioning predicted the development of explicit L2 knowledge in a classroom context. Cognitive ID measures included Turner and Engle's (1989) Operation Span task to measure WMC and a Stroop task to measure inhibitory control. The authors additionally controlled for L2 motivation, measured compositely using thirty-three Likert-type items adapted from Gardner’s (2004) Attitudes/Motivation Test Battery (AMTB) which quantified interest in foreign languages, attitudes toward learning the L2, integrative and instrumental orientation and L2 class anxiety based on learner self-report. Learners were native English speakers enrolled in a first-semester university German introductory class \( (n = 8) \) or third-semester Spanish introductory conversation course \( (n = 16) \). Performance was tested at the beginning and end of the semester (3 months) in an untimed multiple-choice grammar test and a test assessing vocabulary knowledge. Spanish materials were taken from the Diplomas de Español como Lengua Extranjera (DELE) and German materials were fill-in-the-blank items taken from a placement exam. Preliminary analyses found no reliable group differences between the German and Spanish language students on any of the performance measures so both language groups were analyzed together \( (n=24) \).

L2 motivation and learner reported GPA/SAT scores were entered as covariates in a series of multiple regression analyses. Results revealed that the total set of predictors accounted for 60.7 percent of the variance in test scores at Time 1, 59.6 percent of variance at Time 2, and 40.8 percent of variance in test-retest difference scores (i.e., change in proficiency over time). L2 motivation and GPA both reliably predicted initial L2 proficiency (beginning of semester) but

WM truly leads to long-term successful SLA” (p. 143) according to different stages of acquisition.
when L2 motivation and GPA were controlled, WMC, but not inhibitory control, accounted for significant additional variance in performance at Time 1 (10%). Also, crucially, WMC reliably predicted the degree of L2 learning over the three-month period whereas no other predictor accounted for these changes. Also, the repeated measures regression model showed that with every one point increase in WM capacity, proficiency increased around 0.85 percent, indicating that over the course of a semester, an individual with a WM score ten points higher than average should show an increase in L2 proficiency nearly twice that of the average participant (p. 108).

Overall, these results suggest that greater WM resources predict gains in explicit L2 grammatical knowledge over the course of a short learning period for low-proficiency learners.

However, Linck and Weiss administered the same grammar and vocabulary portions of the Spanish *DELE* at test (beginning of semester) and retest time (end of semester). Therefore, the issue of test effects is a concern given that an increase in performance over time could reflect practice effects rather than actual growth in learner knowledge, although a three-month period could be long enough to prevent learners from remembering what they had seen at Time 1. Furthermore, there was no third assessment wave in the design, which would gauge not only learning gains but also retention and in so doing, capture the more complex, variable nature of L2 knowledge trajectories (i.e., positive/negative linear, curvilinear, plateaus, reversals, etc.).

Overall, the difference in findings between studies by Linck and Weiss and Sagarra (i.e., a *more is better* view versus no beneficial effects) regarding the relationship between WMC and L2 performance over time could be due to several factors. While both studies arguably measure explicit L2 knowledge (*DELE*—Linck & Weiss; UI cloze passage and sentence completion—Sagarra), certain methodological differences in how WMC was measured, the length of the period of learning, and the operationalization of beginning proficiency level could shed light on
the different findings. First, Saggarra’s measure of WMC (RSpan) did not tap simultaneous processing and storage ability whereas Linck and Weiss’s measure of WMC (OSpan) required both processing and storage demands, which is more likely to stretch learners’ abilities to their limits and accurately reveal variation in working memory capacity (Logie, 2011). Second, the period of learning was relatively long (1.5 years) in Saggarra’s study and considerably shorter (3 mos.) in Linck and Weiss’ design. It could be that WMC plays a stronger role in short-term rather than long-term development of explicit L2 knowledge for beginning classroom learners, but this remains pure speculation and more empirical research is needed to clarify the impact of the length of the learning period in revealing effects of cognitive IDs. Finally, beginning Spanish learners in Saggarra’s study were in their second and fourth semesters of instruction whereas beginning learners in Linck and Weiss were enrolled in first-semester German and third-semester Spanish conversation courses. Such variation in the operationalization of learner proficiency reduces comparability between the learner samples in each study.

Summary.

In sum, the majority of empirical research investigating the role of WMC in L2 classroom learning has provided evidence in support of the more is better hypothesis (Miyake & Friedman, 1998) as well as Just and Carpenter's (1992) single-resources model that predicts WMC effects on grammatical processing. Specifically, studies have reported a significant role for WMC in learning under different pedagogical conditions (Erlam, 2005; Robinson, 2002, 2005; Sanz, under review), sensitivity to number agreement during online processing in advanced learners (Coughlin & Tremblay, 2013), sensitivity to gender agreement during online processing in intermediate (Sagarra & Hershensohn, 2010) and beginning learners (Sagarra, 2007b), performance on tests likely tapping explicit knowledge in beginning learners (Kormos &
Sáfár, 2008; Linck & Weiss, 2011), and performance on tests designed to tap both implicit and explicit L2 knowledge in advanced learners (Erçetin & Alptekin, 2013), though the relationship with explicit L2 knowledge was more robust. Other empirical evidence indicates that the facilitative effect of WM components, the phonological loop in particular, likely narrows as learners gain proficiency in the target language (Kormos & Sáfár, 2008; O’Brien, Segalowitz, Collentine & Freed, 2006; Winke, 2005). Taken together, it seems that L2 proficiency seems to be a mediating factor as the majority of studies finding a beneficial role for WMC have investigated low proficiency learners, with the exception of Coughlin and Tremblay and Erçetin and Alptekin whose learners were advanced. However, the latter studies both measured WMC in the target language, likely causing a confound with L2 proficiency.

On the other hand, some research has reported that WMC does not impact L2 grammatical development for beginning learners (Sagarra, 2000), sensitivity to morphological agreement errors for advanced learners (Foote, 2011), or performance on tests assessing explicit metalinguistic knowledge by learners estimated to be at an intermediate level (Roehr & Gánem-Gutiérrez, 2009). Importantly, however, studies by Sagarra and Roehr and Gánem-Gutiérrez did not follow complex dual-task methodology requiring a simultaneous processing requirement and Foote scored WMC based on word recall only; such methodological limitations may have obscured true variation in learners' WMC. More empirical research is needed that controls for the above factors in order to clarify whether the relationship between WMC and L2 learning depends on the proficiency level of the learner and the type of L2 knowledge assessed.

**Psychosocial dimension of L2 learning.**

Thus far, I have discussed adult L2 learning from a cognitive, information-processing perspective that emphasizes similarities with learning other complex skills. However, learning
“is also socially and culturally bound, which makes language learning a deeply social event” (Dörnyei, 2003, p. 4). Indeed, many scholars have provided ample evidence that language learning success is not only constrained by cognitive function, but also by the social and psychological complexities underlying its development (e.g., Dörnyei, 2005, 2009; Gardner, 1985, 2001). Within the psychosocial dimension of L2 learning, the most studied learner-internal factor is motivation, with over five decades of research dedicated to theorizing and operationalizing its role in SLA (Ushioda & Dörnyei, 2012, p. 396). In the following sections, I detail relevant models defining L2 motivation and the quantitative approaches taken to measure this construct followed by a review of empirical research conducted within the socio-psychological and socio-dynamic frameworks.

Defining L2 motivation.

In overviews of L2 motivation research, four phases are typically identified: 1) the socio-psychological period (1959-1990); 2) the cognitive-situated period (1990’s); 3) the process-oriented period (late 90’s–early 21st century); and 4) the socio-dynamic period (current). The first framework was pioneered by Robert Gardner and colleagues with a focus on social context and attitudes toward the L2 community whereas the latter three phases have been led by Zoltán Dörnyei and colleagues in an effort to integrate L2 motivation research with mainstream motivational psychology and situate its study within specific learning settings (Phase 2), with a focus on motivational change (Phase 3), and within theories of self and identity (Phase 4).

In terms of successfully linking the construct of L2 motivation to L2 achievement, the earliest (Phase 1) and most current (Phase 4) frameworks are the most comprehensive and have received the most empirical support (Dörnyei, 2009; Ushioda & Dörnyei, 2012). In addition, and perhaps more importantly, there is considerable empirical evidence of significant relationships
between the core concepts driving the two frameworks, specifically, Gardner’s (1985, 2001) influential integrativeness motive and the ideal self in Dörnyei’s (2005) L2 Motivational Self-System. As such, the construct of L2 motivation in phases two and three is only briefly defined whereas its conceptualization and supporting empirical research within the first and last phases are elaborated in more detail.

The concept of motivation in L2 research originates within social psychology, or Gardner’s socioeducational model of SLA (Gardner & Lambert, 1959; Gardner & Lambert, 1972; Gardner, 1985, 2000, 2001, 2006, 2010). Gardner & Lambert (1959) originally defined it as the “effort and enthusiasm students show in their attempt to acquire the language” (p. 267). Gardner and Lambert (1972) later claimed that L2 motivation was a powerful cause of variability in SLA, independent of ability or aptitude, which has spurred on four decades of accumulated empirical data, mainly with L2 learners of French living in Canada, a English and French-speaking bilingual society. Within this model, L2 motivation comprises three main components: 1) Integrativeness (integrative orientation, interest in foreign languages, and attitudes towards the L2 community); 2) Attitudes toward the learning situation (attitudes towards the language teacher and the course); and 3) Motivation (effort, desire, and attitude towards learning). At the core of this conceptualization are individuals’ attitudes toward the target language community and their desire to identify with speakers of the target language, or integrativeness. Levels of motivation are influenced and maintained by attitudes toward the learning situation and integrativeness. Gardner has referred to the joint influence of all three components on outcome measures as the 'integrative motive' (2000, p. 18; see also Gardner & Smythe, 1974; 1981).

In the early 90’s several researchers questioned the relevance of integrative motivation for different learning contexts (e.g., second versus foreign language versus immersion settings)
which caused a shift in thinking about the construct of motivation under Gardner’s model (e.g., Crookes & Schmidt, 1991; Dörnyei, 1990; Skehan, 1989). From this initial call to reorient the research agenda, a more situated analysis of motivation emerged, emphasizing contextual aspects of motivation inherent in L2 instructional environments (e.g., Dörnyei, 1994; Julkunen, 1989, 2001). Cognitively oriented concepts of motivation from educational research came into focus such as intrinsic/extrinsic motivation, self-efficacy and attributions (Williams & Burden, 1997). Other relevant notions to emerge were trait versus state motivation, or enduring patterns of motivation over periods of time characteristic of the learner versus temporary engagement at particular points in time (Tremblay, Goldberg, & Gardner, 1995).

The temporal dynamics of motivation were further highlighted within a process-oriented model of motivation (Dörnyei & Ottó, 1998; Dörnyei, 2000, 2001; Ushioda, 1996a, b, 2001). Essentially, this conceptualization aimed to capture the “the ebb and flow” (Dörnyei, 2001, p. 16) of L2 motivation and better distinguish between the motivation to engage in L2 learning and the motivation to sustain that engagement (or not). The evolution of motivation is viewed gradually in this model and involves several phases such as initial planning and goal setting, intention formation, task generation, action implementation, action control, and outcome evaluation. These different subphases of the motivational process are associated with varying motives which can be influenced by different internal and external factors throughout a learning period. Though the process model of L2 motivation seems promising for studies of change, empirically it is challenged with clearly defining when a learning process begins and ends and accounting for interference from other actional processes (Dörnyei, 2005).

The most current conception of L2 motivation as articulated in the L2 Motivational Self System (Dörnyei, 2005, 2009; Dörnyei & Ushioda, 2009) is rooted in self-discrepancy theory
based on the notion of possible selves or ‘future self-guides’ (Markus & Nurius, 1986). Two of the most relevant self-guides in an academic context are deemed to be the ideal and the ought-to selves, the first being a representation of the attributes that one would ideally like to possess, while the latter refers to the attributes that one believes one ought to possess (Higgins, 1987). In other words, ideal self-guides promote one’s hopes, aspirations, and accomplishments while ought self-guides prevent and regulate negative outcomes associated with failing to meet responsibilities and obligations. Applying this line of thought to language learning, the basic tenet of Dörnyei’s theory is that if proficiency in the target language is integral to one’s ideal or ought-to self, this aspiration will serve as a powerful motivator to learn the language because of a psychological desire to reduce the discrepancy between current and future self states (Ushioda & Dörnyei, 2012, pp. 400-401). A third, less theorized and studied component is L2 learning experience which concerns situated, ‘executive’ motives related to the immediate learning environment and experience (e.g., the impact of the teacher, the curriculum, the peer group) (Dörnyei, 2009).

Measuring L2 motivation.

Following methodological approaches in mainstream cognitive and social psychology, L2 motivation has mainly been studied in a robust tradition of quantitative psychometric measurement, typically through extensive self-report. Other qualitative individual-centered approaches using interviews have been proposed by Ushioda (1996, 2001) in order to capture the non-linear complex relationships between context and individual. While interviews can offer rich insights into learner thought processes, this method is limited in scale and scope, a challenge when studying larger samples of learners. As such, below I detail those measurement efforts that have been quantitative in nature.
Under the socioeducational model, motivation is measured by Gardner’s (1985) *Attitude/Motivation Test Battery* (AMTB) originally developed for use with English-speaking Canadians studying French as a second language. It consists of a series of 11 scales of grouped Likert and multiple-choice items probing the attitudinal-motivational components in the theory, two sections of ‘semantic differentials’ evaluating the teacher and the course, language classroom and use anxiety, and *instrumental* and *integrative orientation* subtests which respectively assess one's practical and interpersonal reasons for language study. It yields composite indices of several variables that can be statistically analyzed in relation to other independent or dependent variables. Overall, the AMTB has proven to be a rigorous assessment with good psychometric properties and remains the only standardized test of motivation. In a meta-analysis of seventy-five empirical studies, it was shown to statistically explain a large amount of variance in L2 achievement in various L2 contexts with a “more than medium” effect size (.29-.39) (Masgoret & Gardner, 2003, p. 151).

This quantitative approach has been applied to constructing self-report scales in order to measure constructs like *ideal* and *ought-to selves* through Yes/No responses to current and possible future self description (e.g., ‘Be a knowledgeable person’) or Likert statement-type items (e.g., ‘If I made the effort, I could learn a foreign language’: 1-Strongly disagree to 6-Strongly agree) and question-type items (e.g., ‘Do you like meeting people from English-speaking countries?’: 1-Not at all to 6-Very much). Variations of these scales have been administered to over 6,000 learners of various ages and proficiency levels studying in five different countries; the validity of the self-system has specifically been demonstrated with L2 learners in China (Taguchi, Magid, & Papi), Hungary (Csizér & Kormos), Iran (Taguchi et al.), Japan (Ryan; Taguchi et al.) and Saudi Arabia (Al-Shehri), all reported in Dörnyei and Ushioda’s...
co-edited volume (2009) and results display good internal consistency reliability (Dörnyei & Ushioda, 2009b, p. 353). In particular, Cronbach's alpha for criterion measures (related to intended effort), ideal L2 self and ought-to L2 self scales are reported to be .83, .89, and .76, respectively (Dörnyei & Taguchi, 2010, p. 123).

Interestingly, however, the ideal L2 self and Gardner’s integrative motive are not separate notions, and actually overlap conceptually and empirically. They not only converge on the core issue of identification but elements of the integrative motive consistently correlate with possible selves (Dörnyei & Csizér, 2002; Csizér & Dörnyei, 2005; MacIntyre, Mackinnon, & Clément, 2009). In a large-scale survey measuring several attitudinal/motivational dimensions related to Gardner’s theory and administered to 13,000 Hungarian adolescents, Dörnyei and Csizér (2002) and Csizér and Dörnyei (2005) found convincing evidence through structural equation modeling that in fact, integrativeness could and should be reinterpreted from an ideal L2 self perspective. Dörnyei (2005) claimed that the L2 self “presents a broader frame of reference with increased capacity for explanatory power” (2005, p. 104) since it can be used to explain variation in a wider range of second and foreign language learning situations and on average explains around 10% more in criterion measures compared to integrativeness (42% versus 32%), according to Dörnyei.

MacIntyre, Mackinnon, & Clément (2009) also found consistently significant correlations between possible selves measures and integrativeness and motivation scales ranging between 0.54 to 0.76 in a study investigating motivation in 135 English-speaking French high school students. Nonetheless, MacIntyre et al. (2009) advise caution with selves research which invites questionable veracity and impartiality. Statistically, there is still no established measurement instrument that compares to the AMTB, and the situated component of the integrative motive,
attitudes toward the learning situation, has no parallel operationalization in the L2 Self System. Furthermore, the correlations between integrativeness and the ideal L2 self “are not so high as to preclude differential predictions of language learning outcomes” (MacIntyre et al., p. 208).

**Empirical research on L2 motivation.**

In this section, I review four relevant empirical studies that implemented general and situation-specific components of the AMTB (Gardner, Masgoret, Tennant, & Mihic, 2004; Yanguas, 2007) and possible selves questionnaires (Dörnyei & Chan, 2012; Ryan, 2009) to study the relationship between L2 motivation and achievement in college-age L2 learners.

One critique of L2 motivation as measured by the AMTB is that it portrays motivation as a static construct at one point in time. However, some research provides evidence that the AMTB is a sensitive measure of motivation along several points during a learning period as well (Gardner, Masgoret, Tennant, & Mihic, 2004; Yanguas, 2007). Gardner, et al. (2004) explored motivational change over the course of a year in university intermediate French learners ($N = 197$) in Canada. Importantly, most of the students had taken French from the fourth grade until the end of secondary school, and a number of them had been enrolled in a French immersion program for some of their schooling (p. 10). The authors measured both general motivation (AMTB) in the first and last testing sessions and state motivation (attitudes toward the course and teacher) and state anxiety (related to the course) at four points in between via a short questionnaire. Fifty-seven learners were present for all six testing sessions whereas ninety-one learners participated in the first and last testing session.

A Repeated Measures ANOVA revealed there was little change in more general areas (e.g., interest in foreign languages, attitudes toward learning French) whereas there was significantly greater change in measures referring to the classroom environment directly (French
class anxiety, motivational intensity, French teacher evaluation, and French course evaluation) (p. 19). In terms of level of achievement (i.e., course grade), it was found that achievement was significantly affected by Motivational intensity (i.e., reported effort expended to learn French), French course evaluation, and French class anxiety.

Yanguas (2007) studied motivational evolution both during task completion and over the course of a semester in university Spanish heritage language (HL) learners ($N = 44$) who were enrolled in an introductory course for Spanish native speakers. Yanguas used both quantitative and quantitative techniques to probe the construct of motivation; he conducted learner interviews at three points in the semester and to assess general motivational variables, he adapted a version of the 78-item AMTB for Hispanic native learners of Spanish which learners completed in the first two weeks of the semester. To address more situational aspects of motivation, semantic differential assessments from the AMTB were used to determine attitudes toward the Spanish course and teacher (ATLS) and this section was administered at the beginning and end of the semester. Situation-specific task motivation questionnaires within the process-oriented model (Dörnyei & Ottó, 1998) were given before and after closed and open-ended tasks during which online verbal reports were also collected.

Results based on ANOVA, correlation and multiple regression analyses revealed that Language Anxiety was the only AMTB variable to significantly account for performance variance in scores on listening and reading tasks whereas no AMTB variable was a significant predictor of writing or speaking scores in this learner population. As expected, the process-oriented model was found to be a more consistent predictor of task performance overall. However, contrary to Gardner's claims, Yanguas found that instrumental, rather than integrative orientation, was more implicated in the relationship with task-specific motivational variables (p.
Overall, Gardner’s socio-educational model was reported to efficiently account for course-related aspects of motivation over the semester.

Ryan (2009) aimed to validate Dörnyei’s concept of an ideal L2 self in relation to integrativeness and replicate methodology and results from his Hungarian studies in a Japanese educational context. A total of 2,397 learners of English enrolled in primary or secondary institutions completed a questionnaire probing seven motivational factors including, but not limited to, attitudes to learning English, ideal L2 self, interest in foreign languages and international contact (an element of international posture-Yashima, 2002). However, contrary to the two studies reported above, L2 motivation was studied in relation to a non-achievement measure operationalized as intended learning effort, a scale the author used to correlate to the main motivational dimensions. Results revealed a strong relationship between integrativeness and the ideal L2 self \((r = 0.59)\), supporting previous claims that the two concepts may be tapping into the same pool of identification learners feel towards the target language and its native speakers. However, for university non-English majors, the ideal L2 self was more closely related to intended learning effort \((r = 0.74)\) than integrativeness \((r = .61)\). For English majors, relationships were similar \((0.71 \text{ vs } 0.54)\). Ryan interprets these findings to empirically support the need to reinterpret L2 motivation from the perspective of projected future self states, arguing, along with others (e.g., Dörnyei, 2003, 2005, 2009) that “integrativeness is simply one local manifestation of a much more complex, powerful construct” (p. 137).

Kormos and Csizér (2008) also provided strong empirical support for the main self construct of Dörnyei’s model relative to integrativeness. They reported that language learning attitudes and the Ideal L2 Self were the main factors affecting L2 motivation for 623 native Hungarian speakers learning English in secondary, university and adult foreign language
learning contexts. Correlations with Integrativeness for each population of learners were substantially lower than those found with the Ideal L2 self (e.g., for university and adult learners, respectively: \( r = .51 \) vs. \( .75 \); \( .38 \) vs. \( .85 \)).

Dörnyei and Chan (2012) also reported strong correlations between the ideal L2 self and L2 English grades (\( r = .24 \)) and L2 Mandarin grades (\( r = .42 \)) for 172 low-intermediate adolescent learners of English and Mandarin in Hong Kong (Dörnyei & Chan, 2012). They also found differences in the strength of the relationship between the Ideal L2 self and *self-reported learning effort* and course grades, with the former relationship having much higher effect sizes (\( r = 0.67-0.68 \)). This supports the argument made by Masgoret and Gardner (2003) that correlations with L2 motivation depend on the nature and type of measurement of achievement. They maintain that self-reported ratings "could be expected to correlate more highly with motivation, since there is more opportunity for motivation to be implicated" (p. 156) versus a more objective criterion of learner ability like course grades.

**Summary.**

As can be seen, L2 motivation is a powerful internal difference factor related to variability in adult language learning success. Despite undergoing several conceptual transformations throughout the years, the core elements driving learner motivation and related behavior seem to converge on the notion of self-identification with the target language and L2 speaker community. While recent research has argued for the superiority of the ideal L2 self over integrativeness in capturing the relationships between L2 motivation and learning (Csizér & Dörnyei, 2005; Dörnyei, 2005; Kormos & Csizér, 2008; Ryan, 2009), some researchers remain cautious about the methodology used in selves research (e.g., MacIntyre et al., 2009) and the veracity of relationships found with self-reported effort ratings (Masgoret & Gardner, 2003) is
questionable. Furthermore, no research to date has reported using possible selves questionnaires to investigate its relationship with L2 achievement over time or in a US language-learning context. The AMTB has proven to be an effective measure of L2 motivational change and its relationship to achievement over a year (Gardner et al., 2004) and a semester of learning (Linck & Weiss, 2011; Yanguas, 2007). Another benefit of the AMTB is a well-developed and tested scale to assess attitudes toward the learning situation whereas the L2 Motivational Self System does not currently offer a clear measure of a situated component in this regard. Overall, it seems for now that both models of L2 motivation should be used to measure this dynamic dimension of L2 learning.

L2 Development over Time

Overview

Learning in any capacity happens over time. As asserted by Ortega and Iberri-Shea (2005), many questions about the rate, route, and outcomes of L2 learning are “fundamentally questions of time and timing” (p. 26). Nonetheless, the bulk of SLA research remains cross-sectional in nature, providing a snapshot of L2 ability at one point in time. Studies that collect two waves of data are only marginally better as they conceptualize change incrementally as the difference between scores measured on two assessment occasions (Willett, 1989), also known as a difference or gain score. To sufficiently probe the construct of change and pursue questions about: 1) within-individual change—How does each person change over time?—and 2) between-individual differences in change—What predicts differences among people in their change?, longitudinal data is necessary (Singer & Willett, 2003, p. 4).

Different conceptions of "longitudinal" L2 research.

Early longitudinal studies of L2 learners (e.g., Huebner, 1983; Schumann, 1978) mainly
focused on the interlanguage development of L2 morphosyntax of one or two participants over a period of at least one year and this small-scale case study approach continued into the 90’s (e.g., Lardiere, 1998; Sato, 1990). These studies typically aimed to document the developmental stages learners pass through and the errors they make along the way, errors that have been shown to follow a universal sequence regardless of instructional or curricular presentation and that notoriously exhibit U-shaped or zig-zag behavior (Long, 2007). Thus, early longitudinal research has been strongly motivated by a universal grammar (UG) perspective with the goal of identifying the developmental stages in L2 acquisition and determining the extent to which they mirror L1 acquisition.

In a review of thirty-eight recent longitudinal studies in SLA published between 2002 and 2004, Ortega and Iberri-Shea (2005) analyze how studies defined and documented change over time and discuss the strengths, weaknesses and challenges in longitudinal research. They report that longitudinal work has been undertaken from several research perspectives including quantitative studies of linguistic features and qualitative studies focused on sociocognitive and sociocultural dimensions of L2 learning. Furthermore, almost half the studies reviewed investigated some aspect of L2 learning by college-level populations and the time scale structuring data collection generally spanned anywhere between three or four months and up to six years, interpreted by the authors to reflect the one-semester and four-year periods of time that structure most educational contexts where SLA research is conducted (p. 37).

Another line of research conducted within a neurocognitive framework has implemented a longitudinal design to study L2 online comprehension processes in adult English-speaking novice French learners after one month (30 hours), four months (80 hours), and eight months (136 hours) of university classroom instruction (e.g., McLaughlin, Tanner, Pitkänen, Frenc-
The goal of this research has been to determine how much L2 exposure is needed before learners can efficiently discriminate between syntactically well-formed and ill-formed L2 sentences, as measured in event-related brain potentials.

**Defining L2 development over time.**

The question that still remains is what constitutes longitudinal data in L2 learning; however, the answers remain vague. Without defining 'period of time', R. Ellis (2008) determines longitudinal studies to be those that “experimentally elicit data from a single learner over a period of time” whereas pseudo-longitudinal studies compare “groups of learners at different stages of development” (p. 588). Ortega and Iberri-Shea's review indicates that more recent studies have chosen to define time periods based on feasibility and practicality for the researcher, rather than theoretically-based definitions. As such, they advocate for refining and broadening our working definition of “longitudinal” SLA research (p. 28) based on a variety of temporal units—months, years, semesters, sessions, etc. Singer and Willett (2003) suggest simply choosing a time metric that makes most sense for one’s research questions and experimental design, but that includes a sufficient number of data collection waves (ideally 3 or more) (p. 11) that allow for modeling complex growth patterns such as "no change", positive and negative "linear change", "quadratic change" and "cubic change" trajectories (pp. 214-216).

Given the issues of feasibility and practicality of studying participants over long periods of time, the current study adopts one academic semester as a meaningful period of time to study L2 learner development while pursuing an innovative statistical approach rarely used in SLA research (Ortega & Iberri-Shea, 2005) but that provides a robust means of analyzing intra- and inter-learner change over time.
Measuring L2 development over time.

Singer and Willett (2003) assert that studying change in the phenomenon of interest requires i) multiple waves of data, ii) a substantively meaningful metric for time, and iii) an outcome that changes systematically, whether that outcome grows or increases over time, or takes a more complex trajectory, including decreases, plateaus or reversals, as is the case in L2 learning. Therefore, statistically, it is essential to employ a meaningful model that is capable of analyzing multiple levels of change such as time-series design (Hatch & Lazaraton, 1991; Mellow, Reeder, & Forster, 1996) or multi-level hierarchical linear models (Singer & Willett, 2003). The latter technique is able to characterize both individual and latent, or group, growth trajectories based on individual growth parameters (i.e., intercept, or baseline performance & slope, or rate of change) over time (Level-1). It is also capable of detecting heterogeneity in change between individuals (Level-2) by studying interindividual variation in the growth parameters and predictors of that variation (Singer & Willett, pp. 53-54).

Hierarchical linear modeling, or latent growth curve analysis (GCA), is a technique that has almost exclusively been used to study factors predicting academic achievement in child populations in the fields of education and psychology but has yet to be consistently applied in the realm of L2 research where repeated-measures analyses of variance dominate the study of change (Heck, Thomas, & Tabata, 2010). Only a handful of published studies found in a linguistics database search have used this technique to chart vocabulary development in bilingual children (e.g., Kan & Kohnert, 2012) and grammatical development in children at risk for developmental disorders (e.g., Rice, Wexler, Marquis, Hershberger, 2000). Only one published study in SLA has employed GCA to investigate what cognitive IDs (PSTM and phonological
awareness) predict vocabulary development from first to sixth grade in English language learners (Farnia & Geva, 2011), based on Farnia's dissertation study (Farnia, 2006).

This is a considerable limitation considering that GCA models have certain benefits over RM ANOVAs including i) the ability to handle missing values in the participant sample, ii) to accommodate not only continuous dependent variables (DV$s$) but ordinal DV$s$ as well (i.e., ‘yes’/’no’), iii) to specify exact time intervals between data collection waves for each subject, iv) to analyze complex sets of data with many independent variables and/or covariates;, and v) to capture (non)linear change (see Heck, Thomas, & Tabata, 2010 for an introduction to multilevel & longitudinal modeling with IBM SPSS).

In sum, GCA represents a promising direction for SLA research that aims to study L2 development over time and includes at least three measurement waves. A second wave of data collection in SLA typically aims to gauge the immediate effects of different instructional treatments while a third testing time intends to measure retention or loss of L2 skills, typically based on performance on delayed post-tests administered one to four weeks after instruction (Norris & Ortega, 2000, p. 457). However, the majority of SLA studies focus on immediate (short-term) development and thus far, there is still little research to date that has provided evidence of what learner ID factors promote retention or prevent loss of L2 knowledge.

**The study of retention of L2 skills.**

Gardner (1982) proposed three different time points in the language acquisition/language attrition process. Time 1 represents the beginning of language study, Time 2 is the termination point of L2 training or instruction, and Time 3 is a later point in time when retention is assessed. He termed the period between Time 1 and Time 2 as the *acquisition phase* and that between Time 2 and Time 3 as the *incubation period*. Frequent periods with limited L2 exposure or
instruction are common during the course of learning and loss of L2 skills likely affect “the development of most (perhaps all) L2 learners” (Bardovi-Harlig & Stringer, 2010, p. 39). However, certain factors may help prevent loss of L2 skills during "incubation" phases such as frequency of L2 use, previous level of proficiency obtained, and level of L2 motivation (Bardovi-Harlig & Stringer, 2010). This research is briefly reviewed below despite little empirical evidence on the effects of a period of limited or no L2 exposure, especially for L2 grammar, as well as the effects of learner IDs on L2 retention, cognitive IDs in particular.

In general, productive skills (speaking & writing) are more prone to loss than receptive skills (listening & reading) following a period of no L2 exposure or instruction (Bardovi-Harlig & Stringer, 2010). However, few studies have investigated retention of grammatical performance (p. 24), excluding those that included a delayed post-test one to four weeks after instruction (e.g., Erlam, 2005; Sanz, under review). However, a few studies have investigated retention of L2 grammar based on performance outcomes in Spanish after a three to five year period of limited exposure (Bahrick, 1984) and in French after a two year delay (Weltens, Van Els, & Shils, 1989). A more recent study assessed both performance and neural processing outcomes around five months after undergoing explicit or implicit training in an artificial language (Morgan-Short, Finger, Grey, & Ullman, 2012). Other studies have measured learners’ perceived loss of their receptive and productive skills in French after a 9-month absence of instruction (Gardner & Lysynchuk, 1990), after summer vacation (Gardner, Lalonde, Moorcroft, & Evers, 1987), or six months after completing an intensive French course (Gardner, Lalonde, & MacPherson, 1985). These studies have found that certain individual differences such as a high level of proficiency (Morgan-Short, et al., 2012; Bahrick, 1984; Weltens et al., 1989; Gardner, et al., 1985) and high level of motivation as measured on the AMTB (Gardner et al., 1985; Gardner et al., 1987;
Gardner & Lysynchuk, 1990) promote language maintenance over loss (or learner-reported perception of loss). However, perception should not be equated with behavioral evidence of L2 knowledge.

Furthermore, controlling for what happens during the 'period of no exposure' is key to demonstrating evidence of retention. Morgan-Short et al. reported retention in terms of behavioral performance and neurocognitive processing patterns after five months of no exposure to an artificial language indicating that a substantial period of no exposure “does not necessarily lead to lower proficiency (use it or lose it) in terms of performance, and in fact can even lead to increased native-like neural processing” (p. 14). This finding held independently of whether learners had undergone explicit or implicit training. However, before being tested after the five-month delay, participants engaged in a warm-up session, which could have activated prior knowledge and as a result, affected learner processing and performance outcomes. Therefore, this is a methodological concern that researchers should take into account when investigating performance after a period of no exposure.

Summary.

To sum up, a key objective in studying change in learners' interlanguage knowledge should be to analyze the developmental trajectory both within and between individual learners according to at least three points in time and to use robust statistical methods that avoid the assumption that change is linear. Overall, SLA research is lacking a long-term perspective of L2 development and of the factors that may predict differences in retention, or loss, of L2 grammar. We need more empirical evidence of the relative ability of several learner IDs in predicting retention, including not only L2 proficiency and motivation, but cognitive capacity as well.
I now turn to the question of what aspects of the interlanguage system are relevant to measure in performance. Given that learners in this study were at three levels of proficiency in L2 Spanish, a range of grammatical features were chosen that vary in difficulty, in line with previous research (Bowles, 2011; Gutiérrez, 2012).

**Development of L2 Spanish**

**Overview**

In this section, I give a brief overview of Spanish grammar and discuss the selected target structures in relation to the developmental stages of acquiring Spanish as a first and second language (Liceras, 1996; López Ornat, 1994, 1997; Montrul, 2004) and factors influencing L2 grammar difficulty in general (DeKeyser, 2005). Theoretical issues in L2 acquisition such as whether or not L2 learners have no access, partial access or full access to Universal Grammar (UG) are outside the scope of this study but the reader is referred to Bley-Vroman (Fundamental Difference Hypothesis—1989, 1990), Montrul (2004), Schwartz and Sprouse (Full Transfer/Full Access Hypothesis—1996) and White (1989, 2003) for full treatment of these questions.

**Spanish grammar.**

In terms of its morphosyntactic characteristics, Spanish is a nominative/accusative S-V-O language with rich nominal and verbal inflectional systems (Montrul, 2004, pp. 26-27). It is a head-initial language where the head of the phrase precedes its complements and exhibits null subject phenomena in which subject pronouns can be expressed overtly or not, which is related to its rich verbal agreement (e.g., Construyeron un puente “(they) built a bridge”—Zagona, 2002, p. 7). Nouns, adjectives, past participles, and personal, relative, and interrogative pronouns are inflected for gender and number, but not for case (except for pronouns). In the noun phrase, there must be gender and number agreement among nouns, determiners, and adjectives (e.g., esos
niños traviesos “those naughty boys”—Montrul, p. 32). Finite verbs are inflected for person, tense (e.g., present-(Yo) hablo, past-hablé, future-hablare), aspect (preterite-hablé, imperfect-hablaba, progressive-estaba hablando), and mood (indicative-hablo, subjunctive-hable, conditional-hablaría).

L2 acquisition of Spanish morphology and syntax.

As asserted in Chapter 1, L2 learning outcomes are distinct for children and adults. Whereas children uniformly reach native or native-like competence, late adult learners vary greatly in the ultimate level of competence they achieve. However, one similarity between children acquiring Spanish as their L1 and adults acquiring Spanish as their L2 is that both groups of learners have been found to make systematic errors in verbal inflection and tense and agreement morphology (López Ornat, 1994; Montrul, 2004), regardless of the source and duration of such errors. The most persistent problems for English-speaking learners of Spanish pertain to its rich inflectional system, or those L2 features not present in their L1 grammatical system such as morphological agreement (e.g., verbal inflection, grammatical gender), aspect, and mood (DeKeyser, 2005).

DeKeyser highlights three general factors in determining grammatical difficulty for adult L2 learners: i) complexity of form, ii) complexity of meaning, and iii) complexity of the form-meaning relationship. He asserts that the actual difficulty of grasping the form-meaning relationship during L2 sentence processing lay in the degree of “transparency of form-meaning relationships to a learner” (p. 3). Aspect and mood express abstract notions, which together with the morphemes inflected to encode such meanings, create opaque form-meaning mappings and cause problems for English L1 speakers, especially in initial stages of learning when morphology remains a weak cue for meaning (DeKeyser, 2005, p. 7; see MacWhinney, 2001). Both external
and internal factors constrain acquisition of these ‘advanced’ areas of L2 grammar such as salience and frequency in the input (external) and storage and cognitive processing efficiency (internal).

Following these considerations, the seventeen targeted grammatical structures used in Bowles (2011) and Gutiérrez (2012) were reduced to ten in order to focus on key problematic structures for adult native English-speaking learners of Spanish and to increase grammatical and ungrammatical exemplars in the assessment tests. They included i) agreement morphology (subject-verb, noun-determiner and noun-adjective gender agreement), ii) constructions with *gustar* “to be pleasing to”, iii) aspect (preterit-imperfect aspectual distinction), iv) the copular contrast, *ser/estar* “to be” (use with locatives and events), and v) mood (present subjunctive in noun clauses, imperfect subjunctive, past hypothetical conditionals) (See Appendix A for grammatical/ungrammatical exemplars of each targeted grammatical structure).

*Subject-verb agreement.*

Spanish-speaking children produce finite, inflected verbs very early, around 1-2 years old, with systematic, but relatively infrequent errors (Bel, 2002; López Ornat, 1997). First and 3rd person singular forms appear first and most frequently, followed by 2nd person singular, then 1st and 3rd personal plural and finally 2nd person plural forms. Children display accurate production of verbal morphology in present, past and future contexts by age 4 (Montrul, 2004, p. 146).

On the other hand, morphological agreement errors in this domain are highly variable for English native speaking learners of Spanish in early and intermediate stages of L2 learning. Learners can either eventually overcome these errors at higher levels of proficiency or they can fossilize. Bruhn de Garavito (2003) studied beginning learners of L2 Spanish and found high error rates of agreement in the elicited production data in terms of person (e.g., 3rd person for 2nd...
person) and use of infinitives for finite verbs (e.g., *Yo hablar* instead of *Yo hablo* ‘I speak’). For L2 Spanish learners with mid-to-advanced proficiency, Bowden, Gelfand, Sanz and Ullman (2010) report frequency effects (based on reaction times (RTs)) for four classes of regular (stem-changing) and irregular (non-stem-changing) verbs in the first-person singular present and imperfect forms. The authors interpreted this result as suggesting that L2 learners store inflected verb forms as memorized representations, and do not necessarily rely on L1 combinatorial processes.

*Gender agreement: Articles and adjectives.*

Gender agreement within the noun phrase emerges early in native Spanish-speaking children, between 1 and 2 years old, followed by number agreement (López Ornat, 1994, 1997, 2003). Around 2 years old, determiner-like elements precede nouns and are termed *protodeterminers* (López Ornat, 1997) in that they seem to mark gender (e.g., *a boca* ‘the mouth’; *e culito* ‘the bootie; *e pie* ‘the foot’—1994, p. 306). Noun-determiner gender agreement is eventually mastered by 3 or 4 years of age. In terms of noun-adjective agreement within the noun phrase, especially between adjectives and inanimate nouns, children go through periods of overgeneralization (e.g., *mota roja* instead of *moto roja* ‘red motorcycle’; *tierra azula* instead of *tierra azul* ‘blue soil’) and errors can persist until around 3 years old (Hernández Piña, 1984). However, noun-adjective agreement is usually mastered by age 4 (Montrul, p. 57).

As English does not have gender and number agreement between nouns, articles and adjectives, it tends to be another area of persistent difficulty for native English-speaking L2 learners of Spanish. Research has typically found that learners with incomplete knowledge of the language tend to overextend the default (masculine) form in Spanish (e.g., *Veo un nariz rojo* instead of *Veo una nariz roja* ‘I see a red nose’) and are typically most inaccurate with feminine
gender (Montrul, Foote, & Perpiñán, 2008, p. 523). However, these issues apply more in the case of oral rather than written production. In terms of processing, it has been found that beginning (Sagarra, 2007b, 2008) and intermediate learners (Sagarra & Hershensohn, 2010) with high working memory capacity are sensitive to noun-adjective gender disagreement and learners with high proficiency of an artificial language can show evidence of L1-like neural patterns of gender agreement processing between articles and nouns, but not between adjectives and nouns (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010). The latter neurocognitive findings held for both implicit and explicit instructed conditions.

_Gustar-type dative experiencer constructions._

To the author’s knowledge, little to no longitudinal empirical work exists on the topic of L1 acquisition of “psych” verbs (Belletti & Rizzi, 1988) in Spanish. On the other hand, L2 acquisition of these verbs denoting mental states is complex for L1 English speakers, primarily due to word order (Sanz, 1999; VanPatten, 1984) and obligatory clitic doubling (Montrul, 1997). English relies on word order for functional assignment of sentential roles whereas Spanish relies on morphology (VanPatten, 1984). Thus, in Spanish these dative ‘experiencer’ constructions can display canonical (S-V-O) or non-canonical word order (O-V-S), with the latter being the more frequent, unmarked order seen in example (1).

1. **A Tomás le gustan las hamburguesas.**
   To Thomas-dat. 3s-dat. likes the hamburgers-nom.
   ‘Thomas likes hamburgers.’

   In the English equivalent to (1), Thomas takes nominative case as the agent of liking the hamburgers (accusative case) whereas in Spanish, Thomas experiences the state of liking and the hamburgers function as the subject of the sentence. The difficulty of acquiring these constructions persists for advanced and even near-native speakers of Spanish (Sanz, 1999, pp. 4-
5); however, some studies report positive effects for instruction and feedback on the acquisition of *gustar*-type verbs for beginning learners (e.g., Bowles, 2005; Hsieh, 2008). In particular, these studies have found that feedback can facilitate development of *gustar*, with a significant advantage for explicit feedback in terms of short-term development and for both implicit and explicit feedback in delayed L2 performance.

*Aspect.*

Aspect enables a sentence to denote a complete or incomplete event representing “different ways of viewing the internal temporal constituency of a situation” (Comrie, 1976, p. 3). As noted earlier, Spanish-speaking children can use more than one tense form contrastively relatively early, between 1 and 2 years old (e.g., Bel, 2002). The [±perfective] distinction between the preterit and imperfect in Spanish is realized through inflectional morphology but the emergence of past morphology is linked to verbal aspect, the interpretive properties of which depend on cognitive development. Thus, it has been found that the preterit precedes the imperfect in child L1 learners of Spanish (Bel, 2001; Hernández Piña, 1986; López Ornat, 1994, 1997).

Verbal aspect is notoriously hard to acquire for natives speakers of English since English does not overtly distinguish aspect in the past tense (Salaberry, 2000, p. 4). In Spanish, the perfective viewpoint is represented by simple past tense morphemes whereas the imperfective viewpoint is represented by the progressive past tense, or imperfective morphemes. For example, in (2) the event was completed and therefore realized in the preterit whereas in (3) the event was habitual which requires inflection of imperfect morphology.

2. Luis compr-ó una casa.
   Luis buy-PRET a house.
   ‘Luis bought a house’
On the other hand, as the above translations show, English uses the simple past tense to convey a completed event and aspectual particles (i.e., modal verbs) and the progressive tense to convey the imperfective, habitual viewpoint.

Much empirical inquiry into this domain has mainly centered on testing the predictions of the Aspect Hypothesis (Andersen, 1991; Andersen & Shirai, 1994, 1996). This hypothesis proposes eight stages of aspectual acquisition based on naturalistic data from two Spanish L2 learners in Puerto Rico. Studies that tested the hypothesis for instructed learners at intermediate and advanced proficiency levels (e.g., Bardovi-Harlig, 1992; Bardovi-Harlig & Reynolds, 1995; Salaberry, 1999, 2000) have generally found support for the proposed stages though there has been less research with beginning learners (Tracy-Ventura, 2008, p. 22). For learners at initial stages of proficiency, a default past tense marker (the preterit) has been proposed (Salaberry, 1999, p. 167), reflecting the earlier emergence of past tense morphology in the perfective, similar to L1 acquisition. Other research has focused on the interpretations that intermediate and advanced L2 learners and Spanish native speakers assign to tense/aspect morphemes (e.g., Montrul & Slabakova, 2000; Slabakova & Montrul, 2002). They report that L2 learners are eventually able to overcome parametric differences and acquire the complex semantic distinctions encoded in Spanish verbal morphology similar to native speakers.

*Ser versus Estar: The Spanish copular contrast.*

The contrast between the two copulas *ser* and *estar* ‘to be’ in Spanish affects several different functional contexts. Locative and preadjectival contexts where both *ser* and *estar* can be used are thought to be the most difficult to acquire for non-native speakers of Spanish (Geeslin,
2002, 2003 p. 706). In the current study, I focus on locative environments in which *ser* is exclusively used to refer to the time, date or location of an event and *estar* is used as the default for all physical locations of animate and inanimate objects. This distinction can be seen in examples (4) and (5), respectively:

4. **Gabriel está en Puerto Rico.**
   Gabriel is-PRES in Puerto Rico
   “Gabriel is in Puerto Rico”

5. **El bautismo es el sábado en la iglesia.**
   The baptism is-PRES the Saturday in the church
   “The baptism is on Saturday at church”

Early research on the L2 acquisition of copula choice identified stages of development (Guntermann, 1992; Ryan & Lafford, 1992; Van Patten, 1985, 1987) whereas more recent research focuses on aspectual properties of *ser* and *estar* with the former unmarked for aspect and the latter marked [+perfective] (VanPatten, 2010; Bruhn de Garavito & Valenzuela, 2008). VanPatten (1985, 1987) first identified five stages through which beginning instructed English-speaking Spanish learners pass as they acquire copula choice in Spanish: 1) Absence of copula in learner speech (*Ella—alegre ‘She—happy’*); 2) Selection of *ser* to perform most copula functions (*Soy Juan ‘I am Juan’*); 3) Appearance of *estar* with the progressive (*Estoy hablando ‘I am speaking’*); 4) Appearance of *estar* with locatives (*Estoy en clase ‘I am in class’*); and 5) Appearance of *estar* with adjectives of condition (*Estoy alegre ‘I am happy’*). Ryan and Lafford (1992) and Gunterman (1992) generalized VanPatten’s stages to naturalistic learning environments but found a reverse order of stages four and five.

*Mood.*

Mood refers to the grammatical expression of modality, a semantic term, and refers to the probability, obligation or necessity of what is stated, according to the point of view of the
speaker (Comrie, 1976). Whereas the indicative mood communicates a commitment to the truth-value of the statement, the subjunctive indicates a lack of such commitment (Palmer, 2001). Modality in Spanish is a morphosyntactic category that distinguishes between present, past, and future forms and choice of indicative or subjunctive is signaled by syntactic and semantic elements in the sentence (Montrul, 2004, p. 100). For example, *ir* “to go” is conjugated below in three contexts requiring the use of the present and imperfect subjunctive in the subordinate clause because the speaker expresses present (6), past (7) or past hypothetical desire (8) in the independent clause.

6. Quiero que **vayas** a la fiesta.
   I want that you go-SUBJ to the party.
   ‘I want you to go to the party.’

7. Quería que **fuerais** a la fiesta.
   I wanted that you go-past-SUBJ to the party
   ‘I wanted you to go to the party.’

8. Estaría muy feliz si **fueras** a la fiesta.
   I would be very happy if you go-past-SUBJ to the party
   ‘I would be very happy if you went to the party.’

Whereas by age 3 or 4, L1 child learners of Spanish have typically mastered a variety of verbal forms, the complexity of the semantics of Spanish subjunctive mood is acquired later, between ages 5 and 7 (Montrul, 2004, p. 163) as it depends on cognitive maturation. On the other hand, cognitive maturity should not be an obstacle for adult L2 learners. Nonetheless, learners at intermediate and advanced levels of proficiency still experience considerable difficulty in achieving accuracy in this domain (e.g., Collentine, 1995, 2003, 2010), which is likely related to the need to process information across clauses, from a cognitive perspective (Johnston, 1995). For example, Collentine (1995) found that problems in morphological inflection stem from learners’ poor control of complex syntax in that they produce simple syntax rather than
subordinate clauses required in most subjunctive contexts. Furthermore, Fernández (2008) and Leow (1993) argue that the salience of the Spanish subjunctive in comprehensible input is low and learners do not consistently notice it, in line with VanPatten’s (1997) primacy of meaning principle. Other researchers such as Bruhn de Garavito (1997) find that highly advanced and near-native speakers can eventually overcome parametric differences in subjunctive mood between Spanish and their native language, at least in terms of accurate interpretation.

Summary.

Overall, tense, aspect, and mood morphology emerges systematically in both child L1 and adult L2 Spanish learners (e.g., singular before plural, masculine before feminine, present before past before future, preterit before imperfect, indicative before subjunctive—Montrul, 2004). However, the duration of nontargetlike use is distinct; children quickly become native users of the language, displaying targetlike use of most morphological and syntactic domains quite early, with the exception of the subjunctive mood. In contrast, errors in agreement morphology and morphosyntax can persist into advanced stages of adult L2 learning, especially in complex domains of aspect and mood with opaque form-meaning relationships. The difficulty of processing complex grammatical features that are not part of the L1 system is likely related to a limited cognitive capacity to process information, which constrains L2 development in adults.

Summary of Gaps in Previous Research Investigating Relationships Between Learner IDs and L2 Development

A review of the literature has revealed several gaps in our understanding of the role learner IDs may play in explaining the highly variable nature of adult L2 development over time. In particular, it is apparent that very little work has jointly investigated cognitive and psychosocial learner ID factors; this is a considerable limitation considering that investigating IDs in isolation could result in overemphasizing the impact of any one variable in particular
(Ackerman, 2003). As noted in Chapter One, researchers in cognitive and educational psychology have long acknowledged a cognition-motivation interface (e.g., Snow, 1989, 1994) and some researchers in SLA (e.g., Hulstijn, 2002; Larsen-Freeman & Cameron, 2008; Segalowitz & Trofimovich, 2012) have begun to emphasize that the cognitive, motivational, social, and experiential aspects of L2 use likely interact in complex ways. However, systematic inquiries into the nature of these potential relationships are currently lacking.

For example, as discussed by Segalowitz and Trofimovich, having more efficient cognitive resources may increase a learner’s motivation to engage in L2 contact compared with a learner whose cognitive processing efficiency is relatively low. Likewise, a learner with high motivation and positive attitudes toward the target language and community may be more likely to make decisions that result in more frequent exposure to the L2 outside the classroom, which in turn may lead to improved cognitive fluency, or processing efficiency. Similarly, Schmidt (2010), referring to Gardner (1988), speculates that motivated learners learn better than unmotivated ones because "they pay attention more and selectively attend to morphosyntactic information, not only content " (p. 732). Thus, high motivation may mean more persistence in paying attention to language input, which leads to more noticing that can enhance learning.

In terms of their relative explanatory power, Juffs and Harrington (2011) maintain that WMC may indeed be a factor in explaining some variability among learners, but other factors like motivation may turn out in the end to be much more powerful explanatory variables in L2 learning (p. 146). Linck & Weiss (2011) is the only study that statistically controlled for L2 motivation in their investigation of the relationship between cognitive capacity (executive function) and L2 learning in low-proficiency learners over a semester (3 months). After factoring out the performance variance explained by motivation, working memory was still identified to be
a reliable significant predictor of L2 performance. However, the relative explanatory power of L2 motivation and WMC was not a focus of the study nor was the possible relationship between learner IDs. Furthermore, the authors acknowledge that the preliminary finding of a causal relationship between WMC and development of L2 grammar obtained in this small-scale study (24 learners) needs cross-validation with a larger sample of participants.

Second, an important gap in the literature pertains to the role of proficiency level in mediating potential relationships between learner IDs and L2 learning outcomes. It is not clear whether components underlying a general ability to learn languages, especially WMC, differentially impacts performance by learners at different stages of interlanguage development given the few studies that have isolated proficiency as an independent variable. Many have focused on two ability levels that are arguably too close to enable meaningful comparison (e.g., early versus late beginners—Sagarra; beginner versus ‘pre-intermediate’ learners—Kormos & Sáfish) or investigated one proficiency level only (e.g., advanced—Erçetin & Alptekin; Foote; Coughlin & Tremblay). In general, studies that have adequately measured both processing and storage components of WMC in the learners' L1 have reported WMC effects on L2 processing and performance for learners at lower proficiency levels. Likewise, in terms of phonological short-term memory, O’Brien et al. (2006) and Winke (2005) provide evidence that this component plays a broader role at earlier stages of proficiency concerning oral fluency, writing, speaking, and reading skills with a more specific role in accurate grammatical performance and listening ability at advanced stages of learning. In the realm of L2 motivation, very little systematic research exists that has investigated a differential role for this psychosocial ID in learners at different stages of L2 development. To efficiently pursue the question of the role of learner proficiency level, we need a developmental perspective in learner ID research, as

A third noticeable gap pertains to the variable of time and the degree to which IDs can predict learning and retention during and after a period of instruction. While the explanatory capacity of some IDs may be consistent or more fluctuating over time both within the same learner as well as between different learners, the cross-sectional design of the majority of SLA research (Ortega & Iberri-Shea, 2005) limits our understanding in this realm. A few studies conducted within a cognitive perspective (Sagarra, 2000; Linck & Weiss, 2011) and a psychosocial perspective of L2 learning (Gardner et al., 2004; Yanguas, 2007) specifically aimed to investigate relationships between their respective ID factors and L2 development over varying periods of time, but these studies differed in several ways and did not aim to gauge retention. We need more research that identifies the intervening factors that can explain loss or maintenance of L2 performance after a period without instruction (Bardovi-Harlig & Stringer, 2010), in addition to those finding a positive role for high proficiency level (Bahrick, 1984; Morgan-Short et al., 2012; Weltens et al., 1989) and high L2 motivation (Gardner et al., 1985; Gardner et al., 1987; Gardner & Lysynchuk, 1990), or learner perception of loss in the latter case.

Relatedly, in addition to analyses of variance, we need to employ more robust statistical approaches to analyze change over time in order to capture the complex, variable nature of learner growth trajectories. Growth curve linear analysis, a powerful statistical approach commonly used in cognitive psychology and educational research (Singer & Willett, 2003), is one approach that can improve and strengthen the interpretation of relationships between IDs and L2 learning over time in SLA research.
A fourth gap pertains to the potentially differential involvement of IDs in implicit and explicit L2 learning and knowledge, keeping in mind that implicit and explicit learning mechanisms and mechanisms used to access implicit and explicit knowledge are not necessarily the same (N.C. Ellis, 2005; Hulstijn, 2005). Studies investigating learning under implicit and explicit instructional conditions (Erlam, 2005, Robinson, 2002, 2005b, Sanz, under review) have contradicted the claim made by researchers in cognitive psychology (Reber, 1993; Reber & Allen, 2000) and in early SLA research (Krashen, 1981) that cognitive ability matters only for explicit learning processes. In contrast, these studies found a meaningful role for cognitive IDs, language analytic ability and working memory in particular, in mediating learning under both implicit and explicit instruction (Robinson, 2005b) or under implicit instruction only (Erlam, 2005; Sanz, under review). The latter finding was interpreted by the authors in terms of a possible 'neutralizing' effect of instruction on the beneficial role for IDs under more structured learning conditions with explicit grammatical explanation. However, it should also be kept in mind with these studies that instructional conditions do not guarantee internal learning processes, or that learners implicitly or explicitly learned the targets of development (i.e., with or without awareness) (Schmidt, 1994, 2010). On the other hand, studies investigating artificial grammar learning have not found correlations with WMC under implicit or explicit instructional conditions (Robinson, 2002, 2005b) and Kaufman et al. (2010) did not find a relationship between implicit (non-linguistic) sequence learning ability and WMC, though they did find a meaningful relationship with processing speed. The latter findings can likely be explained in terms of the difference between artificial and natural languages with the latter requiring complex form-meaning connections.
Pertaining to L2 knowledge, some evidence suggests that WMC, as measured on complex span tasks, underlies access to and development of explicit L2 knowledge in the beginning stages of learning (Kormos & Sáfár; Linck & Weiss) and use of both implicit and explicit knowledge by advanced learners (Erçetin & Alptékin). However, the task demands of measures of implicit and explicit L2 knowledge should also be considered in light of the fact that the nature of tests used to assess L2 development can make different demands on working memory (Sanz, 1997) and different WM components likely interact with different L2 skills (i.e., listening, reading, writing, speaking) (Wen, 2012). Furthermore, when drawing comparisons, future research in this realm should be cognizant of whether tests target knowledge of a range of grammatical structures in the L2 or one or two target structures in order to further refine the extent of the functions of WMC in the development and use of L2 knowledge.

Finally, it is crucial to employ reliable, validated indices of ID constructs and L2 knowledge (Dörnyei & Ushioda, 2009). In particular, the test battery used to tap implicit and explicit linguistic knowledge (e.g., Bowles, 2011; R. Ellis, 2004, 2005, 2009; Elder & Ellis, 2009; Gutiérrez, 2012; Han & Ellis, 1998) merits further validation with learners of target languages other than English and with certain methodological improvements. For example, analyzing learner response speed and variability in the oral imitation task (OIT) could improve the interpretation of the degree to which learners were automatically accessing their language knowledge. Also, collecting judgment source (Feel, Guess, Rule) and (written) rule verbalization in the untimed grammaticality judgement task would strengthen the interpretation of the degree to which learners were accessing their language knowledge more or less explicitly.

Furthermore, the relative explanatory power of the two central notions within L2 motivation research, Integrativeness and the Ideal L2 Self, needs clarification. While several
researchers report that these constructs are strongly correlated (e.g., Csizér & Dörnyei, 2005; MacIntyre, Mackinnon, & Clément, 2009; Ryan, 2009; Taguchi, Magid & Papi, 2009), MacIntyre et al. (2009) have argued that this "does not preclude differential predictions of language learning outcomes” (p. 208). Thus, questionnaires within both dominant frameworks of L2 motivation should be employed to further probe their ability to explain L2 achievement over time for learners at different skill levels.

Rationale for the Current Study

The current study aims to address the gaps outlined above by investigating the role of individual differences in explaining the rate of L2 development, specifically differences in the growth of implicit and explicit knowledge. I specifically investigate the role cognitive and psychosocial learner-internal differences play, both separately and in conjunction, i) at three different points in the developmental continuum (i.e., proficiency level) and ii) at three different times during and after a learning period (i.e., beginning and end of a semester (3 months) and 1 month after instruction). I also contribute to the line of research aiming to improve the validity, reliability, and replicability of assessment techniques measuring knowledge of (implicit) and about (explicit) problematic grammatical structures in a second language, in this case Spanish.

Research Questions (RQs)

The following research questions motivated this dissertation research. Due to the lack of previous literature investigating the independent variables in conjunction over time and by learner proficiency level, no directional hypotheses were made.
RQ1: What is the trajectory of development of implicit and explicit L2 knowledge for classroom learners of Spanish during and after a semester of instruction? Does development of knowledge type differ by proficiency level?

RQ2: Are cognitive and psychosocial ID factors significantly related to the development and retention of *implicit* L2 knowledge?

   RQ2a) If so, to what degree are they related to beginning, intermediate, and advanced levels of L2 Spanish at three points in time?

   RQ2b) To what degree do they predict performance over time?

RQ3: Are cognitive and psychosocial IDs significantly related to the development and retention of *explicit* L2 knowledge?

   RQ3a) If so, to what degree are they related to beginning, intermediate, and advanced levels of L2 Spanish at three points in time?

   RQ3b) To what degree do they predict performance over time?

RQ4: What relationships exist, if any, between the cognitive and psychosocial dimensions of L2 learning in this learner population?
CHAPTER THREE: Research Design and Methodology

Overview

This chapter begins by describing the three learner groups in the participant section including the inclusion criteria for participation in the study (see Table 3). Then all terms relevant to the method and design of the study are operationalized in terms of the measure used to assess them, namely, implicit and explicit knowledge of L2 Spanish (Dependent Variables) and L2 aptitude, WMC, PSTM, processing speed (cognitive IDs) and L2 motivation (psychosocial ID) (Independent Variables). Other materials and questionnaires used are detailed in the subsequent section followed by the research design and procedure (see Figure 3) and scoring and coding procedures for all tasks. In the analysis section, the reliability and statistical analyses are explained. Finally, the results of the pilot phase of the study are discussed along with key revisions to assessment materials.

Description of Participants

Eighty-seven native English speakers (or bilinguals who report English as one of their native languages) between the ages of 18 and 40 years old learning Spanish as a second language in a private North American university participated in this study. Participants were enrolled in Beginning (2nd semester; n = 23), Intermediate (4th semester; n = 33) and Advanced (6th semester; n = 31) Spanish courses. All participants completed a participant biodata questionnaire to control for relevant language learning background variables, which is detailed below.

Participant biodata questionnaire.

The participant biodata questionnaire was developed based on discussions in Mackey and Gass (2005, pp. 124-127) and Dörnyei (2002/2010, chapter 2). The questionnaire elicited basic biodata information about participants such as age, gender, and native language(s). In a more
detailed portion, participants were asked to report information about learning Spanish and any additional languages. Relevant background variables included age of first exposure (AoE) to Spanish and amount and type (formal/informal) of prior study of Spanish. They were also asked to report AoE, amount/type of prior study, self-rated proficiency, and frequency of use of any additional languages they indicated knowing. This section aimed to control for native language, study abroad experience, and knowledge of other romance languages. Inclusion criteria for participation in the study are summarized in Table 3.

Table 3

<table>
<thead>
<tr>
<th>Variables</th>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Age</td>
<td>18-40 years old</td>
</tr>
<tr>
<td>b) Native Language(s)</td>
<td>English (+ Non-Romance language)</td>
</tr>
<tr>
<td>c) Prior Study: Romance Languages</td>
<td>Little to no formal study</td>
</tr>
<tr>
<td>d) Spanish L2 Background</td>
<td>Post-puberty age of first exposure</td>
</tr>
<tr>
<td>e) Study Abroad: Spanish-speaking country</td>
<td>≥ 2 weeks</td>
</tr>
<tr>
<td>f) Placement Test</td>
<td>Within 3 SDs ± mean for proficiency group</td>
</tr>
</tbody>
</table>

As Table 3 shows, to be included in the study participants had to a) fall within the 18-40 years old age range; b) report English as one of their native languages and if bilingual, a non-Romance language (i.e., French, Italian, Portuguese) as their other native language; c) report no formal study or minimal proficiency in a romance language; d) report a late age of first exposure to Spanish at or after puberty (i.e., 10-12 years old); and e) report spending two weeks or less in a Spanish-speaking country. Due to the decline of cognitive function with age, working memory
capacity and processing speed in particular (e.g., Salthouse, 1991, 1996), any participant older than 40 years old was excluded from participation in the study to ensure comparability in cognitive capacity.

For their time, students were awarded five extra credit points on their lowest test grade in their Spanish class. All subject recruitment and compensation procedures were approved by the university's human subjects committee (IRB approval # 2012-27; project expiration – January 2014).

**Operationalizations**

**Implicit and explicit L2 knowledge.**

Learners’ second language, or interlanguage, knowledge in Spanish is operationalized here as performance on tests targeting a range of grammatical structures. Performance measures are based on several criterial features as proposed in previous research (and detailed in Chapter 2) (R. Ellis, 1994, 2004, 2005, 2009; Han & R. Ellis, 1998), in particular: 1) *Time available*, 2) *focus of attention*, 3) *systematicity*, and 4) *metalanguage*. Based on findings in this line of research and logistical feasibility, the two ‘best’ measures of implicit/explicit knowledge were used; namely the elicited oral imitation task (OIT) was used to measure implicit L2 knowledge of Spanish and performance on an untimed grammaticality judgment task (UGJT) was taken as an indicator of their explicit L2 knowledge. However, as previously mentioned, certain methodological improvements were made to strengthen internal validity (detailed below).

Both tests targeted ten grammatical structures centering around tense, aspect and mood morphology that are acquired at both early and late stages of Spanish L1 and L2 acquisition (e.g., López Ornat, 1994, 1997; Montrul, 2004). These were chosen from the seventeen grammatical structures used in Bowles (2011) and Gutiérrez (2012) based on being the most persistent
problems for English-speaking learners in the realm of Spanish morphosyntax (DeKeyser, 2005) given that they are L2 features not present in their L1 grammatical system. The number of targeted structures was reduced from seventeen to ten in order to increase the overall number of grammatical and ungrammatical exemplars of each structure (2 > 4 OIT; 4 > 6 UGJT). Three versions of each measure were created for a total of six test versions (3 testing sessions x 2 tests). They were administered in a counter-balanced, split-block design structured around the three assessment waves (see the section on Research Design for more detail).

*Elicited oral imitation task (OIT).*

In the elicited OI task, learners were instructed to focus on meaning when responding to a series of forty randomized statements in Spanish (two grammatical/two ungrammatical per target structure). Learners were instructed to agree with, disagree with or say they are not sure about the meaning, or content, of the statement and then to immediately repeat the sentences they heard out loud in correct Spanish. Syllable length was controlled in all test versions; sentences were 16.7 (5.8 SD), 17.3 (5.3 SD), and 17.6 (5.9 SD) syllables long on average. The OIT took between fifteen and twenty minutes to complete.

The version of the task used here was administered in SuperLab 4.0 and reaction times (RTs) were measured (in milliseconds) for oral sentence repetitions. In the pilot-testing phase of data collection, RTs were roughly gauged via a key press but a connector used to measure voice onset reaction time in the SuperLab program (SV-1 key) was purchased for the main phase of data collection through a departmental grant. However, it proved to be problematic and unreliable in its ability to accurately measure voice onset RT. Thus, funding from an external grant was used to hire research assistants to measure RTs for each item ($N = 40$) by hand using Praat, the free online software system for acoustic analysis (http://www.fon.hum.uva.nl/praat/).
Untimed grammaticality judgment task (UGJT).

In the adapted version of the UGJT used here, participants were instructed to judge the grammaticality of sixty randomized written sentences (three grammatical/three ungrammatical per target structure) and indicate whether they used ‘Feel’, ‘Guess’ or ‘Rule’ to make their judgment as recommended by R. Ellis (2004). Additionally, participants who indicated using rule knowledge were asked to verbalize that rule (in writing) explaining the ungrammatical element in the sentence stimulus. The UGJT was administered through Survey Monkey, an online questionnaire application (http://www.surveymonkey.com/). Learners had no time limit but took approximately thirty to forty-five minutes to complete.

Each test was created using basic vocabulary from the first six chapters in the textbook used at the beginning level within the non-intensive language program (Vistazos—VanPatten, Lee & Ballman, 2010). Using simple vocabulary was crucial since learners were at three different proficiency levels. As reliability is a property of how a particular sample responds to a testing instrument, the pilot phase of this study was crucial for ensuring reliability of testing instruments for this learner population. All reliability and item analyses as well as the counter-balanced design employed in the main data collection phase are further detailed in the Analyses section.

Cognitive IDs.

L2 aptitude: Modern Language Aptitude Test (MLAT).

The ability to learn second languages is operationalized here in terms of three abilities: i) the capacity to identify distinct sounds, associate the symbols that represent them and retain those associations (phonetic coding ability); ii) to recognize the grammatical functions of words (or other linguistic entities) within sentences (grammatical sensitivity), and iii) the ability to
rapidly learn new sound-meaning associations and retain them (*rote learning ability*) (Carroll, 1962, 1981, 1990). Parts III-Spelling Clues (SC), IV-Words in Sentences (WS) and V-Paired Associates (PA) of the Modern Language Aptitude Test (MLAT—Carroll & Sapon, 1959) were used to operationalize each aptitudinal component.

In the 50-item SC test, the participants’ task is to quickly recognize English words that have been spelled creatively and match each word with the correct synonym among four choices. In the 45-item WS test, a set of stimulus sentences have one word underlined and a second sentence that has four underlined. Learners must choose the right word out of four that serves the same grammatical function as the underlined word in the stimulus sentence. The PA test involves memorizing 24 Kurdish words paired with their English meanings. After a two-minute practice phase, learners must match the Kurdish word with its equivalent English word based on four options in a 24-item multiple-choice (MC) test. Together, all MLAT sections were completed in paper-and-pencil format and take approximately 45 minutes to complete.

Sections in the MLAT-Short form are reported to be highly reliable with Cronbach's alpha ranging from 0.83 to 0.91 (e.g., de Graff, 1997; Winke, 2005) and test-retest reliability between 0.90 and 0.91 (e.g., Sparks, Patton, Ganschow, & Humbach, 2011). This componential measure of language aptitude has proven quite durable throughout the years, but this study follows others (e.g., McLaughlin, 1995; Miyake & Friedman, 1998; Robinson, 2001, 2002a; Sanz, 2005; Sawyer & Ranta, 2001; Skehan, 2002; Wen & Skehan, 2011) in pinpointing working memory capacity as the site, or mental workspace, where all aptitude components are likely integrated (Sawyer & Ranta, p. 342).

*Working memory capacity: Operation Span (OSpan) Task.*
The present study primarily focuses on the functions of the central executive in Baddeley and colleagues’ multicomponent model of working memory (Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Baddeley, 1986, 2000, 2003, 2007, 2010) such as focusing and sustaining attention, updating, encoding, and retrieval. WMC is operationalized here within a dual task paradigm as simultaneous storage capacity and processing efficiency.

An operation span (O-Span) task (Turner & Engle, 1980; Unsworth, Heitz, Schrock, & Engle, 2005) was used to tap working memory capacity and was conducted in the participants’ L1 (English) to avoid a confound between L2 working memory capacity and L2 proficiency level. In the same vein, the choice to employ a nonverbal (i.e., domain-general) measure of WMC was made to avoid any potential confound between L1 working memory capacity and L1 proficiency, especially given that some participants in this study could be bilingual (e.g., report two native languages as long as one is English and the other is not a romance language). While the reading span (R-Span) task (Daneman & Carpenter, 1980; Waters & Caplan, 1996) is arguably the most widely used measure of WMC in SLA research, the verbal nature of the secondary information-processing task (i.e., judging the plausibility of sentences) makes it language dependent while the processing requirement in the O-Span (i.e., solving math equations) is less language dependent (Engle, 2002; Linck & Weiss, 2011).

Nevertheless, Turner and Engle’s (1989) version of the O-Span task is verbal to some degree as learners are instructed to maintain words in memory that appear after each sentence for later recall. Previous research has suggested that some of the shared variance between span tasks that use words and a measure of higher order cognition, such as reading comprehension, is due to word knowledge (Unsworth, Heitz, Schrock & Engle, 2005) or memory strategies based on the semantic content of words (Kane, Hambrick & Conway, 2004). Thus, in the O-Span version used
here and standardized by Unsworth and colleagues, letters replace words as the to-be-recalled items. The modified O-Span significantly correlates with the original O-Span used by Turner and Engle ($r = .45$, $p < .01$) and is reported to have high internal consistency (Cronbach’s $\alpha = .78$) and test-retest reliability ($\alpha = .83$) (Unsworth et al., 2005, p. 501).

In the O-Span task used here, participants solved simple arithmetic equations out loud and were instructed to maintain capitalized letters in memory for later recall. For example, if the participant saw “IS (6 x 2) + 1 = 13?....F,” they would say “Is six times two plus one equal to thirteen…Yes…F.” Once the participant said the letter, the researcher advanced to the next sentence or operation and letter in the set to the amount of rehearsal time. Each participant was presented with twelve sets of varying difficulty, ranging from two to five sentences or mathematical operations in each set. At the end of each set, participants saw three question marks cueing them to write down all of the letters that they could remember from that set in the correct order on an answer sheet provided. Participants first engaged in three practice trials to ensure full comprehension of task instructions.

Individuals’ span scores were calculated according to both processing and storage demands. The limits of WMC were measured as the maximum number of letters accurately recalled in sets of increasing difficulty (two to six sentences in a set) while simultaneously judging calculating simple arithmetic. They were given a processing score based on the percentage of mathematical equations judged correctly and a storage score based on the maximum number of letters accurately recalled. Data for any learner who did not reach 80% or above in the processing task were excluded. The O-Span was individually administered on Superlab 4.0 and took up to twenty minutes to complete. Scoring procedure is detailed further in a later section.
**Phonological short-term memory: Digit span task.**

Working memory capacity is also viewed in terms of the functioning of its phonological loop, responsible for the temporary storage and maintenance of verbally coded information (Baddeley & Hitch, 1974; Baddeley, 1986, 2003). The storage component is responsible for passive immediate store of phonological input while the articulatory system is responsible for actively rehearsing the phonological input in order to prevent it from decay.

A digit span task was used to operationalize this component of WMC as it “provides a useful indication of the capacity of an individual’s phonological loop” (Baddeley, Gathercole, & Papagno, 1998, p. 159) and is one of the most widely used measures of phonological short-term memory (PSTM) in many standard measures of intelligence (e.g., WAIS-III, Weschler, 1997). Other tasks, such as nonword repetition (NWR), are reliable measures of PSTM when they involve minimal or no familiar linguistic information, but are otherwise confounded with existing lexical knowledge or familiarity with word patterns in the language of testing (Baddeley, 2003; French & O’Brien, 2008). Using an English-based NWR would favor monolingual native speakers of English and because some participants in this study are bilingual, the Digit Span task seems more appropriate which only requires familiarity with the numbers 1 through 9 in English.

In this task, participants heard a set of numbers and repeated the numbers aloud in the same order. There were two sets each of two to nine digits. If a participant failed to correctly repeat two sequential number sets, the experimenter terminated the task. Participants were awarded one point for every set remembered correctly until they reached the last set or until they missed two sets in a row. The number of sets repeated accurately is interpreted as the limit of their phonological short-term memory and the maximum number of points possible was 16. This task was designed based on guidelines provided in van den Noort et al. (2006).
Processing speed: Digit symbol coding task (DSCT).

Based on the criterion of time pressure in use of implicit knowledge, the potential importance of basic cognitive processing resources in L2 cognitive fluency (Segalowitz, 2010) and evidence of increased morphological processing efficiency at higher levels of proficiency (e.g., Hopp, 2010; Rodgers, 2011), a simple measure of processing speed was also included as a domain-general cognitive ID in this study to complement measures of working memory. It is operationalized here as the speed and accuracy with which learners search for and code symbols that randomly correspond to the numbers 1-9 on the Digit-Symbol Coding task (DSCT), a part of a standard measure of general intelligence (Wechsler Adult Intelligence Scale (3rd ed) (WAIS-III—Wechsler, 1997). For this task, participants receive a sheet of paper with a key at the top of the page showing the digits from one to nine and a unique symbol assigned to each digit. A seven-row grid beneath the key displays randomly ordered digits with empty boxes beneath each one. The participant’s task is to use the key to copy the symbols that correspond to the digits in the empty boxes. The test consists of 140 items (empty boxes). Each participant has two minutes to complete as much of the grid as possible.

Psychosocial IDs.

L2 motivation.


Attitudes and Motivational Test Battery (AMTB).

L2 motivation was measured through learner self-report on the Attitudes Motivation Test Battery (AMTB) (Gardner & Lambert, 1959; Gardner & Lambert, 1972) that measures three
main components: 1) *Integrativeness* (integrative orientation, interest in foreign languages, and attitudes towards the L2 community); 2) *Attitudes toward the learning situation* (attitudes towards the language teacher and the course); and 3) *Motivation* (effort, desire, and attitude towards learning). *Instrumental* and *integrative orientation* subtests respectively assess one's practical and interpersonal reasons for language study and a series of semantic differentials probe attitudes toward the teacher and course assess more state-like fluctuating aspects of motivation. Components of general L2 motivation are measured by a series of eleven scales of grouped Likert-type items (78 items) (1-Strongly Disagree; 7-Strongly Agree). The eleven subscales are consistently reported to have high reliability with Cronbach's alpa above .80, with the exception of the instrumental orientation scale (Masgoret & Gardner, 2003, p. 125).

State-like aspects of motivation (evaluation of the course and teacher) are based on responses to two sections of semantic differentials consisting of fifty-items total (e.g., ‘My Spanish Teacher’: Efficient _____: _____: _____: _____: _____: _____: _____ Inefficient). Learners were instructed to mark an X in the appropriate blank space that most closely approximated their feeling about the relevant statement and were instructed to work quickly and give their first impression in response to each statement. They were informed that their responses would remain anonymous.

The version of the AMTB implemented in the current study was adapted for instructed L2 learners of Spanish based on the version used in Yanguas (2007) created for Spanish heritage learners. Following Gardner et al. (2004) and Yanguas, learners completed AMTB scales at the beginning and end of the learning period. All portions of the AMTB took up to thirty minutes to complete and were administered online via Survey Monkey.

*Possible selves questionnaire.*
Using the framework of Dörnyei’s L2 Motivational Self System (2005, 2009), the *Ideal L2 self* and *Ought-to L2 self* dimensions of L2 motivation as well as variables related to intended effort (i.e., motivated learning behavior) were measured in a possible selves questionnaire used and adapted for several learning contexts (e.g., Al-Shehri, 2009; Czizér & Kormos, 2005; Kormos & Csizér, 2008; Ryan, 2009; Taguchi, Magid, & Papi, 2009). The questionnaire included thirty statements total with ten statements per component and learner responses were assigned one to seven points per item indicating degree of agreement with the statement (e.g., *I can imagine myself living abroad and having a discussion in English*: 1-Strongly Disagree; 7-Strongly Agree). Dörnyei (2010) reports that reliability coefficients using Cronbach’s alpha for each of the three categories are considerably high based on use in previous empirical studies (*Ideal L2 self*, \( \alpha = 0.88 \); *Ought-to L2 self*, \( \alpha = 0.79 \); *Criterion Measures*, \( \alpha = 0.86 \)) (p. 123). This questionnaire was completed with pencil and paper and took up to 10 minutes to complete.

**L2 proficiency level.**

Proficiency level in L2 Spanish was operationalized as institutional enrollment in beginning, intermediate and advanced Spanish courses at the researcher's home university. Participants were enrolled in the second level within each instructional sequence (i.e., Introductory II (SPAN 004), Intermediate II (022), and Advanced II (104)). Additionally, all participants completed the grammar portion of the Department of Spanish and Portuguese Summer/Fall 2012 Spanish Placement Exam at Georgetown University as a way to ensure comparability within and between proficiency groups. The placement test is designed to reflect the objectives of the different levels of the Spanish language curriculum in the areas of listening, reading, vocabulary, and grammar and has successfully been used for seventeen years as a way to determine appropriate course registration for Spanish students in the non-intensive Spanish
language program at Georgetown. The grammar section consists of thirty-six 4-option multiple-choice items targeting a range of grammatical structures in Spanish such as tense, aspect and mood inflection and gender and number concord. Learners have up to 30 minutes to complete sentences with the correct missing verb, adjective, pronoun, relative conjunction, etc. according to the context of the sentence.

Other Materials

An example of all materials are available in Appendices A-L.

Classroom teaching style questionnaire.

Due to the need to recruit a large number of learners at each proficiency level, L2 instructional exposure for individual learners could not be experimentally controlled in this study. However, analyses of variance were run with Spanish language instructor as an independent variable to test for significant performance differences between learners. Also, Spanish language instructors filled out a simple teaching inventory about their instructional and feedback practices with the goal of locating them on an implicit-explicit continuum. Design of this self-report instrument was based on discussions of classroom-based research in Chaudron (1988, 2001) and an example of a teaching method questionnaire provided by Swaffar, Arens and Morgan (1982). The first section requests general background information in terms of instructors' experience teaching Spanish as well as their self-perceived fluency in Spanish and English. The second section aimed to get a general idea of their teaching style on a 'typical day in their classroom' by asking four questions targeting two main components: 1) Frequency of activity and task type (e.g., meaning-focused versus form-focused; skill focus/modality) and 2) Frequency of type of feedback (e.g., implicit (recasts, prompts, clarification request, etc.) versus explicit (over correction, metalinguistic explanation, etc.); opportunities for repair/self-correction) as well as
general classroom atmosphere (e.g., teacher-centered versus student-centered). For classroom tasks, teachers were instructed to indicate the frequency with which they employ a series of activity types (e.g., paired or group work). With respect to feedback, they indicated the degree to which they agreed with a series of statements related to possible ways of correcting student errors during classroom activities (e.g., repeating the error). The questionnaire was sent to instructors to complete online at their convenience via Survey Monkey (www.surveymonkey.com).

**Outside exposure/use questionnaire.**

At Time 2 (end of instructional period) and Time 3 (one month after instructional period), participants completed a brief questionnaire addressing how many hours a week they reported spending outside of the classroom studying for and preparing assignments for their Spanish course and how frequently they reported using Spanish on their own volition (e.g., interacting with native Spanish-speakers or watching TV and/or movies in Spanish). Learners were given a four-choice response option for the former (i.e., 0-3hrs; 4-6hrs; 7-9hrs; 10-12hrs) and latter dimensions (i.e., Always; Frequently; Sometimes; Never).

At Time 3, learners completed an additional debriefing questionnaire explaining the general goals of the experiment and asking whether they specifically made an effort to study any aspect of what they were being tested on during any phase of their participation in the study.

**Research Design**

The design of this study is quasi-experimental in nature in that learners were not randomly assigned to treatment groups; rather they were placed into one of three proficiency groups based on institutional enrollment and performance on the departmental Spanish placement exam. This study’s design is also correlational and predictive as it is of interest to test
relationships between variables and to make predictions about the likelihood of the presence of one variable from the presence of another (or others) (Mackey & Gass, 2005, p. 145). There are six main independent variables that correspond to cognitive measures - L2 aptitude, WM, PSTM, processing efficiency - L2 motivation, and proficiency level. The two dependent variables for each group of learners are based on their performance scores on a measure of implicit knowledge (sentence repetition accuracy and latency) and a measure of explicit knowledge (grammaticality judgment accuracy).

**Procedure.**

The administration of all measures corresponding to the above variables was structured around three time points during and after a semester of instruction: Time 1 (T1) was the beginning of the period of instruction; Time 2 (T2) occurred 10 weeks later at the end of instruction; and Time 3 (T3) was four weeks after the end of the semester and before the following spring semester (i.e., after the holiday break). Figure 3 summarizes the data collection timeline, materials, and procedure.

Figure 3

*Materials and procedure*
There were two sessions at each testing time for a total of six research sessions. Sessions 1, 3 and 5 were completed individually with the researcher in a soundproof laboratory using
SuperLab or Microsoft PowerPoint. Learner oral repetitions in the elicited oral imitation task and non-word repetition task were audio recorded using Audacity for later transcription. Sessions 2, 4, and 6 were carried out in a group classroom laboratory and learners completed tasks on individual computers using Survey Monkey or Blackboard. The total participation time to complete all sessions depended on each individual learner but took approximately five to six hours on average: Session 1—up to 1 hour; Session 2—1.5 hours; Session 3—20 minutes; Session 4—1 hour; Session 5—20 minutes; and Session 6—30 minutes. All learner questionnaires, ID and performance measures were computerized with the exception of the MLAT and the Selves L2 motivation questionnaire. Instructor teaching style questionnaires were sent via Survey Monkey to be completed at their convenience.

**Counterbalanced design.**

A counter-balanced, or split-block, design was used to order performance tasks differently among all participants to avoid any potential ordering effects (Mackey & Gass, 2005, p. 143). Three different versions of each performance measure was created and one-third of all learners were randomly assigned to one of three groups (A, B, or C), regardless of proficiency level, and completed versions 1, 2 or 3 (V1, V2, V3) at the first, second and third testing times (T1, T2, T3). Table 4 provides an overview of this counterbalanced design.

Table 4

*Counterbalanced design in the administration of tests of implicit and explicit L2 knowledge*

<table>
<thead>
<tr>
<th>Groups</th>
<th>Testing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
</tr>
<tr>
<td>Group A</td>
<td>V1</td>
</tr>
<tr>
<td>Group B</td>
<td>V2</td>
</tr>
</tbody>
</table>
Note. One-third of all learners are randomly assigned to groups A, B or C regardless of proficiency level.

**Scoring and coding procedures.**

The scoring and coding procedures for all measures of L2 knowledge and individual differences are further explained in this section.

**OIT.**

Semantic processing and sentence repetition accuracy and latency data were collected in the adapted version of the OIT task used here. To ensure semantic processing, learners' meaning judgment responses (Si/No) of the forty sentences were converted into percentages in order to analyze the extent of their meaning comprehension of the sentences. In terms of accuracy, two scores were calculated according to previous literature: 1) Score A, a form-based method following that used in Bowles (2011), Gutiérrez (2012) and R. Ellis's studies, and 2) Score B, a form and meaning-based method adapted from the system used in Bowden (2012), Ortega et al. (in preparation) and Ortega (2000). The latter scoring method was employed given the intended meaning-focused nature of this task. 40 total points were possible according to Scoring A (0-1 pt; 40 items total) and 120 points were possible according to Scoring B (0-4 pts; 40 items total).

For each item, learners received 1 point for accurately repeating the obligatory target structure and 0 points for not doing so (A) and were awarded 0-4 points based on both accurate repetition of form and the extent to which they accurately repeat idea units in the sentence stimulus (B). In the content-based scoring method, 0 points were given to minimal (one word) or no repetition; 1 point for repeating up to 25% of the original sentence stimulus but not accurately repeating, or modifying, the target structure; 2 points for 25-50% accurate sentence repetition but without the target structure; 3 points for sentences that accurately included over 50% of idea
units as well as the obligatory target structure, and 4 points for exact repetition and formal accuracy (see Appendix M for examples falling under each score descriptor). Overall raw points according to each scoring system were converted into percentages for each testing time (T1, T2, T3) and change or gain scores were calculated from T1 > T2, T2 > T3 and T1 > T3.

Latency data was also gathered for oral sentence repetitions with the idea that faster reaction times (RTs) should reflect more automatic processes in accessing implicit knowledge. All RTs were filtered following standard data filtering and cleaning procedures in cognitive psychology and SLA (Jiang, 2011; Lachaud & Renaud, 2011). RTs faster than 500 ms or slower than 5000 ms were removed and treated as outliers and based on these values, RTs that were 3 SDs above or below the participant's mean RT were also excluded from the analyses. While Lachaud and Renaud (2011) maintain that eliminating values above and below the ±2 SD limits around the mean of the general distribution is standard procedure (p. 391), this study employed a slightly stricter ±3 SD rule. Additionally, if over a third (33%) of the overall data for each individual participant was missing, their mean RT was not included in the analyses. However, RTs could not be filtered according to error response rates given the low accuracy rates on this challenging task. As an alternative (see the section on Latency in Chapter 3), accuracy scores from the OIT were entered as a covariate in all subsequent analyses run with Latency as a Dependent variable, a procedure recommended by a resident statistician, Dr. Rusan Chen (personal communication, March, 2013).

Also, it is relevant to consider findings reported by Spada (2013) concerning the possibility that this measure of 'implicit knowledge' in fact reflects automatized explicit knowledge. She found in a retrospective questionnaire that a majority of learners were thinking about form or form and meaning while completing the task and concluded that they may have
been rapidly accessing their explicit knowledge on this task. Given that previous studies employing the OIT have not methodologically addressed this possibility, a coefficient of variation (CV) as proposed by Segalowitz and Segalowitz (1993) was also calculated as a way to operationalize automatization of L2 knowledge and distinguish speed from processing efficiency, or a more organized underlying system. This involves a simple calculation of dividing the mean standard deviation by the mean RT (SD/RT) for each individual participant.

**UGJT.**

In terms of accuracy, one point was awarded for sentences correctly marked grammatical or ungrammatical (60 points possible) and raw number of points out of sixty was then converted into a percentage. Learners also received a separate accuracy score for grammatical and ungrammatical sentence stimuli in order to assess performance separately. As with the OIT, overall gain scores were also calculated. Each response on the UGJT was also coded according to Feel, Guess or Rule source attributions. For the source of their judgment, a percentage score was calculated per participant and per proficiency group based on how many items out of sixty they judged based on intuition, a guess or a grammatical rule. The difference between Feel and Guess source attributions was explained to learners in terms of feeling like they knew an answer was (un)grammatical but not knowing why (Feel) and not knowing at all (Guess). Finally, the accuracy of rules provided for sentences correctly judged as ungrammatical in Spanish was calculated individually out of the total rules each participant provided for sentences judged to be ungrammatical. For example, if a participant indicated using rule knowledge for sixteen items she judged to be ungrammatical, and described fourteen accurately, her score would be 88%.

**MLAT (III, IV, V).**
A composite score and separate scores by component were calculated as indices of L2 aptitude. One point was awarded for each correct answer and total raw number of points was converted into a percentage score (out of 119 points possible). Percentage scores were also calculated for each component consisting of fifty items measuring phonetic coding ability (Spelling Clues), forty-five items measuring grammatical sensitivity (Words in Sentences) and twenty-four items measuring associative memory or rote foreign language learning ability (Paired Associates). This procedure is similar to that employed by studies measuring L2 aptitude both compositely and componentially (De Graff, 1997; Robinson, 1996, 1997).

**OSpan task.**

In complex tasks, such as the operation span task, there are two sources of data: one from the processing component of the task and one from the storage component. While many correlational studies do not consider processing performance in the WM span score (Conway et al., 2005, p. 774), the current study includes it to verify learners were in fact engaged in the secondary demand. The storage score in both tasks is based on serial recall of forty-two possible letters and the processing score is composed of forty-two dichotomous responses (Yes/No) based on the accuracy of simple mathematical operations. Following Conway et al. (2005), if accuracy in processing fell below 85%, the entire data set for that subject was discarded (but no participant fell below 85%).

Recall scores were calculated following an all-or-nothing scoring method (Absolute) as well as partial-credit scoring methods (Total & Lenient). In the absolute approach, learners received one point for each letter in sets where all letters were accurately recalled in the right order; in other words, they did not receive any points for a set if any letter is out of order, is missing or is incorrectly added. The total scoring method awards a point for each letter...
accurately recalled but letters out of order are not counted. The lenient method awards a point for any and all letters correctly recalled regardless of the order in which the participant wrote them down. Typically all three scoring methods correlate, but Conway et al. supports using all three methods (2005, p. 776).

**Digit span task.**

On the digit span test of phonological short-term memory, participants were awarded one point for every set of digit sequences (out of eight) remembered correctly until they reached the last set or until they missed two sets in a row. A percentage score was calculated out of a maximum sixteen points possible.

**DSCT.**

Processing speed was calculated as simple percentage score in terms of the number of correctly coded symbols out of a possible 140 points.

**AMTB and Selves questionnaires.**

L2 motivation as measured in the 78-item AMTB yielded an overall composite score as well as separate scores for the eleven subscales tapping all attitudinal-motivational components including *integrativeness* (integrative orientation, interest in foreign languages, and attitudes towards the L2 community), *attitudes toward the learning situation* (attitudes towards the language teacher and the course), *motivation* (effort, desire, and positive affect), and *instrumental* and *integrative* orientation. One to seven points corresponding to Likert-type statements indicated degree of agreement with the statement (1-Strongly disagree; 7-Strongly agree). These were the scales used in items probing interest in learning languages (*n* = 10), attitudes toward the Spanish-speaking community (*n* = 10), motivational intensity (*n* = 10), desire to improve Spanish (*n* = 10), attitudes toward improving Spanish (*n* = 10), Spanish class
anxiety \((n =10)\) and use anxiety \((n = 10)\). An average Likert-response was calculated per component. Attitudes toward the course and teacher were evaluated using semantic differential scales in which learners indicated whether their feeling more closely aligned with one adjective or the other. Points ranged from one to seven for each adjective pair and included twenty-four statements each (e.g., ‘My Spanish Course: Useful _____: _____: _____: _____: _____: _____: _____: Useless).

An overall composite score for L2 motivation was also calculated for answers to the 30-item possible selves questionnaire and separate indices were computed for the Ideal L2 Self scale, the Ought-to L2 Self scale, and motivated learning behavior, or intended learning effort, each composed of ten items. Learner responses were assigned one to seven points per item indicating degree of agreement with the statement (1-Strongly disagree; 7-Strongly agree) and an average Likert-scale response for each component was computed.

Statistical Procedures

Overview.

This section provides an overview of all statistical procedures, beginning with \textit{a priori} power analyses and reliability estimates. Then, the analyses used to address the four research questions guiding the study are discussed in turn. All analyses were run using the Statistical Package for Social Sciences (SPSS, version 20.0).

To decrease the likelihood of committing a Type I error, the alpha \((\alpha)\) level of statistical significance was set to .05 which means that no association is reported significant in the learner population unless it is found to be probable at the .05 level. In other words, there is only a 5% chance of incorrectly rejecting the null hypothesis when it is in fact true (i.e., no association exists) (Howell, 2008, p. 158). To decrease the likelihood that a Type II error occurs (i.e., \textit{not}
identifying a relationship in the population under study when one indeed exists) power analyses were calculated before data collection to pre-determine the number of participants needed for optimal statistical power as recommended by Larson-Hall (2010, 2012). Higher power increases the probability of correctly rejecting a false null hypothesis (Howell, 2008, p. 352), and is principally based on a statistical result’s effect size (e.g., correlation coefficient—Pearson’s $r$ or $R^2$, partial eta$^2$, Cohen’s $d$ or $f$) and the sample size.

**Power analysis.**

As stated above, an experiment’s power is a function of effect size and participant sample size ($N$) as well as the type of statistical test being performed. An effect size measures the degree to which an effect can be attributed to the influence of an independent variable on a dependent variable, or to the relationship between variables (Larsen-Hall, 2010, p. 46). It is typically based on $r$-family measures for associations and $d$-family or eta-squared ($\eta^2$) measures for differences between means. Based on a set of conventions established by Cohen (1988) to test means, effect sizes can be interpreted as small ($d = .20$; $\eta^2 = .01$), medium ($d = .20-.50$; $\eta^2 = .06$) or large ($d = .50-.80$; $\eta^2 = .14$). Cohen's $f$ can also be used to test means ($f = .10$ (small); $f = .25$ (medium); $f = .40$ (large)). For correlations, the correlation coefficient $r$ is an effect size in itself indicating how much of a relationship is explained by the variables of interest. Associations are generally interpreted as small ($r = .10$), medium ($r = .30$) or large ($r = .50$) (Cohen, 1988) and the percentage of variance in the dependent variable explained by the independent variable is expressed as squared $r$ ($R^2$) (as calculated in regression analyses). $R^2$ is likewise considered to be a small (.01), medium (.09) or large (.25) amount of variance explained.

Power analyses conducted prior to data collection can help determine the ideal sample size needed to produce reasonably high power, which should be above .50 and is considered
adequate at .80 (Murphy & Myors, 2004; c.f. Larsen-Hall, 2010, p. 105). A power level of .80 means that four out of five times, a reliable effect in the population will be found whereas a power level of .50 indicates you run a 50% risk of failing to confirm a valid research hypothesis. Power analyses conducted here are guided by discussion in Larson-Hall (pp. 104-111) using a free online software application (http://cran.r-project.org/).

To calculate necessary sample size prior to the main phase of data collection, I set predetermined levels for power (80%), significance level (alpha = .05), and entered a medium effect size. Below Table 5 summarizes the desired overall sample size to obtain a high power level when running correlations, analyses of variance and multiple regressions.

Table 5

Sample size determined by a priori power analyses

<table>
<thead>
<tr>
<th></th>
<th>Effect size</th>
<th>Significance</th>
<th>Power</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation (two-tailed)</td>
<td>r = .3</td>
<td>0.05</td>
<td>0.8</td>
<td>84.7</td>
</tr>
<tr>
<td>Correlation (one-tailed)</td>
<td>r = .3</td>
<td>0.05</td>
<td>0.8</td>
<td>67.2</td>
</tr>
<tr>
<td>ANOVAs &amp; Regression</td>
<td>f = .4</td>
<td>0.05</td>
<td>0.8</td>
<td>21.1 (per group)</td>
</tr>
</tbody>
</table>

What Table 5 shows is that if statistical analyses were to produce a medium effect size, a sample size of 85 or more participants would be sufficient to find the relationship in the population with adequate power in a two-tailed correlational analysis. However, this number decreases to 67 when the analysis specified is one-tailed (positive correlations only). When the effect size is increased to r = .5, the total sample size needed decreases to N = 28.9 because the larger the effect size, the more easily the effect will show up and the less participants you will need to find
For analyses of variance and regressions, a minimum number of 21 learners per proficiency group would be necessary for a medium effect size to emerge at a satisfactory power level.

For the Confirmatory and Exploratory Factor Analyses conducted with mean performance scores on the tests of implicit and explicit knowledge (see below for further detail), I followed logic discussed in Bowles (2011, p. 258) who cites recommendations by Nunnally (1978) and Kass and Tinsley (1979) that one should have ten times as many participants as factors or between five and ten participants per factor. Thus, in order to find meaningful loadings for two factors, an overall sample size of at least twenty is needed (i.e., 10 learners per potential factor).

**Tests of reliability.**

The reliability of a test can be calculated using Cronbach’s alpha (α) which indicates the level of consistency in learner responses as well as how accurately items measure a single construct. First, reliability coefficients using Cronbach’s alpha were calculated for accuracy scores on all three versions of performance measures as well as composite and componential indices of L2 motivation (AMTB; Selves). Those that yielded alphas ≥ 0.80 were interpreted as having high internal consistency. To ensure high reliability in item response patterns across proficiency levels, pilot testing of assessment materials was crucial. Thus, during the pilot testing phase of the study, an item analysis was performed on item response patterns in both performance tests (OIT and UGJT) guided by discussions in Larson-Hall (2010) and Howell (2008). Overall average performance on each item as well as group averages per item were initially calculated and those items that were determined to be too easy or too difficult were examined and subsequently modified, or replaced completely. Results and revisions from the pilot study are more specifically discussed at the end of this chapter.
Research question 1: Development & retention of implicit/explicit knowledge over time by proficiency level.

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS-20.0) and the alpha level was set at 0.05 throughout. To answer RQ1, descriptive statistics (mean, standard deviation (SD), minimum, maximum) were initially calculated for accuracy at all testing times on the tests of implicit and explicit knowledge as well as for latency data (reported in milliseconds) on the implicit measure. The procedure used to filter RT data was previously detailed in the Scoring section. Accuracy for grammatical (GR) and ungrammatical (UGR) exemplars on the OIT and UGJT were separately calculated to analyze performance differences and to enter the scores separately in factor analyses (see below), following suggestions in previous studies that grammatical items may elicit implicit knowledge and ungrammatical items are more likely to elicit explicit knowledge (R. Ellis, 2005, p. 161).

Prior to conducting factor analyses, Pearson Product Moment correlations were run between test scores at Time 1 in order to ensure that tests were significantly intercorrelated with coefficients (r) that reached statistical significance at the .05 level or higher (Field, 2009). Overall mean accuracy scores as well as GR and UGR accuracy scores for both performance measures were then entered into a Confirmatory Factor Analysis for all proficiency groups, which is appropriate when testing a priori hypotheses (see Isemonger, 2007 for a discussion). This was done in order to attempt to replicate the two-factor solution found in R. Ellis (2005), Bowles (2011), Gutiérrez (2012) and Erçetkin and Alptekin (2013) and confirm that the tests tap different knowledge sources in this learner population. However, in order to analyze the psychometric properties of the tests by proficiency level, three exploratory factor analyses were also performed with each group separately to see what model emerged.
Previous studies have also reported model fit indices to test goodness of fit which indicates how consistent the empirical data are with a proposed model. Chi-square is typically used to estimate goodness of fit, but this statistic is dependent on sample size and is also affected by the size of the correlations in the model. A nonsignificant p value for Chi-square indicates an acceptable fit for the model (Byrne, 2001). However, a good-fitting model is not necessarily a valid model. Other fit indices such as the Tucker Lewis index (TLI), or non-normed fit index (NNFI), the Comparative fit index (CFI) and the root square error of approximation (RMSEA) are not dependent on sample size. TLI and CFI values greater than .90 and a RMSEA value less than .05 both indicate a good fit for the model. I calculated all indicators for the sake of comparability.

Next, mean accuracy and latency scores were submitted to a series of one-way Analyses of Variance (ANOVAs) at each testing time (n = 3) with Group (Beginning, Intermediate, Advanced) as the between-subjects factor. Scheffé post hoc analyses were performed in the case that group differences emerged.

To compare intra- and inter-learner achievement over time, univariate analyses of variance were conducted. First, three 3x3 Repeated Measures ANOVAs were conducted with Time as the Within-subjects variable (Time 1, 2, 3) and Group as the Between-subjects variable (3 proficiency levels). Outcome, or dependent variables, are mean performance scores on the tests of i) implicit and ii) explicit knowledge as well as iii) mean latency scores on the test of implicit knowledge. Development is interpreted as the change in accuracy and latency from Time 1 to Time 2 and retention from Time 2 to Time 3 as well as from Time 1 to Time 3. A Time x Group interaction would reveal differential development and retention patterns of implicit and explicit knowledge for learners at beginning, intermediate and advanced proficiency levels.
As this experimental design measures performance at three points in time, linear or curvilinear development was also investigated using a growth curve linear analysis (GCA) as recommended by statisticians (Chen, 2012) and researchers in the field of cognitive psychology and education (e.g., Heck, Thomas & Tabata, 2010; Singer & Willett, 2003) and SLA (e.g., Ortega & Iberri-Shea, 2005). In this statistical approach, the logic is based on the slope as an indicator of change over time. Thus, a slope and intercept were calculated for each participant during the time interval with the intercept being the estimated baseline of performance and the slope being the person’s change score. A higher or faster rate of learning would be indicated by a more pronounced slope. To see whether there are differences in change over time between proficiency groups, relative slope and intercept averages are calculated within a Level-1 unconditional (within-person) and conditional model (between-level), with the latter being used with fixed predictors (i.e., proficiency level). Between-group predictors of variability in intercepts can also be entered as explanatory variables in a Level 2 model (Heck, Thomas & Tabata, 2010, p. 12), which is essentially the equivalent of performing a multiple regression (applicable to RQ2 and 3).

Finally, in order to ameliorate concern that the individual teaching style of instructors impacted learner performance, one-way ANOVAs were performed with mean gain scores (Time 1 to Time 2) on each test as a continuous dependent variable and instructor as an independent variable (coded as 1, 2, 3, etc.). No significant differences were found on either measure of knowledge for any proficiency group (see Chapter 4 Results for RQ1).

**Research questions 2 & 3:** Relationship between learner IDs & development/retention of implicit (RQ2) & explicit (RQ3) knowledge by proficiency level and predictive ability over time.
To investigate whether cognitive IDs (i.e., L2 aptitude, L1 WM, L1 PSTM, & processing speed) and a psychosocial ID (i.e., L2 motivation) significantly relate to performance variance on tests of implicit and explicit knowledge, a Pearson Product moment correlation matrix was run with composite and componential indices for all ID variables and mean accuracy and latency scores at each separate testing session as well as calculated difference scores. Finally, to analyze the differential ability of learner IDs to predict development of implicit/explicit knowledge by proficiency level, a series of hierarchical multiple regression analyses were run. For relationships between ID measures and performance scores identified to be robust based on power and effect size, those ID scores were entered as explanatory, or predictor variables in a series of hierarchical regression models. Cognitive and affective IDs with the strongest correlations were entered in the first step in the regression model to examine the amount of performance variance explained. Then, ID scores with significant, though relatively weaker, relationships with performance were entered into the second and third steps of the regression to identify any additional unique achievement variance explained. Level 2 GCA models were also run with proficiency as a fixed predictor and learner ID measures entered as covariates.

**Research question 4: Relationships between cognitive & psychosocial IDs.**

The fourth and final research question inquires about the nature of interaction, if any, between the cognitive and affective dimensions of learning a second language and how that relationship potentially varies over time and at increasing proficiency levels. To statistically address this question, a Pearson product moment correlation matrix was run between all learner ID measures across groups and by group as well as at each testing time. The strength of relationships that emerge will be interpreted based on their correlation coefficient ($r$), or effect size and power levels ($> .50 - > .80$).
Pilot Study

Overview.

The pilot study phase of this dissertation was carried out in spring semester 2012. The main goal of pilot testing was to ensure reliability and validity of performance measures, to ensure learners understood instructions, and to gauge the feasibility of completing materials (e.g., timing, use of technology, testing location, individual versus group administration, etc.). Thus, only reliability estimates, descriptive statistics, and item difficulty across groups are reported here. Any ID measures that had not been previously administered and timed by the researcher were also completed by pilot participants (i.e., background questionnaire; L2 motivation questionnaires).

Pilot participants.

After receiving university institutional review board (IRB) human subjects approval in early February (#2012-27), the researcher asked permission from the non-intensive language program director and all instructors to visit classes at each proficiency level and recruit participants who were offered five points extra credit in exchange for their time. The researcher visited six total classes of beginning (SPAN-003-Intro I), intermediate (SPAN 021-Int I) and advanced (SPAN 103-Adv I) learners and explained the goals and required time for participation in the study. While several students initially signed up, seventeen completed all four pilot study sessions (12 female; 5 male) with seven beginning learners, five intermediate learners and five advanced learners (Mean age-20.05 years; Max-29 years/Min-18 years). All participants reported normal or corrected vision and were right-handed, with the exception of one left-handed participant.
As previously described, the main background criteria used for inclusion in the study required that learners be native English-speakers with no reported knowledge of other romance languages or study abroad experience in a Spanish-speaking country over two weeks. After applying these criteria to biodata reported in the background questionnaire, data from four participants were excluded from analysis. Two beginning learners were excluded due to reporting French as a native language and deafness on her left side, as well as extensive formal study of French and Italian. One intermediate learner was excluded because he reported that he had been exposed to Spanish in the home growing up and one advanced learner was not eligible based on two months study abroad experience in Costa Rica and Spain. Thus, the overall sample size in the pilot study was reduced to thirteen ($n = 17 > 13$) (beginning: $n = 5$; intermediate: $n = 4$; advanced: $n = 4$).

**Method.**

As the goal of the pilot study was not to measure learning over time but rather to assess the internal consistency of assessment materials, and because IRB human subjects approval did not come through until late February, the timeline of data collection during the spring 2012 semester was condensed (late February – mid-April). Also, all learners completed the oral imitation and untimed grammaticality judgment tasks but the number of learners who completed each of the three versions varied between thirteen (Version 1 (T1)) and six to seven (T2, T3). The first version of each task was created and administered in late February and the second and third versions were subsequently created and administered in mid-April. Thus, the latter versions were split between the total sample size. In the main phase of data collection, test versions will be administered following a counterbalanced, or split-block design as detailed previously.

*Some methodological issues.*
A few issues arose in the administration of the OI task that merit discussion and revision. The OI task was administered individually on SuperLab (4.0). However, gathering oral latency data proved to be a challenge. Latency data measured in milliseconds based on voice onset time requires a special connector (SV-1 key) for SuperLab. At the time of the pilot study, the researcher did not have access to or funding to purchase the device but a rough measure of latency was implemented. Learners were instructed to press the space bar each time they judged the meaning of the sentence they heard as well as when they began repeating it aloud. However, this method proved to be problematic and somewhat complicated. Many learners in each group forgot to press the space bar and continued speaking while others pressed it too quickly. Ultimately, it was determined that the ‘rough’ measure of latency data was not a true reflection of what it intended to measure, the degree of automatic processes, and thus, is not reported here. Oral latency data will be gathered using the SV-1 key in the fall; however, it too may be subject to technical drawbacks (e.g., verbal false starts).

Another issue with the OI task became apparent in the design of meaning judgments. Following Bowles (2011), who did not calculate meaning processing scores, several sentences were ‘personal’ such as ‘Yo cocino todos los días’ ‘I cook every day’. Whether a learner responded ‘Sí’ or ‘No’ to this statement depended on their cooking habits, and not on an objective judgment. Other sentences are not personal, but could be subjectively judged as well, such as ‘La universidad de Georgetown es una escuela respetada’ or ‘Los profesores ganan mucho dinero’. Even judgment of more objective exemplars, such as ‘Muchos estudiantes tomaron clases de español el año pasado’ or ‘Los estudiantes de primer año viven en la residencia estudiantil’ might vary according to learners’ opinions. In fact, very few sentences were determined to be strictly objective (e.g., A los bebés les gusta la leche (Sí); El Gran Cañon
está en Pensilvania (No)). Overall, all learners seemed to demonstrate comprehension of the content of sentences and confirmed at the end of the session that they had understood what the sentences were saying. Also, since the number of idea units is included in the more content-based scoring method (0-4pts), this could be taken as a reflection of meaning comprehension.

Finally, some learners spontaneously corrected sentences to match their meaning judgment, which sometimes resulted in avoidance of the target structure. For example, in response to an ungrammatical exemplar of noun-adjective agreement ‘*Los diccionarios son libros pequeñas’, several learners responded ‘No’ and modified the sentence as ‘Los diccionarios son libros grandes’, avoiding a targetlike production of ‘pequeños’. Based on learner feedback, this misinterpretation will be addressed by including more practice trials to clarify this aspect of the instructions.

Compared to the OI task, the administration of the UGJT was relatively free of issues. It was completed via Survey Monkey online in a group classroom language laboratory. Learners took up to 45 minutes to complete it and demonstrated full comprehension of instructions. A couple technological issues arose in terms of disabling cookies but those were quickly resolved through consultation with the lab assistant.

**Pilot study results.**

**OIT.**

All audio recordings were transcribed and scored following the two scoring procedures previously described (Score A (0-1 pts); Score B (0-4pts)). Descriptive statistics by group are summarized for the three versions of the elicited oral imitation task in Table 6.

Table 6

_Pilot study: Descriptive statistics on the oral imitation task_
<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pts</td>
<td>10.5 (2.4)</td>
<td>15.5 (.7)</td>
<td>13 (3)</td>
</tr>
<tr>
<td>%</td>
<td>25</td>
<td>39</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>73 (19.3)</td>
<td>97 (7.1)</td>
<td>88.3 (14)</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>61</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pts</td>
<td>19 (4.2)</td>
<td>17.7 (2.5)</td>
<td>16 (1.4)</td>
</tr>
<tr>
<td>%</td>
<td>48</td>
<td>44</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>98.6 (10.5)</td>
<td>94 (13.1)</td>
<td>97.5 (7.8)</td>
</tr>
<tr>
<td></td>
<td>61</td>
<td>59</td>
<td>61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>3</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pts</td>
<td>25 (1.6)</td>
<td>26.5 (6.4)</td>
<td>24 (4.2)</td>
</tr>
<tr>
<td>%</td>
<td>63</td>
<td>66</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>123 (3.6)</td>
<td>126 (18.4)</td>
<td>119.5 (13.4)</td>
</tr>
<tr>
<td></td>
<td>77</td>
<td>79</td>
<td>75</td>
</tr>
</tbody>
</table>

Note. Score A (40 points possible); Score B (160 points possible); Group 1 (Beginning); Group 2 (Intermediate); Group 3 (Advanced)

The OIT discriminated well among proficiency levels using both scoring procedures, though Score B provides more variability (0-4 pts) and thus higher overall percentages across groups. Based on Score A (0-1 pts), beginning learner scores range between 25-30%, intermediate learner scores between 40-48% and advanced learner scores between 60-66%. Similar, but higher ranges are obtained with Score B, with the exception of the T2 version of the task for beginning and intermediate learners in which beginners slightly outperformed their intermediate counterparts (61% vs. 59%). Furthermore, within each proficiency group, raw points achieved were similar on each task version indicating comparable difficulty.

Table 7 below summarizes the reliability of the form-based (A) and form and meaning-based (B) scoring procedures along with the strength of the relationship between them.
As can be seen in Table 7, both scoring procedures have high reliability coefficients above .80 and are strongly correlated with one another. Score B (0-4 pts) seems to be the most reliable across testing versions ($\alpha = .899 - .967$) and approximates reliability estimates reported in Ortega and Bowden’s work ($\alpha = .95 - .985$). While Score A (0-1 pts) is also internally consistent ($\alpha = .830 - .872$), it is less reliable than Score B and Bowles’ reported reliability ($\alpha = .96$) which could be related to the low overall sample size.

**UGJT.**

Descriptive statistics by group are summarized for the three versions of the UGJT task in Table 8 and reliability estimates are given in Table 9.
Overall accuracy (SD) | 37 (7.1) | 62 | 43 (5.3) | 71 | 41 (4.7) | 68

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35 (6.1)</td>
<td>59</td>
<td>38 (2.8)</td>
</tr>
<tr>
<td>2</td>
<td>35 (6.8)</td>
<td>59</td>
<td>43 (5.5)</td>
</tr>
<tr>
<td>3</td>
<td>45 (3.8)</td>
<td>76</td>
<td>46 (5.7)</td>
</tr>
</tbody>
</table>

Note. 60 points possible; 1 (Beginning); 2 (Intermediate); 3 (Advanced)

Overall performance on the UGJT was comparable across test versions, ranging from 37-43 raw points (62-71%). Not surprisingly, test versions 2 and 3 are slightly higher due to being administered a little later in the semester (1.5 mos.). The latter two test versions also discriminated well between group performance, with beginners scoring in the low 60’s (60-63%), intermediate learners in the low 70’s (70-72%) and those at an advanced stage of proficiency in the upper 70’s (75-77%). However, on the first test version, beginning and intermediate learners performed the same, reaching 59% accuracy while advanced learners reached 76% accuracy.

Performance on grammatical (n = 30) versus ungrammatical (n = 30) items across groups indicated higher accuracy on grammatical exemplars. This is to be expected based on Bowles and R. Ellis’ findings that ungrammatical exemplars seem to elicit explicit knowledge more effectively. In all three versions, average raw points earned were 20 (3.3 SD), 26 (2.1) and 22 (1.8) points on grammatical items versus 17 (6.2), 16 (4.6) and 19 (5.3) points on ungrammatical items. Ungrammatical exemplars also revealed more variable performance as indicated by the higher standard deviations from the group mean.

Below Table 9 summarizes reliability estimates for the three versions of the UGJT.

Table 9

Pilot study: Untimed grammaticality judgment test reliability

<table>
<thead>
<tr>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
</table>
Note. $\alpha$ (Cronbach’s alpha); V1, Version 1; V2, Version 2; V3, Version 3.

These reliability estimates indicate that while the first version was internally consistent ($\alpha = .83$), the reliability of the other two versions was lower. A likely reason could be related to the fact that thirteen learners completed version 1 whereas seven and six learners completed the latter versions, respectively. Another cause could be the vocabulary used to create items, which was carefully reviewed before making revisions for the main data collection phase.

**Item difficulty.**

The easiest and most difficult items are calculated by overall mean performance by item. For the OI task, I report mean score based on Score B which was slightly more reliable than Score A. For both tasks, the easiest items were grammatical exemplars of subject-verb agreement and gender agreement between nouns and adjectives. Learners scored above three points on average in the OI task which indicates that the majority of learners were able to preserve over half the idea units in the sentence and provide the target structure accurately. In the UGJT, the mean score was one for these exemplars, indicating that 100% of learners judged these accurately.

However, the most difficult items varied somewhat by task, which is not surprising considering the difference in task demands, that is, verbally repeating a sentence with the obligatory target structure (production) versus judging grammaticality (comprehension). The most difficult items on the OI task were grammatical (GR) and ungrammatical (UGR) past hypothetical conditional sentences followed by the imperfect subjunctive (GR & UGR), present subjunctive in noun clauses (UGR) and aspect (UGR), all with average scores of 2 or below. This means that the majority of learners had trouble accurately producing form and preserving
meaning in these items. The average syllable length of past hypothetical conditional sentences, 24 syllables, was also longer than the overall average, between 16 and 17 syllables, which likely affected repetition.

The most difficult items in the UGJT task also involved past and present subjunctive and aspect with average scores less than .5 (out of 1), meaning that they were judged accurately less than half the time. However, past hypothetical conditional items were not among the lowest scoring items despite arguably being the most complex morphologically, syntactically and cognitively. Such chance performance is likely due to guessing and after taking a closer look at the source of learners’ judgment, it was found that around 90% of learners judged these sentences based on feel or guess and no learner in any group provided an accurate rule underlying use. This reinforces the need to analyze judgments in light of their reported source; in the main phase of data collection, answers judged based on guessing will be excluded.

**Summary of changes resulting from pilot study.**

Overall, the pilot-testing phase of assessment materials was very useful in pinpointing several areas for improvement in order to increase reliability and validity of these measures of L2 knowledge, briefly summarized in turn below.

**Oral Imitation Task**

1. Refine latency measures through use of voice onset reaction times.
2. Add practice trials to clarify instructions about meaning judgment and subsequent sentence repetition.
3. Modify meaning of sentences to be as objective as possible in order to calculate a semantic processing score.
4. Shorten syllable length of past hypothetical conditionals.
Untimed Grammaticality Judgment Task

5. Review exemplars in second and third test versions to ensure comparable use of vocabulary.

6. Analyze judgment accuracy in conjunction with reported source of knowledge (Feel, Guess, Rule) and accuracy in verbalizing a rule.
CHAPTER 4: Results

In order to investigate the developmental trajectories of L2 knowledge over time and to identify which cognitive and psychosocial individual differences significantly relate to and can explain variation in the shape and speed of this development, a series of quantitative statistical analyses were performed within and between proficiency groups. Descriptive statistical results are first provided followed by results based on inferential statistical procedures used to answer the four research questions guiding the study. Final summaries of the results according to each research question are provided at the end of the chapter along with corresponding figures.

Descriptive Statistics

In this section, descriptive information about the learning sample is provided including information about instructors who were teaching learners who participated in the study. Then, reliability coefficients (Cronbach's alpha) and summary item statistics are reported for ID measures as well as reliability estimates for the two performance measures. The subsequent section is organized by research question (RQ). Descriptive statistics are first reported for performance measures, followed by the results of inferential analyses used to test for statistical differences in the variables of interest.

Biodata and Language Background Information

Participants.

Eighty-seven native English-speaking L2 learners of Spanish enrolled in non-intensive Spanish courses \((n = 11)\) at Georgetown University participated in the study. Twenty-three learners were enrolled in a second-semester Beginning course (Intro II); thirty-three were enrolled in a second-semester Intermediate course (Int II); and thirty-one were enrolled in a second-semester Advanced course (Adv II) in the Fall semester of 2012. Participant numbers
decreased slightly from the beginning of the semester (Time 1; \(N = 87\)) to the end of the semester (Time 2; \(N = 83\)). A little over a third returned one month after instruction (Time 3; \(N = 33\)), mainly due to Intermediate II learners not continuing in the Spanish curricular sequence the following spring semester and Advanced II learners going abroad in the spring/summer. The participant sample was 32% male (\(n = 28\)) and 68% female (\(n = 59\)) and learners were 19 years old on average (0.8 standard deviation (SD); Max-22; Min-18).

All subjects reported English as their native language and less than one percent of the total sample (\(n = 6\)) reported speaking an additional non-romance native language which included Arabic (1), Mandarin Chinese (1), Geo (1), Urdu/Hindi (2), and Russian (1). All bilingual participants reported native language dominance in English. Participants also reported having some degree of exposure to a third language (L3) including French (5), Portuguese (1), Latin (2), German (2), Japanese (1), Mandarin Chinese (2), Arabic (2), Ukrainian (1), and Russian (1). Those that reported having been exposed to French or Portuguese were contacted to further discuss the length and nature of their exposure and they were determined to have little to no proficiency (e.g., a semester of instruction in junior high or high school).

In terms of background information learning Spanish, thirty-two participants reported having spent two weeks or less in a Spanish-speaking country (\(n = 7\) Beg; \(n = 10\) Int; \(n = 15\) Adv). Average age of exposure (AoE) to Spanish for the beginning group was 16 years old (2.7 SD; Max-22 yrs; Min-10 yrs), 13 years old for the intermediate group (1.7 SD; Max-16 yrs; Min-9 yrs) and 12 years old for the advanced group (1.9 SD, Max-16 yrs; Min-8 yrs). The length of prior classroom instruction in Spanish was reported in years and the majority of beginning learners reported up to one year of previous instruction in Spanish whereas the majority of
intermediate learners reported 3.5-4 years or up to five years. The majority of advanced learners reported 4.5-5 years or more than five years of previous Spanish instruction.

While subjects were primarily grouped based on institutional enrollment, a section of an in-house department placement exam focused on grammar and vocabulary was also administered to confirm group assignment and ensure that there were significant differences between proficiency groups. Table 10 displays mean performance scores, standard deviation from the mean, and minimum and maximum scores on the placement test by proficiency group. All scores are expressed in percentages.

Table 10

Descriptive statistics (%) of the Departmental Spanish Placement Exam (36 items) by group

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td>1) Beg (n = 22)</td>
<td>0.41</td>
</tr>
<tr>
<td>2) Int (n = 33)</td>
<td>0.63</td>
</tr>
<tr>
<td>3) Adv (n = 31)</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note. Beg, Beginning group; Int, Intermediate; Adv, Advanced.

Performance on the placement exam lends support to the decision to group learners based on course enrollment. Intermediate learners scored 22% higher than beginners and advanced learners scored 33% higher than beginners and 11% higher than intermediate learners. Results of a one-way analysis of variance (ANOVA) revealed significant differences between groups (F_{2,85} = 67.96, p = 0.000**). Post hoc Scheffé analyses determined each group to be significantly different from one another (p = 0.000**). While it could be argued that learners at the high and
low end of their respective groups could be placed one group above or below their actual placement, groups are maintained for the purposes of this dissertation.

It was also important to ensure similar amounts of exposure to and interaction in Spanish outside of the classroom over the course of the semester (Time 1 to Time 2) and to control for outside exposure over the holiday break (Time 2 to Time 3). Therefore, learners completed questionnaires at the second and final testing times. On the first questionnaire, they provided estimates of the amount of hours spent preparing for class (i.e., doing homework, etc.) and using Spanish for personal reasons outside the classroom. They also used a four-point Likert scale (Always, Frequently, Sometimes, Never) to rate the frequency of using Spanish according to six activities (+ 'Other' category): 1) Watch TV/movies; 2) Listen to music; 3) Interact with native speakers; 4) Interact online; 5) Read newspapers or magazines; 6) Receive one-on-one tutoring. The questionnaire at Time 3 only included frequency ratings for activities.

At the end of the semester, 79% of learners reported 0-3 or 4-6 hours of Spanish class preparation and 88% reported 0-3 or 4-6 hours of personal use outside class. Less than 10% of learners reported frequently watching TV/movies, interacting online, reading newspapers or magazines or getting tutored during the semester while between 15-20% reported frequently listening to music in Spanish and interacting with Spanish native speakers. Thus, a clear majority of learners (> 85%) answered 'Sometimes' or 'Never' for all activities. Five learners (6% of the overall sample at Time 2) reported another activity outside of those listed (e.g., "Occasionally attended Spanish club meetings"; "I set my computer to Spanish"; "I attempted to understand Spanish written on signs").

Learners reported similar responses one month after instruction with over 90% reporting only 'Sometimes' or 'Never' participating in all listed activities. Nine participants (27% of sample
at Time 3) reported another activity such as reading maps at Disney world, briefly speaking with a ski instructor in Spanish and attempting to have a conversation at Christmas with a cousin who is a professional interpreter. Three advanced learners also reported trying to practice speaking with best friends from home who are Latin American, but when asked for more detail, they said they almost always interacted in English. Importantly, no participant reported intentionally studying target structures that they heard or saw in the assessment tasks completed in the study at either the end of the semester or after the holiday break, but reported studying grammar for class assignments (i.e., "I studied verb conjugations in order to write essays for class").

Instructors.

Participants were recruited from several Spanish courses in order to reach numbers needed to meet statistical standards (see Table 5 in the a priori Power Analysis section of Chapter 3). Consequentially, type of instruction learners were exposed to on a daily basis during the period of learning (i.e., a semester) could not be strictly controlled in this study. Nonetheless, as previously detailed, instructors in the Spanish language program are educated from a psycholinguistic perspective and are encouraged to implement the task-based curriculum by engaging learners in communicative activities with a focus on meaning in the classroom and a focus on formal grammar outside the classroom (Leow, 1994, 1995). An effort was also made by the researcher i) to gather language background and self-reported teaching style information about the eight instructors and ii) to statistically control for any differences in learner performance according to instructor.

Beginning instructors \((n = 3)\) reported 2-6 years previous Spanish language teaching experience whereas intermediate instructors \((n = 3)\) reported 3-4 years and one lecturer reported 23 years of experience. Advanced instructors \((n = 2)\) reported a similar number of years' prior
experience, two reporting 4 years and a lecturer reporting 25 years. Overall, there were 3 native English speakers and 5 native Spanish speakers including one Spanish/Czech native bilingual speaker. All instructors rated their second language proficiency (Spanish or English) to be near-native or high.

Additionally, instructor responses to the 21-item Likert-type teaching style survey revealed many similarities among the eight instructors. This brief survey gauged frequency and type of classroom activities employed, including listening, speaking, reading and writing tasks, tasks with technology, and individual, small group/paired tasks and teacher-led discussions. Instructors were also asked to rate type of feedback provided (implicit/explicit) and percentage of time spent speaking on a typical day in their classroom. Likert-type items were rated on a 1-4 scale ranging from Never (1), Sometimes (2), Frequently (3) or Always (4). Based on their average Likert responses, it was found that instructors reported frequently employing a similar range of activity types and a balance of implicit and explicitly focused feedback, as well as promoting a student-centered classroom.

While the self-reported questionnaire responses were useful in confirming similar teaching styles among the instructors, it was most important to ensure that any performance differences found at the end of the semester were not attributable to different instructors. Thus, a series of one-way ANOVAs were run for each performance measure (Dependent variable), the oral imitation task (OIT) and the untimed grammaticality judgment task (UGJT), with Instructor as the Independent variable. No significant differences by instructor were found in the gain score (Time 1 to Time 2) on the oral imitation task for beginning learners ($F_{2, 20} = .214, p = .809$), intermediate learners ($F_{2, 31} = 1.79, p = .184$), or advanced learners ($F_{2, 29} = .003, p = .954$). Similarly, performance on the untimed grammaticality judgment task did not differ by instructor
for any of the three groups ($F_{2,20} = .214, p = .944; F_{2,31} = .443, p = .646; F_{1,29} = .02, p = .905$, respectively). Thus, any significant performance differences found can be attributed to the independent variables of interest.

**Test Reliabilities and Summary Item Statistics: Individual Difference (ID) Measures**

**Cognitive IDs.**

The various measures of cognitive ability in this study have been extensively used in previous experimental research in psychology and SLA with similar college-age participant populations. Their test reliabilities (Cronbach's alpha) are consistently reported to be high, in the range of .83-.91 for sections of the MLAT (e.g., de Graff, 1997; Robinson, 2002; Winke, 2005) and .70-.90 for digit and complex span tasks tapping phonological and executive working memory (Conway et al., 2005, p. 776). In Table 11 below, the summary item statistics, expressed in percentage scores, are provided for all measures of cognitive capacity, namely, the Modern Language Aptitude Test (MLAT-Short Form), the Operation Span (O-Span) task, the Digit Span task, and the Digit Symbol Coding Task (DSCT).

**Table 11**

*Summary item statistics (%) on measures of cognitive capacity for all participants (N = 87)*

<table>
<thead>
<tr>
<th>Measures</th>
<th>Summary Item Statistics</th>
<th>M (SD)</th>
<th>Min.</th>
<th>Max.</th>
<th>N of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling Clues (III)</td>
<td></td>
<td>0.81 (0.15)</td>
<td>0.00</td>
<td>1.00</td>
<td>50</td>
</tr>
<tr>
<td>Words in Sentences (IV)</td>
<td></td>
<td>0.56 (0.14)</td>
<td>0.22</td>
<td>0.89</td>
<td>45</td>
</tr>
<tr>
<td>Paired Associates (V)</td>
<td></td>
<td>0.62 (0.23)</td>
<td>0.13</td>
<td>1.00</td>
<td>24</td>
</tr>
<tr>
<td>Composite</td>
<td></td>
<td>0.68 (.11)</td>
<td>0.35</td>
<td>0.95</td>
<td>119</td>
</tr>
</tbody>
</table>
Performance on the OSpan task was scored following three methods (Absolute (A), Total (T), Lenient (L)) that varied in strictness of scoring criteria, and composite scores for A and L were also calculated as a measure of processing and storage capacity. Since reliability of WMC span scores can in part be derived from correlations among them (Unsworth et al., 2005; Waters & Caplan, 2003), bivariate correlations using Pearson Product moment were run between scoring methods. All scores were strongly inter-correlated with high effect sizes ($r$ values) and $p$ values < .001** (OSpan-A & OSpan-T, $r = 0.90, p = 0.00**$; OSpan-A & OSpan-L, $r = 0.84, p = 0.00**$; OSpan-T & OSpan-L, $r = 0.88, p = 0.00**$). Accuracy in the processing component of the OSpan task did not fall below 85% for any participant; thus no WM data were discarded (see Conway, Kane, Bunting, Hambrick, Wilhelm, and Engle, 2005, p. 775).

It was also important to ensure comparability of subjects' working memory capacity before running inferential statistical analyses on WMC scores and performance measures to
address research questions. Table 12 displays the descriptive statistics for performance on the OSpan task by group.

Table 12

*Summary item statistics (%) on OSpan measure by group for all participants (N = 87)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Score</th>
<th>Group</th>
<th>Summary Item Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M (SD)</td>
</tr>
<tr>
<td>OSpan</td>
<td>Absolute</td>
<td>Beg (n=23)</td>
<td>0.38 (0.14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Int (n=33)</td>
<td>0.53 (0.15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adv (n=31)</td>
<td>0.50 (0.20)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Beg</td>
<td>0.67 (0.12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Int</td>
<td>0.73 (0.12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adv</td>
<td>0.73 (0.14)</td>
</tr>
<tr>
<td></td>
<td>Lenient</td>
<td>Beg</td>
<td>0.81 (0.08)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Int</td>
<td>0.85 (0.08)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adv</td>
<td>0.85 (0.11)</td>
</tr>
</tbody>
</table>

*Note*. Beg, Beginning group; Int, Intermediate group; Adv, Advanced Group.

Three one-way ANOVAs were conducted with Group as the Independent variable and OSpan scores (Absolute, Total, Lenient) as Dependent variables. Levene's test revealed homogenous variance among the groups. No significant differences between proficiency groups were found for the OSpan-Total (OSpan-T) or OSpan-Lenient (OSpan-L) scores; however the main effect of group was statistically significant for OSpan-Absolute (OSpan-A), with the strictest order and accuracy criteria \((F_{2,84} = 4.82, p = 0.01^*\). Post hoc Scheffé comparisons showed that scores in the Beginning group were significantly lower than the Intermediate group (38% vs 53%; \(p = 0.01^*\) and marginally lower than the Advanced group (38% vs 50%; \(p = 0.07\)). However, the beginning group also had the lowest number of participants \((n = 23)\), a smaller overall range of
scores (12% to 64%), and thus less variation from the mean (i.e., lower standard deviation) than the other groups. Based on these considerations, OSpan-A scores are maintained as an ID variable in subsequent statistical analyses.

A one-way ANOVA was also run with the measure of phonological short-term memory (Digit Span) as the Dependent variable. No significant group differences were found between proficiency levels ($F_{2,84} = 0.91, p = 0.83$). Group means ranged from 67% (0.14 SD) for beginners to 68% (0.11) and 69% (0.12) for intermediate and advanced learners, respectively. Overall, this learner population is interpreted to possess comparable cognitive resources.

**Psychosocial IDs.**

While all measures of cognitive capacity were administered once in the learning period, this study's research design measured the construct of motivation at two points in time, at the beginning and end of the semester with 2.5 months between testing times. In Table 13 below, reliability estimates (Cronbach's alpha) and summary item statistics, reported in terms of their raw-Likert scale mean responses (1-7), are given for the two measures of motivation, the Attitudes and Motivation Test Battery (AMTB) based on Gardner's socio-educational model and the Selves Questionnaire derived from Dörnyei's Self-system model of L2 motivation.

Cronbach's alpha ($\alpha$) measures how well the items on a test measure a single construct and it is calculated using the number of items and the average inter-item correlation. According to most authors (e.g., Cohen, 1988; Larson-Hall, 2010), a measure can be considered acceptable if Cronbach's alpha is between 0.70-0.80. Reliability estimates are calculated by using scores from the total participant population, which varies at each testing time.

Table 13

*Reliability estimates and summary item statistics on L2 motivation measures (avg. likert response 1-7) from all participants at Time 1 (N = 86) and Time 2 (N = 82-83)*
<table>
<thead>
<tr>
<th>Time</th>
<th>Motivational Components</th>
<th>Cronbach's Alpha (α)</th>
<th>Summary Item Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M (SD)</td>
</tr>
<tr>
<td>AMTB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Integrativeness (IG)</td>
<td>0.82</td>
<td>5.53 (0.65)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.89</td>
<td>5.56 (0.76)</td>
</tr>
<tr>
<td>1</td>
<td>Motivation (MOT)</td>
<td>0.92</td>
<td>5.07 (0.64)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.92</td>
<td>5.06 (0.56)</td>
</tr>
<tr>
<td>1</td>
<td>Instrumental Orientation (IO)</td>
<td>0.77</td>
<td>5.21 (1.02)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.73</td>
<td>5.13 (0.96)</td>
</tr>
<tr>
<td>1</td>
<td>Language Anxiety (LgAnx)</td>
<td>0.93</td>
<td>3.70 (1.00)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.94</td>
<td>3.68 (1.32)</td>
</tr>
<tr>
<td>1</td>
<td>Attitudes Toward Learning Situation (ATLS)</td>
<td>0.96</td>
<td>5.41 (0.73)</td>
</tr>
<tr>
<td>2 (n=83)</td>
<td></td>
<td>0.95</td>
<td>5.54 (0.64)</td>
</tr>
<tr>
<td>1</td>
<td>AMTB Composite</td>
<td>0.94</td>
<td>4.99 (0.52)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.95</td>
<td>5.04 (0.53)</td>
</tr>
<tr>
<td>SELVES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Ideal Self (IS)</td>
<td>0.92</td>
<td>5.24 (1.16)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.89</td>
<td>4.99 (0.98)</td>
</tr>
<tr>
<td>1</td>
<td>Ought-to Self (OS)</td>
<td>0.89</td>
<td>3.02 (1.16)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.83</td>
<td>3.04 (0.99)</td>
</tr>
<tr>
<td>1</td>
<td>Motivated Learning Behavior (MLB)</td>
<td>0.89</td>
<td>4.70 (1.04)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.85</td>
<td>4.58 (0.93)</td>
</tr>
<tr>
<td>1</td>
<td>Selves Composite</td>
<td>0.92</td>
<td>4.32 (0.84)</td>
</tr>
<tr>
<td>2 (n=82)</td>
<td></td>
<td>0.89</td>
<td>4.20 (0.70)</td>
</tr>
</tbody>
</table>
Note. Time 1 (Beginning of semester, September 2012); Time 2 (End of semester, December 2012).

As can be seen in Table 13, all motivational components tested on the AMTB and Selves questionnaire were internally consistent with Cronbach's alpha reaching above 0.70. As with previous studies confirming the validity of the self-system (e.g., Csizér & Kormos, 2005; Dörnyei & Ushioda, 2009b; Ryan, 2009), results here display good internal consistency for this learner population. Cronbach's alpha for ideal L2 self, ought-to self and motivated learning behavior, or intended effort, have previously been reported as .89, .76, and .83, respectively (Dörnyei & Taguchi, 2010, p. 123), in line with alphas reported here though slightly lower. This is also the first empirical study to administer the Selves questionnaire at two points in time, indicating good test-retest reliability given the similar alphas for T1 and T2.

As the version of the AMTB used here was adapted from Yanguas (2007), the procedure he used to further analyze reliability of the AMTB according to the eleven total scales was also followed, and is reported in the table below.

| Table 14 |

Reliability estimates for all scale categories on AMTB measure for all participants

<table>
<thead>
<tr>
<th>Scale</th>
<th>AMTB Component</th>
<th>N of Items</th>
<th>Cronbach's Alpha T1 (N = 86)</th>
<th>Alpha (α) T2 (N = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrative orientation</td>
<td>1) IG</td>
<td>4</td>
<td>0.70</td>
<td>0.80</td>
</tr>
<tr>
<td>Interest in learning Languages</td>
<td>1) IG</td>
<td>10</td>
<td>0.74</td>
<td>0.79</td>
</tr>
<tr>
<td>Attitudes toward Spanish-speaking Community</td>
<td>1) IG</td>
<td>10</td>
<td>0.76</td>
<td>0.83</td>
</tr>
<tr>
<td>Motivational Intensity</td>
<td>2) MOT</td>
<td>10</td>
<td>0.74</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Cronbach's alpha coefficients obtained for the eleven variables are similar to those reported by Yanguas (2007, p. 312) and Gardner et al. (2004, pp. 12-14), with only the latter study reporting reliability at the beginning and end of a learning period as well. As reported by Yanguas ($\alpha = .63$) and Gardner et al. ($T1-\alpha = .73$, $T2-\alpha = .74$), Instrumental Orientation was one of the components that yielded the lowest reliability ($T1-\alpha = .77$, $T2-\alpha = .73$), though alpha was still above .70. Otherwise, all reliability coefficients for the eleven scales comprising the five main components were comparable to previous research.

**Test Reliabilities: Performance Measures**

**Reliability: Oral Imitation Test (OIT) (Implicit Knowledge).**

In this section, I report the reliability estimates (Cronbach's alpha) of the two scores (A-Form-based, 40 points total; B-Form & meaning-based, 160 points total) on each of the three versions of the implicit knowledge test (OIT).

Cronbach's alpha and Pearson's correlation coefficients are given in Table 15 for the three comparable test versions used to carry out a counter-balanced, split-block design. As a reminder, learners received two scores based on formal accuracy only (Score A, 0-1 pts) and form and
meaning-based criteria (Score B, 0-4 pts), with meaning referring to what degree learners were able to repeat the meaning, or idea units, present in the sentence stimuli they heard aurally.

Table 15

*Overall reliability on the oral imitation test (40 items) and correlation between scoring methods on each test version (V1, V2, V3)*

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score A</td>
<td>.828</td>
<td>.848</td>
<td>.859</td>
</tr>
<tr>
<td>Score B</td>
<td>.930</td>
<td>.910</td>
<td>.927</td>
</tr>
<tr>
<td>α</td>
<td>.969**</td>
<td>.955**</td>
<td>.960**</td>
</tr>
</tbody>
</table>

*Note. V1, Version 1; V2, Version 2; V3, Version 3.
α (Cronbach’s alpha); r (Pearson’s correlation coefficient)
*p ≤ .05*; *p ≤ .001***

As Table 15 shows, both scoring procedures proved highly reliable with alpha coefficients above .80 for each testing version. Scoring methods were also strongly correlated with one another, providing further reliability. However, Score B (0-4 pts) was the most reliable across testing versions (α = .91 - .93) and approximates reliability estimates reported in Ortega and Bowden’s research (α = .95 - .985). While Score A (0-1 pts) is also highly internally consistent (α = .828 - .859), it is slightly less reliable than Score B and the alpha coefficient reported in Bowles’ study (α = .96). Based on these comparisons and the focus on meaning as a criterial feature of implicit knowledge, Score B (Form & Meaning based scoring) will be used in statistical analyses carried out to address the first and second research questions.

Reliability: Untimed Grammaticality Judgment Task (UGJT) (Explicit Knowledge).
Next, in the table below I report the reliability estimates (Cronbach's alpha) of overall performance and performance on (un)grammatical items ((U)GR) for each of the three versions of the test tapping explicit L2 knowledge (UGJT).

Table 16

Overall reliability on the untimed grammaticality judgment test (N = 60 items) and by item grammaticality (N = 30 GR/30 UGR) on each test version (V1, V2, V3)

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.749</td>
<td>0.820</td>
<td>0.718</td>
</tr>
<tr>
<td>GR</td>
<td>0.557</td>
<td>0.560</td>
<td>0.590</td>
</tr>
<tr>
<td>UGR</td>
<td>0.736</td>
<td>0.831</td>
<td>0.699</td>
</tr>
</tbody>
</table>

Note. GR, Grammatical; UGR, Ungrammatical; V1, Version 1; V2, Version 2; V3, Version 3.

Reliability estimates for the UGJT were high for all three test versions but alpha was above .80 for version 2 only and only for UGR items. Reliability for versions 1 and 3 are slightly lower than coefficients reported in R. Ellis (α = .83) and Bowles (α = .94) but higher than that reported in Erçetin and Alptekin (α = .61). No previous study employing the UGJT has reported reliability estimates according to grammaticality of items. For this learner population, it was evident that responses to UGR items were more consistent than responses to GR items.

The descriptive statistics and reliability estimates presented in this section aimed to give background information about participants in the study, their performance on ID measures, and to report the internal consistency of materials used. The section that follows presents descriptive performance statistics on measures of L2 knowledge and results of the statistical analyses used to answer each research question.

Descriptive and Inferential Statistics: Performance Measures
Research question 1: Development & Retention of Implicit/Explicit Knowledge over Time by Proficiency Level

To address the first research question (What is the trajectory of development of implicit and explicit L2 knowledge for classroom learners of Spanish at increasing proficiency levels during and after a semester of instruction?), I first report descriptive statistical information about learner performance on the tests of implicit and explicit knowledge. Then, a series of One-way and Repeated Measures Analyses of Variance (ANOVAs) were conducted with mean scores on each measure to test for significant performance differences within and between groups at each point in time and over the course of 3.5 months. A Level-1 growth curve linear analysis (GCA) was also run to further analyze within and between-learner growth trajectories and to confirm the interpretation of results obtained from the ANOVAs. Finally, a correlation matrix and principal component factor analysis were run with performance scores at Time 1 to provide further evidence that the tests used were indeed tapping separable knowledge constructs in this learner population, as reported in previous literature (e.g., Bowles, 2011; R. Ellis, 2005; Erçetin & Alptekin, 2013).

Descriptive and inferential statistics: Oral imitation test (OIT).

Three sources of performance data were gathered from the test of implicit knowledge at each testing time: i) meaning processing accuracy, ii) sentence repetition accuracy and iii) sentence repetition latency.

Meaning comprehension: OIT.

First, I report mean accuracy scores for meaning judgments (Si/No) in Table 17 given by learners before repeating the sentences they heard aurally, indicating whether they thought the content of the sentences was true or false.
Table 17

Descriptive statistics for meaning processing score on the oral imitation task by group and testing time

<table>
<thead>
<tr>
<th>Group</th>
<th>Testing Time</th>
<th>OIT Meaning Processing Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
</tr>
<tr>
<td>All</td>
<td>1 (n = 87)</td>
<td>0.86 (0.09)</td>
</tr>
<tr>
<td></td>
<td>2 (n = 83)</td>
<td>0.91 (0.08)</td>
</tr>
<tr>
<td></td>
<td>3 (n = 33)</td>
<td>0.95 (0.05)</td>
</tr>
<tr>
<td>Beg</td>
<td>1 (n = 23)</td>
<td>0.78 (0.12)</td>
</tr>
<tr>
<td></td>
<td>2 (n = 21)</td>
<td>0.83 (0.10)</td>
</tr>
<tr>
<td></td>
<td>3 (n = 11)</td>
<td>0.90 (0.06)</td>
</tr>
<tr>
<td>Int</td>
<td>1 (n = 33)</td>
<td>0.86 (0.06)</td>
</tr>
<tr>
<td></td>
<td>2 (n = 32)</td>
<td>0.92 (0.06)</td>
</tr>
<tr>
<td></td>
<td>3 (n = 8)</td>
<td>0.96 (0.03)</td>
</tr>
<tr>
<td>Adv</td>
<td>1 (n = 31)</td>
<td>0.92 (0.05)</td>
</tr>
<tr>
<td></td>
<td>2 (n = 30)</td>
<td>0.94 (0.03)</td>
</tr>
<tr>
<td></td>
<td>3 (n = 14)</td>
<td>0.98 (0.03)</td>
</tr>
</tbody>
</table>

Note. Time 1 (Beginning of semester, September 2012); Time 2 (End of semester, December 2012); Time 3 (End of holiday break, January 2013); Beg, Beginning group; Int, Intermediate group; Adv, Advanced Group.

Table 17 shows that all groups improved over time in their ability to accurately comprehend sentences they heard aurally. From the beginning to the end of the fall semester (2.5 months), beginning and intermediate learners' scores increased around 5-6% and advanced learners improved 2%. One month after the end of instruction (i.e., Time 3), scores continued to improve, between 4% (Int, Adv) and 7% (Beg) with net gains in meaning comprehension (T1 > T3) ranging between 6% (Adv), 10% (Int) and 12% (Beg). Thus, 3.5 months after the start of the semester, all groups improved in meaning comprehension accuracy in this task. The continued
improvement from Time 2 to Time 3 cannot be attributed to outside exposure to Spanish after the end of instruction, given the almost complete lack of contact with Spanish self-reported by learners on the language contact questionnaire.

**Sentence repetition accuracy: OIT.**

Table 18 below provides the descriptive performance statistics according to both form (A) and form/meaning-based (B) methods of scoring sentence repetition accuracy. All scores are reported in percentages.

Table 18

*Descriptive statistics for sentence repetition accuracy on the oral imitation task by group and testing time*

<table>
<thead>
<tr>
<th>Group</th>
<th>Testing Time</th>
<th>Accuracy</th>
<th>Score A</th>
<th></th>
<th>Score B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M (SD)</td>
<td>Min.</td>
<td>Max.</td>
</tr>
<tr>
<td>All</td>
<td>1 (n=87)</td>
<td>0.46 (0.15)</td>
<td>0.08</td>
<td>0.95</td>
<td>0.63 (0.14)</td>
</tr>
<tr>
<td></td>
<td>2 (n=83)</td>
<td>0.54 (0.16)</td>
<td>0.20</td>
<td>0.98</td>
<td>0.70 (0.12)</td>
</tr>
<tr>
<td></td>
<td>3 (n=33)</td>
<td>0.59 (0.17)</td>
<td>0.23</td>
<td>0.88</td>
<td>0.74 (0.13)</td>
</tr>
<tr>
<td>Beg</td>
<td>1 (n=23)</td>
<td>0.31 (0.11)</td>
<td>0.08</td>
<td>0.55</td>
<td>0.49 (0.13)</td>
</tr>
<tr>
<td></td>
<td>2 (n=21)</td>
<td>0.36 (0.10)</td>
<td>0.20</td>
<td>0.58</td>
<td>0.56 (0.10)</td>
</tr>
<tr>
<td></td>
<td>3 (n=11)</td>
<td>0.44 (0.10)</td>
<td>0.23</td>
<td>0.55</td>
<td>0.62 (0.11)</td>
</tr>
<tr>
<td>Int</td>
<td>1 (n=33)</td>
<td>0.44 (0.09)</td>
<td>0.25</td>
<td>0.65</td>
<td>0.63 (0.08)</td>
</tr>
<tr>
<td></td>
<td>2 (n=32)</td>
<td>0.54 (0.11)</td>
<td>0.28</td>
<td>0.80</td>
<td>0.71 (0.07)</td>
</tr>
<tr>
<td></td>
<td>3 (n=8)</td>
<td>0.63 (0.17)</td>
<td>0.35</td>
<td>0.85</td>
<td>0.77 (0.10)</td>
</tr>
<tr>
<td>Adv</td>
<td>1 (n=31)</td>
<td>0.58 (0.14)</td>
<td>0.25</td>
<td>0.95</td>
<td>0.74 (0.09)</td>
</tr>
<tr>
<td></td>
<td>2 (n=30)</td>
<td>0.66 (0.14)</td>
<td>0.43</td>
<td>0.98</td>
<td>0.78 (0.09)</td>
</tr>
<tr>
<td></td>
<td>3 (n=14)</td>
<td>0.68 (0.13)</td>
<td>0.43</td>
<td>0.88</td>
<td>0.81 (0.08)</td>
</tr>
</tbody>
</table>

*Note.* Score A: Form-based (40 points possible); Score B: Form & Meaning-based (160 points possible); Time 1 (Beginning of semester, September 2012); Time 2 (End of semester, December
2012); Time 3 (End of holiday break, January 2013); Beg, Beginning group; Int, Intermediate group; Adv, Advanced Group.

As seen in Table 18, sentence repetition accuracy shows a similar pattern to meaning comprehension with gains in accuracy at the end of instruction and continued improvement one month later. Across groups, there was a 7% gain in accuracy for Score B from the start to the end of the semester and a 4% improvement one month later (11% net increase over 3.5 months). By proficiency group, advanced learners descriptively outperformed intermediate learners who outperformed beginning learners at each testing time, as expected. However, as proficiency increases performance gains decrease over time. Beginners and intermediate learners increased 7-8% in accuracy while advanced learners gained 4% over the two and a half month period. One month after the end of instruction, beginners and intermediate learners both improved an additional 6% (for a 13-14% net increase, respectively), whereas advanced learners showed a 3% gain in accuracy at Time 3 (7% net increase).

A significant positive relationship was found between meaning processing accuracy and sentence repetition accuracy on the oral imitation task at all testing times ($r = .829**; r = .672**; r = .792**$) with power reaching 100%. In other words, learners with higher meaning comprehension were also more accurate in this task at each point in the learning period.

To investigate whether there were significant performance differences by group on the test of implicit knowledge, participants' accuracy scores on the meaning processing and sentence repetition components were submitted to three one-way ANOVAs as Dependent variables with Group as the Independent variable. The results of the one-way ANOVAs are shown below in Table 19.

Table 19

*Results of one-way ANOVAs with test scores (OIT) by group and testing time*
As Table 19 shows, there were significant performance differences between groups at each testing time. At the beginning of the semester, both intermediate and advanced groups scored significantly higher than beginning learners and advanced learners significantly outperformed intermediate learners on both meaning processing and sentence repetition accuracy. This pattern persisted at Time 2 for sentence repetition accuracy but not for meaning comprehension. While both intermediate and advanced groups outperformed beginners, there were no significant differences in meaning comprehension between the higher proficiency groups by the end of the semester, which also held true one month after the end of instruction for both meaning comprehension and sentence repetition accuracy.

To test for significant differences in improvement over time by proficiency group (i.e., Time*Group interactions), mean scores were submitted to two Repeated Measures ANOVAs (RM-ANOVAs) with Time as the within-subjects factor with two (T1-T2) and three levels (T1-T2-T3) and Group as the between-group factor with three levels (Beg, Int, Adv).
For meaning comprehension, both Time ($F_{2,80} = 33.30, p = 0.00$, partial eta-squared = .294, Power = .969) and Group ($F_{2,80} = 26.13, p = 0.00$, partial eta-squared = .395, Power = .969) were statistically significant from the beginning to the end of the semester (2.5 mos.), but no Time*Group interaction was found indicating that all groups uniformly improved over time. Over 3.5 months ($T_1 > T_3$), both Time ($F_{2,30} = 18.30, p = 0.00$, partial eta-squared = .497, Power = 1.00) and Group ($F_{2,30} = 9.57, p = 0.00$, partial eta-squared = .390, Power = .969) again were significant, with no Time*Group interaction. The line graph in the figure below shows mean gains in meaning comprehension (%) by group over the total investigation period (3.5 mos.).

Figure 4

*Gains in meaning comprehension (OIT) over time by group*

Similar results were obtained for the RM-ANOVAs run with mean sentence repetition accuracy scores. Both Time ($F_{2,80} = 55.89, p = 0.00$, partial eta-squared = .411, Power = 1.00) and Group ($F_{2,80} = 44.04, p = 0.00$, partial eta-squared = .524, Power = 1.00) were statistically significant from the beginning to the end of the semester (2.5 mos.), but no Time*Group interaction was found indicating that all groups uniformly improved over time. Over 3.5 months ($T_1 > T_3$), both Time ($F_{2,30} = 17.32, p = 0.00$, partial eta-squared = .544, Power = .999) and Group ($F_{2,30} = 16.97, p = 0.00$, partial eta-squared = .531, Power = .999) again were significant, with no Time*Group interaction.
interaction. The line graphs below visually display mean gains in oral imitation accuracy by group over the total investigation period (Figure 5) as well as group means with error bars showing +/- 2 SDs from the mean (Figure 6). As a reminder, only 33 learners participated in all sessions at all three testing times \((n = 11, \text{Beg}; n = 8, \text{Int}; n = 14, \text{Adv})\).

Figure 5

*Gains in sentence repetition accuracy (OIT) over time by group*

![Figure 5](image)

Figure 6

*Gains in sentence repetition accuracy (OIT) over time by group with error bars*

![Figure 6](image)
Figure 5 visually shows the differences in means for the beginning, intermediate and advanced proficiency groups, as previously described, with more marked distance between performance by beginning and advanced learners. Although gains over time appear to be similar, a steeper increase is evident for the beginning group, especially from the beginning to the end of the semester. Given that advanced learners have more knowledge of Spanish grammar, it is expected to see less pronounced development over time. The SD error bars in Figure 6 also portray higher performance variability for the beginning group, especially at the outset of the learning period.

**Linear growth curve analyses: OIT.**

Given that traditional analyses of change like Repeated Measures ANOVAs assume that individuals follow similar trajectories in their change over time and eliminate subjects who have not participated in each wave of data collection, thus decreasing power, a linear growth curve analysis (GCA) was also performed as a robust means of investigating the intra- and interindividual variation found in performance on the test designed to tap implicit knowledge (OIT). As discussed in depth in Chapter 3, this analysis does not assume the same exact trajectory for all subjects in a sample or group because individual growth parameters (intercepts and slopes) are calculated for each person (Singer & Willett, 2003, pp. 53-54). Thus, GCA accounts for the fact that different people have distinct change trajectories based on differences in their intercept (i.e., the constant and standard from which change is measured) and slope (i.e., mean rate of growth). A latent growth curve is the average of all individual growth curves.

As there were three waves of measurement in this study, it was possible to model nonlinear growth in L2 development over the 3.5 month period. First, a Level-1 unconditional growth curve model was posited to establish a baseline for sentence repetition accuracy across
time (intraindividual variation) and to find out whether variance in the slope, or rate of growth, was significant. Then, a Level-2 conditional growth curve model was run to account for interindividual variation in growth between groups. Table 20 displays the unstandardized coefficients (i.e., in the original metric (%)), for latent growth curve parameters on the oral imitation task. Significant coefficients for fixed and random effects indicate that effects and variances associated with intercepts and slopes deviate from 0.

Table 20

*Unstandardized coefficients for growth curve parameters and variances: OIT*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td><strong>Level-1 Unconditional Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>58.5**</td>
<td>1.6</td>
</tr>
<tr>
<td>Slope</td>
<td>4.92**</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Level-2 Conditional Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>70.3</td>
<td>2.1</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs. 2</td>
<td>-11.06</td>
<td>3.3</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>-25.25</td>
<td>3.2</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>-14.20</td>
<td>3.0</td>
</tr>
<tr>
<td>Slope</td>
<td>3.56</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs. 2</td>
<td>1.93</td>
<td>2.1</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>1.33</td>
<td>2.0</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>3.26</td>
<td>2.0</td>
</tr>
</tbody>
</table>

*Note.* 1, Beginning group; 2, Intermediate group; 3, Advanced group.

A $t$ statistic is used to make statistical comparisons. $p < .05^*,$ $p < .01^{**}.$
The coefficients in the Unconditional (Level-1) and Conditional (Level-2) models provide information similar to the results of the RM-ANOVA: there is significant improvement within individuals over time and significant differences exist between individuals in their intercepts. However, groups do not significantly differ in the slope of their development, or their rate of change over time, as indicated previously in the nonsignificant Time*Group interaction found with RM-ANOVA. Moreover, there was little difference in estimated overall gains found with the RM-ANOVA and GCA analyses between T1 and T2 (+4%, +4.92%, respectively) with a slightly higher estimate of slope variance between T1 and T3 for GCA (+8% versus +9.8%).

Importantly, the relationship between the intercept and slope was negative in the intraindividual unconditional model ($r = -.763^{**}$) indicating that the higher one's intercept, or baseline of development, the lower, or less pronounced the slope, or change in their development in this task. This makes sense in light of group performance differences by proficiency level; advanced learners had higher intercepts at the outset and thus, less growth in knowledge over time, in contrast to beginners whose performance baseline was lower which allowed for a steeper latent growth curve. Nonetheless, the less pronounced slope for advanced learners was not statistically different than the lower proficiency groups given the nonsignificant Slope*Group coefficients.

While the latent growth curve (i.e., mean group trajectory) averages all of the individual growth curves together and shows a positive linear increase over time, it is important to consider the variation in individual developmental trajectories. To visually display this variation, accuracy scores were randomly selected from five participants from each proficiency group who had completed each of the three assessment waves. Figure 7 shows that the shape of individual learner development within each group clearly differs from an upward linear trajectory. Some
learners have a more pronounced positive or negative slope between each assessment wave while others appear to show very little change. Advanced learners, like participant 69, whose intercept (initial level) is relatively high (around 90%) from the outset, show a flatter, less pronounced trajectory while beginning learners, like participant 18, convey a steeper upward slope with a much lower intercept.

Figure 7

*Individual growth curve trajectories: Implicit knowledge.*

![Individual Growth Curve Trajectories](image)

*Sentence repetition latency: OIT.*

In the version of the oral imitation task implemented here, oral reaction time (RT) data were also gathered, or the time it took participants to repeat a sentence stimulus heard aurally after judging its meaning (in milliseconds). Automatic language processing is though to reflect quick and effortless performance because it does not require conscious control or additional attentional resources (Segalowitz 2003). However, no previous research study using the OIT has collected oral reaction time data despite its potential to indicate the degree to which learners were automatically accessing their knowledge (i.e., more or less rapidly).
Before calculating mean RTs by individual and group, the data were screened for outlier RTs on all items \((n = 40)\) following standard data filtering and cleaning procedures in cognitive psychology and SLA (Jiang, 2011; Lachaud & Renaud, 2011) with one exception; given the low accuracy rates on this challenging task, RTs could not be filtered according to error response rates. As an alternative, accuracy scores from the OIT were entered as a covariate in all subsequent analyses run with latency as a dependent variable (see the section on Latency in Chapter 3), a procedure recommended by a resident statistician at the researcher's home institution (Dr. Rusan Chen, personal communication, March, 2013).

Overall, 8,120 RT data points were analyzed across testing times (T1-3,480, 40 items x 87 participants; T2-3,320, 40 items x 83 participants; T3-1,320, 40 items x 33 participants) and RTs faster than 500 ms or slower than 5000 ms were removed and treated as outliers. Based on these values, RTs that were 3 SDs above or below the participant's mean RT were also excluded from the analyses. Additionally, if over a third (33%) of the overall data for each individual participant was missing, their mean RT was not included in the analyses following Rodgers (2011). Overall, 10.8% of data (T1-T2-T3) were excluded as outliers. 70% of the total participant sample was included at Time 1 \((n = 61)\), 55% at Time 2 \((n = 55)\) and 52% at Time 3 \((n = 21)\).

Additionally, a coefficient of variation (CV) was calculated for each participant at each testing time following Segalowitz and colleagues' formula \((\text{SD divided by mean RT})\). This enables one to differentiate between speedup and automatization longitudinally with a significant decrease in CV over time indexing the degree of the latter. Table 21 displays both mean latency accuracy \((\text{in milliseconds (ms)})\), standard deviation from the mean, and mean CVs by group and testing time. These data were found to meet the linearity assumption that mean RTs and SDs are positively correlated (Hulstijn et. al, 2009, p. 563) and this relationship was significant at each
testing time (T1−r = .844**; T2−r = .711**; T3−r = .751**) with power reaching 100% at T1 and T2 and 96% at T3.

Table 21

Descriptive statistics for latency (ms) on the oral imitation task by group and testing time

<table>
<thead>
<tr>
<th>Group</th>
<th>Testing Time</th>
<th>Latency M (SD)</th>
<th>Min.</th>
<th>Max.</th>
<th>Coefficient of Variation (CV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1 (n=76)</td>
<td>963 (341)</td>
<td>514</td>
<td>1877</td>
<td>.39 (.13)</td>
</tr>
<tr>
<td></td>
<td>2 (n=59)</td>
<td>1082 (384)</td>
<td>545</td>
<td>2075</td>
<td>.38 (.12)</td>
</tr>
<tr>
<td></td>
<td>3 (n=21)</td>
<td>937 (243)</td>
<td>579</td>
<td>1408</td>
<td>.35 (.10)</td>
</tr>
<tr>
<td>Beg</td>
<td>1 (n=21)</td>
<td>1156 (415)</td>
<td>635</td>
<td>1877</td>
<td>.42 (.15)</td>
</tr>
<tr>
<td></td>
<td>2 (n=16)</td>
<td>1267 (413)</td>
<td>707</td>
<td>2075</td>
<td>.46 (.11)</td>
</tr>
<tr>
<td></td>
<td>3 (n=7)</td>
<td>968 (261)</td>
<td>608</td>
<td>1408</td>
<td>.40 (.08)</td>
</tr>
<tr>
<td>Int</td>
<td>1 (n=29)</td>
<td>893 (333)</td>
<td>514</td>
<td>1759</td>
<td>.39 (.14)</td>
</tr>
<tr>
<td></td>
<td>2 (n=25)</td>
<td>1042 (397)</td>
<td>545</td>
<td>1832</td>
<td>.36 (.11)</td>
</tr>
<tr>
<td></td>
<td>3 (n=6)</td>
<td>938 (324)</td>
<td>579</td>
<td>1357</td>
<td>.40 (.15)</td>
</tr>
<tr>
<td>Adv</td>
<td>1 (n=26)</td>
<td>887 (211)</td>
<td>603</td>
<td>1499</td>
<td>.38 (.12)</td>
</tr>
<tr>
<td></td>
<td>2 (n=18)</td>
<td>968 (277)</td>
<td>582</td>
<td>1576</td>
<td>.33 (.10)</td>
</tr>
<tr>
<td></td>
<td>3 (n=8)</td>
<td>910 (184)</td>
<td>644</td>
<td>1215</td>
<td>.32 (.06)</td>
</tr>
</tbody>
</table>

Note. 40 items; Time 1 (September 2012); Time 2 (December 2012); Group 3 (January 2013); Beg, Beginning group; Int, Intermediate group; Adv, Advanced Group

In terms of learner proficiency level, Table 21 shows that at each testing time intermediate and advanced learners responded faster than beginners while average oral reaction times for the two higher proficiency groups were similar. On the other hand, SD from the mean indicated less variable responses with increasing proficiency with the exception that beginners had lower SD than their intermediate peers 3.5 months after the start of instruction. Thus, while
advanced and intermediate learners were similar in response speed compared to beginners, response variability distinguished more clearly between each of the three levels. Over time, the pattern that emerges for all groups is a slight increase in oral latency from the beginning to the end of instruction (2.5 mos.) (i.e., slower mean RTs) with a slight decrease in latency one month later (i.e., faster mean RTs). From the first to the last testing time, intermediate and advanced learners were slightly slower in their oral sentence repetition (45 and 23 ms, respectively), while the lowest proficiency group was faster by 188 ms. SD shows a similar pattern to that found for oral latency with an increase after a semester of instruction, with the exception of the beginning group whose SD essentially remains the same, and a subsequent decrease one month after instruction. From the first to the last testing time, a reduction in response variability is evident for all groups. Overall, the variable pattern observed for both oral latency and SD (i.e., increase followed by a decrease) was less pronounced for advanced learners, indicating more stable response latencies over time for this group.

Moreover, the fact that the standard deviation from the mean was the lowest for advanced learners at all three testing times was reflected in the index of automatization (CV). While the advanced group displayed the lowest CVs and was the only group to show gradually decreasing CVs over time (.38 > .33 > .32), beginners displayed the highest CVs with a more variable pattern of response efficiency over time (.42 > .46 > .40). On the other hand, while intermediate learners initially seem to become more efficient in their oral sentence repetition by the end of the semester (.39 > .36), they show an increase in average CV one month later (.40). Thus, this pattern seems to indicate that the lower proficiency L2 learners are at a more variable point in their development of L2 knowledge of Spanish grammar.
Taken together, there is some evidence for faster and more efficient use of language knowledge at higher proficiency levels and automatization of knowledge over time across the overall sample of instructed learners with gradually decreasing CVs (.39 > .38 > .35). On the other hand, all groups responded more slowly and more variably after a semester of instruction, which is somewhat surprising given that we expect learners to become faster with more L2 exposure. Thus, it may be that learners were actually processing more of the target structures present in the aural sentence stimuli at the end of the semester, resulting in slower and more variable oral repetition. Nonetheless, when the total investigation period (T1 > T3) is taken into account, SD data indicate that all groups became less variable in their responses. These results are further considered in the Discussion.

To investigate whether there were significant differences in response latency (Mean RT) and efficiency (Mean CV) by group, mean RTs and mean CVs at each testing time were submitted to three one-way ANOVAs with Group as the Independent variable and latency scores as the Dependent variable. The results of the one-way ANOVAs are shown below in Table 22.

Table 22

Results of one-way ANOVAs with latency scores (OIT) by group and testing time

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{2,60}$</td>
<td>$F_{2,44}$</td>
<td>$F_{2,16}$</td>
</tr>
<tr>
<td>Mean RT</td>
<td>5.70**</td>
<td>3.16*</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Adv, Int &gt;</td>
<td>Adv &gt; Beg</td>
<td></td>
</tr>
<tr>
<td>Beg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean CV</td>
<td>0.898</td>
<td>4.17*</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>Adv &gt; Beg</td>
<td></td>
</tr>
</tbody>
</table>
Note. OIT, Oral Imitation Task; The symbol > indicates that the first group(s) responded significantly faster than the second group (Mean RT) or were more efficient in their RTs (Mean CV). *p < .05, **p < .01.

As Table 22 shows, the difference in mean response times for beginners and the higher proficiency groups was statistically significant at the beginning of the semester; specifically, the advanced group was 330 ms faster than beginners and intermediate learners were 255 ms faster. However, the difference between advanced and intermediate groups did not reach statistical significance, as the mean RT for the advanced group was only slightly lower. At the end of the semester, the advanced group was again descriptively faster than the other two groups, but this difference was statistical only compared to beginners (336 ms). In terms of the index of automatization calculated for each group (CV), only the advanced group was significantly more efficient (i.e., lower CVs) than beginners in orally repeating the target sentences at the end of the semester but did not statistically differ from intermediate learners. At Time 3, no significant differences were found, which is likely related to the small sample size (n = 21).

To test for significant changes in response latency over time by proficiency group (i.e., Time*Group interactions), mean RT and CV scores were submitted to two Repeated Measures ANOVAs (RM-ANOVAs) with Time as the within-subjects factor with two (T1 > T2) and three levels (T1 > T2 > T3) and Group as the between-group factor with three levels (Beg, Int, Adv). Analyses of covariance (RM-ANCOVAs) were also run with mean accuracy scores on the OIT entered as covariates in order to account for not filtering RTs by incorrect responses (see Chapter 3 for a more detailed description).

For mean RTs, significant main effects emerged for both Time ($F_{2,52} = 8.19, p = 0.006$, partial eta-squared = .136, Power = .802) and Group ($F_{2,52} = 3.59, p = 0.035$, partial eta-squared = .121, Power = .641) from the beginning (T1) to the end of the semester (T2), but no
Time*Group interaction was found indicating that no group significantly differed from the other in their change in oral latency. However, when latency scores are adjusted according to accuracy performance in the RM-ANCOVA, neither within- or between-subjects effects reach significance, indicating that Time and Group did not significantly explain variance in oral latency when response accuracy was accounted for. Over 3.5 months (T1 > T3), there were no significant effects found, which again could be related to the sample size at the final testing time.

The line graph in the figure below visually displays the change in mean response latency (ms) from the beginning to the end of the semester (2.5 mos.) by group.

Figure 8

Mean RTs over time from the beginning to the end of the semester (2.5 mos.) by group

Note. 1, Beginning group; 2, Intermediate group; 3, Advanced group.

For response efficiency (mean CVs), the within-subjects factor (Time) was not significant whereas a main effect for Group was found (F_{1, 53} = 6.32, p = 0.003, partial eta-squared = .193, Power = .881) (T1 > T2), but no significant interaction was found. However, when CVs are adjusted according to accuracy performance in the RM-ANCOVA, the between-subjects effect does not reach significance, indicating that Group did not significantly explain variance in response efficiency when accuracy was accounted for. Over 3.5 months (T1 > T3), there were no
significant effects found. The line graph in the figure below visually displays the change in mean response efficiency over the semester by group.

Figure 9

*Mean CVs from the beginning to the end of the semester (2.5 mos.) by group*

![Graph showing changes in mean response efficiency over the semester by group.]

*Note.* 1, Beginning group; 2, Intermediate group; 3, Advanced group.

To investigate the nature of the relationship between accuracy and response latency on the test of implicit knowledge, bivariate Pearson product moment correlations were also conducted. Accuracy was negatively related to mean RTs at both the outset (T1−$r = -.507^{**}$, $N = 61$, Power = 0.99) and end of the learning period (T2−$r = -.411$, $N = 45$, Power = 0.81) but no relationship was found one month later. In other words, learners who responded faster in this task (i.e., lower mean RTs) at the beginning and end of the semester were also more accurate.

Finally, a mean RT was calculated for each participant in their native language (English) based on a reduced English version of the oral imitation task. Given that participants learned their native language implicitly and use it automatically, the idea was to assess the difference (ms) between participants' mean L1 English RT (Eng RT) and their mean L2 Spanish RT (Span RT) in order to evaluate the extent to which learners were accessing their L2 knowledge more or
less automatically. In other words, the lower the mean difference between Eng and Span RTs, the more likely learners were accessing their L2 knowledge implicitly, or similarly to their L1. Table 23 displays this information in milliseconds (ms).

Table 23

*Descriptive statistics for difference (milliseconds) between L2 mean RT (Spanish) and L1 mean RT (English) by group and testing time*

<table>
<thead>
<tr>
<th>Group</th>
<th>Testing Time</th>
<th>Latency: Span RT - Eng RT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
</tr>
<tr>
<td>All</td>
<td>1 (n=45)</td>
<td>260 (291)</td>
</tr>
<tr>
<td></td>
<td>2 (n=37)</td>
<td>447 (279)</td>
</tr>
<tr>
<td></td>
<td>3 (n=12)</td>
<td>322 (259)</td>
</tr>
<tr>
<td>Beg</td>
<td>1 (n=11)</td>
<td>460 (345)</td>
</tr>
<tr>
<td></td>
<td>2 (n=12)</td>
<td>617 (276)</td>
</tr>
<tr>
<td></td>
<td>3 (n=5)</td>
<td>339 (275)</td>
</tr>
<tr>
<td>Int</td>
<td>1 (n=19)</td>
<td>253 (256)</td>
</tr>
<tr>
<td></td>
<td>2 (n=17)</td>
<td>408 (262)</td>
</tr>
<tr>
<td></td>
<td>3 (n=3)</td>
<td>410 (270)</td>
</tr>
<tr>
<td>Adv</td>
<td>1 (n=15)</td>
<td>121 (212)</td>
</tr>
<tr>
<td></td>
<td>2 (n=18)</td>
<td>274 (189)</td>
</tr>
<tr>
<td></td>
<td>3 (n=4)</td>
<td>234 (280)</td>
</tr>
</tbody>
</table>

We see in Table 23 that descriptively there is a much smaller difference (ms) at each testing time between advanced learners' L2 mean RT and L1 mean RT compared to their lower proficiency peers. Likewise, the 'distance' between intermediate learners' L2 mean RT and L1 mean RT is smaller than that found for beginners. Thus, there seems to be evidence that learners
at higher proficiency levels were accessing their L2 knowledge of Spanish more similarly to the way they use their native language in a parallel task than novice learners.

To determine whether there were statistical differences in L1 and L2 RTs according to group, three one-way ANOVAs performed at each testing time revealed that the advanced group's mean RT difference was significantly smaller than that of the beginning group at Time 1 ($F_{2, 44} = 5.12^*, p = 0.01$) and Time 2 ($F_{2, 36} = 4.74^*, p = 0.02$). However, while descriptively smaller, the mean difference (Span RT-Eng RT) for the intermediate group did not significantly differ from the beginning group.

As Table 23 shows, the 'distance' between L1 and L2 latency grows over 2.5 months of instruction (Time 1 to Time 2) for all groups, indicating longer Span RTs at Time 2 (i.e., slower response times). When Span RT-Eng RT difference scores from Time 1 to Time 2 were submitted to a Repeated Measures-ANOVA, both Time ($F_{1, 48} = 7.11, p = 0.01$, partial eta-squared = .186, Power = .827) and Group ($F_{1, 48} = 5.48, p = 0.007$, partial eta-squared = .193, Power = .881) were statistically significant with no Time*Group interaction found. The main effect for Time could also be related to the fact that learners were exposed to more complex grammar in their classrooms over the course of the learning period, resulting in slower response times compared to the start of the semester, but this remains purely speculative. No significant effects were found from Time 1 to Time 3.

**Descriptive and inferential statistics: Untimed grammaticality judgment task (UGJT).**

Three sources of performance data were gathered from the test of explicit knowledge as measured on the untimed grammaticality judgment task at each testing time: i) grammaticality judgment accuracy, ii) judgment source attribution and iii) rule accuracy (when rules were provided for items judged to be ungrammatical in Spanish).
Judgment accuracy: UGJT.

Below, I first report mean accuracy scores (%) for grammaticality judgments (Table 24) overall and by item type ((U)GR) and then present inferential statistical analyses used to investigate significant differences in learner performance. Subsequent sections present descriptive and inferential statistics for judgment source attributions (Feel, Guess, Rule) and mean judgment accuracy by source attributions.

Table 24

Descriptive statistics for judgment accuracy on the untimed grammaticality judgment task (UGJT) by group and testing time

<table>
<thead>
<tr>
<th>Group</th>
<th>Testing Time</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GR M (SD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
</tr>
<tr>
<td>All</td>
<td>1 (n=86)</td>
<td>0.70 (0.12)</td>
</tr>
<tr>
<td></td>
<td>2 (n=83)</td>
<td>0.74 (0.12)</td>
</tr>
<tr>
<td></td>
<td>3 (n=33)</td>
<td>0.77 (0.09)</td>
</tr>
<tr>
<td>Beg</td>
<td>1 (n=22)</td>
<td>0.61 (0.09)</td>
</tr>
<tr>
<td></td>
<td>2 (n=21)</td>
<td>0.64 (0.09)</td>
</tr>
<tr>
<td></td>
<td>3 (n=11)</td>
<td>0.71 (0.07)</td>
</tr>
<tr>
<td>Int</td>
<td>1 (n=33)</td>
<td>0.71 (0.08)</td>
</tr>
<tr>
<td></td>
<td>2 (n=32)</td>
<td>0.74 (0.10)</td>
</tr>
<tr>
<td></td>
<td>3 (n=8)</td>
<td>0.83 (0.06)</td>
</tr>
<tr>
<td>Adv</td>
<td>1 (n=31)</td>
<td>0.76 (0.10)</td>
</tr>
<tr>
<td></td>
<td>2 (n=30)</td>
<td>0.80 (0.10)</td>
</tr>
<tr>
<td></td>
<td>3 (n=14)</td>
<td>0.79 (0.10)</td>
</tr>
</tbody>
</table>

Note. 60 points possible; Time 1 (September 2012); Time 2 (December 2012); Group 3 (January 2013); Beg, Beginning group; Int, Intermediate group; Adv, Advanced Group; GR, grammatical; UGR, ungrammatical
Overall, learners across groups showed a 4% gain in accuracy from the start to the end of the semester and a 3% improvement one month later (7% total increase over 3.5 months). By proficiency group, advanced learners descriptively outperformed intermediate learners who outperformed beginning learners at the start and end of the semester, as expected; however, intermediate learners descriptively outperformed both beginners and advanced learners at Time 3, in contrast to the test of implicit knowledge (OIT). Overall, beginners and intermediate learners increased 3% in accuracy and advanced learners gained 4% over the two and a half month period while one month after the end of instruction, beginners improved an additional 7% and intermediate learners improved an additional 9% (for a 10-12% net increase, respectively). In contrast, advanced learners showed a -1% decrease in accuracy at Time 3 (3% net gain) which could be interpreted as a plateau in the development of explicit knowledge for advanced learners.

In comparison to accuracy scores on the OIT, learners scored higher on average on the UGJT than on the OIT with percentages across groups ranging from 70% (T1) to 77% (T3) on the UGJT versus 63% (T1) to 74% (T3) on the OIT. By group, superior performance on the UGJT versus the OIT was consistent. This is unsurprising, given that this test is designed to tap explicit knowledge and these are instructed learners who have learned Spanish in a classroom setting. However, overall net gains on the UGJT were lower than those on the test of implicit knowledge (OIT) over time (i.e., 7% net accuracy gain (UGJT) compared to an 11% net gain (OIT)), which is likely related to the higher mean scores on the UGJT at the outset of instruction.

Finally, Table 24 shows that all groups more accurately judged grammatical items compared to ungrammatical items as evidenced in the lower percentage scores for the UGR items, which were around 20% less accurate across groups.
To investigate whether there were significant performance differences by group on the test of explicit knowledge, three one-way ANOVAs were run with Group as the Independent variable and mean grammaticality judgment scores the Dependent variable. The results of the one-way ANOVAs are shown below in Table 25.

Table 25

Results of one-way ANOVAs with test scores (UGJT) by group and testing time

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{2,85}$</td>
<td>Post hoc</td>
<td>$F_{2,82}$</td>
</tr>
<tr>
<td><strong>UGJT</strong></td>
<td>17.62**</td>
<td>Adv, Int &gt; Beg</td>
<td>16.70**</td>
</tr>
<tr>
<td></td>
<td>$p = .059$</td>
<td>Adv &gt; Int</td>
<td>$p = .069$</td>
</tr>
</tbody>
</table>

Note. UGJT, Untimed Grammaticality Judgment Task; The symbol > indicates that the first group scored significantly higher than the second group.

* $p < .05$, ** $p < .001$.

At both Times 1 and 2, the higher proficiency learners (Int, Adv) were significantly more accurate in their grammaticality judgments compared to beginners. The statistical difference between Int and Beg (Int > Beg) persisted one month after instruction as well. However, the mean difference in accuracy between advanced learners and lower proficiency groups only approached significance at later assessment waves.

To test for significant differences in improvement over time by proficiency group (i.e., Time*Group interactions), mean scores were submitted to two Repeated Measures ANOVAs (RM-ANOVAs) with Time as the within-subjects factor with two (T1-T2) and three levels (T1-T2-T3) and Group as the between-group factor with three levels (Beg, Int, Adv).

Both Time ($F_{2,80} = 13.94$, $p = 0.00$, partial eta-squared = .148, Power = .958) and Group ($F_{2,80} = 22.22$, $p = 0.00$, partial eta-squared = .357, Power = 1.00) were statistically significant
from the beginning to the end of the semester (2.5 mos.), but no Time*Group interaction was found indicating that there were comparable gains over time. Over 3.5 months (T1 > T3), again both Time ($F_{2,30} = 6.61, p = 0.004$, partial eta-squared = .313, Power = .880) and Group ($F_{2,30} = 9.44, p = 0.001$, partial eta-squared = .386, Power = .967) were statistically significant with no Time*Group interaction. The line graph in Figure 10 below shows mean gains in judgment accuracy (%) by group over time (3.5 mos.) with error bars showing +/- 2 SDs from the mean.

Figure 10

_Gains in grammaticality judgment accuracy over time (3.5 mos.) by group_

![Graph showing gains in grammaticality judgment accuracy over time (3.5 mos.) by group.](image)

*Note.* 1, Beginning group; 2, Intermediate group; 3, Advanced group.

While Figure 10 shows seemingly higher standard deviation from the mean for all groups compared to SD on the oral imitation task, performance on each task actually has similar intragroup variance (see Tables 18 and 24). Also the lower proficiency groups show continued gains at Time 3 while the advanced group seems to plateau, with a 1% decrease in accuracy.

**Linear growth curve analyses: UGJT.**

As with the test of implicit knowledge, a Level-1 and Level-2 GCA were also performed in order to confirm the intra- and interindividual variation found in judgment accuracy.
performance on the test of explicit knowledge (UGJT). Table 26 displays the unstandardized coefficients for the growth curve parameters and variances associated with the intercepts and slopes on the UGJT.

Table 26

*Unstandardized coefficients and variances for growth curve parameters: UGJT*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Level-1 Unconditional Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>67.2**</td>
<td>1.5</td>
</tr>
<tr>
<td>Slope</td>
<td>3.22**</td>
<td>.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
<th>Variance (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-2 Conditional Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>73.8</td>
<td>2.2</td>
<td>34.2**</td>
</tr>
<tr>
<td>Group 1 vs. 2</td>
<td>-11.20</td>
<td>3.4</td>
<td>3.29**</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>-17.04</td>
<td>3.3</td>
<td>-5.14**</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>-5.85</td>
<td>3.1</td>
<td>-1.86</td>
</tr>
<tr>
<td>Slope</td>
<td>2.60</td>
<td>1.4</td>
<td>1.91**</td>
</tr>
<tr>
<td>Group 1 vs. 2</td>
<td>0.82</td>
<td>2.2</td>
<td>.705</td>
</tr>
<tr>
<td>1 vs. 3</td>
<td>1.48</td>
<td>2.1</td>
<td>.713</td>
</tr>
<tr>
<td>2 vs. 3</td>
<td>0.66</td>
<td>2.0</td>
<td>.349</td>
</tr>
</tbody>
</table>

*Note. 1, Beginning group; 2, Intermediate group; 3, Advanced group. A t statistic is used to make statistical comparisons. p < .05*, p <.01**.*

As found with the test of implicit knowledge, the coefficients in the Unconditional (Level-1) and Conditional (Level-2) models support results obtained from the RM-ANOVA: while there is significant improvement within individuals over time and significant differences
between individuals in their intercepts (with the exception of the advanced and intermediate groups), no significant interindividual variation was found in the change in development of explicit knowledge by group. There was again little difference in estimated overall gains across groups found in the RM-ANOVA and GCA analyses between T1 and T2 (+3% versus +3.22%) or between T1 and T3 (+6% versus +6.4%).

The relationship found in the unconditional model between the intercept and slope again was negative ($r = -.801^{**}$) showing that the higher one's intercept, or baseline of development, the lower, or less pronounced the slope of their development in accuracy on this task.

Variation in individual growth curve trajectories compared to mean group trajectories, or latent growth curves, is shown in Figure 11 using accuracy scores from the same participants randomly selected in the test of implicit knowledge. The figure below shows considerably different shapes of development over time for individual learners in comparison with the latent growth curve for their respective groups. Some learners show positive or negative curvature in their linear change between assessment waves, while others only show continuous increases or decreases in performance. Certain advanced learners like subject 58 whose intercept (initial level) is relatively high (around 90%) from the outset, show a flatter, less pronounced trajectory whereas low-level learners like participant 19 have a sharper slope given a much lower intercept. Compared to individual growth curves on the test of implicit knowledge, these are clearly more variable in nature.

Figure 11

*Individual growth curve trajectories: Explicit knowledge.*
In addition to judging the grammaticality of the sentences they read, learners were also asked to indicate the source of their judgment (Feel, Guess, Rule), principally to assess the extent to which learners were basing their judgments on knowledge of Spanish grammar rules. Table 27 shows the descriptive statistics for learner-reported source attributions.

Table 27

<table>
<thead>
<tr>
<th>Grammaticality judgment source attributions (UGJT) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>All</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Beginning</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Advanced</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Note.* Time 1, T1; Time 2, T2; Time 3, T3.

As can be seen in the table above, the majority of learners in all groups at all testing times reported basing their grammaticality judgments on Rule knowledge whereas around a third reported using 'Feel', or intuition. A very low percentage of learners reported to be guessing with the exception of the beginning group who guessed at more than double the rate of learners at more advanced levels of proficiency. On the other hand, the advanced group reported using their
intuition to make their judgments more frequently than learners at lower proficiency levels, specifically 4-8% more frequently at Time 1 and 2 and 12-32% more frequently at Time 3. Nonetheless, more than a quarter of beginning and intermediate learners reported using 'Feel' to judge the sentences. What this actually means for late adult L2 learners who have had relatively limited exposure to and naturalistic language learning experience is difficult to interpret, which should be kept in mind as a methodological limitation when interpreting these results.

In terms of judgment accuracy according to reported source attributions, learners in all groups were most accurate when they used a grammatical rule in Spanish, as seen in Table 28.

Table 28.

<table>
<thead>
<tr>
<th>Group</th>
<th>Feel T1</th>
<th>T2</th>
<th>T3</th>
<th>Guess T1</th>
<th>T2</th>
<th>T3</th>
<th>Rule T1</th>
<th>T2</th>
<th>T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>.59 (.49)</td>
<td>.65 (.48)</td>
<td>.69 (.46)</td>
<td>.49 (.50)</td>
<td>.56 (.49)</td>
<td>.61 (.41)</td>
<td>.79 (.40)</td>
<td>.81 (.37)</td>
<td></td>
</tr>
<tr>
<td>Beginning</td>
<td>.54 (.49)</td>
<td>.53 (.5)</td>
<td>.61 (.49)</td>
<td>.46 (.50)</td>
<td>.46 (.5)</td>
<td>.59 (.46)</td>
<td>.70 (.44)</td>
<td>.73 (.41)</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>.57 (.50)</td>
<td>.65 (.48)</td>
<td>.73 (.45)</td>
<td>.50 (.50)</td>
<td>.58 (.45)</td>
<td>.62 (.45)</td>
<td>.80 (.40)</td>
<td>.80 (.36)</td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>.63 (.48)</td>
<td>.71 (.46)</td>
<td>.73 (.45)</td>
<td>.52 (.50)</td>
<td>.67 (.47)</td>
<td>.63 (.49)</td>
<td>.84 (.36)</td>
<td>.88 (.33)</td>
<td></td>
</tr>
</tbody>
</table>

Note. T1, Time 1; T2, Time 2; T3, Time 3.

As Table 28 shows, learners accurately judged sentences 70-88% of the time when they reported using Rule knowledge whereas they were accurate 53-73% of the time when they 'felt' they were making the correct judgment, but did not explicitly know why, and 46-67% accurate when they guessed. These differences were shown to be statistically significant based on the results of a one-way ANOVA ($F_{2,5159} = 170.87, p = .000$). Post hoc Scheffé tests showed that judgments based on Guessing were significantly less accurate than those made according to intuition (Feel) or rule knowledge (Rule) ($p = .000$) and judgments based on Feel were
significantly less accurate than those based on Rule ($p = .000$). The same results were found at Time 2 ($F_{2, 4979} = 107.04, p = .000$) and Time 3 ($F_{2, 1979} = 36.46, p = .000$) with the exception that mean judgment accuracy did not significantly differ according to Feel and Guess for Time 3. To test for above-chance accuracy performance according to guessing, a non-parametric binomial test was run. Results revealed that mean judgment performance based on guessing was not significantly above chance at Time 1 ($.49, p = .55$), but was significantly above chance at the end of the learning period ($.56, p = .03$) and one month later ($.61, p = .02$), suggesting that a semester of instruction benefitted learners' ability to accurately judge sentence exemplars even when they reported not knowing whether the sentence was grammatically correct in Spanish.

The bar chart in the figure below visually relays performance results for judgment accuracy according reported source attributions.

Figure 12

*Judgment accuracy by source attributions (%)*

Given that the UGJT used here intended to tap explicit knowledge of a range of grammatical structures of Spanish, the fact that learners performed significantly higher when they reported using rule knowledge lends support to its validity; however, accuracy improved
over time for judgments based on Feel as well, and also slightly increased for self-reported guessing. These results will be further considered in the Discussion in terms of methodological issues in interpreting self-reported source attributions and underlying L2 knowledge.

**Rule knowledge accuracy (UGR items): UGJT.**

The third source of data collected from this test was accuracy in describing rules for items judged to be ungrammatical based on rule knowledge. As the total number of rules provided varied by participant, mean rule accuracy percentages were individually calculated by averaging the number of rules accurately described out of the total number of rule source attributions for UGR items each participant reported (e.g., 15 (rules correctly described) out of 18 (total rules provided for UGR items) = 83% rule accuracy). At the beginning of the semester, both intermediate and advanced learners provided rules for around half of the UGR items ($n = 30$) (Int: 15.85 (3.5 SD); Adv: 15.97 (5.9 SD)) and were accurate in describing those rules 94% and 88% of the time, respectively. Beginning learners on average provided rules for almost a third (31%) of the UGR items (9.32 (5.9 SD)) and accurately described them 84% of the time.

At the end of the semester 2.5 months later, intermediate and advanced learners provided slightly more rules on average (16.84 (5.5 SD), 17.43 (7.0 SD) respectively) than at Time 1. While advanced learners increased in their ability to accurately describe rules (from 88% to 92%), intermediate learners were slightly less accurate (from 94% to 91%). Beginners also provided more rules on average and, like the advanced group, were more accurate in describing them than at the outset of the semester (10.81 rules (5.9 SD), from 84 to 88% accuracy). One month after the end of the instructional period, all groups continued to increase in their ability to accurately describe rules explaining the ungrammatical exemplars (90-98%) and beginners and intermediate learners provided more rules overall than at previous testing times (Beg: 13.45 (5.8
SD); Int: 19.38 (3.4 SD); however, advanced learners on average reported less rule knowledge than at previous testing times (14.43 (9.5 SD)), which is perhaps related to the higher rate of 'Feel' source attributions compared to the other groups.

Statistically, intermediate and advanced learners provided significantly more rules to explain ungrammatical items than did beginning learners at Time 1 ($F_{2,85} = 13.61, p=.000, \eta^2 = .247$) and Time 2 ($F_{2,82} = 8.33 p=.001, \eta^2 = .172$), but not at Time 3. There were few group differences in rule description accuracy with the exception that the intermediate group was significantly more accurate in describing rules than were beginners at Time 3 only ($F_{2,30} = 6.22, p=.006, \eta^2 = .308$). In other words, when learners reported knowing a grammatical rule, they were comparably accurate in verbalizing those rules regardless of proficiency level.

**Test version comparability.**

As this research was structured around a counter-balanced, split-block design, three comparable test versions (A, B, C) of each test were distributed equally among the three proficiency groups at each of the three testing times. For example, if a learner completed test version A at Time 1, she completed version B at Time 2 and version C at Time 3. To ensure that tests were indeed comparable in their level of difficulty, a one-way ANOVA was performed for each testing time with test version as the Independent variable. On the Oral Imitation task eliciting implicit knowledge, there were no significant performance differences on any of the three test versions for Score A ($T1- F_{2,84} = 0.13, p = 0.88; T2- F_{2,80} = 1.15, p = 0.32; T3- F_{2,30} = 2.09, p = 0.14$) or Score B ($T1- F_{2,84} = 0.04, p = 0.96; T2- F_{2,80} = 0.85, p = 0.43; T3- F_{2,30} = 1.33, p = 0.28$). For the Untimed Grammaticality Judgment measure of explicit knowledge, there were no significant mean performance differences between test versions at Time 1 ($F_{2,83} = 3.08, p = 0.05$), Time 2 ($F_{2,80} = 1.17, p = 0.32$) or Time 3 ($F_{2,80} = 1.15, p = 0.32$); the borderline significant
difference at the first testing time can be explained by a slightly lower mean percentage score for test version C (A - 72% (0.11); B - 73% (0.10); C - 66% (0.10)). Despite extensive pilot testing, the 6-7% lower mean score on version C at Time 1 could be due to more difficult vocabulary. Nonetheless, the difference only approaches significance and therefore, all test versions are interpreted to have a comparable level of difficulty. Any significant performance differences found between groups can be attributed to independent variables of interest.

**Test validity.**

**Correlation between placement exam & performance.**

To provide further validity of the psychometric properties of the two tests of implicit and explicit knowledge, Pearson product moment bivariate correlations were run between mean scores on the internal Department placement exam (grammar section) used to confirm proficiency grouping and the mean scores on the oral imitation test and untimed grammaticality judgment test at the beginning of the semester (Time 1). The resulting relationships were significantly positively correlated for the OIT ($r = 0.753$ (Score A), 0.754 (Score B), $p < .001**$) and the UGJT ($r = 0.725$, $p < .001**$), providing further validity that all assessments used reliably discriminated between proficiency levels. In other words, the higher (or lower) participants scored on the placement exam, the higher (or lower) their performance was on the OIT and UGJT.

**Factor analyses.**

Finally, as done in previous empirical studies using the original and adapted versions of the implicit/explicit test battery, a Pearson Product Moment correlation matrix and principal component factor analyses were run with performance scores at Time 1 to evaluate the validity of the tests used with this learner population. Analyses were run with both form-based (A) and
form and meaning-based (B) scores on the test of implicit knowledge with the latter reported in all subsequent analyses unless otherwise noted. Table 29 below shows the correlation matrix for overall learner accuracy scores on the two tests at Time 1 as well as performance on GR and UGR items separately. Each possible pair of tests was significantly intercorrelated with coefficients ($r$) that reached statistical significance at the .05 level or higher, indicating that associations were large enough to conduct the factor analysis (Field, 2009).

Table 29

Pearson Product Moment correlational matrix for tests of implicit (OIT) and explicit (UGJT) L2 knowledge (Time 1)

<table>
<thead>
<tr>
<th>Test</th>
<th>OIT</th>
<th>OIT (GR)</th>
<th>OIT (UGR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGJT</td>
<td>.62**</td>
<td>.60**</td>
<td>.58**</td>
</tr>
<tr>
<td>UGJT (GR)</td>
<td>.47**</td>
<td>.48**</td>
<td>.51**</td>
</tr>
<tr>
<td>UGJT (UGR)</td>
<td>.54**</td>
<td>.41**</td>
<td>.52**</td>
</tr>
</tbody>
</table>

Note. OIT, Oral Imitation Task; UGJT, Untimed Grammaticality Judgment Task; GR, Grammatical exemplars; UGR, Ungrammatical exemplars.

*p < .05, **p < .001 (2-tailed).

Next, a confirmatory principal component factor analysis was run with test scores across proficiency groups (following Bowles, 2011 and R. Ellis, 2004, 2005) in order to inspect eigenvalues for extracted components and to replicate a two-factor model. In the confirmatory factor analysis run using SPSS version 20.0, a two-factor model was specified a priori as done in R. Ellis (2005), R. Ellis and Loewen (2007) and Bowles (2011). Accuracy scores for GR and UGR items were entered separately for both measures of knowledge, which has not been previously done but was recommended by R. Ellis and Loewen (2007) and Isemonger (2007). The two factors accounted for 81.7% of the total performance variance across groups. Factor 1

200
accounted for the majority of the shared variance (67%); although the eigenvalue for the second factor was slightly below 1.0 (.903), it accounted for a substantial increase in the shared variance (i.e., 15.1%). Table 30 shows the eigenvalues of the two factors and Table 31 shows the respective factor loadings for overall accuracy scores as well as performance by item (U/GR).

The oral imitation test (GR & UGR items) loaded heavily on Factor 1 at .80 or higher whereas the untimed grammaticality judgment test (GR & UGR items) loaded heavily on Factor 2 at .70 or higher.

Table 30

*Principal component factor analysis across groups (Time 1)*

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Variance</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.00</td>
<td>66.69</td>
<td>66.69</td>
</tr>
<tr>
<td>2</td>
<td>.903</td>
<td>15.05</td>
<td>81.73</td>
</tr>
</tbody>
</table>

Table 31

*Loadings for principal component factor analysis across groups (Time 1)*

<table>
<thead>
<tr>
<th>Test</th>
<th>Factor</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>OI</td>
<td>.939</td>
<td>.337</td>
</tr>
<tr>
<td>OI-GR</td>
<td>.806</td>
<td>.359</td>
</tr>
<tr>
<td>OI-UGR</td>
<td>.862</td>
<td>.232</td>
</tr>
<tr>
<td>UGJT</td>
<td>.345</td>
<td>.933</td>
</tr>
<tr>
<td>UGJT-GR</td>
<td>.202</td>
<td>.780</td>
</tr>
<tr>
<td>UGJT-UGR</td>
<td>.343</td>
<td>.760</td>
</tr>
</tbody>
</table>

Note. Rotated factor matrix with Varimax rotation and Kaiser Normalization.
As done in previous studies, I calculated model fit indices to test goodness of fit between the empirical data and a proposed model including Chi-square, the Tucker Lewis index (TLI), or non-normed fit index (NNFI), the Comparative fit index (CFI) and the root square error of approximation (RMSEA) (detailed in Chapter 3). Mplus (Version 6), a latent variable modeling program similar to AMOS, was used to calculate goodness of fit statistics. Summary statistics for the two-factor model using form and meaning-based scores on the test of implicit knowledge are $\chi^2 = 1014.8$, $df = 6$, $p = 0.00$, $CFI = 0.11$, $TLI = -.049$, $RMSEA = 1.39$. Goodness of fit summary statistics for the two-factor model using form-based scores on the test of implicit knowledge are $\chi^2 = 1515.5$, $df = 10$, $p = 0.00$, $CFI = 0.15$, $TLI = -.23$, $RMSEA = 1.32$. Based on these indicators, a two-factor model using overall and disaggregated (U/GR) test scores across groups for OIT and UGJT is statistically unlikely to occur with both form-based scores (used in previous research) and form and meaning-based scores calculated for the OIT. This is contrary to fit indices reported for the two-factor model for intermediate Spanish learners in Bowles' study (2011) ($\chi^2 = 2.9$, $df = 4$, $p = 0.90$, $NFI = 0.974$, $RMSEA = 0.00$). One likely reason is that the model for different proficiency levels could be distinct and combining the data in one analysis does not produce a good model that fits all proficiency groups (Rusan Chen, P.C., May 11, 2013).

Therefore, the current study aimed to further probe the psychometric properties of the tests according to learner proficiency level. In line with this objective, three exploratory factor analyses were run by proficiency group without specifying the structure of the model, or the number of factors to be extracted.

Two components were extracted based on eigenvalues for intermediate and advanced Groups while three components were extracted for the beginning group. Table 32 shows that the two-factor model for the intermediate and advanced groups respectively accounted for 75% and
81% of the cumulative variance while the three-factor model for the beginning group accounted for 87% of the total performance variance. Factor loadings also differed by proficiency group as seen in Table 33. Whereas the loadings for the advanced group mirror those of the confirmatory factor analysis across groups, this is not the case for the lower proficiency groups. For beginners the UGR items on the OIT are split between Factors 1 and 2 and the GR items on the UGJT loaded heavily on a third factor (.947). For the intermediate group, factor loadings were similar to the advanced group with the exception that GR items on the UGJT were split between Factors 1 and 2.

Table 32

**Exploratory factor analysis by group (Time 1)**

<table>
<thead>
<tr>
<th>Group</th>
<th>Component</th>
<th>Eigenvalue</th>
<th>Variance</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beg</td>
<td>1</td>
<td>2.89</td>
<td>48.08</td>
<td>48.08</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.31</td>
<td>21.89</td>
<td>69.97</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.04</td>
<td>17.38</td>
<td><strong>87.36</strong></td>
</tr>
<tr>
<td>Int</td>
<td>1</td>
<td>3.26</td>
<td>54.39</td>
<td>54.39</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.22</td>
<td>20.29</td>
<td><strong>74.68</strong></td>
</tr>
<tr>
<td>Adv</td>
<td>1</td>
<td>3.58</td>
<td>59.73</td>
<td>59.73</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.29</td>
<td>21.44</td>
<td><strong>81.17</strong></td>
</tr>
</tbody>
</table>

*Note. Beg, Beginning; Int, Intermediate; Adv, Advanced.*

Table 33

**Loadings for rotated factor matrix by group (Time 1)**

<table>
<thead>
<tr>
<th>Group</th>
<th>Test</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Proficiency</td>
<td>Task</td>
<td>Beg</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>OIT</td>
<td>.950</td>
</tr>
<tr>
<td></td>
<td>OIT-GR</td>
<td>.935</td>
</tr>
<tr>
<td></td>
<td>OIT-UGR</td>
<td>.408</td>
</tr>
<tr>
<td></td>
<td>UGJT</td>
<td>.136</td>
</tr>
<tr>
<td></td>
<td>UGJT-GR</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>UGJT-UGR</td>
<td>.118</td>
</tr>
</tbody>
</table>

Note. OIT, Oral Imitation Task; UGJT, Untimed Grammaticality Judgment Test; GR, Grammatical exemplars; UGR, Ungrammatical exemplars

What these results show is that the validity of the tests used to measure implicit and explicit second language knowledge appears to vary by proficiency level. The factor loadings for the advanced group indicate a clearer separation of knowledge constructs whereas there is considerable overlap in factor loadings for UGR performance scores on the OIT with GR performance scores on the UGJT loading on a third factor for the beginning group. Also, GR performance scores on the UGJT were split between factors for the intermediate group. These results call into question whether the tests are measuring the intended knowledge constructs for all learners, particularly on the OIT as it is very plausible that novice learners are accessing knowledge explicitly, rather than implicitly, for UGR items. This issue will be considered further in the Discussion (Chapter 5).
RQ 2: Relationship between Learner IDs & Development/Retention of Implicit Knowledge by Proficiency Level (2a) and Predictive Ability over Time (2b)

While RQ1 was concerned with identifying significant intra and inter-individual variation in performance on tests of implicit/explicit knowledge, the second research question extends this inquiry to what relationships exist between learner characteristics and performance variation. Specifically, RQ2 asks which learner IDs are related to performance on the measure of implicit knowledge as well as the extent to which they can predict development and retention of implicit knowledge in instructed learners at increasing proficiency levels over a 3.5 month period. First, Pearson Product moment bivariate correlations were run both across and by proficiency group to identify the cognitive and psychosocial IDs that were significantly related to performance at each testing time; these IDs were then entered into a series of sequential (hierarchical) multiple regressions to gauge their relative predictive ability. Building on the results of the regressions, a Level-2 GCA was run to see which ID predictors accounted for group differences in performance as well as rate of growth across the three testing waves.

**Correlations: IDs & implicit knowledge (OIT).**

Correlations between IDs and performance at each testing time are presented, as opposed to relationships with gain scores (T1 --> T2; T1 --> T3), given that few relationships were found with the latter which might be related to low variation in gain scores. The correlation matrix between cognitive ID components (N = 6) and performance on the oral imitation task by grammaticality (All/GR/UGR) and testing time is found in the table below. For ease of presentation, Pearson's r values are provided for relationships found for the total learner sample but relationships by group are noted in the summary below.
Table 34

Correlation matrix: Cognitive IDs and performance on the OIT at Time 1 (n = 87), Time 2 (n = 83) and Time 3 (n = 33)

<table>
<thead>
<tr>
<th>Time</th>
<th>Items</th>
<th>L2 Aptitude (MLAT)</th>
<th>WM</th>
<th>PSTM</th>
<th>PROC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>III-PC IV-GS V-RL</td>
<td>OSpan</td>
<td>Digit Span</td>
<td>DSCT</td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.357** --</td>
<td>--</td>
<td>.255*</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>.359** --</td>
<td>--</td>
<td>.224*</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>UGR</td>
<td>.319** --</td>
<td>--</td>
<td>.231*</td>
<td>.264*</td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.237* --</td>
<td>--</td>
<td>.207+</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>-- .210+ --</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>UGR</td>
<td>.255* --</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>-- -- .364*</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>-- -- --</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>UGR</td>
<td>-- -- .413*</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. MLAT, Modern Language Aptitude Test (III-Phonemic coding ability; IV-Grammatical sensitivity; V-Rote Learning); WM, Working memory; OSpan, Operation Span; PSTM, Phonological short-term memory; PROC, Processing speed; DSCT, Digit symbol coding task; GR, Grammatical exemplars; UGR, Ungrammatical exemplars. Correlation (r) is significant at the 0.01** level (2-tailed). Correlation (r) is significant at the 0.05* level (2-tailed).

As shown in Table 34, several relationships were found between cognitive IDs and accuracy on the measure of implicit knowledge. At the beginning of the semester, phonemic coding (PC) ability, a component of L2 aptitude, and phonological short-term memory (PSTM) were positively related to performance though the former relationship had higher effect sizes (r values) and higher power, at 93% for all items. For the positive association between PSTM and performance on the OIT, power reached 67% across items, 55% for GR items and 70% for UGR items. However, correlational analyses by group revealed that positive relationships with PC and PSTM, particularly for performance on UGR items, were only significant for the lower
proficiency groups (beginning & intermediate). Weak but significant relationships between the Grammatical Sensitivity (GS) component of L2 aptitude as tested on MLAT-IV and performance on UGR items ($r = .378^*, N = 31, p = .036$) and between Rote Learning (RL) and all items ($r = .364^*, N = 31, p = .044$) emerged for the advanced group only.

The executive component of working memory capacity (WMC), as scored on the OSpan-Composite-L (Lenient), also positively related to performance on UGR items at the beginning of the learning period but with relatively low power (58%). Relationships with WMC by group did not reach significance, likely due to the low $r$ value (UGR-.231) found across groups and small sample sizes for each individual group. In sum, a superior ability to encode phonemes in Spanish and more efficient phonological memory resources were related to higher accuracy scores for all items, and UGR items in particular, at the outset of instruction.

At the end of the instructional period 2.5 months later, PC was again significantly related to overall performance and UGR items but relationships were weaker than at Time 1 with lower power across items (56%) and for UGR items (65%). This relationship only reached significance for beginning learners and was more robust than the correlation across groups ($r = .562^{**}, N = 21, p = .008, \text{Power} = 77\%)$. Relationships between grammatical sensitivity (GS) and GR items, and between PSTM and overall performance both approached significance at the end of the semester. However, as found at Time 1, when correlations were run by group, positive relationships with PC and PSTM emerged only for beginning and intermediate groups. PSTM significantly correlated with UGR items only for the intermediate group ($r = .359^*, N = 32, p = .043, \text{Power} = 52\%)$ and PC significantly correlated with UGR items only for the beginning group ($r = .562^{**}, N = 21, p = .008, \text{Power} = 77\%)$. No significant relationships emerged
between performance on this task and any of the measures of cognitive capacity for advanced learners.

One month after the end of instruction, working memory capacity was the only cognitive ID to significantly correlate with overall performance and UGR items, and the latter relationship reached 67% power. This means that there was a 33% chance of not finding a statistical correlation when it in fact existed. Given the low sample size at T3 (n = 33), this relationship would likely reach high power (> .80) with more learners. When correlational analyses were run by group, significant positive relationships were found for beginning learners between PSTM and both GR (r = .669*, N = 11, p = .024, Power = 63%) and UGR items (r = .626*, N = 11, p = .039, Power = 54%). Slightly stronger relationships were found between PC and UGR items (r = .693*, N = 11, p = .018) and WMC and UGR items (r = .688*, N = 11, p = .019) for the beginning group as well with power close to 70% in both cases. For the intermediate group a strong relationship was found between PSTM and UGR items (r = .853**, N = 8, p = .007, Power = 81%). A relationship between rote learning ability and performance on UGR items approached significance for advanced learners. Overall, the higher a learner's ability to encode sounds and to process, manipulate and store language information, the higher their ability to accurately repeat sentences on this task 3.5 months after the start of the semester, especially in the case of UGR exemplars. However, these cognitive abilities were only related to performance for lower proficiency learners.

The correlation matrix between psychosocial ID components (N = 8) and performance on the oral imitation task by item (All/GR/UGR) and testing time is organized in Table 35 below. Constructs of L2 motivation measured at T1 were correlated with performance at T1 and motivational constructs measured at T2 were correlated with performance measured at T2.
Correlations with performance scores at T3 were run with both T1 and T2 mean Likert ratings for all motivational components, and all significant relationships that emerged were related to the T2 measure of motivational components, with the exception of ATLS (T1 ATLS & OIT approached significance).

Table 35

*Correlation matrix: Psychosocial IDs and performance on the OIT at Time 1 (n = 86), Time 2 (n = 83) and Time 3 (n = 33)*

<table>
<thead>
<tr>
<th>Time</th>
<th>Items</th>
<th>AMTB</th>
<th>Selves</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IG</td>
<td>MOT</td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>UGR</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.341**</td>
<td>.274*</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>.297**</td>
<td>.270*</td>
</tr>
<tr>
<td></td>
<td>UGR</td>
<td>.349**</td>
<td>.250*</td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>--</td>
<td>.345*</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>--</td>
<td>.341*</td>
</tr>
<tr>
<td></td>
<td>UGR</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. AMTB, Attitude/Motivation Test Battery; IG, Integrativeness; MOT, Motivation; IO, Instrumental orientation; LgAnx, Language anxiety; ATLS, Attitudes toward learning situation; IS, Ideal self; OS, Ought-to self; MLB, Motivated learning behavior; GR, Grammatical exemplars; UGR, Ungrammatical exemplars

Correlation (r) is significant at the 0.01** level (2-tailed).
Correlation (r) is significant at the 0.05* level (2-tailed).

As Table 35 shows, language anxiety (LgAnx) and situational motivation components (i.e., attitudes toward the learning situation-ATLS) on the AMTB were significantly correlated with overall performance on the test of implicit knowledge, with the strongest relationship between ATLS and UGR items (Power = 80%). The positive relationship with ATLS indicates
that superior performance is related to a more positive evaluation of one's Spanish course and teacher reported at the outset of instruction. Also, a higher mean Likert rating for LgAnx signifies lower anxiety; thus, the positive relationship found with LgAnx indicates that superior performance on the measure of implicit knowledge is related to lower reported Spanish class and use anxiety. Correlational analyses by group showed that the relationship with language anxiety was significant only for beginning learners, for UGR items in particular ($r = .719^{**}$, $N = 21$, $p = .000$, Power = 97%), and the relationship with ATLS was significant only for the intermediate group for both GR ($r = .527^{**}$, $N = 33$, $p = .001$, Power = 87%) and UGR items ($r = .571^{**}$, $N = 33$, $p = .001$, Power = 94%). A moderate relationship between reported Motivational Intensity (MOT) and performance on GR items was also found for intermediate learners. The only relationship to emerge at Time 1 for advanced learners was between Instrumental orientation (IO) and performance on UGR exemplars ($r = .372^*$, $N = 31$, $p = .039$, Power = 58%).

The Ideal L2 self and motivated learning behavior (MLB) components in the Selves questionnaire were also significantly, positively related to performance on the OIT across groups at the beginning of the semester meaning the more positively learners reported viewing their Ideal self in Spanish and the more effort they reported to learn Spanish, the higher their sentence repetition accuracy. The power of the relationship with IS found in the total participant sample approached 100% while the relationship with MLB was relatively weaker. By group, the positive relationship with IS was only found for the intermediate group and for GR exemplars in particular ($r = .435^*$, $N = 33$, $p = .011$, Power = 71%). Relationships with MLB by group did not reach significance, likely due to the low $r$ values (.223-.224) across groups.
At the end of the semester, the strong relationship with LgAnx remained significant with comparable effect sizes and power levels. Again, this relationship was primarily accounted for by the beginning group and was found to be strongly associated not only with performance on UGR items \((r = .594**, N = 21, p = .005, \text{Power} = 90\%)\) but also with GR items \((r = .640**, N = 21, p = .002, \text{Power} = 83\%)\). A weak association between LgAnx and performance on UGR items for intermediate learners also emerged at Time 2 \((r = .376*, N = 32, p = .034, \text{Power} = 57\%)\).

Two other AMTB components, Integrativeness (IG) and Motivational intensity (MOT) were also found to be significantly related to performance across groups at Time 2. Power for the relationship with IG was above 80\% for overall performance and UGR items, with slightly lower power for GR items (78\%). The power of the relationship with MOT was lower, ranging from 62\% (GR) to 69\% (UGR) to 71\% (overall). By group, correlational analyses revealed that the intermediate group accounted for both relationships found with IG and MOT and these were significant for UGR items only (Power = 64\% (IG), 74\% (MOT)).

The relationship with IS at Time 2 remained strong across groups reaching 100\% power for overall performance and for UGR items (96\% for GR items). By group, this relationship was found to be significant only for beginning learners \((UGR - r = .633**, N = 21, p = .002, \text{Power} = 89\%)\) with a relationship approaching significance for the intermediate group. The positive association between performance across groups and reported learning effort (MLB) reached higher power at the end of the learning period (88\%) compared to 2.5 months earlier (55\%). However, MLB was significantly related to performance on both GR and UGR exemplars only for the beginning proficiency group, and this association was stronger for UGR items (90\% vs. 72\% power). For the first time, the Ought-To L2 self component, or one's sense of responsibility
to learn the foreign language, significantly correlated with performance on UGR items on the OIT at the end of the learning period. This relationship was significant for both beginners \((r = .529^*, N = 21, p = .014, \text{Power} = 71\%)\) and advanced learners \((r = .438^*, N = 30, p = .018, \text{Power} = 69\%)\), but not for intermediate learners.

After one month without instruction, LgAnx was the only AMTB component to significantly correlate with performance, but only for UGR items and this relationship had low power (53%), likely due to a smaller sample size at Time 3 \((n = 33)\). Again, this relationship only reached significance for beginning and intermediate groups, for both GR (Power = 76%) and UGR items (Power = 67%) in the case of beginners but only UGR items (Power = 87%) for intermediate learners. Across groups, relationships between MOT and GR items, and ATLS and UGR items approached significance but these associations reached significance for intermediate learners only (Power = 73% and 86%, respectively).

Positive relationships with the Ideal L2 self found across groups remained strong at Time 3 with high power for overall performance and UGR items (> 0.80). However, the latter relationship was explained by advanced group only \((\text{UGR} - r = .635^*, N = 14, p = .015, \text{Power} = 70\%)\). A higher MLB rating was again positively associated with overall performance and UGR items but with low power (52%), meaning that there was about a one in two chance of finding a statistical relationship. For intermediate learners relationships with MLB were stronger, reaching 70% for GR items and above 90% for UGR items but no relationships were found with reported learning effort for the other groups.

**Correlations: Cognitive IDs and latency (OIT).**

Correlational analyses were also run between cognitive IDs and the latency data collected on the test tapping implicit knowledge (OIT). The correlation matrix between cognitive ID
components ($N = 6$) and mean latency (RT) and efficiency (CV) on the oral imitation task by item (All/GR/UGR) and testing time is organized in the table below.

Table 36

*Correlation matrix: Cognitive IDs, latency (RT) and CV (OIT) at Time 1 ($n = 76$), Time 2 ($n = 59$) and Time 3 ($n = 21$)*

<table>
<thead>
<tr>
<th>Cognitive IDs</th>
<th>Time 1</th>
<th></th>
<th></th>
<th>Time 2</th>
<th></th>
<th></th>
<th>Time 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT ($ms$)</td>
<td>CV</td>
<td></td>
<td>RT ($ms$)</td>
<td>CV</td>
<td></td>
<td>RT ($ms$)</td>
<td>CV</td>
<td></td>
</tr>
<tr>
<td>L2 Aptitude (MLAT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III-PC</td>
<td>--</td>
<td>-.320*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>IV-GS</td>
<td>--</td>
<td>-.255*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>V-RL</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-.638**</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Composite</td>
<td>--</td>
<td>-.328*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Working Memory (WM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ospan-Composite-L</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-.487*</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological Short-term Memory (PSTM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>-.309*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Processing Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>--</td>
<td>--</td>
<td>-.318*</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Coding Task (DSCT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. MLAT, Modern Language Aptitude Test (III-Phonemic coding ability; IV-Grammatical sensitivity; V-Rote Learning); WM, Working memory; OSpan, Operation Span (Composite-L); PSTM, Phonological short-term memory; PROC, Processing speed; DSCT, Digit symbol coding task. Correlation ($r$) is significant at the 0.01** level (2-tailed). Correlation ($r$) is significant at the 0.05* level (2-tailed). $^*p = .06$; $^{**}p = .07$*

As Table 36 shows, all significant relationships between cognitive capacity and latency were negative, indicating that faster and more efficient responses (i.e., lower mean RTs & CVs) on the OIT are related to having higher cognitive resources in this learner population. At Time 1,
lower CVs were significantly related to higher phonological working memory resources, as measured by the digit span task (Power = 68%); but when correlational analyses were run by group, the negative association between CVs and PSTM only approached significance for the advanced learners and a significant negative relationship was found between RTs and PSTM for beginners (Power = 61%). The index of automatization (CV) was also significantly negatively related to scores on the composite measure of L2 aptitude, and to phonemic coding and grammatical sensitivity components individually. However, relationships with PC and the MLAT-Composite were much stronger (Power = 71-73%) compared to the robustness of the negative association with the GS component (50%). The former relationships (CVs & PC; CVs & MLAT-Composite) were found to be significant for the intermediate group only (Power = 64% and 59%, respectively).

At the end of the semester, the only cognitive variable to significantly correlate with latency was processing speed, indicating that learners with faster oral reaction times also had a higher processing speed, as measured on the digit symbol coding task (DSCT). However, this relationship had relatively low power (57%) and approached significance for the intermediate group only. One month later, the rote memory component of L2 aptitude was strongly, negatively correlated with mean RTs (Power = 81%) as was the composite (Lenient) measure of working memory capacity (processing + storage), though to a far lesser degree (Power = 51%). No relationships reached significance by group, given the very small sample sizes at Time 3. Overall, these results seem to suggest that learners with more memory resources and a higher ability to encode phonemes in the foreign language also responded faster and more efficiently at the first and final testing time 3.5 months later.

**Multiple regressions: IDs and implicit knowledge (OIT).**
To test the extent to which explanatory variables (IDs) can explain variation in the response variable (accuracy scores on the OIT), a series of sequential, or hierarchical, multiple regressions were run both i) across groups and ii) by proficiency group. This type of regression allows one to identify the unique variance accounted for by an explanatory variable when there are multiple variables that overlap with the response variable. The order in which ID variables were entered were based on the strength of correlations (r) with performance on the OIT. Regression, or explanatory models are provided for cognitive and psychosocial explanatory variables separately, followed by a regression model that includes IDs with the strongest relationships in both dimensions. In each model the multiple correlation coefficient is reported ($R^2$) as well as the squared semipartial correlations, or change in $R^2$ ($\Delta R^2$). $R^2$ provides an indication of how much variance in scores of the response variable can be explained by the variance in the explanatory variable(s) while $\Delta R^2$ expresses the unique contribution of each variable (Larson-Hall, 2011, p. 198).

The cognitive ID model for performance on the test of implicit knowledge at the first testing time constitutes two explanatory variables based on strength of relationships found in correlational analyses: phonological short-term memory and phonemic coding ability, a component of L2 aptitude. The linearity of each relationship was first examined using a scatterplot. For example, the relationship between phonemic coding and performance on the OIT is visually displayed in the figure below. When entered as the only explanatory variable in the regression model, PC explains almost 13% of the variance in the response variable ($R^2$), or performance on the OIT, with a couple notable outliers.

Figure 13

*Relationship between performance on the OIT and phonemic coding ability (T1)*
When entered after PSTM, the unique variance that PC accounts for is essentially cut in half (6%) but remains significant, as seen in Table 37 below. In addition to total $R^2$ and $\Delta R^2$, the unstandardized regression coefficients ($\beta$) are given for each cognitive variable added to the model.

Table 37

*Cognitive ID explanatory model: Implicit knowledge (T1)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>PSTM $\beta$</th>
<th>PC $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 All</td>
<td>.07*</td>
<td>.07*</td>
<td>30.48*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.17*</td>
<td>.17*</td>
<td>40.12*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.20*</td>
<td>.20*</td>
<td>35.23**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.00</td>
<td>.00</td>
<td>2.61</td>
<td></td>
</tr>
<tr>
<td>2 All</td>
<td>.13*</td>
<td>.06*</td>
<td>12.83</td>
<td>28.24*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.20</td>
<td>.03</td>
<td>25.63</td>
<td>14.19</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.20</td>
<td>.00</td>
<td>35.67*</td>
<td>-1.01</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.09</td>
<td>.09</td>
<td>-7.77</td>
<td>30.14</td>
</tr>
</tbody>
</table>

*Note.* PSTM, Phonological short-term memory; PC, Phonemic coding ability.

What Table 37 shows is that by itself, PSTM is a powerful predictor of performance at the beginning of the semester across groups, explaining 7% of the overall variance. When
regressions were run separately for proficiency groups, PSTM accounted for an even higher percentage of variance for the lower proficiency groups (17-20%) whereas this cognitive ID did not explain any variance in the advanced group's scores. In the second model when the phonemic coding component of L2 aptitude is entered, we see that it is a significant predictor overall but predicts little variance for lower proficiency groups (0-3%). PC accounts for 9% unique variance in the advanced group's scores, but this was not statistically significant. When entered together, PSTM remains the only significant predictor for the intermediate learners.

To explain variance in scores at the end of the semester (T2), the order in which the two cognitive ID variables were entered was switched based on the strength of relationships found. PSTM emerged again as a significant predictor for performance by intermediate learners, explaining 13% of unique performance variance. On the other hand, PC ability was the only significant predictor for the beginning learners, accounting for 24% of the variance, though it lost significance when entered with PSTM. Neither cognitive ID significantly explained performance variance for advanced learners.

Table 38

*Cognitive ID explanatory model: Implicit knowledge (T2)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total R²</th>
<th>Δ R²</th>
<th>PC B</th>
<th>PSTM B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
<td>.06*</td>
<td>.06*</td>
<td>22.91*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.24*</td>
<td>.24*</td>
<td>32.93*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.01</td>
<td>.01</td>
<td>-3.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.01</td>
<td>.01</td>
<td>7.47</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.07</td>
<td>.01</td>
<td>17.35</td>
<td>12.59</td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.26*</td>
<td>.02</td>
<td>23.71</td>
<td>14.49</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.14</td>
<td>.13*</td>
<td>-11.86</td>
<td>26.91*</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.01</td>
<td>.00</td>
<td>7.47</td>
<td>-.58</td>
</tr>
</tbody>
</table>
Note. PC, Phonemic coding ability; PSTM, Phonological short-term memory.

One month after the end of instruction (T3), working memory capacity, as measured on the Operation Span task (Composite-Lenient), was the only cognitive ID variable significantly related to performance on the OIT. It should be remembered that at Time 3, the total participant sample consisted of 33 learners and thus, any relationships that do exist in the population should be harder to discern. Nonetheless, WMC significantly predicted 13% of performance variance, or retention, at the final testing time, and this estimate was far higher for beginning (30%) and intermediate learners (47%) than for advanced (2%). However, relationships are borderline significant given the small sample sizes by group ($n=11$ (Beg), $n=8$ (Int)). Again, cognitive capacity was not a significant explanatory variable for the advanced group on this task.

Table 39

*Cognitive ID explanatory model: Implicit knowledge (T3)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>WMC $B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 All</td>
<td>.13*</td>
<td>.13*</td>
<td>93.60*</td>
<td></td>
</tr>
<tr>
<td>Beg</td>
<td>.30&quot;</td>
<td>.30&quot;</td>
<td>117.89&quot;</td>
<td></td>
</tr>
<tr>
<td>Int</td>
<td>.47+</td>
<td>.47+</td>
<td>-283.06+</td>
<td></td>
</tr>
<tr>
<td>Adv</td>
<td>.02</td>
<td>.02</td>
<td>21.58</td>
<td></td>
</tr>
</tbody>
</table>

Note. WMC, Working memory capacity; $p = .06+$, $p = .07$.

Several components of L2 motivation were significantly related to performance on the test of implicit knowledge at the beginning of the semester and were entered in the regression model according to the strength of associations found. Table 40 below shows the results of the psychosocial ID explanatory model for performance at the outset of the learning period.
### Table 40

**Psychosocial ID explanatory model: Implicit knowledge (T1)**

<table>
<thead>
<tr>
<th>Model Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>MLB $B$</th>
<th>LgAnx $B$</th>
<th>ATLS $B$</th>
<th>IS $B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 All</td>
<td>.03</td>
<td>.03</td>
<td>2.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beg</td>
<td>.09</td>
<td>.09</td>
<td>3.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int</td>
<td>.06</td>
<td>.06</td>
<td>1.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adv</td>
<td>.00</td>
<td>.00</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 All</td>
<td>.07*</td>
<td>.04*</td>
<td>1.56</td>
<td>2.83*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beg</td>
<td>.37*</td>
<td>.28*</td>
<td>.31</td>
<td>6.68*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int</td>
<td>.08</td>
<td>.03</td>
<td>1.70</td>
<td>1.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adv</td>
<td>.05</td>
<td>.05</td>
<td>-.65</td>
<td>2.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 All</td>
<td>.11*</td>
<td>.04*</td>
<td>.32</td>
<td>2.90*</td>
<td>4.38*</td>
<td></td>
</tr>
<tr>
<td>Beg</td>
<td>.37*</td>
<td>.00</td>
<td>.32</td>
<td>6.69*</td>
<td>-.047</td>
<td></td>
</tr>
<tr>
<td>Int</td>
<td>.40**</td>
<td>.31**</td>
<td>-.71</td>
<td>1.77</td>
<td>6.98**</td>
<td></td>
</tr>
<tr>
<td>Adv</td>
<td>.13</td>
<td>.09</td>
<td>-.10</td>
<td>2.05</td>
<td>-6.65</td>
<td></td>
</tr>
<tr>
<td>4 All</td>
<td>.18**</td>
<td>.07*</td>
<td>-2.18</td>
<td>2.40</td>
<td>3.51</td>
<td>4.26*</td>
</tr>
<tr>
<td>Beg</td>
<td>.40#</td>
<td>.03</td>
<td>2.78</td>
<td>7.12*</td>
<td>-1.74</td>
<td>-3.83</td>
</tr>
<tr>
<td>Int</td>
<td>.40**</td>
<td>.01</td>
<td>-1.24</td>
<td>1.44</td>
<td>6.48**</td>
<td>.99</td>
</tr>
<tr>
<td>Adv</td>
<td>.17</td>
<td>.03</td>
<td>-1.68</td>
<td>1.95</td>
<td>-6.45</td>
<td>2.16</td>
</tr>
</tbody>
</table>

*Note.* MLB, Motivated learning behavior; LgAnx, Language anxiety; ATLS, Attitudes toward the learning situation; IS, Ideal self; $^*p = .06$; $^#p = .07$.

Table 40 shows that with all four explanatory variables entered, 18% of the overall performance variance was accounted for but the Ideal L2 self was the only significant unique predictor of variance (7%) with all variables entered. However, the picture changes by proficiency group with far more overall variance accounted for in the final model for the lower ability groups (40%). Also, language anxiety (LgAnx) accounted for 28% of variance for the beginners and was the only significant explanatory variable in the final model whereas attitudes toward the learning situation (ATLS) reported by intermediate learners accounted for 31% of the
variance and was the only significant predictor in the final model for this group. Although the four psychosocial IDs together accounted for 17% of variance for advanced learners, the model was not significant.

Integrativeness (IG) and Motivational intensity (MOT) were added to the psychosocial ID model with T2 scores given the significant associations found with learner performance at the end of the semester. Slightly more overall variance was accounted for with all five variables entered (25% versus 18% for the T1 psychosocial model) and the Ideal Self was again the only significant predictor of variance across groups in the final regression model. By group, language anxiety ($\Delta R^2 = 46\%$) and reported motivated learning behavior ($\Delta R^2 = 20\%$) explained a significant amount of unique performance variance for the beginning group whereas no relationships were found for the higher proficiency groups. These five variables together explained a much higher amount of overall variance for the beginners (68%) compared to their higher proficiency peers (16%). The psychosocial ID model for performance on the test of implicit knowledge (T2) is found in Table 41.

Table 41

*Psychosocial ID explanatory model: Implicit knowledge (T2)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>LgAnx</th>
<th>MOT</th>
<th>MLB</th>
<th>IG</th>
<th>IS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.06*</td>
<td>.06*</td>
<td>2.84*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.46**</td>
<td>.46**</td>
<td>6.00**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.07</td>
<td>.07</td>
<td>1.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.03</td>
<td>.03</td>
<td>1.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.10*</td>
<td>.04*</td>
<td>2.02</td>
<td>4.42*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.46**</td>
<td>.00</td>
<td>5.72**</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.14</td>
<td>.07</td>
<td>1.35</td>
<td>3.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.09</td>
<td>.07</td>
<td>.63</td>
<td>3.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In the psychosocial explanatory model run with performance scores 3.5 months after the first testing session, LgAnx (as measured at T2) persisted as a very powerful predictor of performance variance for beginners (56%) and emerged as a significant predictor for the intermediate learners as well (57%). This means that more than half of the performance variance on the oral imitation task one month after instruction was accounted for by learners' reported language anxiety at the end of the semester. Reported MLB (as measured at T2) also persisted as a significant predictor for lower level groups, accounting for 21-29% of unique variance in the response variable in the final model. In contrast to previous points in the semester, IS (T2) was not a significant predictor across groups when all explanatory variables were entered, but did predict a unique amount of variance for beginners and 23% of the variance for advanced learners in Model 2, though this only approached significance. Overall, the three psychosocial IDs significantly predicted over 80% of performance variance for beginning and intermediate learners as shown in Table 42.

Note. LgAnx, Language anxiety; MOT, Motivational intensity; MLB, Motivated learning behavior; IG, Integrativeness; IS, Ideal self; \( ^* p = .06; ^** p = .07 \).
Table 42

**Psychosocial ID explanatory model: Implicit knowledge (T3)**

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>LgAnx</th>
<th>IS</th>
<th>MLB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R^2$</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.07</td>
<td>.07</td>
<td>2.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.56**</td>
<td>.56**</td>
<td>8.68**</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.57*</td>
<td>.57*</td>
<td>7.61*</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.07</td>
<td>.07</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.23*</td>
<td>.16*</td>
<td>.78</td>
<td>6.77*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.62*</td>
<td>.06</td>
<td>10.83**</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.57</td>
<td>.00</td>
<td>7.41</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.30</td>
<td>.23</td>
<td>-1.27</td>
<td>9.55</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>.24*</td>
<td>.01</td>
<td>.88</td>
<td>5.27</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.83**</td>
<td>.21**</td>
<td>13.39**</td>
<td>-10.80*</td>
<td>7.83*</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.85*</td>
<td>.29*</td>
<td>2.58</td>
<td>.10</td>
<td>14.66*</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.31</td>
<td>.01</td>
<td>-1.09</td>
<td>8.64</td>
<td>.74</td>
</tr>
</tbody>
</table>

*Note.* LgAnx, Language anxiety; IS, Ideal self; MLB, Motivated learning behavior; $^*p = .06$; $^#p = .07$.

Finally, up to two significant cognitive and psychosocial explanatory variables were entered together in a sequential regression model to gauge their *relative* predictive ability in performance variance across groups. Based on the strength of relationships, cognitive variables were either entered together in the first step of the regression and psychosocial variables in the second step, or vice versa. As the Tables below show, together cognitive and psychosocial explanatory variables significantly accounted for 22%, 25% and 29% of variance in accuracy on the test of implicit knowledge at each testing time, respectively. Psychosocial IDs consistently predicted comparatively more variance than cognitive IDs (2-3% more at T1 and T3; 13% more at T2). Importantly, the Ideal L2 self was the only explanatory variable to uniquely account for
additional variance on its own at T1 and T2 when all ID components were entered into the regression model.

Table 43

*Cognitive & psychosocial ID combined explanatory model: Implicit knowledge (T1)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>1) PC B</th>
<th>2) ATLS B</th>
<th>IS B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.10*</td>
<td>.10*</td>
<td>28.07*</td>
<td>12.91</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.22**</td>
<td>.12**</td>
<td>20.48</td>
<td>6.22</td>
<td>2.27</td>
</tr>
</tbody>
</table>

*Note.* PC, Phonemic coding; PSTM, Phonological short-term memory; ATLS, Attitudes toward the learning situation; IS, Ideal self; $^+p = .06; ^#p = .07.$

Table 44

*Cognitive & psychosocial ID combined explanatory model: Implicit knowledge (T2)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>1) PC B</th>
<th>2) MLB B</th>
<th>IS B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.06*</td>
<td>.06*</td>
<td>22.91*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.25**</td>
<td>.19**</td>
<td>12.88</td>
<td>.376</td>
<td>5.61**</td>
</tr>
</tbody>
</table>

*Note.* PC, Phonemic coding; MLB, Motivated learning behavior; IS, Ideal self; $^+p = .06; ^#p = .07.$

Table 45

*Cognitive & psychosocial ID combined explanatory model: Implicit knowledge (T3)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>1) WMC B</th>
<th>2) MLB B</th>
<th>IS B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.13*</td>
<td>.13*</td>
<td>93.60*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note. WMC, Working memory capacity; MLB, Motivated learning behavior; IS, Ideal self; $^+ p = .06; ^# p = .07.$

**Level-2 linear growth curve analysis (OIT).**

Multiple regression analyses reported above were used to identify cognitive and psychosocial measures of individual differences that predicted a significant amount of variance in performance at each assessment wave. In addition, person-level (Level 2) Conditional linear models were formulated to account for variability in individual growth parameters (i.e., intercept and slope) in the development of implicit knowledge as found in the Level 1 GCA. Interindividual factors (IDs) found to be associated with performance were entered as covariates, first separately in simple conditional models and then together in combined models to fit the data and assess unique contributions by predictors to variance in initial intercepts and rate of growth.

First simple conditional models were fit to the data by separately entering each ID measure into the model (based on multiple regressions). Results of these analyses revealed that 2 cognitive ID components (PSTM, PC) and 5 motivational components (ATLS, MOT, LgAnx, IS, MLB) each significantly predicted average performance across time. Then, cognitive measures were entered together and psychosocial measures were entered together to fit combined models for predicting intra and interindividual change in performance on the OIT. Both PSTM and PC remained significant independent predictors while only LgAnx and ATLS remained significant in the psychosocial ID model. Thus, PSTM, PC, LgAnx and ATLS were entered in the final combined conditional model to examine whether these IDs made unique contributions to the slope, or rate of growth, on this task and whether they played a differential role by group. No IDs significantly predicted the degree of change in development over time but three IDs significantly predicted group performance differences and are reported below in Table 46.
Table 46

*Final model fitted for implicit knowledge (OIT) with cognitive & psychosocial ID predictors*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Level-1 Unconditional Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>65.6**</td>
<td>13.6</td>
<td>.39**</td>
<td>.05</td>
</tr>
<tr>
<td>Slope</td>
<td>3.50**</td>
<td>1.3</td>
<td>.13**</td>
<td>.03</td>
</tr>
<tr>
<td>Level-2 Conditional Combined Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>26.1</td>
<td>9.1</td>
<td>2.87**</td>
<td>.35** (.05)</td>
</tr>
<tr>
<td>Slope</td>
<td>3.33</td>
<td>1.3</td>
<td>2.69**</td>
<td>.11** (.03)</td>
</tr>
<tr>
<td>Group 1 vs. 2</td>
<td>-11.61</td>
<td>3.5</td>
<td>-3.80**</td>
<td></td>
</tr>
<tr>
<td>Group 1 vs. 3</td>
<td>-25.25</td>
<td>3.2</td>
<td>-8.26**</td>
<td></td>
</tr>
<tr>
<td>Group 2 vs. 3</td>
<td>-12.98</td>
<td>2.9</td>
<td>-4.52**</td>
<td></td>
</tr>
<tr>
<td>Group 1 vs. 2</td>
<td>2.02</td>
<td>1.9</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>Group 1 vs. 3</td>
<td>1.55</td>
<td>1.9</td>
<td>.823</td>
<td></td>
</tr>
<tr>
<td>Group 2 vs. 3</td>
<td>3.57</td>
<td>1.8</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>IDs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LgAnx</td>
<td>2.9</td>
<td>.75</td>
<td>3.90**</td>
<td></td>
</tr>
<tr>
<td>ATLS</td>
<td>3.5</td>
<td>1.3</td>
<td>2.73**</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>14.1</td>
<td>7.0</td>
<td>2.01*</td>
<td></td>
</tr>
<tr>
<td>PSTM</td>
<td>2.8</td>
<td>7.3</td>
<td>.380</td>
<td></td>
</tr>
<tr>
<td>LgAnx*Group 1 vs. 2</td>
<td>5.78</td>
<td>1.9</td>
<td>3.12**</td>
<td></td>
</tr>
<tr>
<td>LgAnx*Group 1 vs. 3</td>
<td>6.03</td>
<td>1.7</td>
<td>3.47**</td>
<td></td>
</tr>
<tr>
<td>LgAnx*Group 2 vs. 3</td>
<td>.26</td>
<td>1.7</td>
<td>.148</td>
<td></td>
</tr>
<tr>
<td>ATLS*Group 1 vs. 2</td>
<td>0.92</td>
<td>2.8</td>
<td>.327</td>
<td></td>
</tr>
<tr>
<td>ATLS*Group 1 vs. 3</td>
<td>13.20</td>
<td>4.0</td>
<td>3.28**</td>
<td></td>
</tr>
<tr>
<td>ATLS*Group 2 vs. 3</td>
<td>12.29</td>
<td>3.8</td>
<td>3.23**</td>
<td></td>
</tr>
</tbody>
</table>
The results of the Level-2 GCA show that with four IDs entered as covariates, variance in individual growth parameters (intercept and slope) for both within (Time) and between-person (Group) effects remained significant in the final model but change over time was not significantly different by group. This finding is supported in the initial Level-2 GCA run without covariates and also reflects the nonsignificant Time*Group interaction found in repeated measures ANOVAs. With the two motivational ID components (LgAnx, ATLS) entered into the final model, phonemic coding ability is the only cognitive ID that remains a significant predictor of higher average performance while PSTM loses significance. Moreover, the psychosocial IDs, LgAnx and ATLS, were stronger predictors of performance on the OIT relative to PC.

By group, LgAnx was a significant predictor of performance on the OIT for beginners but not for the higher proficiency groups whereas ATLS was a significant predictor for both lower proficiency groups but not for the advanced learners. PSTM significantly accounted for performance by beginners compared to the advanced group and this relationship approached significance for intermediate learners (vs. advanced).

**Research question 3: Relationship between Learner IDs & Development/Retention of Explicit Knowledge by Proficiency Level (3a) and Predictive Ability over Time (3b)**

Similarly to RQ2, the third research question asks what learner factors, if any, are related to performance on the measure of *explicit* knowledge as well as the extent to which they can predict development and retention of explicit knowledge in instructed learners at increasing proficiency levels over time. First, Pearson Product moment bivariate correlations were run both
across and by proficiency group to identify the cognitive and psychosocial IDs that were significantly related to performance at each testing time; these IDs were then entered into a series of sequential (hierarchical) multiple regressions to gauge their relative predictive ability. Building on the results of the regressions, a Level-2 GCA was run to see which ID predictors accounted for group differences in performance as well as rate of change across the three testing waves.

**Correlations: IDs and explicit knowledge (UGJT).**

As done with the test of implicit knowledge, correlations for IDs and judgment accuracy performance on the UGJT at each testing time are presented, as opposed to relationships with gain scores (T1 -- T2; T1 -- T3). The correlation matrix between cognitive ID components (N = 6) and performance on the untimed grammaticality judgment task by item (All/GR/UGR) by testing time is organized in the table below. For ease of presentation, Pearson's $r$ values are provided for relationships for the total learner sample but relationships by group are noted in summary below.

Table 47

*Correlation matrix: Cognitive IDs and performance on the UGJT at Time 1 (n = 86), Time 2 (n = 83) and Time 3 (n = 33)*

<table>
<thead>
<tr>
<th>Time</th>
<th>L2 Aptitude</th>
<th>WMC</th>
<th>PSTM</th>
<th>PROC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>III-PC</td>
<td>IV-GS</td>
<td>V-RL</td>
<td>OSpan</td>
</tr>
<tr>
<td>1 All</td>
<td>.240*</td>
<td>.209*</td>
<td>--</td>
<td>.310**</td>
</tr>
<tr>
<td>GR</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>UGR</td>
<td>.301**</td>
<td>.234*</td>
<td>--</td>
<td>.326**</td>
</tr>
<tr>
<td>2 All</td>
<td>--</td>
<td>.226*</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>GR</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>UGR</td>
<td>--</td>
<td>.269*</td>
<td>--</td>
<td>.215*</td>
</tr>
</tbody>
</table>
Table 47 shows that significant, positive relationships found between cognitive abilities and performance on the test of explicit knowledge across groups can almost exclusively be explained by performance on UGR items at the beginning and end of the semester. Two components of L2 aptitude, phonemic coding ability (PC) and grammatical sensitivity (GS), were significantly, positively related to overall performance and UGR items at the beginning of the semester (Power = 81% and 58%, respectively). Working memory capacity, as measured on the OSpan-Composite-A (Absolute), was also strongly related to overall performance and UGR items at Time 1 (Power > 0.80). By group, the relationship between PC and UGR items was significant for beginning learners ($r = .544^{**}$, $N = 22$, $p = .009$, Power = 76%) and a relationship between PSTM and UGR items was revealed for intermediate learners ($r = .453^{**}$, $N = 33$, $p = .008$, Power = 76%) and also approached significance for beginners.

The relationship between performance on the UGJT and GS across groups persisted at Time 2 and was slightly stronger than at Time 1 (Power = 69% vs. 58%). A weak association was found between WMC and performance on UGR exemplars at the end of the semester (Power = 50%) but with only a one in two chance of finding a relationship if it in fact existed. A relationship between phonological memory resources (PSTM) and performance on GR items

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>GR</th>
<th>UGR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. MLAT, Modern Language Aptitude Test (III-Phonemic coding ability; IV-Grammatical sensitivity; V-Rote Learning); WMC, Working memory capacity; OSpan-Comp-A, Operation Span-Composite-Absolute (A); PSTM, Phonological short-term memory; PROC, Processing speed; DSCT, Digit symbol coding task; GR, Grammatical exemplars; UGR, Ungrammatical exemplars
Correlation is significant at the 0.01^{**} level (2-tailed).
Correlation is significant at the 0.05^{*} level (2-tailed).
$p = .06; #p = .07$
across groups approached significance at Time 2; by group PSTM correlated significantly with overall performance and UGR items by the intermediate group only (Power = 62% and 52%, respectively). No significant relationships were found between cognitive IDs and performance by beginning and advanced groups at the end of instruction.

In the last testing session, WMC, as measured on the OSpan-Comp-A, was the only cognitive ID to significantly relate to performance on all items regardless of grammaticality (Power = 86%), though the relationship was stronger for GR items (Power = 78%) compared to UGR items (Power = 55%). Compared to the relationship between WMC and overall performance on the test of implicit knowledge, power was much higher for WMC and performance on the test measuring explicit knowledge at the first (55% - IK vs. 83% - EK) and last testing sessions (67% - IK vs. 86% - EK). No correlations by group were found at Time 3, with the exception of a strong positive correlation between the rote memory component of L2 aptitude and GR items for beginners ($r = .792^*, N = 11, p = .016$, Power = 86%). No significant relationships were found between cognitive capacity and performance on the UGJT for the advanced group at any of the three testing times.

In Table 48 below, the correlation matrix between psychosocial ID components ($N = 8$) and performance on the untimed grammaticality judgment task by item (All/GR/UGR) and testing time is given.

Table 48

| Correlation matrix: Psychosocial IDs and performance on the UGJT at Time 1 ($n = 86$), Time 2 ($n = 83$) and Time 3 ($n = 33$) |
|---|---|---|---|---|---|---|---|---|
| Time | AMTB | Selves |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| * | IG | MOT | IO | LgAnx | ATLS | * | IS | OS | MLB |
Few components of L2 motivation as measured by the AMTB were significantly related to performance on the test of explicit knowledge over time with the exception of the situational motivation component, or attitudes toward the learning situation (ATLS). That is, a more favorable evaluation of Spanish course and teacher reported by learners at the beginning of the learning period were significantly related to higher performance for all items across groups at Time 1 (Power = 75%) and Time 2 (81%) but only with GR items at Time 3 (55%). By group, the relationship with ATLS was significant with performance only for UGR items and only for the intermediate group at the beginning and end of the semester (Power = 68% and 75%, respectively).

Two other components of L2 motivation measured on the AMTB, Integrativeness (IG) and Motivational Intensity (MOT), were significantly related to performance on GR items across groups at the end of the semester though these associations had relatively low power (56% and 59%, respectively). By group, significant relationships were found between MOT measured at
Time 1 and performance on both GR and UGR items for intermediate learners at the beginning and end of instruction, though the relationship with UGR items was again more powerful (Power = 67% vs. 57% (T1); 84% vs. 70% (T2)). At the beginning of instruction, a significant relationship also emerged between LgAnx and UGR items for beginning learners (Power = 61%); however, this was far less powerful than the relationship with UGR items on the test of implicit knowledge (97%) identified at Time 1 for this group. Furthermore, the relationship between LgAnx and UGR items on the UGJT disappeared by the end of instruction and 1 month later whereas it continued to be strongly related to performance on the OIT at the two later testing times for beginning learners.

The Ideal L2 self (IS) component of L2 motivation was significantly related to overall performance at the beginning of the learning period with relatively high power (72%) and this relationship persisted 2.5 months later with power approaching 0.80 (78%). This relationship was more robust for GR items at Time 1 but for UGR items at Time 2. Correlational analyses by group revealed that relationships between IS and GR items at Time 1 and IS and UGR items at Time 2 were significant for the intermediate learners but not for their lower or higher proficiency peers. A correlation between the motivated learning behavior (MLB) component of the Selves questionnaire, or reported learning effort, and overall performance approached significance at Time 1 and reached significance at Time 2 across groups, but with GR items only and with relatively low power (56%). Again, performance by the intermediate group explained this relationship except MLB significantly related to overall performance (Power = 79%) and the relationship was slightly stronger for UGR items (Power = 64%-UGR vs. 53%-GR) for this group.
The only statistical relationships found one month after the learning period had concluded was between ATLS (measured at T2) and GR items (Power = 55%) while a relationship between ATLS and overall performance reached significance for beginners only (Power = 63%). There were also weak associations between ATLS and IO measured at Time 1 and performance on GR items 3.5 months later (Power = 54% and 55%, respectively) for beginners.

**Multiple regressions: IDs and explicit knowledge (UGJT).**

Once significant relationships were identified between ID variables and scores on the UGJT at each testing time, a series of sequential (hierarchical) multiple regressions were run both i) across groups and ii) by proficiency group in order to test the extent to which explanatory variables (IDs) can explain and predict variance in performance. The order in which ID variables were entered were based on the strength of correlations (r) with performance on the UGJT at each testing time. As reported in RQ2, each regression model provides the total R^2, the change in R^2 (Δ R^2) and the unstandardized regression coefficients for explanatory variables in each model.

As done with the test of implicit knowledge, the linearity of each relationship was first examined using a scatterplot. For example, the relationship between working memory capacity (as scored on the OSpan-Composite-Absolute) and performance on the UGJT at the first and last testing times is shown in the following scatterplots. When entered as the only explanatory variable in the regression model at Time 1, WMC significantly explains almost 10% of the variance in judgment accuracy scores on the UGJT as seen in Figure 14 whereas 3.5 months later the relationship is clearly more linear (Figure 15), accounting for 25.5% of variance in the response variable, with some notable outliers.

Figure 14

*Relationship between performance on the UGJT and working memory capacity (T1)*
When the grammatical sensitivity (GS) and phonemic coding (PC) components of L2 aptitude are added into the cognitive ID explanatory model posited for performance on the UGJT at Time 1, WMC emerges as the only significant predictor, accounting for 6% additional variance. Table
Table 49

Cognitive ID explanatory model: Explicit knowledge (T1)

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>GS</th>
<th>PC</th>
<th>WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.04*</td>
<td>.04*</td>
<td>15.53*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.07</td>
<td>.07</td>
<td>16.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.00</td>
<td>.00</td>
<td>1.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.01</td>
<td>.01</td>
<td>9.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.07*</td>
<td>.03</td>
<td>9.34</td>
<td>15.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.26*</td>
<td>.19*</td>
<td>-9.02</td>
<td>35.61*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.00</td>
<td>.00</td>
<td>.79</td>
<td>3.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.02</td>
<td>.01</td>
<td>5.10</td>
<td>9.92</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>.13*</td>
<td>.06*</td>
<td>7.93</td>
<td>9.76</td>
<td>28.73*</td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.33*</td>
<td>.07</td>
<td>-7.48</td>
<td>32.12*</td>
<td>30.15</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.01</td>
<td>.01</td>
<td>.89</td>
<td>1.99</td>
<td>4.80</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.04</td>
<td>.02</td>
<td>4.96</td>
<td>4.28</td>
<td>14.51</td>
</tr>
</tbody>
</table>

**Note.** GS, Grammatical sensitivity; PC, Phonemic Coding; WMC, Working memory capacity; *p = .06; #p = .07.

While the final regression model across groups identifies WMC as the only significant predictor of performance variance with all IDs entered, phonemic coding ability seems to be a more important variable for the beginning group relative to WMC, explaining 19% unique variance in test scores. This cognitive ID regression model overall explains very little of the variance for the upper level groups but explains a third of the performance variance for beginners (33%).

Based on correlations, grammatical sensitivity (GS) was the only variable entered in the regression model at the second testing time and both GS and WMC were entered at the third testing time. GS explained little overall variance by itself (5%) at the end of the semester and
even less variance relative to WMC one month later (2%), though GS accounted for more
variance for beginning learners (9%). Together, GS and WMC significantly predicted 25% of the
variance in UGJT test scores at Time 3 but WMC was the only significant unique predictor,
accounting for 23% of the total $R^2$. Unique variance, or $\Delta R^2$, predicted by WMC was also high,
though not significant, for the beginning (18%) and advanced groups (17%). With a higher
sample size, this predictive power would likely reach significance.

Table 50

*Cognitive ID explanatory model: Explicit knowledge (T2, T3)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>GS B</th>
<th>WMC B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.05*</td>
<td>.05*</td>
<td>18.38*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.02</td>
<td>.02</td>
<td>8.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.03</td>
<td>.03</td>
<td>11.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.01</td>
<td>.01</td>
<td>8.81</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.02</td>
<td>.02</td>
<td>10.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.09</td>
<td>.09</td>
<td>-13.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.04</td>
<td>.04</td>
<td>12.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.06</td>
<td>.06</td>
<td>18.02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.25*</td>
<td>.23**</td>
<td>1.70</td>
<td>47.25**</td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.27</td>
<td>.18</td>
<td>-19.05</td>
<td>39.61</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.05</td>
<td>.01</td>
<td>16.00</td>
<td>-22.60</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.23</td>
<td>.17</td>
<td>12.18</td>
<td>34.61</td>
</tr>
</tbody>
</table>

*Note.* GS, Grammatical sensitivity; WMC, Working memory capacity; $^+p = .06$; $^*p = .07$.

The explanatory model posited for performance on the test of explicit knowledge at Time
1 is provided below with three psychosocial ID components entered: Motivated Learning
Behavior (MLB), Attitudes Toward the Learning Situation (ATLS) and the Ideal L2 Self (IS).
Table 51

*Psychosocial ID explanatory model: Explicit knowledge (T1)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total $R^2$</th>
<th>$\Delta R^2$</th>
<th>MLB B</th>
<th>ATLS B</th>
<th>IS B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>.04</td>
<td>.04</td>
<td>1.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Beg</td>
<td>.13*</td>
<td>.13*</td>
<td>2.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.08</td>
<td>.08</td>
<td>2.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.02</td>
<td>.02</td>
<td>-1.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.09*</td>
<td>.05*</td>
<td>.734</td>
<td>3.58*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.15</td>
<td>.02</td>
<td>1.56</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.22*</td>
<td>.14*</td>
<td>.634</td>
<td>4.09*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.03</td>
<td>.01</td>
<td>-1.88</td>
<td>1.89</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>.10*</td>
<td>.01</td>
<td>-3.33</td>
<td>3.13*</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.22</td>
<td>.07</td>
<td>3.59</td>
<td>1.68</td>
<td>-3.06</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.27*</td>
<td>.05</td>
<td>-3.60</td>
<td>3.03*</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.03</td>
<td>.00</td>
<td>-1.26</td>
<td>1.81</td>
<td>-82</td>
</tr>
</tbody>
</table>

*Note.* MLB, Motivated learning behavior; ATLS, Attitudes toward the learning situation; IS, Ideal self; $^+ p = .06; ^# p = .07.$

As Table 51 shows, reported MLB was the most important explanatory variable for performance by beginning learners, accounting for 13% of the variance, whereas learners' reported attitudes (ATLS) mattered most for the intermediate group's performance, significantly explaining 14% of the unique variance in scores. In contrast to the psychosocial regression models posited for implicit knowledge, the unique variance accounted for by the Ideal self component of L2 motivation was not significant when all variables were entered. Overall, these three variables together explained very little variance in the response variable for advanced learners (3%).

2.5 months after the start of the learning period, ATLS as measured at Time 1 remained a powerful predictor of variance for the intermediate learners only but IS again was not a powerful
explanatory variable when entered together with ATLS.

Table 52

*Psychosocial ID explanatory model: Explicit knowledge (T2)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>Total R²</th>
<th>Δ R²</th>
<th>IS</th>
<th>ATLS (T1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>All</td>
<td>.09**</td>
<td>.09**</td>
<td>3.87**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.05</td>
<td>.05</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.09</td>
<td>.09</td>
<td>4.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.03</td>
<td>.03</td>
<td>-2.95</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>.13**</td>
<td>.04*</td>
<td>2.73*</td>
<td>3.57*</td>
</tr>
<tr>
<td></td>
<td>Beg</td>
<td>.18</td>
<td>.13*</td>
<td>.34</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>Int</td>
<td>.24*</td>
<td>.15*</td>
<td>2.11</td>
<td>6.13*</td>
</tr>
<tr>
<td></td>
<td>Adv</td>
<td>.06</td>
<td>.03</td>
<td>-1.63</td>
<td>-4.15</td>
</tr>
</tbody>
</table>

*Note.* IS, Ideal selves; ATLS, Attitudes toward the learning situation; *p = .06; #p = .07.

At Time 3, the only significant relationship found was between ATLS (T1) and performance on the UGJT (GR items). The total R² was significant across groups, accounting for 13% of the variance, but when the model was run by proficiency group, this prediction was only significant for beginning learners (39%).

Finally, up to two significant cognitive and psychosocial explanatory variables were entered together in a sequential regression model to gauge their relative predictive ability for performance variance across groups. Cognitive variables were entered together in the first step of the regression and psychosocial variables were entered in the second step, based on the strength of associations. Together cognitive and psychosocial explanatory variables respectively accounted for 22% (T1), 16% (T2) and 30% (T3) of the performance variance on the test of explicit knowledge. In contrast to the cognitive and psychosocial explanatory model posited for performance on the test of implicit knowledge, cognitive, rather than psychosocial, IDs were the
only explanatory variables to significantly account for unique variance when all ID components were entered into the regression model, specifically, WMC at the beginning of the semester and 3.5 months later and GS at the end of the semester (approached significance).

Table 53

*Cognitive & psychosocial ID combined explanatory model: Explicit knowledge (T1)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Total R²</th>
<th>Δ R²</th>
<th>1) GS</th>
<th>2) ATLS IS</th>
<th>WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.13**</td>
<td>.13**</td>
<td>11.82</td>
<td>32.40**</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.22**</td>
<td>.09*</td>
<td>9.83</td>
<td>31.75**</td>
<td>2.89#</td>
</tr>
</tbody>
</table>

Note. †p = .06; ‡p = .07.

Table 54

*Cognitive & psychosocial ID combined explanatory model: Explicit knowledge (T2)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Total R²</th>
<th>Δ R²</th>
<th>1) GS</th>
<th>2) ATLS IS</th>
<th>WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.05*</td>
<td>.05*</td>
<td>18.42*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.16**</td>
<td>.11#</td>
<td>15.45#</td>
<td>3.19</td>
<td>2.71#</td>
</tr>
</tbody>
</table>

Note. †p = .06; ‡p = .07.

Table 55

*Cognitive & psychosocial ID combined explanatory model: Explicit knowledge (T3)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Total R²</th>
<th>Δ R²</th>
<th>1) WMC</th>
<th>2) ATLS (T1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.26**</td>
<td>.26**</td>
<td>47.84**</td>
<td></td>
</tr>
</tbody>
</table>
Multiple regression analyses reported above were used to identify cognitive and psychosocial factors that predicted a significant amount of variance in performance at each assessment wave. In addition, person-level (Level 2) Conditional linear models were formulated to account for the significant variability in individual growth parameters (i.e., intercept and slope) in the development of explicit knowledge as found in the Level 1 GCA. Interindividual factors associated with performance were entered as covariates, first separately in simple conditional models and then together in combined models to fit the data and assess unique contributions by predictors to variance in initial intercepts and rate of growth.

First simple conditional models were fit to the data by separately entering each ID measure into the model (based on multiple regressions). Results of these analyses revealed that 3 cognitive ID components (WMC, PSTM, PC) and 3 motivational components (ATLS, IS, MLB) each significantly predicted average performance across time. Then, cognitive measures were entered together and psychosocial measures were entered together to fit combined models for predicting intra and interindividual change in performance on the UGJT\(^\text{iv}\). There were no independent significant cognitive ID predictors when WMC, PSTM and PC were entered jointly and the only motivation component to independently account for variance was ATLS in the psychosocial ID model. WMC, PSTM, PC, and ATLS were entered in the final combined conditional model to examine whether IDs made unique contributions to the slope, or rate of growth, on the UGJT and whether they played a differential role by group. As found for the test of implicit knowledge, no IDs significantly predicted the degree of change in development but
four IDs significantly predicted group performance differences and are reported in the table below.

Table 56  

*Final model fitted for explicit knowledge (UGJT) with cognitive & psychosocial ID predictors*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
</tr>
<tr>
<td>Level-1 Unconditional Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>67.2**</td>
<td>1.5</td>
</tr>
<tr>
<td>Slope</td>
<td>3.22**</td>
<td>.96</td>
</tr>
<tr>
<td>Level-2 Conditional Combined Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>37.1</td>
<td>10.5</td>
</tr>
<tr>
<td>Group 1 vs. 2</td>
<td>-12.30</td>
<td>3.4</td>
</tr>
<tr>
<td>Group 1 vs. 3</td>
<td>-15.96</td>
<td>3.3</td>
</tr>
<tr>
<td>Group 2 vs. 3</td>
<td>-3.66</td>
<td>3.1</td>
</tr>
<tr>
<td>Slope</td>
<td>2.57</td>
<td>1.3</td>
</tr>
<tr>
<td>Group 1 vs. 2</td>
<td>1.17</td>
<td>2.1</td>
</tr>
<tr>
<td>Group 1 vs. 3</td>
<td>1.50</td>
<td>2.0</td>
</tr>
<tr>
<td>Group 2 vs. 3</td>
<td>.33</td>
<td>2.0</td>
</tr>
<tr>
<td>IDs</td>
<td>ATLS</td>
<td>3.83</td>
</tr>
<tr>
<td>WMC</td>
<td>11.58</td>
<td>9.1</td>
</tr>
<tr>
<td>PSTM</td>
<td>4.85</td>
<td>7.6</td>
</tr>
<tr>
<td>PC</td>
<td>3.91</td>
<td>6.6</td>
</tr>
<tr>
<td>ATLS*Group 1 vs. 2</td>
<td>.76</td>
<td>2.9</td>
</tr>
<tr>
<td>ATLS*Group 1 vs. 3</td>
<td>9.62</td>
<td>4.1</td>
</tr>
<tr>
<td>Group</td>
<td>1 vs. 2</td>
<td>1 vs. 3</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>PSTM</td>
<td>10.34</td>
<td>15.8</td>
</tr>
<tr>
<td>IS</td>
<td>2.96</td>
<td>5.36</td>
</tr>
<tr>
<td>MLB</td>
<td>1.37</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. 1, Beginning group; 2, Intermediate group; 3, Advanced group.
A t statistic is used to make statistical comparisons. $p < .05^*$, $p < .01^{**}$.

The results of the Level-2 GCA show that with four IDs entered as covariates, variance in individual growth parameters (intercept and slope) for both within (Time) and between-person (Group) effects remained significant in the final model but change across time was not significantly different by group. This is mirrored by the results of the initial Level-2 GCA run without covariates as well as the nonsignificant Time*Group interaction found in repeated measures ANOVAs. ATLS seems to be the strongest predictor of performance on the UGJT as no cognitive ID remains significant when it is entered into the final model.

By group, ATLS significantly differentiated performance between beginning and intermediate learners on the one hand, and advanced learners on the other. In other words, a positive evaluation of Spanish course and teacher mattered more for lower proficiency learners in explaining performance variance. Also, IS and MLB components of the Selves construct of motivation were significant predictors for lower proficiency groups but not for the advanced learners, though the MLB*Group difference only approached significance for beginning versus advanced learners. PSTM was the only significant cognitive predictor of group performance on the UGJT and was found to play a significantly more important role for intermediate learners.
when compared to advanced learners; this relationship approached significance for beginners (vs. advanced learners).

**Research question 4: Relationships between Cognitive and Psychosocial IDs**

The fourth and final research question aims to explore the nature of relationships between the cognitive and psychosocial dimensions of learning in this population of Spanish L2 learners. A series of Pearson product moment bivariate correlations were run i) between cognitive ID components, ii) between psychosocial ID components, and iii) between both cognitive and psychosocial ID components.

**Correlations: Cognitive ID components.**

Table 57 displays the correlation matrix for the four cognitive ID variables, and the relationships among the six components measured.

**Table 57**

*Correlation matrix: Relationship between cognitive ID components (N = 87)*

<table>
<thead>
<tr>
<th></th>
<th>L2 Aptitude</th>
<th>WMC</th>
<th>PSTM</th>
<th>PROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-PC</td>
<td>--</td>
<td>.457**</td>
<td>.276*</td>
<td>.299**</td>
</tr>
<tr>
<td>IV-GS</td>
<td>--</td>
<td>--</td>
<td>.183</td>
<td>.113</td>
</tr>
<tr>
<td>V-RL</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.282**</td>
</tr>
<tr>
<td>OSpan</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.265*</td>
</tr>
<tr>
<td>Digit Span</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>DSCT</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

...
As Table 57 shows, eight relationships emerged between the different aspects of cognitive capacity measured here. The phonemic coding ability component of L2 aptitude (PC) as measured on the MLAT-III was found to be strongly, positively correlated with grammatical sensitivity (GS, MLAT-IV), phonological short-term memory as measured on the Digit Span, working memory capacity as measured on the OSpan-Comp-L (Operation Span-Composite-Lenient) and, to a lesser degree, with rote learning, or memorization (RL, MLAT-V). WMC was also positively correlated with RL and PSTM. Processing speed, as measured on the DSCT (Digit Symbol Coding Task), was strongly related to PSTM and weakly associated with RL. Few relationships emerged between the grammatical sensitivity component of L2 aptitude and other aspects of cognitive capacity measured here with the exception of phonemic coding ability.

Importantly, the relationships that were found in this learner population were positive, indicating that different aspects of cognitive resources seem to reinforce one another. For example, the higher one's working memory capacity, the higher his or her ability to successfully encode L2 phonemes, and the more efficient one's processing speed, the more efficient his or her phonological memory resources are likely to be.

**Correlations: Psychosocial ID components.**

Within the psychosocial dimension of learning, the construct of L2 motivation was measured on two questionnaires and at the first two testing times. Table 58 shows the correlation matrix for the two motivational constructs, and relationships among the eight components measured.
Table 58

*Correlation matrix: Relationship between psychosocial ID components at Time 1 (N = 86) and Time 2 (N = 82)*

<table>
<thead>
<tr>
<th></th>
<th>AMTB</th>
<th>Selves</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>IG</td>
<td>MOT</td>
</tr>
<tr>
<td>IG</td>
<td>--</td>
<td>.668**</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>.721**</td>
</tr>
<tr>
<td>MOT</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>IO</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>LgAnx</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>ATLS</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>IS</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>OS</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>T2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>MLB</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note.* AMTB, Attitude/Motivation Test Battery; IG, Integrativeness; MOT, Motivation; IO, Instrumental orientation; LgAnx, Language anxiety; ATLS, Attitudes toward learning situation; IS, Ideal self; OS, Ought-to self; MLB, Motivated learning behavior.

Correlation is significant at the 0.01** level (2-tailed).
Correlation is significant at the 0.05* level (2-tailed).

Not surprisingly, several strong significant positive correlations were found between components of L2 motivation as measured within Gardner's socioeducational model of L2 motivation (AMTB) and Dornyéi's L2 motivational Self-system. At the beginning of the semester, the
central construct in Gardner's model, Integrativeness (IG), was significantly related to all other motivational components on the AMTB with the exception of Language Anxiety (LgAnx). IG also significantly correlated with the central construct in Dörnyei's model, the Ideal Self (IS), as well as with the Motivated Learning Behavior (MLB) component of the Selves questionnaire, though not with the Ought-to Self (OS). These relationships persisted 2.5 months later, with the exception of IG and learner reported attitudes toward the learning situation (ATLS). At the end of the learning period, a significant relationship emerged between IG and LgAnx.

The central component in Dornyëi's model, the Ideal L2 Self, was significantly correlated to all AMTB components at both testing times, with the exception of ATLS at the end of the semester. IS also significantly related to MLB and, to a lesser extent, to the Ought-To L2 Self (OS), at the beginning and end of the learning period, but OS and MLB were not significantly correlated at any point in time.

Other significant relationships found at both testing times include those between MOT and IO, LgAnx, ATLS, IS, MLB; IO and IS, MLB; LgAnx and IS; and ATLS and MLB. Both MOT and IO significantly correlated with OS at Time 1, but not Time 2.

**Correlations: Cognitive & psychosocial ID components.**

While relationships reported above are informative, it is not surprising that ID components tapping different aspects of a similar overall construct share a significant amount of variance. An arguably more interesting inquiry, and one less frequently investigated in SLA, is to jointly analyze both the cognitive and psychosocial underpinnings of foreign language learning. This is a necessary step in L2 ID research in order to account for the complex myriad of ways that IDs may interact in facilitating or inhibiting L2 development over time. As L2 motivation
was measured twice, relationships between cognitive and psychosocial ID variables are reported at both the beginning (T1) and end of the learning period (T2) in the table below.

Table 59

*Correlation matrix: Relationship between components of cognitive and psychosocial ID variables at Time 1 (n = 86-87) and Time 2 (n = 82-83)*

<table>
<thead>
<tr>
<th>IDs</th>
<th>III-PC</th>
<th>IV-GS</th>
<th>V-RL</th>
<th>WMC</th>
<th>PSTM</th>
<th>PROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>MOT</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-.257*</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>IO</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>LgAnx</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.287**</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.229* .275*</td>
<td>--</td>
</tr>
<tr>
<td>ATLS</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.224*</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-.227* .217*</td>
<td>--</td>
</tr>
<tr>
<td>IS</td>
<td>T1</td>
<td>.228*</td>
<td>--</td>
<td>--</td>
<td>.216*</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>.232*</td>
<td>--</td>
<td>--</td>
<td>.202+</td>
<td>--</td>
</tr>
<tr>
<td>OS</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>.346**</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>MLB</td>
<td>T1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.239*</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note.* The Absolute-Composite measure of Working Memory Capacity is reported here. Correlation is significant at the 0.01** level (2-tailed). Correlation is significant at the 0.05* level (2-tailed).

+p = .06.

As Table 59 shows, overall, few relationships were found between the cognitive and psychosocial dimensions of L2 learning given the several ID components measured (N = 14);
nonetheless, the relationships that emerged were informative. Whereas Integrativeness (IG) and Instrumental Orientation (IO) as measured on the AMTB did not significantly relate to cognitive ability at any point in time, reported Motivational intensity (MOT) was negatively related to WMC at the beginning of the semester (T1) and learner attitudes toward the learning situation (ATLS) were negatively related to WMC 2.5 months later (T2). In other words, the lower a learner's working memory capacity, the higher her reported learning effort and evaluation of the learning situation, perhaps in order to compensate for less efficient processing and storage capacity. However, these relationships were relatively weak (Power = 66% and 54%, respectively), thus increasing the chance of not finding a statistical correlation if it in fact existed.

The language anxiety (LgAnx) component of L2 motivation was positively related to PSTM at the beginning of the semester (Power = 76%) and to both PSTM and WMC at the end of the semester (T2) with power approaching 80% for PSTM (71%) but not for WMC (54%). This finding indicates that the more efficient memory resources learners possess, the lower their reported Spanish class and use anxiety given that a higher mean Likert rating (Strongly Agree) for the LgAnx component of motivation indicates less anxiety. A higher PSTM also significantly correlated with a more positive Ideal L2 self at the beginning of the learning period and with more favorable attitudes toward the learning situation (ATLS) 2.5 months later, though power in both cases was around 50% meaning that there was only a one in two chance of finding these statistical relationships. Also, higher processing speed was weakly associated with a more positive evaluation of one's Spanish course and teacher (ATLS) at the beginning of the semester (Power = 54%).

The Ideal Selves component within Dornyëi’s model of L2 motivation was also significantly positively correlated with both the phonemic coding ability of L2 aptitude (MLAT-
III) and phonological working memory at the beginning and end of the learning period. Thus, the higher a learner's phonemic coding ability and phonological short-term memory capacity, the more favorably they reported envisioning their Ideal L2 self. However, again power did not reach 80%.

While the Ought-To L2 Self did not correlate with cognitive abilities at Time 1, a strongly significant positive relationship emerged with PC at the end of the semester with high power (.90), indicating that a learner's sense of responsibility to learn the language was related to a higher ability to accurately encode sounds in the foreign language. The third motivational component within Dornyéi's model, motivated learning behavior, significantly correlated with processing speed only at the beginning of the learning period (Power = 60%). This could be interpreted as the more efficient a learner's processing speed, the more effort he or she reported learning Spanish, and vice versa. While it may be intuitive to think that less efficient processing skills would necessitate more learning effort on the part of the learner, as indicated by the negative association between MOT and WMC at T1 and ATLS and WMC at T2, it is also possible that higher processing speed positively influences effective learning behavior both inside and outside the classroom.

As with all correlational analyses, it is difficult to discern the direction of the relationships, and whether there is a reciprocal influence between IDs or a directional influence. Nonetheless, the relationships found here between the cognitive and psychosocial dimensions of learning L2 Spanish in a classroom setting are instructive and provide evidence that ID variables do interact, and likely do not exert influence on L2 development in isolation. Moreover, relationships with psychosocial ID components of L2 motivation were found to be dynamic,
rather than static, as they changed based on the point at which they were assessed during the learning period.

Summary of the Results

In this chapter, descriptive and inferential analyses and the results pertaining to each of the four research questions were presented. Below I provide summaries of the results found for each research question followed by their corresponding figures.

Research question one.

One-way and repeated measures ANOVAs as well as within and between-person GCAs revealed that all proficiency groups significantly improved in their ability to accurately comprehend and repeat sentences in Spanish they heard aurally (operationalized as implicit knowledge) and to judge whether sentences they read were acceptable or not in Spanish (operationalized as explicit knowledge). Positive linear improvement on both measures of L2 knowledge was found at the end of a semester of classroom instruction and one month later, but inspection of individual growth trajectories within groups showed more variable patterns of development, with evidence of U-shaped behavior (i.e., dips in accuracy at Time 2).

By group, advanced learners significantly outperformed intermediate and beginning learners and intermediate learners significantly outperformed beginners on the test of implicit knowledge. Furthermore, the highest proficiency group not only significantly outperformed the lower proficiency groups on this test at each point in time but also accessed their knowledge of L2 grammar faster and more efficiently. However, at the end of the semester of instruction, slower, rather than faster, reaction times were found for all groups which likely reflects that learners were processing more of the target structures present in sentence stimuli 2.5 months
after the outset of the learning period. Thus, while there was evidence for increasing automatization of L2 knowledge with increasing proficiency, this was not true over time.

On the test of explicit knowledge, the higher proficiency groups significantly outperformed beginners but, while the advanced group scored descriptively higher than the intermediate group, they were not significantly different from one another. Learners in all groups reported basing their judgments of ungrammatical sentence exemplars on rule knowledge significantly more than 'Feel' or 'Guess' options, though beginners guessed relatively more frequently and the advanced group reported intuitively making judgments ('Feel') more frequently than their peers. All learners generally reported knowing more rules over time and intermediate and advanced learners provided significantly more rules than beginners at the beginning and end of instruction. Learners were also significantly more accurate in judging the grammaticality of sentences when they reported using knowledge of a grammatical rule in Spanish and were comparably accurate in their ability to verbalize those rules regardless of proficiency level.

Finally, while the internal consistency of the tests used to tap L2 knowledge constructs was high for this learner population and comparable to previous research, the validity of the tests was further investigated and critically considered in terms of the proficiency level of the learner and item grammaticality. Figure 16 summarizes the significant results of the one-way and repeated measures ANOVAs and Figures 17 and 18 present the results of the confirmatory and exploratory factor analyses conducted to answer RQ1.
**Figure 16**

*RQ1 Results: Summary of significant one-way and Repeated Measures ANOVAs for performance on tests of implicit and explicit knowledge (T1, N = 87; T1 ➔ T2, N = 83; T1 ➔ T3, N = 33)*

![Graph showing ANOVA results for implicit and explicit knowledge]

**Figure 17**

*RQ1 Results: Test validity based on Principal Component Confirmatory Factor Analysis across groups*

![Diagram of factor analysis]

*Note: GR, Grammatical items; UGR, Ungrammatical items; Two-factor model reported with form-based scores (OIT).*
**Research questions two and three.**

Correlational and regression analyses as well as Level-2 GCAs run with IDs as covariates revealed several relationships between the cognitive and psychosocial ID components measured and learner performance on both tests of implicit and explicit knowledge of L2 Spanish grammar. The role of what different learners bring cognitively and psychologically to the task of learning a foreign language was found to interact with several factors, specifically, i) the proficiency level of the learner, ii) the point in time when learners were assessed, iii) the type of assessment task used to measure L2 knowledge, and iv) the grammaticality of test items. Results are briefly summarized according to these four factors.
In terms of proficiency level, cognitive resources and L2 motivation were strongly related to performance by low proficiency learners whereas their explanatory capacity was found to be limited for advanced learners. In fact, no relationships were found between IDs and performance on the measure of explicit knowledge for the advanced group and weak relationships were found with grammatical sensitivity, rote learning, and instrumental orientation on the test of implicit knowledge for this group. On the other hand, phonemic coding ability, phonological working memory and language anxiety were consistently related to and predicted performance on the measure of implicit knowledge for the beginning group whereas phonemic coding ability, attitudes toward the learning situation and, to a lesser extent, language anxiety, explained performance variance on the test of explicit knowledge for beginners. For intermediate learners, phonological working memory, attitudes toward the learning situation and motivational intensity were the most consistent predictors of performance on both measures of L2 knowledge. Across proficiency groups, psychosocial IDs, specifically the Ideal L2 self, consistently predicted comparatively more performance variance on the test of implicit knowledge over time compared to cognitive IDs whereas cognitive IDs, specifically working memory capacity, were found to be stronger explanatory variables for performance on the test of explicit knowledge.

Nonetheless, the strength and durability of relationships found between IDs and performance over time generally varied; for example, the role of phonemic coding ability for beginning learners was more robust at the end of the semester whereas the role of phonological working memory was strongest at the beginning of the learning period for the intermediate learners. Furthermore, relationships with working memory capacity were found to be strongest across groups at the third testing time, one month after the end of instruction, but this relationship was more robust for performance on the test tapping explicit knowledge. Finally, no
ID was significantly related to or predicted the rate of growth over time but rather performance at each point in time.

In relation to item grammaticality, cognitive ID components were overall more strongly related to performance on ungrammatical exemplars of the ten target structures present in the stimuli for both measures of knowledge whereas L2 motivation generally had a more global impact on performance regardless of grammaticality, although certain psychosocial components had stronger relationship with UGR items (i.e., motivational intensity, Ideal L2 self). These results suggest that grammaticality interacts with learner differences in determining how learners accessed their knowledge on both tests.

A summary of the significant Pearson Product moment correlations found between ID variables and performance on the tests of implicit (RQ2) and explicit knowledge (RQ3) over time is presented in Figures 19 and 20, respectively. Figure 21 highlights those IDs, or explanatory variables that significantly predicted performance variance on both measures at each assessment wave (RQ2 & RQ3).
**Figure 19**

**RQ2 Results:** Summary of Pearson Product Moment correlations between cognitive and psychosocial ID variables and performance on test of implicit knowledge

![Diagram of implicit knowledge correlations](image)

*Note: All relationships found were positive and relationships with $r^{**}$ have power $\geq .80$; UGR, Ungrammatical items only.*

**Figure 20**

**RQ3 Results:** Summary of Pearson Product Moment correlations between cognitive and psychosocial ID variables and performance on test of explicit knowledge

![Diagram of explicit knowledge correlations](image)

*Note: All relationships found were positive and relationships with $r^{**}$ have power $\geq .80$; GR, Grammatical items only; UGR, Ungrammatical items only.*
Research question four.

Correlational analyses used to answer the final research question indicated that several relationships exist between the cognitive and psychosocial dimensions of learning a foreign language. Generally, relationships between the six cognitive and eight motivational ID components measured suggest that IDs both reinforce and compensate one another in this sample of adult instructed L2 learners. Unsurprisingly, significant positive relationships found within each dimension of L2 learning suggest that various aspects underlying a central ID construct reinforce one another. For example, the higher one's ability to process, store and retrieve information (working memory capacity), the higher her capacity to make sound-symbol associations (phonemic coding ability) and ability to memorize connections between stimuli and responses (rote learning). Likewise, the stronger one's interpersonal reasons for studying Spanish
and attitudes toward the Spanish native-speaking community (Integrativeness), the more favorably she is likely to report viewing her ideal future self-image as a foreign language learner (Ideal L2 Self).

More importantly, cognitive and psychosocial IDs were significantly associated with one another suggesting that internal differences do not act in isolation and likely exert a joint influence on L2 development. For example, negative relationships found between working memory capacity and reported motivational intensity suggest that a learner with low working memory resources is likely to report more effort to learn Spanish. Similarly, low phonological short-term memory was significantly related to higher language anxiety. Together, these relationships suggest that learners with less efficient processing and storage capacity may attempt to compensate for low cognitive resources with an increased effort to learn the language and they may experience more language use and classroom anxiety.

The final figure below summarizes the significant relationships among cognitive and psychosocial individual differences that exist in this instructed L2 learner population (RQ4).
RQ4 Results: Summary of Pearson Product Moment correlations between cognitive and psychosocial ID variables

Figure 24

<table>
<thead>
<tr>
<th>Cognitive ID variables</th>
<th>Psychosocial ID variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemic Coding Ability</td>
<td>Motivational Intensity</td>
</tr>
<tr>
<td>1) $r = .23^*$, Power = .56</td>
<td>Language Anxiety</td>
</tr>
<tr>
<td>2) $r = .23^*$, Power = .56</td>
<td>Attitudes toward the Learning Situation</td>
</tr>
<tr>
<td>3) $r = .35^{**}$, Power = .90</td>
<td>Ideal L2 Self</td>
</tr>
<tr>
<td>Working Memory Capacity</td>
<td>Motivated Learning Behavior</td>
</tr>
<tr>
<td>4) $r = -.26^*$, Power = .66</td>
<td>Language Anxiety</td>
</tr>
<tr>
<td>5) $r = .23^*$, Power = .54</td>
<td>Attitudes toward the Learning Situation</td>
</tr>
<tr>
<td>6) $r = -.23^*$, Power = .54</td>
<td>Ideal L2 Self</td>
</tr>
<tr>
<td>Phonological Short-Term Memory</td>
<td>Language Anxiety</td>
</tr>
<tr>
<td>7) $r = .25^{**}$, Power = .76</td>
<td>Attitudes toward the Learning Situation</td>
</tr>
<tr>
<td>8) $r = .28^*$, Power = .71</td>
<td>Ideal L2 Self</td>
</tr>
<tr>
<td>9) $r = .22^*$, Power = .54</td>
<td>Ought-to L2 Self</td>
</tr>
<tr>
<td>10) $r = .22^*$, Power = .54</td>
<td></td>
</tr>
<tr>
<td>Processing Speed</td>
<td></td>
</tr>
<tr>
<td>11) $r = .22^*$, Power = .54</td>
<td></td>
</tr>
<tr>
<td>12) $r = .24^*$, Power = .60</td>
<td></td>
</tr>
</tbody>
</table>

Time 1
(n = 86-87)

2.5 months

Time 2
(n = 82-83)
CHAPTER FIVE: Discussion and Conclusions

Overview

This study aimed to examine the extent to which the developmental trajectories of linguistic knowledge during and after a semester of instruction are similar in second language classroom learners of Spanish at increasing levels of proficiency. It was also of interest to examine what cognitive and psychosocial factors drive development of and mediate access to L2 knowledge of Spanish morphology and syntax in order to account for individual variation in the shape and speed of growth over time.

In the previous chapter, the results of the analyses used to answer each of the four research questions were presented and summarized. In this chapter, I discuss empirical findings pertaining to each research question alongside current second language acquisition research relevant to the independent variables (learner IDs) and dependent variables (implicit and explicit L2 knowledge) investigated here. I then draw summary conclusions based on my findings and outline the theoretical and pedagogical implications prompted by this research. Finally, I consider limitations of the study and identify productive avenues for future research in this area.

Research question 1: Development and retention of implicit/explicit L2 knowledge over time by proficiency level.

The first research question asked: *What is the trajectory of development of implicit and explicit L2 knowledge for classroom learners of Spanish during and after a semester of instruction? How do learner knowledge trajectories differ by proficiency level?* Most notably, a semester of instruction significantly enhanced learners' ability to accurately comprehend and produce sentences they heard in Spanish, as measured by the test of implicit knowledge, as well as to receptively judge if sentences they read in Spanish were grammatical or not, as measured by the test of explicit knowledge. This finding supports the general consensus that instruction
positively impacts L2 development (e.g., Norris & Ortega, 2000) and provides further evidence for the facilitative role of classroom instruction in promoting the development of implicit and explicit knowledge of L2 grammar (Akakura, 2010; R. Ellis, 2002; R. Ellis, Loewen & Erlam, 2006; Kim, 2012). On the other hand, this finding constrasts views espoused by Krashen (1982) and others (e.g., Zobl, 1984) who have claimed a limited role for explicit instruction in SLA.

Importantly, this study made a valuable contribution to the study of change over time in SLA by usign both traditional inferential statistical analyses and innovative linear modelling techniques to map change and growth in L2 knowledge over time. While use of repeated measures analyses of variance has been identified as a methodological shortcoming in previous quantitative longitudinal research (Ortega & Iberri-Shea, 2005, pp. 40-41), both RM-ANOVA and linear growth curve analyses yielded similar results in the current study. Namely, they provided evidence of significant improvement over time and significant performance differences between proficiency groups but no significant differences between groups in their rate of growth over time.

However, while the trajectories of development by group showed positive linear upward slopes for performance on both measures of L2 knowledge, inspection of individual growth trajectories revealed more variable development with "no change" and "linear" increases and/or declines (Singer & Willett, 2003, p. 217) over the total 3.5-month period of investigation. Specifically, individual learner performance reflected U-shaped behavior (McLaughlin, 1987) and 'delayed rule restructuring' (Gass, 1984, p. 121), with dips in accuracy at the end of the learning period and gains in accuracy one month later. This is in line with the permeable, dynamic nature of change in interlanguage systems (Ortega, 2009; Selinker, 1972, 1992; Tarone, 1983), which is not always linear and does not proceed from one stage of development to the
next. Comparatively, however, individual knowledge trajectories were less variable for advanced learners, which supports the notion that interlanguage variability and use of rules in performance becomes more stable at advanced stages of development (Selinker, 1972, 1992).

Relatedly, continued gains in accuracy were observed for all groups after one month without instruction, particularly on the measure of implicit knowledge, whereas advanced learners seemed to hit a plateau in their explicit knowledge. Continued improvement at the third assessment wave, especially for low proficiency learners, contrasts the loss of L2 grammatical skills that typically occurs after a period of disuse or reduced input (Bardovi-Harlig & Stringer, 2010; Gardner, 1982; Ortega & Iberri-Shea, 2005). Similar to individual learner performance, maintenance of L2 knowledge with increased accuracy may also be explained in terms of restructuring processes, or an internal (re)organization of rules governing the learner's interlanguage system (Gass, 1984). Given that learners had been exposed to several new grammatical forms related to tense, aspect and mood morphology during instruction, beginning and intermediate groups in particular, they may have needed time to consolidate and accommodate new grammatical forms into their interlanguage knowledge of Spanish. This interpretation reflects what has been found in research on L2 instruction reporting delayed improvement for implicit instruction two weeks later (e.g., Erlam, 2005; Robinson, 2002; Sanz, submitted; see Norris & Ortega, 2000) and increased native-like neural processing five months later (Morgan-Short, et al., 2012). Nonetheless, findings related to retention are tentative as performance at the final assessment wave only reflects a third of the total learner sample \( n = 33 \). It is possible that those learners who returned at the third testing time had relatively higher levels of motivation than their peers and therefore, may have been in the top range of performance for their respective groups.
Further analysis of performance according to type of L2 knowledge revealed that the test of implicit knowledge more effectively discriminated between the three proficiency levels at each point in time whereas statistical differences were more evident for learners on opposite ends of the proficiency spectrum on the test of explicit knowledge. This finding provides further validation of the elicited oral imitation task not only as a measure of knowledge of L2 Spanish (Bowles, 2011; Gutiérrez, 2012) and English (e.g., Erlam, 2006; Tomita, Suzuki, & Jessop, 2009) but also as a measure of global L2 proficiency in Spanish (Bowden, 2012; Ortega, 2000; Ortega, Iwashita, Norris & Rabie, 2002). Thus, this study supports the use of the OIT as a systematic, efficient option for researchers to independently assess proficiency in their participant samples, a practice that is currently lacking (Tremblay, 2011).

Nonetheless, the superior ability to differentiate between proficiency levels could also reflect the nature of task demands in each measure of knowledge. The oral imitation task requires online aural comprehension and elicits an actual utterance from the learner. The logic behind this task is that a person's ability to repeat the sentences is a reflection of his or her internal grammatical system (Mackey & Gass, 2005, p. 56) as learners can likely only repeat what they have comprehended and parsed (Bley-Vroman & Chaudron, 1994). It is reconstructive in nature, its focus is on meaning and it is performed under time pressure, which reduces the likelihood that learners will explicitly focus on linguistic form and access explicit language knowledge in performance (e.g., Erlam, 2006), though these task characteristics do not ensure they will not focus on form while completing the task (Spada, 2013).

On the other hand, a grammaticality judgment task requires successful written comprehension in the version used here, and elicits a learner's beliefs about the language being learned (receptive) (Mackey & Gass, 2005). Thus, explicit knowledge is receptively measured in
the UGJT based on learners' judgments and is not performed under time pressure given that learners are more likely to gain controlled access to explicit knowledge without time constraints (Sorace, 1996). While these criterial features are meant to distinguish between use of implicit and explicit knowledge, the varying difficulty of each task according to the language skills involved (listening, speaking vs. reading, receptive judgment) likely played a role in effectively distinguishing between learners at each level of proficiency.

The less demanding nature of the explicit knowledge measure is also reflected in the higher average accuracy on this task for all groups at each point in the learning period. However, superior performance could also reflect the nature of learners' previous language learning experience (Bowles, 2011; Philps, 2009) given that these are instructed learners who started learning Spanish in a classroom setting after puberty. Bowles (2011) and Gutiérrez (2012) also found that their instructed Spanish L2 learners scored higher on measures of explicit knowledge relative to implicit knowledge, regardless of proficiency. Overall, superior performance on the tests of explicit knowledge support the criterion of 'learnability' in distinguishing between knowledge constructs (i.e., learners who acquire the L2 as a child are more likely to display high levels of implicit knowledge, while those who are first exposed as adolescents or adults, especially if reliant on instruction, are more likely to display high levels of explicit knowledge) (R. Ellis, 2009, pp. 38-9).

On the other hand, the criterion of 'systematicity' (i.e., whether learners are consistent or more variable in their response to a task) based on standard deviation from the mean did not distinguish between implicit and explicit knowledge in these learner data. SD was found to be comparable between tasks indicating that use of knowledge was similarly variable despite the hypothesis that use of explicit knowledge should be relatively more variable. Nonetheless, this is
is similar to what previous studies have reported for performance by instructed learners at an intermediate level (Bowles, 2011), advanced level (Erçetin & Alptekin, 2013), high and low proficiency (Gutiérrez, 2012), and at a mixed range of proficiency levels (R. Ellis, 2005, 2009). Thus, the criterion of systematicity, as reflected across several studies, highlights the methodological challenge in establishing clearly separable knowledge constructs and underscores the likelihood that learners draw on multiple sources of interlanguage knowledge in performance (DeKeyser, 2009).

**Test reliability and validity.**

In light of this challenge, this study attempted to methodologically improve several aspects of each measure of L2 knowledge in order to strengthen the interpretation of behavioral evidence of knowledge. Reliability estimates indicated that tests were internally consistent for this learner population and comparable to coefficients reported in previous empirical research with similar learner populations, but the psychometric properties of the tests were found to vary according to proficiency level.

First, in terms of the oral imitation task, this is the first study in the line of research prompted by Han and Ellis's study (1998) to verify that learners were not only focused on meaning but also that they had indeed comprehended the input based on an objective meaning processing score that was calculated for each participant. Strong significant positive correlations found between meaning processing accuracy and sentence repetition accuracy at all testing times indicated that learners with higher meaning comprehension were also more accurate. This supports the hypothesis that learners can only accurately repeat what they have parsed and comprehended in this task (Bley-Vroman & Chaudron, 1994).
Second, this study improved the operationalization of time inherent in the definition of implicit knowledge by collecting oral latency data to assess the degree to which learners were accessing their L2 knowledge more or less automatically, as indicated by mean response times (RTs), and efficiently, as indicated by coefficients of variation (CV) (Segalowitz & Segalowitz, 1993). Retrieval of L2 grammatical knowledge was found to be not only faster but more efficiently accessed at increasing levels of proficiency. Moreover, the standard deviation from the mean indicates a gradual decrease over the total 3.5 month of investigation (from Time 1 to Time 3) for all groups. Nonetheless, beginners displayed the longest RTs and the highest SDs from the mean but intermediate and advanced groups were not significantly different from one another in terms of oral reaction time or response variability, indicating that these indices were useful for distinguishing between use of L2 knowledge by the higher proficiency groups on the one hand and beginners on the other.

More rapid and efficient mental processing in learners who have had more exposure to and practice using the L2 is in line with Segalowitz's (2010) view of cognitive fluency and skill acquisition theory (DeKeyser, 2007) as well as the view of automaticity as a continuum (DeKeyser, 1997, p. 196). The mechanisms underlying a particular skill not only become faster with practice, but a reorganization of underlying components occurs, rendering their execution more efficient, and more accurate, overall (DeKeyser, 2001; Hulstijn, Van Gelderen & Schoonen, 2009). In this case, task components involved semantic comprehension of sentences heard aurally, and subsequent oral repetition, or production, of several grammatical structures within them. Thus, advanced learners were more efficient in their execution of these subskills, as reflected in their lower coefficients of variation.
This finding also supports the few empirical studies that have investigated a role for automaticity in processing natural language grammar (Hopp, 2010; Rodgers, 2011), though only the latter study used the CV as an index of automatization. Rodgers found increasing automatization of L2 grammar on a comprehension task (picture identification) in eighty-two L2 Italian learners at three proficiency levels with faster, but not more efficient, performance on a production task (picture description) for advanced learners. On the other hand, the current study did find evidence of significantly more efficient, or automatized use of verbal morphology in oral production for advanced learners, in comparison to beginners. This difference in findings may reflect a difference in task demands given that the picture description task used by Rodgers may have been less controlled and more spontaneous than the oral imitation task used here.

Overall, the collection of latency data on the OIT provided evidence of more automatic use of knowledge with increasing proficiency (i.e., more practice) (DeKeyser, 1997, 2007). Nonetheless, at the end of the learning period, all groups responded slightly more slowly on average, contradicting the central claim proposed by Segalowitz and colleagues that learners should speed up in their use of knowledge over time. Slower processing in this context could be interpreted as a symptom of the dynamic, changing state of the interlanguage knowledge system (Selinker, 1992) given that learners had been exposed to new and more complex grammatical forms in their classrooms over the course of the semester; thus, it is possible that learners, were processing more of the target structures present in the sentence stimuli at the end of the learning period, causing slower response times.

Overall, these results question the construct validity of the OIT as reported in previous research (e.g., Bowles, 2011; R. Ellis, 2005, 2009; Erçetin & Alptekin, 2013; Gutiérrez, 2012), leading one to further consider whether the OIT is measuring implicit knowledge or rapidly
accessed explicit knowledge in instructed learners, as pointed out by Spada (2013, p. 9). Indeed, DeKeyser (2003) claims that automatized knowledge can be 'functionally equivalent' to implicit knowledge (DeKeyser, 2003) yet neurally and anatomically distinct (N.C. Ellis, 2005; Paradis, 1994, 2009).

In terms of the measure of explicit knowledge (UGJT), judgment source attributions and written rule verbalizations were elicited to provide useful information regarding the extent (and limits) of learners' explicit knowledge. Many researchers have suggested using source attributions or certainty indices (e.g. R. Ellis, 2004; Gass, 1994; Sorace, 1985, 1996) to strengthen the interpretation of non-native speaker judgments of grammaticality though previous studies in this line of research have not reported source attribution data or fully probed the verbalizability criterion of explicit knowledge (see Isemonger, 2007), including those studies investigating knowledge of L2 Spanish grammar (Bowles, 2011; Gutiérrez, 2012).

This methodological improvement proved fruitful as learners in all groups reported basing their judgments on rule knowledge significantly more often than intuition ('Feel') or guessing, and were significantly more accurate when they used a rule to make their decision. Accuracy in verbalizing, or describing, grammatical rules was also very high across groups at all testing times (> 80%) indicating that learners were comparably accurate in verbalizing rule knowledge regardless of proficiency level though higher proficiency learners provided significantly more rules at each testing time. Learners, even at the beginning and intermediate levels of proficiency, were also significantly more accurate in their judgments reportedly based on intuition than when they reported to be guessing. However, as previously mentioned, the extent to which beginning and intermediate instructed learners have developed a reliable 'intuition' of L2 grammar is questionable and highlights a key construct validity issues in
interpreting self-reported source attribution data for non-native speaker judgments (e.g., Leow & Hama, 2013; Leow, 1996; Sorace, 1996); in other words, "what kind of knowledge do learners draw on when they judge the grammaticality of a sentence: explicit knowledge, implicit knowledge, or some kind of mixture of both?" (R. Ellis, 2004, p. 254). The goal of using source attributions in previous research has generally been intended as a way to demonstrate (un)awareness of some target aspect of the language or investigate if incidental exposure can result in implicit, or unconscious, knowledge of L2 grammar (e.g., Rebuschat & Williams, 2012). This has been critiqued by Leow and Hama (2013) given that it is doubtful that learner-reported intuition of L2 grammar is equivalent to a native speaker's (i.e., implicit knowledge) (p. 12).

In the context of the current study, source attribution data were elicited to demonstrate the extent to which learners used explicit linguistic knowledge to judge grammatical and ungrammatical L2 sentence exemplars. Given that these participants are late adult L2 learners who have had limited exposure to naturalistic L2 input, it may be that 'Feel' actually indicates use of explicit, though unverbalizable, knowledge. Thus, when employing GJT measures with classroom learners who have relatively limited exposure to L2 data, which are often presented in an explicit format in language textbooks, it may be more productive to interpret source attributions in terms of underlying degree of explicit knowledge, with Feel designating use of explicit, though unverbalizable, knowledge and Rule indicating use of explicit knowledge that the learner is able to describe, with or without using metalingual terminology (R. Ellis, 2004).

Finally, in terms of test validity, no empirical study to date has analyzed the psychometric properties of the implicit/explicit test battery according to the proficiency level of the learner, despite the view that this variable likely affects how learners access and use their L2 knowledge (DeKeyser, 2001, 2007) as well as how it is represented and processed in the brain (Paradis,
Bowles (2011) and Erçetin and Alptekin (2013) investigated one proficiency level, specifically, performance by twenty L2 and heritage intermediate Spanish learners and fifty-one advanced L2 English learners, respectively. Gutiérrez (2012) compared performance by fifty-three L2 Spanish learners at high and low proficiency, but did not report conducting factor analyses with test scores by proficiency level. In R. Ellis's studies (2005, 2009) conducted with ninety-one L2 English learners at mixed skill levels, proficiency level was an uncontrolled variable. While the majority of learners reported learning English in a foreign language context for ten years on average, L2 learners seemed to vary widely in terms of course enrollment, with some reportedly taking low-level university courses and others in advanced courses in an English for Speakers of Other Languages (ESOL) undergraduate program.

To further verify that tests were measuring the intended constructs in this population of instructed L2 Spanish learners at three levels of proficiency, three exploratory factor analyses were conducted by group without specifying the number of factors (Byrne, 2001; Thompson, 2004) (see R. Ellis & Loewen, 2007 & Isemonger, 2007), in addition to the confirmatory factor analysis conducted across groups to test the two-factor model previously reported. Given the possibility that GR and UGR items could encourage the use of different knowledge types, as confirmed empirically by Gutiérrez, and the minimum statistical requirement to have at least two measures per factor for a two-factor model (Klein, 2005), overall scores on both tests as well as disaggregated scores (U/GR) were entered into factor analyses.

An implicit-explicit factor solution was replicated across groups with these learners' data as all test scores (overall/GR/UGR) on the test of implicit knowledge loaded heavily on one factor and test scores on the test of explicit knowledge loaded heavily on a second factor. The cumulative variance accounted for by the two-factor model (82%) was comparable to that
reported in Bowles (87%) and R. Ellis (75%) though these studies investigated performance on all five tests (i.e., elicited oral imitation, oral narration, timed/untimed grammaticality judgment tasks, metalinguistic knowledge test). The total cumulative variance is also similar to that reported in studies employing three of the five original measures (75%–Gutiérrez; 67%–Erçetin & Alptekin).

Importantly, however, several model fit indices indicated that the confirmatory factor analysis did not produce a two-factor model that was statistically likely to occur, in contrast to previous research. This may be due to the fact that only two measures of implicit and explicit knowledge were used rather than the entire test battery. As detailed above, previous research reporting a two-factor model has used three to five tests. However, two measures per factor are the minimum required in a two-factor model and three are recommended (Klein, 2005). Given that this study entered overall scores on each performance measure along with GR and UGR scores, both the minimum and preferred statistical recommendations were met.

Another plausible explanation is that combining the data in one analysis for all proficiency groups did not produce a good fit for the performance data. The three separate exploratory factor analyses conducted by proficiency group revealed that while two components were extracted for intermediate and advanced groups based on eigenvalues, three factors were extracted for beginners. The loadings on each factor for the advanced group mirrored those obtained in the confirmatory factor analysis across groups. However, considerable overlap in factor loadings was found for performance on UGR items on the implicit knowledge measure for the beginning group and GR items on the explicit knowledge measure for the intermediate group. GR items on this task loaded heavily on a separate third factor for the beginning group. These results support the hypothesis that UGR items likely encourage use of explicit knowledge
whereas GR items may elicit implicit knowledge (R. Ellis, 2004, 2005; Gutiérrez, 2012), and also provide evidence for the modularity of learners’ knowledge in relation to the grammaticality of sentences (Hedgcock, 1993, c.f., R. Ellis, 2005).

Overall, the almost evenly split factor loadings for UGR items on the OIT and GR items on the UGJT for the lower proficiency groups indicate that tests could be measuring something different in learners at different proficiency levels. In particular, the split factor loading for UGR items for the beginning groups may reflect the difficulty in performing a language task requiring both aural comprehension and oral repetition, resulting in "the learner attempting to exploit explicit knowledge" (R. Ellis, 2009, p. 13). As noted by N.C. Ellis (2005), "a lack of automaticity in L2 can affect the amount of attention available for L2 error detection; a relative lack of L2 exposure entails that there are fewer relevant correct exemplars available in memory and that these come less readily and less accurately to mind" (p. 331). Thus, it seems plausible that processing UGR exemplars on the oral imitation task was more effortful for beginning learners who have the least amount of exposure to and practice in the L2, and that they were consciously monitoring their speech in a controlled fashion (DeKeyser, 2009; Kormos, 1999). On the other end of the proficiency spectrum, the replicated two-factor model as well as the significantly faster and more efficient response times support the view that performance by advanced learners, who have had relatively more exposure and practice in the L2, has been sufficiently automatized such that their behavior on the oral imitation task may be functionally equivalent to implicit knowledge (DeKeyser, 2003). Thus, higher proficiency learners were able to make form-meaning connections more automatically and efficiently produce accurate utterances on this task, resulting in more clearly separated implicit and explicit factors.
Overall, the split factor loadings question the separability of knowledge constructs, at least based on behavioral evidence, and highlight their likely interaction in performance (Bialystok, 1982, 1994). R. Ellis (2004) acknowledges a reasonable degree of interaction at the level of performance and DeKeyser (2003) has claimed that we can only aim to elicit knowledge "under conditions that are more or less conducive to the retrieval of implicit and explicit knowledge" (p. 320). Taken together, these findings refute a dichotomous view of implicit/explicit interlanguage knowledge, and reinforce the notion that learners likely complete tasks using some combination of implicit and explicit L2 knowledge, depending on task demands, item grammaticality, and proficiency level, among other factors (e.g., target structure).

In sum, this study is a continuation of previous SLA research seeking to verify "how we know what we know" (Selinker & Lakshmanan, 2001, p. 324) by providing concrete indices of the reliability and validity of testing instruments. In line with a basic, yet fundamental goal driving SLA theory and research (R. Ellis, 2005, 2009; Norris & Ortega, 2003, 2012), this study provided behavioral evidence of the long-term development of adult learners' linguistic knowledge, which contributes to our understanding of the successive states of form-function relationships in interlanguage (Selinker, 1992).

The current study has also moved this line of research forward by not only methodologically improving testing methods but also by further analyzing the psychometric properties of the tests. This effort has led to questioning the theoretical implicit/explicit knowledge distinction for late instructed L2 learners and provided support for the 'learnability' constraints proposed for the development of implicit and explicit language knowledge (Bialystok, 1994; DeKeyser, 1994, 2003). As an alternative, this study underscores the relevance of assessing the degree of underlying explicit knowledge in adult classroom learners, as reflected
via judgment source attributions and rule verbalizations, and the degree to which they have
*automatized* their explicit knowledge, as reflected in oral response latency and variability.

Pedagogically, this provides "a more realistic and attainable goal for L2 instructors [...] to create
conditions that will help learners to proceduralize their explicit knowledge" (Spada, 2013, p. 10).

**Research questions 2 and 3: Relationships between learner IDs and development
and retention of implicit (RQ2) and explicit (RQ3) L2 knowledge.**

The second and third research questions (RQs) made the following inquiry: *What ID
factors are related to and predict the development and retention of implicit (RQ2) and explicit
(RQ3) grammatical L2 knowledge of Spanish over time in instructed learners at varying levels of
proficiency?* Findings in relation to both implicit and explicit L2 knowledge will be discussed
here to facilitate comparison.

Overall, inferential statistical analyses revealed some similarities and several differences
in the cognitive and psychosocial factors that facilitate and constrain growth of linguistic
knowledge for adult learners in their 2nd, 4th, and 6th semesters of Spanish instruction.
Importantly, the strength and relative explanatory power of relationships found with IDs were
found to vary according to i) the proficiency level of the learner, ii) the point in time when
performance was assessed during or after instruction, iii) the task demands of the measures of L2
knowledge, and iv) item grammaticality. In light of these factors, the main findings pertaining to
relationships found with cognitive abilities are considered first followed by those found with
motivational components; finally, their relative explanatory power is discussed.

**Cognitive ID components.**

Different components of ID constructs in both domains were consistently found to be
differentially involved in performance according to proficiency level, supporting previous calls
for a componential and developmental perspective in learner ID research (Robinson, 2001, 2005,
In particular, it was evident that cognitive resources played an important role in explaining L2 performance by low proficiency learners of Spanish whereas their explanatory capacity was limited for learners with advanced proficiency. The differential involvement of cognitive IDs according to proficiency level in these data supports Robinson's (2013) claim that learning a language involves different abilities at different stages of development (p. 2) and also underscores the need to develop measures of aptitude that tap abilities drawn on in more advanced stages of L2 learning.

_L2 aptitude._

In terms of L2 aptitude, beginning learners with a superior ability to discriminate between and code sounds in a foreign language so that they can be retained and later recalled (phonemic coding ability) performed significantly better on the elicited oral imitation task. This robust relationship is in line with Skehan's (1989, 1998) claim that phonemic coding is important for processing auditory input in real time and "is likely to be particularly important at the beginning levels of language learning" (1998, p. 203). Successful performance on this task requires an ability to quickly process acoustic stimuli aurally; thus, beginning learners with an efficient capacity to make sound-symbol associations were more likely to accurately access their knowledge of L2 Spanish grammar on this task. However, this ability did not distinguish between performance for intermediate and advanced groups, confirming the idea that the association between L2 ability and phonetic coding ability decrease as L2 proficiency increases (Skehan, 2002).

On the other hand, the ability to recognize the grammatical functions of words in sentences (grammatical sensitivity) and to memorize associations between sounds and meanings (rote learning) was not significantly related to performance on the implicit knowledge task for
lower level learners, though they were weakly correlated with performance by the advanced group on this task at Time 1 and a moderate role was found for these components on the test tapping explicit knowledge at Time 1 (GS) and Time 3 (RL) for beginners. As Robinson (1997) notes, the words-in-sentences test of grammatical sensitivity and the paired-associates test of memory tap abilities that "control access to the learned, but not the acquired, system" (p. 54). In the untimed grammaticality judgment task, learners read, rather than heard, sentences and were not under time pressure to make a judgment about the sentences' grammaticality. As this task was offline and provided time for conscious reflection, language analytic ability would be expected to be relatively more involved compared to the OIT performed under time pressure and with a focus on meaning.

These findings support a changing role for aptitudinal constructs as L2 proficiency increases as found by Winke (2005). In her study with L2 Chinese learners, she found that phonetic coding ability, grammatical sensitivity and memory measures of aptitude were significantly related to performance on a Chinese proficiency test for beginning, but not advanced learners. Additionally, relationships between explicit knowledge and grammatical sensitivity for beginning learners partially support Robinson (1997) and Roehr and Gánem-Gutiérrez (2009). Robinson found a significant correlation between learning easy and hard rules in English under rule-search and instructed (explicit) conditions and grammatical sensitivity for intermediate L2 English learners; Roehr and Gánem-Gutiérrez also found moderate significant correlations between grammatical sensitivity and performance on metalinguistic knowledge tests by L1 English learners of L2 Spanish and L2 German who had studied the language for 4 years on average, likely placing them at a high intermediate level.
However, the relatively nonexistent role for grammatical sensitivity on the implicit knowledge test (with the exception of the advanced group) contrast findings reported by Robinson (1997), deGraff (1997) and Erlam (2005) who all reported positive relationships with grammatical sensitivity and performance by learners who had received implicit instruction [-grammatical explanation]. The contrast in findings could be related to the difference between learning under less explicit instructional conditions and performance on a measure tapping the extent to which learners can implicitly access and use their interlanguage knowledge of L2 Spanish grammar. This also lends support to the idea that implicit learning mechanisms and mechanisms used to access knowledge implicitly are not necessarily the same (N.C. Ellis, 2005; Hulstijn, 2005). Namely, language analytic ability may aid learners more in 'less structured' (Skehan, 1989) instructional conditions in which they must make form-meaning connections relatively automatically rather than playing a prevalent role in accessing L2 knowledge automatically. The difference could also be related to learning one to two rules in a natural language (Robinson) or an artificial language (deGraff) versus performance on a test meant to tap implicit knowledge of a broad range of grammatical structures in a natural language.

Overall, these results confirm previous assertions that the Modern Language Aptitude Test's explanatory power decreases as target language proficiency increases. While composite and componential scores on the MLAT are typically reported to correlate highly with instructed language learning success (Dörnyei & Skehan, 2003; Sawyer & Ranta, 2001; Skehan, 1989, 1998, 2002), these relationships have been found especially for learners in the early stages of development. Robinson (2005, 2013) argues that no current test of aptitude operationalizes a developmental perspective of L2 aptitude, which is needed in order to accurately capture the abilities drawn on in the initial as well as in the later stages of learning a foreign language.
Working memory capacity and phonological short-term memory.

First, results pertaining to phonological short-term memory (as measured on a nonverbal digit span task) are discussed followed by the executive component of working memory capacity (as measured on a nonverbal operation span task).

PSTM emerged as a consistently strong and durable predictor of performance on the test of both implicit and explicit knowledge for intermediate learners, and was also significantly related to performance on the test of implicit knowledge for beginners. These findings support general assertions that variation in phonological loop function is an important determinant of L2 learning (Baddeley, Gathercole & Papagno, 1988) underlying variation in grammatical rule learning (N.C. Ellis, 2012; N.C. Ellis & Martin, 2012; N.C. Ellis & Sinclair, 1996; Williams & Lovatt, 2003). More specifically, the ability to store and rehearse verbal information (Baddeley, 2003) was particularly relevant for explaining performance variation by lower proficiency groups, which confirms the general importance of phonological processing for earlier stages of foreign language learning (e.g., Skehan, 1989, 1998; Sparks & Ganschow, 2001) and contributes to accumulating empirical evidence showing that the facilitative effect of the phonological loop in L2 learning varies across the proficiency spectrum (Kormos & Sáfár, 2008; O’Brien, Segalowitz, Collentine & Freed, 2006, 2007; Winke, 2005).

Kormos and Sáfár found that PSTM, as measured via a non-word repetition (NWR) task (i.e., ability to repeat word-like non-words), significantly correlated with overall performance on a standardized English language exam at the end of a one-year intensive language program for adolescent L2 English learners at an upper-intermediate proficiency level, though not for beginners. Relationships were strongest with the subcomponents thought to measure explicit knowledge (writing, Use of English) but also with oral fluency and vocabulary. O’Brien et al.
(2006, 2007) found a changing role for PSTM over time, as measured on a serial non-word recognition task (SNWR). Based on regression analyses, phonological memory was a significant predictor in narrative abilities (oral fluency and vocabulary learning) for low proficiency L2 Spanish learners whereas PSTM was related to accurate use of complex grammar (subordinate clauses) at more advanced stages of learning. Relatedly, Winke (2005) reported that the role of PSTM, as measured in an aural sentence span task, played a more general role for beginning L2 Chinese learners, correlating with writing, speaking, and reading abilities, whereas it was only correlated with listening ability for advanced learners.

Overall, the fact that PSTM clearly had more durable, robust effects at the intermediate level compared to the beginning group supports findings by Kormos and Sáfár, which could be explained by a relatively stronger role of the executive function of WMC for beginners. However, the significant role of PSTM in mediating complex grammatical accuracy (O'Brien et al.) and listening ability (Winke) for advanced learners was not supported in this study as PSTM did not appear to underlie performance on either task for the advanced group at any point in time. This difference could be explained in terms of the target languages (L2 Spanish versus L2 Chinese) investigated in each study as "learning Chinese requires different abilities than learning other languages" (Winke, p. 158, c.f. Carroll, 1962, p. 102) such as the acquisition of characters; thus, the role of phonological processing ability could persist at more advanced stages of learning Chinese as an adult whereas its role may diminish with increasing L2 proficiency in a romance language like Spanish. The difference in findings could also be explained in terms of the different measures used to tap PSTM. The serial non-word repetition task and aural sentence span task respectively used by O'Brien et al. and Winke may be confounded with verbal language skills or existing lexical knowledge and familiarity with word patterns in the language.
of testing (Baddeley, 2003; French & O’Brien, 2008) whereas the digit span provides a relatively nonverbal measure of the capacity of an individual's phonological loop (Baddeley, Gathercole & Papagno, 1998).

The central executive component of WMC was also measured on a non-linguistic complex span task administered in learners' L1 (Operation Span) in order to avoid a confound between L2 working memory capacity and L2 proficiency level (e.g., Gass & Lee, 2011; Service, Simola, Metsanheimo, & Maury, 2002; Van den Noort et al., 2006; Wen, 2012). Measured in this way, WMC was found to play a significant role in explaining performance variance for the lower proficiency groups, beginning learners in particular. This finding supports previous empirical research that employed complex span tasks tapping both processing and storage components of WMC in the L1 who reported positive WMC effects on overall foreign language performance on a standardized achievement test for beginning learners, especially on the explicit 'Use of English' section (Kormos & Sáfár, 2008), the degree of L2 learning on a test of explicit knowledge over a semester for beginning learners even when L2 motivation was accounted for (Linck & Weiss, 2011), and online sensitivity to gender/(number) agreement violations in beginning (Sagarra, 2007b, 2008) and intermediate learners (Sagarra & Hershensonoh, 2010).

On the other hand, these findings contrast studies that reported nonsignificant relationships between L1 WMC and performance on standardized achievement tests over time for early and late beginning Spanish learners (Sagarra, 2000), on a test of metalinguistic knowledge by intermediate Spanish learners with an average four years of previous instruction (Roehr & Gánem-Gutiérrez, 2009), sensitivity to L2 agreement morphology for advanced Spanish learners (Foote, 2011), and sensitivity to number agreement violations in L2 French for 'mid' and 'high' proficiency learners (Coughlin & Tremblay, 2013). A likely reason for the
difference in findings between the current study and studies by Sagarra and Roehr and Gánem-Gutiérrez could be related to the way WMC was measured; both the latter studies tapped storage only in Daneman and Carpenter's version of the L1 reading span task rather than including a simultaneous processing task, which is thought to reveal the true nature of the limits of an individual's working memory span (Logie, 2011). Miyake (2001) has argued that complex dual-task methodology should always be employed in research investigating the role of WMC in complex cognition in order to accurately reveal whether any effects exist. The discrepancy in findings between this study and Foote and Coughlin and Tremblay's studies may be due to the proficiency level of each learner sample as their learners were advanced and WMC effects are most consistently reported for learners at low proficiency (Sagarra, 2013).

Importantly, the ability to code, store and retrieve materials was found to play a facilitative role not only in accessing explicit grammatical knowledge on the untimed grammaticality judgment task but also "implicit" knowledge as measured on the oral imitation task, for beginning learners. However, the strength of the relationship was consistently more robust with explicit knowledge, which is theoretically expected, given that WMC is thought to be more involved in regulating actions under controlled, conscious processing (e.g., Baars, 2003; Baars & Franklin, 2003; N.C. Ellis, 2005) and the degree of correlation with complex cognition could depend on the demand for controlled attentional processing (Engle, Kane and Tuholski, 1999). This finding partially supports Erçetin and Alptekin's (2013) study who found a significant role for WMC in performance on both measures of L2 knowledge and a relatively stronger relationship with use of explicit knowledge. However, learners were at an advanced level in their study, whereas significant relationships with WMC were only found for lower proficiency groups in the current study. This may be explained by the fact that the authors
measured WMC via a verbal task (RSpan) in learners' L2 (Turkish), arguably causing a confound with L2 proficiency. While their learners were reported to be at an advanced skill level, WMC measured in the L2 is still likely to interact with variation in target language proficiency, and thus obscure whether or not L2 performance correlated with L2 proficiency or WMC.

However, the significant relationship found with performance on the test of implicit knowledge could also have resulted from the interaction between working memory components and the nature of tests used to measure language development (Robinson, 2005b; Sanz, 1997; Sanz, under review). The elicited oral imitation task is reconstructive in nature (Erlam, 2006; Spada, 2013), requiring learners to aurally process incoming input in Spanish and to immediately produce, or repeat, what they heard accurately. Given this dual online requirement, working memory limitations are likely to be "pushed and exceeded" (Sawyer & Ranta, 2001, p. 341) on this task, especially for beginning learners.

Significant relationships found with WMC also underscore the hypothesized retrieval function of WMC in accessing stored long-term knowledge relevant to the tasks in hand (Baddeley & Logie, 1999). While previous research has primarily focused on the involvement of WMC at the initial input processing stages in the learning process, related to an "encoding-dominated" perspective (Skehan, 1998), this study provides instructive evidence about the role WMC plays in mediating how efficiently learners retrieve their L2 knowledge of Spanish grammar, particularly for beginning learners before undergoing a semester of instruction and after a one-month period without L2 exposure. This finding lends support to research investigating the effectiveness of different instructional interventions and who have reported that WMC mediated retention of L2 learning on delayed posttests (Erlam, 2005; Robinson, 2002, 2005; Sanz, under review). Interestingly, these studies also found that more explicit instructional
conditions neutralized the predictive ability of WMC, which is partially supported in the context of the current study given that WMC effects were least robust at the conclusion of the semester of instruction. On the other hand, WMC was a more robust predictor of performance variance on the test tapping explicit versus implicit knowledge, which does not support the stronger role found for learning under more implicit conditions. However, learning processes and retrieval of L2 knowledge are not the same; thus, it is plausible that cognitive mechanisms like WMC are differentially involved in mediating learning and knowledge retrieval (e.g., N.C. Ellis, 2005).

Finally, it is also interesting to note that all relationships found with the test of implicit knowledge were based on OSpan task performance scored most leniently (OSpan-Composite-L (Lenient)) whereas relationships between explicit knowledge and WMC were based on the scoring method with the strictest serial order and accuracy criteria (OSpan-Composite-A (Absolute)). The logic behind scoring WM span tasks is a relatively neglected topic in SLA studies despite the fact that several ways of measuring the same WM span measure "might yield different "span" values for the same person" (Conway et al., 2005, p. 774). Though Conway et al. argue that absolute scoring methods may limit variation in span values, the fact that scoring method interacted with the way L2 knowledge is measured merits further investigation in order to refine our conceptualization of the functions of WM, retrieval of L2 knowledge in particular.

Overall, these findings support a more is better view of WMC (Miyake & Friedman, 1998) and the prediction made by a single-resource model of WMC (Just & Carpenter, 1992; Just, Carpenter & Keller, 1996) that low WMC reduces one's ability to efficiently process L2 grammar. However, both the executive and phonological components of working memory were found to play differential roles according to the proficiency level of the learner. In other words, a limited WMC capacity constrained L2 performance for low proficiency classroom learners to a
greater extent than it did at higher proficiency levels, similar to what was found for the more traditional components of L2 aptitude, phonetic coding ability in particular. However, WMC and PSTM were more consistently related to and predicted L2 development compared to L2 aptitude as measured on the MLAT, providing further evidence for the dominant view that working memory capacity defines a general ability to learn languages (e.g., Miyake & Friedman, 1998; Sanz, 2005; Sawyer & Ranta, 2001; Wen, 2012; Wen & Skehan, 2011). Also, the effects of WMC were most robust at the first and final testing time, indicating that a semester of exposure to Spanish in a classroom setting may have neutralized the facilitative role of WMC. Finally, stronger relationships between WMC and performance on the UGJT on the one hand, and between PSTM and performance on the OIT on the other, support the view that processing (central executive) and storage (PSTM) WM components are fractionated and operate independently to some extent (see Sagarra, 2013; Wen, 2012, forthcoming).

*Processing speed.*

While several meaningful relationships were found between accuracy performance and other cognitive abilities measured here, no associations were evident with processing speed, a component of general intelligence operationalized as the speed at which learners can accurately code symbols corresponding to numerical digits (Weschler, 1997). Nevertheless, a significant negative correlation was found between oral latency and processing speed at the end of the semester for intermediate learners. This indicates that 4th-semester Spanish learners who were able to comprehend and repeat sentences faster on this task, also had higher processing speed scores. In terms of L2 sentence comprehension, Roberts (2012) rightly points out that it is not entirely clear what is meant by "processing speed" as it could relate to efficiency in various processes from "decoding of orthography/speech sounds, to lexical access and selection, to
integration with syntactic and other knowledge, as well as to the prediction or anticipation of up-and-coming input" (p. 181). In any case, processing speed was found to mediate how quickly, but not how accurately, instructed L2 learners at an intermediate proficiency level comprehended input and accessed their knowledge of Spanish grammar.

*Cognitive IDs and oral latency.*

In addition to the negative relationship found between processing speed and oral latency at the end of the semester for intermediate learners, six negative associations were found between cognitive capacity and oral latency on the elicited oral imitation task, indicating that learners who responded faster and more efficiently (i.e., lower mean RTs & CVs) on this task also had more efficient basic cognitive processing resources. For example, at the start of the semester, beginning learners who had lower mean RTs (i.e., responded faster) and advanced learners who had lower mean CVs (i.e., responded more efficiently) also had higher phonological memory resources. Intermediate learners with lower CVs were also found to have high phonemic coding ability and grammatical sensitivity, though the former relationship was more robust. One month after the end of instruction, learners with higher rote learning ability and higher working memory capacity were found to have lower mean RTs but these relationships did not reach significance by group, likely due to small sample sizes.

These findings underscore "the potential importance of individual differences in basic cognitive processing resources, such as speed of processing and working memory capacity" (Segalowitz, 2010, p. 168) in the development of L2 cognitive fluency. In terms of L2 sentence processing, previous research has found that with increasing proficiency, learners can become more efficient in their comprehension of verbal morphology (Hopp, 2010; Rodgers, 2011). However, systematic investigation into the effects of ID variables on automaticity in L2 sentence
processing in learners at varying proficiency levels is still lacking, which Roberts (2012) claims is because researchers typically "attempt to control for–rather than to assume a central role or even to independently investigate–the effects of individual differences" (p. 178) in this realm. This approach is related to a modular view of the L2 processing system as separate from other cognitive systems (Fodor, 1983). In contrast, results here provide tentative evidence in support of a domain-general view of language as beginning and advanced learners with more (phonological) memory resources and intermediate learners with a higher ability to encode phonemes also accessed their knowledge of L2 Spanish faster and more efficiently on the oral imitation task, which involved both semantic and syntactic comprehension and production.

Nonetheless, these results should be interpreted with caution given that the low accuracy scores on the oral imitation task prevented the option of filtering RTs according to error rates, which would have artificially reduced the oral latency data available for analysis. Overall, more research is needed that measures latency on different comprehension and production tasks that are likely to yield higher accuracy rates and not affect RT filtering procedures.

*Cognitive IDs and item grammaticality.*

Finally, the relationships found between cognitive capacity and performance at each testing time were almost exclusively accounted for by *ungrammatical* items, particularly on the UGJT. For example, phonetic coding ability, grammatical sensitivity, working memory capacity and phonological short-term memory *only* significantly correlated with performance for ungrammatical items at the beginning and end of instruction whereas one month later working memory capacity significantly related to all items. This supports claims that L2 learners may respond differently to the grammatical and ungrammatical sentences in a GJT (R. Ellis, 2005, p. 160) and that they may rely on different L2 data bases or *cognitive processes* in accepting well-
formed strings and in rejecting ungrammatical strings (Hedgcock, 1993, c.f. R. Ellis, 2005). However, to my knowledge, no previous study has empirically examined a potentially differential role for cognitive IDs in mediating performance according to item grammaticality; thus, this study provides initial evidence in this realm and merits further research.

**Psychosocial ID components.**

The development of knowledge of L2 Spanish grammar over time in this learner population was not only found to be facilitated and constrained by cognitive function, but also by several aspects of L2 motivation, supporting many SLA researchers emphasizing the social and psychological complexities of learning another language as an adult (e.g., Dörnyei, 2005, 2009; Ehrman & Oxford, 1995; Gardner, 1985, 2001). Similar to the cognitive side of the learner ID equation, the explanatory power of several motivational components diminished at the highest level of proficiency and the effects of L2 motivation depended on testing time and item grammaticality, though not to the same extent as cognitive IDs.

The construct of L2 motivation was quantified in two self-report questionnaires based on two dominant frameworks of motivation, Gardner's Socioeducational Model of SLA (Attitudes & Motivation Test Battery) and Dörnyei's L2 Motivational Self-system (adapted L2 Selves questionnaire). These Likert-type questionnaires were administered to learners at two times before and after instruction to attempt to quantitatively account for the diachronic nature of motivation (Dörnyei, 2001). Both questionnaires were found to be internally consistent and have high test-retest reliability for these instructed L2 learners of Spanish.

The main findings are discussed below in terms of each measure of motivation and the relative explanatory power of central constructs in both models of motivation (i.e., Integrativeness & Ideal L2 Self) is considered for these learners in this university context.
L2 motivation as measured on the AMTB.

Overall, the magnitude of correlations with both measures of L2 knowledge was in line with the medium effect size of relationships previously reported in a comprehensive meta-analysis (Masgoret & Gardner, 2003), at least in relation to objective measures of L2 achievement (versus proficiency self-ratings or course grades). The strongest, most pervasive components of L2 motivation as measured on the AMTB were language use and classroom anxiety and attitudes toward the learning situation (ATLS), with consistently high effect sizes and power levels (> .80) for beginning and intermediate groups. Relationships with reported language anxiety were particularly evident on the test of implicit knowledge for beginning learners with a significant but far less powerful, relationship with use of explicit knowledge. For intermediate learners, language anxiety played a moderate role in mediating performance on the implicit knowledge measure at the last two testing times, but a more positive evaluation of their Spanish course and instructor (ATLS) was clearly the most robust ID predictor of performance variance over time for this group with medium to high effect sizes for measures of both implicit and explicit knowledge.

Considering the definition of language anxiety as "worry and negative emotional reaction aroused when learning or using a second language" (MacIntyre, 1999, p. 27), it is intuitive to think that anxiety would negatively influence attitudes, motivation and performance (Scovel, 1978, 2001). In fact, it has been proposed to be "quite possibly the affective factor that most pervasively obstructs the learning process" (Arnold & Brown, 1999, p. 8). While it can be viewed as a stable predisposition, or character trait (trait anxiety), in the context of this study it was operationalized as a situation-specific construct pertaining to foreign language learning (Horwitz, Horwitz & Cope, 1986). Under this view, there is both "good" anxiety that actually
may facilitate performance, and "bad" anxiety that may debilitate, or harm L2 performance (Horwitz, 2001). Given that a higher mean Likert rating (Strongly agree) on language anxiety scales signifies lower anxiety on the AMTB, positive relationships indicate that superior performance was related to lower reported Spanish class and use anxiety by beginners (i.e., positively skewed item, 'I don’t usually get anxious when I have to respond to a question in my Spanish class'; negatively skewed item, 'It worries me that other students in my class seem to speak Spanish better than I do').

Many kinds of L2 activities can trigger performance anxiety "depending on the students' learning style preferences and skill levels" (Oxford & Ehrman, 1993, p. 193), speaking and listening in particular, but also reading and writing. Within cognitive psychology many researchers have suggested that more anxiety causes reduced efficiency in cognitive processing during task performance (e.g., Eysenck, 1979; Schwarzer, 1986) and this "cognitive interference" has been found in relation to language tasks (MacIntyre & Gardner, 1994). The task demands on the elicited oral imitation task tapping implicit knowledge required learners to attend to meaning while listening to a series of sentences in Spanish and to then accurately repeat the sentences out loud. This dual task requirement likely produced high levels of situational language anxiety that negatively 'interfered' with performance for low proficiency learners whose processing system is likely to be strained under these conditions, not to mention that they are the group with the least amount of exposure to the target grammatical forms. The weak, but significant relationship between language anxiety and performance on the test of explicit knowledge found at the outset of instruction indicates that reading sentences also triggered language anxiety to some degree in beginning learners, but this negative effect did not persist at later testing times.
The strong role found for the situated component of the AMTB, conceptualized as an evaluative reaction to the language learning context, reinforces the fluctuating nature of motivation as a learner-internal difference that is highly dependent on the context in which learning takes place (Dörnyei, 1994). This component played an especially strong role for intermediate learners who are at an interim stage in the proficiency spectrum. It could be that the positive or negative impact of learners' reaction to the learning context matters more for learners 'in the middle' who need to maintain a higher level of motivation in order to decide to move beyond fulfilling a language course requirement and reach the next stage of L2 development.

This interpretation is complemented by the fact that differences in attitudes did not significantly explain performance for the advanced learners at any point in time. This group has reached a high level of proficiency in Spanish, and thus, they are likely to display high levels of trait motivation, or enduring patterns of motivation over periods of time (Tremblay, Goldberg, & Gardner, 1995), and less variable levels of state motivation, or temporary engagement in the learning situation at different points in time.

Taken together, these findings support previous research taking a diachronic perspective of L2 motivation (Gardner et al., 2004; Yanguas, 2007). These studies have also reported a strong role for language anxiety and course and teacher evaluation components of the AMTB. Gardner et al. (2004) found that achievement at the end of a year of instruction (i.e., course grades) was most significantly impacted by French course evaluation, French class anxiety and motivational intensity as reported by ninety-one university French learners at an intermediate level. Yanguas (2007) found language anxiety to be the only AMTB variable to significantly account for performance variance in listening and reading activities over a semester in forty-four Spanish heritage language learners enrolled in introductory native speaker university courses.
However, he reported that no AMTB variable was a significant predictor of writing or speaking scores. The absence of a relationship between language anxiety and speaking contrasts results found here which may be explained by the difference in language learning background for L2 instructed and heritage language learners and their resulting strengths and weaknesses in terms of L2 skills (Valdés, 2001).

Another motivational component on the AMTB that significantly related to L2 development was Motivation (MOT) (i.e., effort, desire, & affect towards learning). MOT was a consistent significant predictor of performance variance on the test of implicit and explicit knowledge for intermediate learners only at the start and end of the semester but relationships with MOT (as reported at Time 1) were far more robust at the end of the 2.5 month learning period. This component in Gardner's theory was originally conceived as how hard a student was willing to work to learn the language, similar to the notion 'goal-directed behavior' (Heckhausen, 1991). Therefore, the Motivational Intensity scale assesses the amount of effort the individual reports investing to learn the target language; conceptualized in this way, it is not surprising that MOT would play a bigger role for explaining performance at the end rather than the start of the semester and for intermediate learners who need to exert significant effort to excel in 4th semester courses focused on a comprehensive review of all grammatical aspects covered up to that point in the curriculum. Nevertheless, reported learning effort can only be interpreted as being indirectly related to a learning outcome or achievement as it precedes behaviors that learners report engaging in and must be inferred, which is a methodological issue with all studies on L2 motivation (R. Ellis, 2008, p. 116).

Furthermore, the 'Motivation' component not only encompasses reported learning effort but also two scales that measure the desire to learn the target language (i.e., extent to which the
individual desires to reach a high level of target language competence) and attitudes toward learning the language (i.e., the affect experienced while learning) (Gardner, 2000). Therefore, 'the motivated individual' is not only someone who expends effort, but also "has goals, desires, and aspirations, [and] enjoys the activity" (Masgoret & Gardner, 2003, p. 128). This characterization seems to be especially pertinent for learners at an interim stage in their linguistic development, an interpretation supported by findings by Gardner et al. (2004) whose L2 French learners were also at an intermediate proficiency level.

The Motivation component has also been reported to have the highest effect sizes for correlations with course grades, objective measures of L2 achievement and proficiency self-ratings for university learners relative to all other AMTB components according to a meta-analysis of 75 independent samples of data (Masgoret & Gardner, 2003). However, this finding was not borne out here as attitudes toward the learning situation and language use and class anxiety components were found to be the most relevant variables of L2 motivation measured within Gardner and colleagues' framework. While Masgoret and Gardner reported on age and learning context, they did not report the proficiency level of learners, which may explain this discrepancy.

Very few overall relationships were found with the instrumental orientation (IO) scale assessing one's practical or professional reasons for learning Spanish. Performance on the implicit knowledge measure was weakly correlated with IO for advanced learners at the outset of instruction and 3.5 months later for beginning learners on the test tapping explicit knowledge, but these associations were unreliable (power < .80). This finding supports previous research reporting that instrumental orientation is typically less reliable and produces correlations with smaller effect sizes compared to other AMTB components (Masgoret & Gardner, 2003, p. 140).
Researchers have also argued that Instrumentality lacks theoretical clarification within Gardner's core theory (Dörnyei, 2005, p. 70). Nonetheless, the concrete benefits of learning the target language, such as getting a job, have been shown to lead to successful foreign language learning (e.g., Dörnyei, 1990).

While it seems logical that learners in a university context may have largely practical or professional motives for learning Spanish, one reason this was not found here may be related to growing interest in studying other lesser learned languages for professional gain. According to the 22nd language enrollment survey published by the Modern Language Association (Furman, Goldberg, & Lusin, 2009), Spanish remains one of the top three most studied languages at U.S. colleges and universities. However, there has been considerable enrollment increases in other foreign languages such as Arabic and Chinese since 2006 (p. 3). Perhaps living and studying in an international community like Washington, DC emphasizes the visibility and importance of lesser learned languages that are deemed 'critical' by the state department and whose learning is incentivized by scholarships offered by programs such as the Critical Languages Scholarship Program, National Security Education Program and U.S. Student Fulbright Program, to name a few. Therefore, it is possible to conjecture that these 'critical' languages are perceived by learners in this context to be more 'instrumental' in terms of being professionally competitive on the job market, somewhat usurping the role that Spanish and French have traditionally held in this regard.

On the other hand, learners in this study were not found to be integratively motivated to learn Spanish either, with only one robust relationship found at the end of the semester with performance on the test of implicit knowledge by intermediate learners. A significant relationship with explicit knowledge found at the same testing time had very low power and did
not reach significance by group. Therefore, the traditional content validity of the concept of integrative motivation, or "a willingness to be like valued members of the language community" (Gardner & Lambert, 1959, p. 271) was not confirmed by the Spanish learner data analyzed here, supporting critiques of the relevance of integrative motivation for different learning contexts (e.g., Au, 1988; Crookes & Schmidt, 1991; Skehan, 1989).

Among others, Oxford (1996) and Oxford and Shearin (1994) have argued that integrative motivation is likely to have more relevance for students of a second language rather than for students learning a foreign language. Indeed, the cumulative empirical evidence in support of integrativeness has primarily been found with Francophones learning English in second language or bilingual environments where both English and French are spoken and have official status (i.e., Canada) (Gardner, 1985, 2000, 2001, 2006, 2010; Masgoret & Gardner, 2003). However, the current learning environment cannot clearly be categorized as a foreign or second language learning context based on Oxford's criterion of target language availability outside the classroom. Spanish and English are spoken in the wider local community, but only English has official status, and while there are ample opportunities to 'experience' Spanish outside the classroom, this largely depends on the students' awareness of and willingness to take advantage of those opportunities. Furthermore, Masgoret and Gardner analyzed studies according to both second and foreign language learning contexts and found that language learning environment had little effect on the correlations of achievement with attitude, motivation and orientation measures (p. 151).

Another possible explanation for the lack of support for integrativeness in this learning context is its traditional conceptualization. Dörnyei (2003, 2005, 2009) has argued that integrativeness should be reinterpreted in terms of the attributes that a person would ideally like
to possess, or an imagined ideal self in the target language, drawing on possible selves theory in social psychology (e.g., Markus & Nurius, 1986; Higgins, 1987). Indeed the view of L2 motivation as a function of the learner's perception of her projected future self-states has received accumulating validation over the last five years (Dörnyei & Chan, 2012) and the ideal L2 self and integrativeness seem to overlap in terms of psychological self-identification, or the identification learners feel towards the target language and its native speakers. This theoretical convergence has been supported empirically by consistently strong correlations between the two constructs (e.g., Csizér & Dörnyei, 2005; MacIntyre, Mackinnon, & Clément, 2009; Ryan, 2009; Taguchi, Magid & Papi, 2009). Robust relationships with high effects sizes were also identified between Integrativeness and the Ideal L2 Self at the start and end of instruction in the Spanish learner data analyzed here (see the discussion for RQ4).

Nevertheless, MacIntyre et al. (2009) have argued that the strong positive association between these two core concepts "does not preclude differential predictions of language learning outcomes” (p. 208), which was upheld in this study given that the Ideal L2 self was a far more powerful ID predictor of L2 performance over time than Integrativeness, as detailed below. This is in line with Dörnyei and colleagues' claims that the L2 self “presents a broader frame of reference with increased capacity for explanatory power” (2005, p. 104) for a wider range of second and foreign language learning situations.

L2 motivation as measured on adapted Selves questionnaire.

Overall, effect sizes (r values) found for the adapted L2 selves questionnaire were lower than those generally reported for selves scales (Dörnyei & Taguchi, 2010), likely related to the far smaller sample size than those typically reported in self-motivation research (e.g., 13,000 Hungarian adolescents in Dörnyei and Csizér, 2002, and Csizér and Dörnyei, 2005). However,
they are comparable to the magnitude of correlations reported in more recent studies between the ideal L2 self and course grades in L2 English and L2 Mandarin (Dörnyei & Chan, 2012).

Motivational components with the most robust explanatory power were beginning and intermediate learners' projected ideal L2 self (IS) states in Spanish, and reported learning effort, or motivated learning behavior (MLB), whereas very few relationships emerged with the ought-to L2 self (OS) scale.

The positive relationship between performance and the Ideal L2 Self was significant for the intermediate group on the test of implicit knowledge at Time 1 with a medium effect size whereas 2.5 months later this relationship was significant for beginning learners only, with a high effect size. At the final testing time 3.5 months after the semester began, the positive relationship with the Ideal L2 Self remained powerful, but this time for the advanced group only. On the test of explicit knowledge, relationships with IS had medium to high effect sizes for the intermediate learners only at the start and end of the learning period but not for their lower or higher proficiency peers.

The magnitude of these correlations were comparable but slightly lower than relationships reported for university English foreign language learners who were both English and non-English majors in Japan (Ryan, 2009). However, the learner sample in Ryan's study was far larger ($N = 2,397$), thus increasing the power of any existing associations. More importantly, L2 motivation was not studied in relation to an objective achievement or performance measure but was correlated with a self-reported intended learning effort scale. Masgoret and Gardner (2003) have maintained that correlations with L2 motivation depend on the nature and type of measurement of achievement, arguing that grades and self-reported ratings "could be expected to correlate more highly with motivation, since there is more opportunity for motivation to be
implicated" (p. 156). Dörnyei and Chan (2012) empirically confirmed this expectation in their study with 172 low-intermediate adolescent learners of English and Mandarin in Hong Kong. Correlations between the learners' ideal English and Mandarin self guides and *self-reported effort* to learn each language had higher effect sizes than relationships with their L2 English grades and L2 Mandarin grades.

As found with the instrumental orientation scale on the AMTB, the Ought-to L2 Self, or one's sense of responsibility to learn Spanish, seemed to play a limited role in this context. OS significantly correlated with performance on the implicit knowledge measure at the end of the learning period only for beginners and advanced learners but power was below 80% in both cases. Dörnyei and Chan (2012) also failed to find significant relationships between course grades and Ought-to L2 English and Mandarin selves in their learner sample, but did report substantial correlations between ought-to self guides and *self-reported effort* to learn English and Mandarin, again sustaining Masgoret and Gardner's observation.

Overall, the Ideal L2 Self seems to be a more powerful motivator compared to the Ought-to L2 Self for these native English-speaking L2 learners of Spanish, which is consistent with past research conducted in several learning contexts (e.g., Ryan, 2009; Kormos & Csizér, 2008; Taguchi, Magid & Papi, 2009). While ideal and ought selves both stem from *self-discrepancy theory* (Higgins, 1998) and involve the desire to reach a certain end-state by reducing the discrepancy between one's actual and projected future self, they are hypothesized to be distinct from one another. Namely, ideal self-guides have a *promotion* focus "concerned with hopes, aspirations, advancements, growth and accomplishments" (e.g., "I can imagine myself living abroad and having a discussion in Spanish") whereas ought-self guides have a *prevention* focus aimed at avoiding or "regulating the absence or presence of negative outcomes" (Dörnyei, 2005,
p. 101) (e.g., "I have to study Spanish because if I do not study it, I think my parents will be disappointed in me"). With this distinction in mind, it may be that ought-to L2 selves are less accessible or less amenable to 'imagery manipulations' (Ruvolo & Markus, 1992) for L2 learners than the ideal L2 self which is more positive and forward-looking. Dörnyei (2005) has also speculated that motives generated by a sense of duty or a fear of punishment, as represented by the ought-to L2 self, may have a short-term effect that falls short of fueling the continued commitment needed for successful L2 learning.

In terms of the scale intended to measure motivated learning behavior (MLB), or the amount of effort learners intend to put into learning, positive associations reached higher power at the latter two testing times compared to 2.5 months earlier, similar to the pattern of relationships found with the motivational intensity subcomponent on the AMTB that also gauges expended effort. Specifically, this component was significantly correlated with performance on the test of implicit knowledge for beginning learners at Time 2 and for intermediate learners at Time 3 with robust power in both cases. MLB was also significantly positively related to performance by the intermediate group on the test of explicit knowledge at the end of the semester with power approaching 80%. These findings indicate that the more effort that low proficiency learners reported learning Spanish (e.g., "I am prepared to expend a lot of effort in learning Spanish"), the better they performed at the end of the semester, on both tests for intermediate learners, and on the test of implicit knowledge for beginners.

The MLB scale used here is conceptually similar to the MOT scale in the social psychological framework in that it gauges how hard a student reports to be willing to work to learn the language. Thus, it is logical that this self-reported quantified amount of effort played a bigger role for explaining performance by low proficiency learners at the end rather than the start
of the semester. However, as previously mentioned, relationships with both effort scales are indirect in that the behaviors that learners report engaging in are not directly observable, and can only be indirectly related to the development of L2 linguistic knowledge. Also the MLB scale is essentially equivalent to intended effort scales that have typically been used as 'criterion measures' (or outcome variables) in the majority of correlational studies conducted within the selves framework starting with Dörnyei and Csizér (2001) and Dörnyei and Csizér (2002). In this study, however, it was analyzed as a correlate of two objective measures of L2 knowledge, rather than as the outcome or achievement variable itself.

Overall, the version of the L2 Selves questionnaire adapted for L2 learners studying Spanish in a university setting in the United States supports the validity of the principal construct within Dörnyei’s theory of the L2 Motivational self-system. This contributes to accumulating evidence provided by studies conducted in several different linguistic and cultural contexts, namely, Hungary (e.g., Csizér & Dörnyei, 2005; Dörnyei & Clement, 2001; Dörnyei & Csizér, 2002; Kormos & Csizér, 2008), China (e.g., Taguchi, Magid & Papi, 2009; Dörnyei & Chan, 2012), Japan (Ryan, 2009; Taguchi et al.), Iran (Taguchi et al.), and Saudi Arabia (Al-Shehri, 2009). However, the validity of the self-system has mainly been probed for adolescent learners of English as a foreign language in countries other than the U.S. and needs further validation with adults learning other target languages in a variety of second and foreign language learning situations.

Furthermore, the third dimension in Dörnyei’s self-system concerning situation-specific motives, or the L2 Learning Experience, still has no direct operationalization comparable to the situated component of the AMTB (attitudes toward the learning situation), which has led MacIntyre et al. (2009) to argue that statistically, there is still no established measurement
instrument of L2 motivation equivalent to the AMTB. Dörnyei (2005, 2009) himself has acknowledged that this component is less theorized and lacks empirical validation within the self-system framework. Also, no previous study, to my knowledge, has administered a Selves questionnaire at two points in time despite Dörnyei's general critique of theories that imply motivation to be a relatively stable emotional or mental state that is "measurable by tapping into it at one point of time (e.g., by administering a questionnaire)" (2001, p. 16). While he and Ushioda (2009) argue for the use of qualitative interviewing techniques to study motivational evolution, the constructs measured in the selves scales themselves need more evidence of longitudinal validity and indices of test-retest reliability for varying learner samples.

_Psychosocial IDs and item grammaticality._

Finally, the relationships found between the eight psychosocial ID components measured and performance at each testing time generally applied in the case of both grammatical and ungrammatical exemplars, indicating a more global impact compared to cognitive capacity which was almost exclusively implicated in performance on _ungrammatical_ items (i.e., five out of six relationships found). However, some relationships found with UGR exemplars had slightly higher effect sizes on the oral imitation task (e.g., attitudes toward the learning situation; ought-to L2 self).

This contrasts the pervasive role found for cognitive IDs in mediating performance _only_ for ungrammatical sentences on the UGJT. It provides preliminary evidence that learners' underlying psychosocial reasons motivating them to learn the target language do not constrain performance in the same way as cognitive ability and seem to more globally impact performance, regardless of grammaticality. To my knowledge, no empirical study has previously investigated this topic but it merits further research in order to clarify the potential interaction between the
role of cognitive and psychosocial IDs and access to learner knowledge according to item grammaticality.

**Relative predictive ability of cognitive and psychosocial IDs.**

The correlational analyses carried out have identified several interesting relationships between learner-internal differences and implicit and explicit L2 knowledge. However, correlations can only identify the extent to which two variables are related, or covary, and cannot show causation. As one of the goals of scientific research is to try to explain and predict phenomena (Larson-Hall, 2010, p. 177), this study aimed to fill an empirical gap by investigating the relative capacity of cognitive and psychosocial learner differences in predicting L2 development over time.

In the combined regression models, cognitive and psychosocial explanatory variables together generally accounted for up to a quarter or a third of the performance variance on the test of implicit knowledge (22%, T1; 25%, T2; 29%, T3) and the test of explicit knowledge across groups (22%, T1; 16%, T2; 30%, T3). According to Larson-Hall (2010), these squared $r$ ($R^2$) values explain a medium (.09) to large (.25) percentage of variance in the dependent variables (pp. 118-119). However, there is still a substantial amount of residual variance left unexplained in the combined regression models. This residual variance indicates that other ID factors not considered here likely exert a substantial influence in the development of linguistic knowledge in instructed learners over time such as use of language learning strategies, cognitive style, personality, sex and previous language learning experience (e.g., Bowden, Sanz & Stafford, 2005; Dörnyei, 2005; Ehrman & Oxford, 1993; Oxford & Ehrman, 1995).

Also, it is interesting that IDs explained almost 10% less performance variance at the end of the semester on the measure of explicit knowledge compared to implicit knowledge (16%
versus 25%). This suggests that a semester of instruction may have neutralized the facilitative effect of IDs in mediating explicit knowledge, supporting empirical studies in L2 instruction reporting that explicit learning conditions may level the playing field in terms of the facilitative effect of higher language aptitude and working memory resources (Erlam, 2005; Robinson, 2002, 2005b; Sanz, under review).

In terms of relative explanatory power, the combined model posited for explicit knowledge across groups revealed cognitive ability, specifically working memory capacity, to be the only explanatory variable to significantly account for unique variance in scores at the beginning of the semester and 3.5 months later when all ID components were entered into the model. On the other hand, L2 motivation, conceptualized as the Ideal L2 self, was the only significant explanatory variable in the final model posited for implicit knowledge across groups at the beginning and end of the semester. These findings indicate that cognitive capacity explains relatively more variance on the measure of explicit knowledge and L2 motivation explains relatively more variance on the measure of implicit knowledge, supporting the view that IDs may differentially mediate implicit and explicit learning and knowledge (Hulstijn, 2005; N.C. Ellis, 2005).

By proficiency group, cognitive and psychosocial IDs had far stronger explanatory power for the lower level groups than across groups. For example, psychosocial IDs such as language anxiety accounted for up to half of the variance in the outcome variable (implicit knowledge) for beginning learners and attitudes toward the learning situation accounted for around a third of the unique variance for intermediate learners. Also, cognitive IDs like phonological short-term memory were strong predictors of performance on the test of implicit knowledge for both beginning and intermediate learners, especially at the start of instruction. The latter finding
indirectly refutes the claim within cognitive psychology that differences in cognitive ability matter less for implicit learning than for explicit learning processes (Reber & Allen, 2000, p. 236; see Reber, 1993), although the same claims do not directly apply to the product (knowledge) of learning.

Overall, results pertaining to the predictive ability of IDs indicate that individual differences not only mediate L2 development but also that predict its variation. This supports Hulstijn (2005) who argues that "if language aptitude, intelligence and working memory can be conceptually related to the constructs of implicit and explicit learning and knowledge, their status might change from correlational and peripheral to central and causal" (p. 136). However, this research has shown that L2 development is not only constrained by cognitive ability but by several components of L2 motivation as well, and thus, should not be neglected when investigating the impact of learner-internal differences on linguistic development.

Research question 4: Relationships among ID components within the cognitive and psychosocial dimensions of L2 learning.

The final research question asked What relationships exist among the cognitive and psychosocial dimensions of L2 learning in this learner population? This research question aimed to further investigate what learners bring cognitively and psychologically to the task of learning a second or foreign language and the potential interaction among these dimensions.

Traditionally, five to six core individual difference variables are examined in relation to indices of L2 achievement in SLA research (Dörnyei, 2005, p. 197; R. Ellis, 2008, p. 10): age, aptitude, cognitive style, language learning strategies, motivation, and personality. However, despite the fact that they likely "do not operate independently of one another" (Gardner, Tremblay & Masgoret, 1997, p. 347), SLA studies have rarely investigated existing empirical relationships among these ID variables in the same sample of learners. This is a necessary step in
L2 learner ID research in order to fully account for the many ways that IDs may interact in facilitating or inhibiting L2 development.

Unsurprisingly, the relationships among the ID components comprising the same general construct of 'cognitive capacity' or 'L2 motivation' were all positive with a magnitude expected of conceptually similar variables (Cohen, 1988; Larson-Hall, 2010). This indicates that specific aspects underlying each general dimension of L2 learning seem to reinforce one another. In terms of cognitive ability, eight positive relationships were identified among the six subcomponents. For example, the higher one's ability to temporarily code, store, and retrieve information (working memory capacity), the higher his or her capacity to successfully discriminate between L2 sounds and retain associations with symbols that represent them (phonemic coding ability). The cognitive ID that was least related to all other cognitive abilities was grammatical sensitivity, a component of L2 aptitude. This cognitive variable was also found to play a very limited role in explaining learner performance variance.

While the central executive (WMC) and storage/articulatory components (PSTM) in Baddeley's multi-component model (Baddeley & Hitch, 1974; Baddeley, 1986, 2000, 2003, 2007, 2010) were significantly associated, the effect size was small to medium, even though a larger effect size would be expected for conceptually overlapping variables. In conjunction with the different relationships found between WMC and PSTM and L2 performance, this weak to moderate association supports the relatively independent operations of components within Baddeley's model of working memory (Sagarra, 2013).

In terms of L2 motivation, twenty positive relationships were found among the eight subcomponents measured at Time 1 and fifteen of those relationships persisted at Time 2, as well as one additional relationship not present at Time 1. For example, at the beginning of the
semester, the stronger a learner's desire to identify with native Spanish speakers (integrativeness), the more favorably they evaluated their instructor and course (attitudes toward the learning situation); however, this relationship was not evident 2.5 months later. Also the more effort learners reported to learn Spanish (motivated learning behavior), the less classroom and language use anxiety they reported experiencing at the beginning but not the end of the semester. The psychosocial ID that was least related to all other motivational components was the Ought-to L2 self which was also found to play a very limited role in explaining learner performance variance.

Most notably, robust relationships were identified between Integrativeness and the Ideal L2 Self at the start and end of instruction, the two central constructs in both Gardner's Socioeducational Model of SLA and Dörnyei's L2 Motivational Self-System. This positive association reflects the strength of correlations found in previous empirical research and supports the claim that the two concepts may be tapping into the same pool of identification that learners feel towards the target language and its native speakers (e.g., Csizér & Dörnyei, 2005; MacIntyre, Mackinnon, & Clément, 2009; Ryan, 2009; Taguchi, Magid & Papi, 2009). For example, MacIntyre et al. (2009) found significant correlations between the ideal L2 self and integrativeness in a study with native English-speaking L2 French high school students and Ryan (2009) reported a strong relationship for L1 Japanese L2 English and non-English majors in a university context.

While these core concepts were found to converge both conceptually and empirically, the Ideal L2 self was found to be a far more powerful ID predictor of L2 performance over time than Integrativeness for this learner population (see the preceding discussion section addressing RQs 2 and 3 for further detail). Therefore, the current study underscores the view that Integrativeness may be "one local manifestation of a much more complex, powerful construct" (Ryan, 2009, p.
centering on a basic identification process within the individual’s self-concept, rather than a metaphorical, integration into an L2 community.

While ID components tapping different aspects of a similar overall construct can be expected to share a significant amount of variance (Gardner, 2000), it was of more interest to analyze the potential interaction between the cognitive and psychosocial underpinnings of foreign language learning, an inquiry less frequently made in SLA research. Conceptually similar ID variables are more likely to reciprocally influence one another, but it is more difficult to establish the direction of the relationship between cognitive ability and attitudes and motivation. However, it seems more rational to think that stable traits such as cognitive ability would influence fluctuating, state-like affective variables like learner attitudes, rather than the other way around.

Ten positive relationships and two negative relationships were identified between the components underlying cognitive capacity and L2 motivation in these instructed L2 learners of Spanish. Six relationships emerged at the beginning of instruction, three of which persisted 2.5 months later, and three that emerged for the first time. The positive and negative nature of these associations indicates that processes of both mutual influence and compensation were at play and contradicts claims that motivation and aptitude operate relatively independently (Gardner, 1985; Gardner, Tremblay & Masgoret, 1997). The two negative relationships found both involved working memory capacity. Learners with lower working memory resources also reported a higher effort and desire to learn Spanish (i.e., motivational intensity) at the outset of instruction as well as a more favorable evaluation of their Spanish course and teacher (ATLS) at the end of the semester. These relationships are interpreted as an attempt to compensate for less efficient processing and storage capacity.
Interestingly, learners with more efficient phonological and executive working memory resources also reported experiencing less foreign language anxiety (as indicated by a higher mean Likert rating) at the beginning and end of the semester (PSTM) and one month later (WMC). A logical interpretation of these relationships is that a diminished ability to code, store, and process language information is likely to cause more anxiety or apprehension upon task completion rather than more language anxiety reducing memory function. However, it has been suggested that high levels of anxiety can also interfere with basic cognitive processing on specific tasks (Eysenck, 1979; MacIntyre & Gardner, 1994). Therefore, it is impossible to unequivocally discern the direction of the relationships found here but this research provides preliminary evidence that the cognitive and psychosocial dimensions of learning L2 Spanish as an adult in a classroom setting do interact in complex ways, as some researchers have suggested (Hulstijn, 2002; Segalowitz, 2010; Segalowitch & Trofimovich, 2012), and that these relationships are not static over time. Nonetheless, the majority of relationships were not very reliable with power below 80%, which decreases the probability of detecting a true statistical relationship between variables when it exists (i.e., Type II error) (Larson-Hall, 2012, p. 465).

Overall, these findings support a broader notion of 'cognition' in L2 individual differences research, as a construct that not only encompasses cognitive ability, information and knowledge "but also emotion and motivation" (Hulstijn, 2002, p. 195). A cognition-motivation interface has long been acknowledged in neurocognitive science (e.g., Pinker, 1997; Schumann, 1998) and cognitive and educational psychology (e.g., Ackerman, 1989, 2003; Snow, 1989, 1994). Researchers in SLA have also begun to increasingly adopt the view that the factors underlying SLA are dynamic, open to change, and mutually affect one another, reflecting a complex or dynamic systems approach (e.g., De Bot, Lowie, & Verspoor, 2007; Dörnyei, 2005, 2009;
Larsen-Freeman, 1997; Larsen-Freeman & Cameron, 2008). Dörnyei (2009) specifically advocates for an approach that integrates cognition, affect, and motivation in longitudinal studies (p. 189) in order to account for the multidimensional, temporal interactions between factors that do not exert influence on learning outcomes in isolation, and cannot always be explained by a simple cause-effect relationship. The correlational data presented here among several ID variables point to the need to move toward a multidimensional, dynamic theoretical framework that is capable of capturing and predicting the complex, changing interactions among the different dimensions underlying L2 development.

**Summary Conclusions of the Study**

The current study attempted to explain the highly variable nature of L2 development in adults by focusing on the many ways learners differ from one another, following cognitive or cognitive-interactionist approaches to adult SLA (Long, 2007; Norris & Ortega, 2012; Sanz, 2005). In line with fundamental goals driving SLA theory and research, the primary goals of this dissertation were to first describe linguistic knowledge as measured during and after a semester of instruction in adult second language classroom learners of Spanish at increasing levels of proficiency, and to then identify what cognitive and psychosocial factors drive their development of and access to L2 knowledge of Spanish grammar over time. A secondary goal was to improve the validity, reliability, and replicability of tests measuring knowledge of and about second language grammar, prompted by previous empirical research (e.g., Bowles, 2011; R. Ellis, 2004, 2005, 2009; R. Ellis & Loewen, 2007; Erçetin & Alptekin, 2013; Erlam, 2006; Gutiérrez, 2012; Han & Ellis, 1998; Philp, 2009; Shiu & Spada, 2012; Spada, 2013).

This research was carried out with eighty-seven native-English speakers learning Spanish in a private Northeastern university in the United States. Participants were split into three
proficiency groups based on course enrollment and performance on a general proficiency test. Though these are classroom learners of Spanish, the learning context is considered to have elements characteristic of both a second and foreign language learning situation based on the criteria of availability of Spanish in the wider community (Oxford & Shearin, 1994; Oxford, 1996). Learners completed four measures of cognitive ability at one point in time, two Likert-type questionnaires tapping L2 motivation at two points in time, and two tests designed to tap implicit and explicit knowledge of ten targeted grammatical structures in L2 Spanish at three points during the 3.5 month investigation period. Participants also completed questionnaires at the end of the learning period and one month later to control for exposure to and use of Spanish outside the classroom.

A multi-faceted statistical approach was taken to analyze relationships and differences in the independent variables (proficiency level & learner ID components) and dependent variables (implicit & explicit L2 knowledge) of interest. Traditional statistical methods used in SLA research (e.g., one-way/repeated measures ANOVAs, correlations, multiple regressions, factor analyses) and innovative hierarchical linear modeling techniques used in cognitive psychology and education (e.g., intra and inter-individual linear growth curve analyses) provided complex and informative results which are summarized below.

First, all groups significantly improved after 2.5 months of classroom instruction in their ability to accurately repeat sentences they heard in Spanish after judging their meaning (implicit knowledge) and to accurately judge whether a sentence they read was grammatical or not in Spanish (explicit knowledge), and rate of improvement was comparable across groups. Nonetheless, higher proficiency groups significantly outperformed lower proficiency groups at each point in time but performance differences were more evident between learners on extreme
ends of the developmental spectrum (i.e., Adv, Int > Beg) on the test of explicit knowledge. In general, there was not only retention but continued gains in accuracy one month later with the exception that advanced learner performance seemed to hit a plateau in terms of explicit knowledge. Compared to the positive linear growth trends found by group, individual learner trajectories of development were nonlinear and variable. Analyses of oral latency data collected on the oral imitation task also indicated some evidence of automatization of linguistic knowledge with increasing proficiency, though caution is warranted in interpreting this finding given the methodological limitation of not filtering by error rate.

Second, the measures of L2 knowledge had high internal consistency for this learner population but the validity of a two-factor model was not uniform for all proficiency levels. Specifically, the separability of L2 knowledge constructs was questioned based on split factor loadings for ungrammatical items on the implicit knowledge measure and grammatical items on the explicit knowledge measure for low proficiency groups.

Third, several meaningful relationships were identified between performance and the six cognitive ID components and eight motivational components measured in this instructed L2 learner population. However, the explanatory power of IDs were found to minimally depend on several factors, namely, i) the proficiency level of the learner, ii) the point in time when knowledge was assessed, iii) the nature of the tests designed to tap L2 knowledge, and iv) test item grammaticality.

Finally, certain components of cognitive and psychosocial ID constructs measured here were found to both reinforce and compensate one another and their relationships were found to be dynamic over time.
The principal findings in this study support, expand, and modify the following conclusions in SLA and learner ID research: i) interlanguage knowledge is a continuum rather than a clear implicit/explicit dichotomy; ii) L2 interlanguage development over time is variable and nonlinear; iii) both cognitive and psychosocial IDs play an important role for L2 development at lower proficiency levels whereas their explanatory power is limited for advanced learners; iv) working memory and phonological working memory are the key constructs defining foreign language aptitude; v) the construct of integrativeness defining L2 motivation should be reinterpreted in terms of the ideal L2 self for certain learning contexts; and vii) a broader notion of cognition is needed in research investigating the role of learner IDs in L2 learning.

In the section below, I outline the implications for SLA theory, assessment, and pedagogy based on the results of this study. I conclude by considering several limitations and recommending future directions for research.

**Implications**

**Theoretical implications.**

The main findings of this study have refined and expanded our understanding of the development of L2 knowledge and the factors that facilitate and constrain its growth over time. Below I elaborate on five principal theoretical implications of the study.

1. *Implicit and explicit L2 knowledge is a continuum.*

This study was a continuation of previous SLA research, originally prompted by Han and Ellis (1998), with the goal of basing testing instruments on operational definitions of implicit and explicit L2 knowledge constructs. However, this was the first empirical study to critically consider the validity of the most effective measures of implicit and explicit knowledge (adapted from Bowles, 2011) for learners at varying levels of proficiency and the second to analyze
performance separately by item grammaticality for both measures, following Gutiérrez (2012). Tests were highly reliable for this learner population and a two-factor solution was replicated across groups and for the advanced group, but the picture was more complicated for lower proficiency learners with less exposure to and practice in the language.

Specifically, this study suggests that a theoretical dichotomy between implicit and explicit knowledge is best viewed as a continuum (DeKeyser, 2003) for instructed L2 learners who learned the target language after puberty and who likely draw on both types of knowledge in performance (Bialystok, 1981, 1994). Analysis of oral latency data by advanced learners also indicated that with practice, instructed learners can proceduralize, or automatize their ability to retrieve knowledge quickly such that it is indistinguishable from implicit knowledge (DeKeyser, 1997, 2007), insofar as behavioral performance reflects the underlying interlanguage system. Overall, more theoretical and empirical investigation into the psychometric properties of the implicit/explicit test battery is needed in order to test theories of implicit and explicit learning and knowledge (Hulstijn, 2005; Norris & Ortega, 2012), particularly the debate over how linguistic knowledge changes as proficiency increases.

2. The role of IDs diminishes at increasing proficiency levels.

Different ID components in both cognitive and psychosocial domains were found to be differentially involved in performance by classroom learners at three stages of interlanguage development, supporting previous calls for a componential and developmental perspective in learner ID research (Robinson, 2001b, 2005a, 2007, 2013; Skehan, 1989, 1998, 2002, 2012). Specifically, the explanatory power of cognitive capacity and L2 motivation was limited for adult learners in their 6th semester of Spanish instruction. In contrast, cognitive and motivational resources, particularly working memory capacity, phonological working memory/phonetic
coding ability, language anxiety, attitudes toward the course and instructor, and reported learning effort, were found to facilitate and constrain Spanish L2 development for adults in their 2nd and 4th semesters.

These findings support the hypothesized importance of phonological processing for earlier stages of foreign language learning (e.g., Skehan, 1989, 1998; Sparks & Ganschow, 2001) and contribute to accumulating empirical evidence indicating that the facilitative effect of working memory components on L2 learning, the phonological loop in particular, varies across the proficiency spectrum, (Kormos & Sáfár, 2008; O'Brien, Segalowitz, Collentine & Freed, 2006, 2007; Winke, 2005). Findings have also refined our understanding of the relevance of different motivational components for learners at different proficiency levels, a topic that has received little to no empirical attention in the literature. In particular, language anxiety and situated components measured on the AMTB within Gardner's Socioeducational Model (Gardner & Lambert, 1972; Gardner, 1985, 2000, 2010) were of most value in explaining performance by low proficiency learners in this study. In Dörnyei's theory of the Motivational L2 Self-System (2005, 2009; Dörnyei & Ushioda, 2009), the ideal L2 self meaningfully captured the relationship between motivation and development across groups and reported learning effort was found to be a useful construct for low proficiency learners.

The fact that L2 development involves different components of cognitive abilities and motivational orientations at different stages of learning has implications for SLA theory, which still lacks a systematic account of the role of individual differences (Skehan, 1998, 2002, 2012). In particular, this research shows that umbrella predictions pertaining to IDs should be avoided and an effort should be made to systematically relate different ID components to stages of
development in order to contribute to a unified framework that is capable of making testable predictions about the rate and quality of adult L2 learning outcomes.


This research has also contributed to further systematic investigation into the traditional aptitude construct and shown that it is indeed most accurately viewed in terms of learners' working memory capacity, at least for learners in the early stages of learning. For low-proficiency groups, several durable relationships were found with both the executive and phonological loop component of WMC whereas only phonemic coding ability was found to play a substantial role in performance for beginners with little involvement for grammatical sensitivity and rote memory components, the latter being an outdated, associative conceptualization of memory. This is interpreted as support for the now dominant view in SLA that working memory capacity is foreign language aptitude as it more efficiently captures "the essence of the three important components of language aptitude suggested by Skehan (1989)" (Miyake & Friedman, 1998, p. 361) (e.g., McLaughlin, 1995; Robinson, 2001a, b, 2002, 2005a; Sanz, 2005; Sawyer & Ranta, 2001; Skehan, 1998, 2002, 2012).

More recently, researchers have pursued this claim further and argued that WMC meets three necessary (pre)conditions to be viewed as foreign language aptitude (Wen, 2012; Wen & Skehan, 2011): 1) There are variations in WM that are specific to individual L2 learners and such variations are measurable; 2) the effects of WM in various SLA stages are constant and pervasive; and 3) different components of WM (phonological & executive components in particular) are highly correlated with different aspects of L2 performance and development (vocabulary, grammar acquisition) and with specific L2 skills and development (listening, reading, speaking, writing) (Wen, 2012, p. 17).
4. Learner IDs mediate the use of both implicit and explicit knowledge.

Cognitive IDs played a meaningful role in how learners implicitly and explicitly accessed their knowledge of L2 Spanish grammar, somewhat contrasting claims that differences in cognitive capacity matter less for implicit learning than for explicit learning processes (Reber, 1993; Reber & Allen, 2000). This finding supports studies in SLA that have found a meaningful role for aptitude under implicit and explicit instructional conditions (e.g., deGraff, 1997; Robinson, 1997) and for working memory capacity under more implicit pedagogical conditions (e.g., Erlam, 2005; Robinson, 2002, 2005b; Sanz, under review).

Nonetheless, IDs were somewhat differentially involved according to the task demands in the tests used to measure L2 knowledge. For example, relationships with working memory capacity were stronger and more robust with the measure of explicit knowledge at the beginning of the semester and 3.5 months later than with implicit knowledge, which is theoretically expected, given that WMC is thought to regulate actions under controlled, conscious processing (Baars, 2003; Baars & Franklin, 2003). On the other hand, phonological short-term memory was more robustly related to the measure of implicit knowledge for both beginning and intermediate learners. The latter finding could also be explained in terms of the interaction between different WM components and the skill assessed given that the measure of implicit knowledge required successful listening comprehension and spoken production (vs. reading comprehension on the test tapping explicit knowledge). Sanz (1997; under review) has emphasized that tests make different demands on working memory in particular and this should be considered when choosing tests to measure language development.
Overall, this research moves forward with a call to analyze general ID concepts in terms of their sub-functions in order to conceptually connect them to the constructs of implicit and explicit learning and knowledge (Hulstijn, 2005, p. 136; see also N. Ellis, 2005).

5. The utility of learner internal IDs varies over time.

The strength of relationships with cognitive and psychosocial IDs fluctuated at different points in the learning period, which provides a more informative look at the changing nature of relationships with IDs over time given that the majority of learner ID research has been cross-sectional in nature (Ortega & Iberri-Shea, 2005).

For example, the role of phonological short-term memory and attitudes toward the learning situation was relatively stable for intermediate learners on the tests of implicit and explicit knowledge but reported motivational intensity and motivated learning behavior were more relevant at the end of the semester. Also, while language anxiety was a durable predictor of performance on the test of implicit knowledge for beginners throughout the whole investigation period, it only mediated performance on the test of explicit knowledge at the outset of instruction.

Also, more aspects of cognitive capacity significantly related to performance at the beginning of the semester compared to 2.5 months and 3.5 months later whereas the role of L2 motivation was more constant over time, even increasing in importance at later testing times. Moreover, IDs explained less performance variance on the test of explicit knowledge at the end of the semester relative to implicit knowledge, which could be due to the neutralizing effects of receiving a semester of classroom instruction. This is similar to findings that explicit instruction with grammatical explanation 'levels the playing field' and neutralizes the beneficial role of individual differences related to aptitude (e.g., Erlam, 2005; Sanz, under review). In terms of L2 motivation, an increasingly positive role over time may reflect the need for L2 learners to
maintain a high level of motivation in order to reach the next stage of development, which supports the intuitive notion that L2 success requires sustained motivation and effort over long periods of time (Dörnyei & Ushioda, 2009; Gardner, 2010; Ushioda & Dörnyei, 2012).

6. The cognitive and psychosocial dimensions of L2 learning interact.

This study has shown that different cognitive abilities and motivational orientations facilitate L2 development, lending support to the idea that there is no single profile characterizing the 'good language learner' (Naiman, Frohlich, Todesco, & Stern, 1978; Griffiths, 2008), but rather, learners can take many different routes to success (Skehan, 1986, 1989). For example, a learner may have low L2 motivation but efficient cognitive resources or high motivation to learn the target language but low cognitive function. In either case, these IDs may compensate one another, rather than functioning in isolation. The stronger predictive power of 'combinations of traits' has long been recognized in cognitive psychology (Ackerman, 2003; Snow, 1989, 1994) and now by researchers in SLA advocating for the adoption of a broader notion of cognition in learner ID research (e.g., Hulstijn, 2002; Larsen-Freeman & Cameron, 2008; Segalowitz & Trofimovich, 2012).

Under this view, researchers on both ends of the theoretical spectrum have validity. On the one hand, L2 learning is similar to learning other complex cognitive skills (e.g., DeKeyser, 2007; McLaughlin, 1987; Mclaughlin, Rossman, & MacLeod, 1983) and on the other, L2 learning is quite different from learning other subjects given the social, cultural, and psychological context in which it takes place (e.g., Dörnyei, 2003, 2005, 2009; Gardner, 2000). In sum, this study has made progress in moving toward a unified framework in SLA that can potentially explain both the cognitive and affective/motivational influences on human behavior (Dörnyei, 2001), in this case learning a second or foreign language.
Practical implications for assessment of L2 performance and learner IDs.

This study also has valuable practical implications for SLA research in terms of refining our measurements of L2 knowledge and learner differences, which are detailed below.

1. Verifying how we know what our learners know.

In order to methodologically improve the construct validity of each measure of knowledge used, additional behavioral evidence was gathered and reported. The modifications on the elicited oral imitation task included i) collecting oral reaction time data to improve upon the operationalization of time inherent in definitions of implicit knowledge, ii) the calculation of an objective meaning processing score to ensure that learners were not only focused on meaning but that they indeed comprehended the input, and iii) the use of a scoring method evaluating the production of both form and meaning. The main modifications on the untimed grammaticality judgment consisted of eliciting i) judgment source attributions (Feel, Guess, Rule) and ii) written rule verbalization accuracy to not only better identify the source of non-native speaker judgments of L2 grammaticality but to also further probe the notion that explicit knowledge is potentially verbalizable, which has been lacking in previous research (Isemonger, 2007).

These proved to be fruitful endeavors and also served to highlight the challenges of measuring type of L2 knowledge. Overall, the time it took learners to orally repeat sentences they heard after judging their meaning was a useful indicator of the degree to which learners were accessing their L2 knowledge more or less automatically (and efficiently). The oral latency data provided some evidence that the retrieval of L2 knowledge is not only faster but more efficiently accessed with increasing proficiency, supporting previous research on automatization (e.g., DeKeyser, 1997; Hopp, 2010; Rodgers, 2011; Segalowitz, 2003; Segalowitz & Segalowitz,
1993; Segalowitz, Segalowitz, & Wood, 1998); though automatization of learner knowledge did not hold true over time.

Overall, this study provides further evidence that the oral imitation task is a valid test of L2 proficiency (e.g., Bowden, 2012; Ortega, 2000; Ortega, Iwashita, Rabie & Norris, 2002) as it effectively distinguished between what learners know at three levels of L2 proficiency in this study. However, its validity as a test of implicit knowledge was questioned. As suggested by other researchers (Shiu & Spada, 2012; Spada, 2013), it seems that when used with adult classroom learners, it is more accurately viewed as a measure of the degree to which learners have automatized their ability to use their (explicit) linguistic knowledge.

In terms of the UGJT, learner reported source attributions and rule verbalization increased its validity as a measure of explicit knowledge (e.g., R. Ellis, 2004; Gass, 1994; Sorace, 1985, 1996). The main goal in eliciting these data was to examine the extent to which learners were basing their judgments on explicit rule knowledge; results revealed that the majority of learners in all groups at all testing times reported basing their judgments of ungrammatical exemplars on a Spanish grammar rule and were significantly more accurate in their judgments when using rule knowledge, regardless of proficiency group. Nonetheless, the interpretation of what 'Feel' (i.e., intuition) means, especially when reported by low proficiency learners, remains problematic (Leow & Hama, 2013). As previously argued, it may be that 'Feel' actually indicates use of explicit, though unverbalizable, knowledge in participants who are late adult L2 learners with limited exposure to naturalistic L2 input. Overall, this study highlights the need to realign the use and interpretation of source attribution data in GJTs in terms of reflecting underlying degree of explicit knowledge in adult classroom learner populations.
Finally, this study supports analyzing disaggregated scores based on item grammaticality. Researchers should separately calculate accuracy scores for both grammatical and ungrammatical exemplars in order to continue investigation into the hypothesis that GR items more likely elicit implicit knowledge while UGR items more effectively elicit explicit knowledge of L2 grammar (R. Ellis, 2004, 2005; Gutiérrez, 2012).


This research also has implications for the way we currently measure L2 aptitude, working memory capacity, and L2 motivation. First, a componential view of L2 aptitude has already been argued for in order to identify which component is capable of capturing associations and predicting L2 performance. In this study, componential scores on the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959) effectively measured individual differences in L2 aptitude for early proficiency learners, mainly in terms of phonemic coding ability and to a lesser extent, grammatical sensitivity. High correlations with instructed L2 learning is typically demonstrated for early proficiency learners (Carroll, 1981; Dörnyei & Skehan, 2003; Sawyer & Ranta, 2001; Skehan, 1989, 1998, 2002), but Robinson (2005, 2013) argues that there are no current tests of aptitude available that operationalize a developmental perspective of L2 aptitude. This is needed in order to accurately capture the abilities drawn on not only in the initial phases but in the later stages of learning a foreign language as well.

In terms of measuring working memory capacity, it was argued that nonverbal, complex span tasks should be administered in the L1 (e.g., Engle, 2012, Sagarra, 2013) in order to avoid a confound with verbal skills and proficiency in the L2, and to efficiently capture the 'limits' of an individual's working memory span (Logie, 2011). Furthermore, attention was called to the way we evaluate performance in complex span tasks which is a relatively neglected topic in SLA.
studies (Conway et al., 2005, p. 774). In this study, relationships between working memory capacity (as measured on the Operation Span task) and performance on implicit and explicit knowledge measures depended on accuracy and order scoring criteria. Whereas the most lenient scoring method related to use of implicit knowledge, the strictest scoring method related to the use of explicit knowledge. These findings are revealing and merit further investigation in order to refine our conceptualization of the functions of WM, retrieval of L2 knowledge in particular. For now, the use of all three scoring methods (Absolute, Total, Lenient) is recommended in future empirical work with the goal of building evidence in this realm.

Finally, administering the two L2 motivation questionnaires at two points in time was helpful in providing evidence that relationships with performance fluctuate depending on the point in time when L2 motivation is measured. However, studies investigating motivational evolution over a period of learning should also incorporate more qualitative techniques such as frequent participant interviews (Dörnyei & Ushioda, 2009; Yanguas, 2007; Ushioda, 1996, 2001; Ushioda & Dörnyei, 2012). Moreover, researchers should carefully consider characteristics of the learning context before deciding which measure of L2 motivation to employ and adapt. In this study, probing the effectiveness of the central constructs in dominant models of L2 motivation proved instructive, given that Integrativeness, as measured on the AMTB with over four decades of cumulative research and empirical support (Gardner & Lambert, 1972; Gardner, 1985, 2000, 2001, 2006, 2010), was not found to capture relationships with L2 motivation in this learner population. However, a related construct centering on the learner’s perception of her ideal future self that has received increasing support over the last five years (Dörnyei, 2005, 2009; Dörnyei & Chan, 2012) was found to be a powerful motivator for these instructed learners.
Nonetheless, the AMTB still remains the only standardized test of motivation and the validity of Selves questionnaires can be improved by incorporating a situated component similar to the AMTB’s semantic differentials section assessing attitudes toward the learning situation, which currently has no parallel operationalization in the L2 Self System (Dörnyei, 2010; MacIntyre et al., 2009).

**Pedagogical implications for the classroom.**

The results of the current study can also be used to provide worthwhile information for the classroom practitioner in two main areas. The first pertains to being informed about the way learners differ from one another and applying that knowledge to the classroom. The second area concerns the use of motivation questionnaires in order to encourage self-awareness and use of effective motivational strategies among students.

1. **Know how your students differ from one another and apply what you know in the classroom.**

To provide the most effective or optimal instruction possible, teachers of a second or foreign language should understand the impact of student individual differences on learning (Ehrman & Oxford, 1995; Oxford & Ehrman, 1993; Oxford & Shearin, 1994) and how they may positively or negatively interact with instructional conditions and task demands (Robinson, 2001, 2002, 2007, 2011). While many good teachers intuitively identify differences in their students to improve their instructional practice, an explicit understanding of IDs that impact the development of their learners can enhance their effectiveness in making informed decisions while teaching (Gurzynski-Weiss, 2010; Leow, 1995) as well as aid in identifying learners who may need certain types of intervention (Roberts & Meyers, 2012).

According to the research carried out in this university context with adult L2 learners of Spanish, teachers should be especially aware of the following cognitive factors: i) The heavy
reliance of low proficiency learners on general cognitive resources to process phonological input and ii) the debilitating effects of language anxiety, especially in the context of listening and speaking tasks for beginners.

In communicative classrooms, there is a heavy emphasis on oral input and production, thus increasing the processing demands on phonological working memory (Hummel & French, 2010). Knowing that learners with low proficiency are likely to feel cognitively burdened, especially those with low phonological memory function, teachers should adapt their instruction to include written text and visual support in activities involving listening comprehension and oral production in order to avoid communication breakdowns and reduce the negative effects of language anxiety on L2 performance.

Peter Robinson has drawn heavily on Snow's (1989, 1994) notion of 'aptitude-complexes' and a dynamic person-situation perspective in his research investigating the interaction between L2 aptitude, instructional treatment, and task complexity (2001, 2011). This research is of direct relevance to the L2 classroom in terms of how teachers can attempt to match learner aptitudes to optimal conditions of instructional exposure. Being aware of what factors are likely to facilitate and constrain learning in their students can lead to more informed and effective decisions in classroom task design and sequencing (see Révész, 2011 for an example of a classroom task-based study examining the interaction of affective variables & task complexity). The current research valuably contributes to L2 pedagogy in this way by providing foreign language teachers with a comprehensive look at the learner-internal part of the L2 learning equation.

2. Use L2 Motivation questionnaires to encourage self-awareness and use of motivational strategies among your students.

This study also provides useful evidence for L2 teachers regarding the role of L2 motivation, particularly in terms of i) the facilitative effects of positive attitudes toward the
learning situation, especially for learners 'in the middle' or interim stages of L2 development, and how this may impact their decision to continue to the next level of language study; ii) the facilitative effects of a positive projected future self-state in guiding learning; and iii) the statistically high association between learner-reported motivated learning behavior and successful learning outcomes.

Findings reported here support the practice of administering motivation questionnaires in order to identify the different attitudes and orientations students have toward learning the target language and to promote continuous self-awareness of the attributes learners would ideally like to possess as a L2 learner of Spanish, Chinese, Portuguese, Arabic, etc. Thus, teachers should not only encourage learners to reflect on the way they view native speakers of the target language but also how they view themselves as speakers of that language. Self-imagery, or self-visualization has been shown to be a trainable capacity that can have a positive effect on learning in many areas (Dörnyei & Chan, 2012).

Taken together, these findings can be useful in L2 teaching methodology courses and teacher training workshops in educating instructors new to teaching a second language as well as those with long-term experience in the L2 classroom.

Limitations of the Study and Directions for Future Research

The results presented and discussed here have made valuable contributions to SLA theory, assessment, and practice. Nevertheless, it is important to discuss the limitations of this research in order to accurately interpret and situate the findings. In this section, I briefly discuss eight limitations and consider how they might inform and improve future research in this area.

The first limitation is the relatively short investigation (3.5 months) compared to recent descriptive-quantitative longitudinal research lasting several months to five or six years (Ortega
& Iberri-Shea, 2005). While three assessment waves permitted the use of hierarchical linear modeling techniques to analyze nonlinear L2 development, four or five assessment waves over several months would enable the study of even more complex growth trajectories (Singer & Willett, 2003) and the ability to identify those relationships with learner IDs that are more fluctuating and those that are consistent and durable over time.

A second limitation is that only two measures of L2 knowledge were used and were arguably mismatched in terms of task demands (i.e., modality of input; receptive versus productive skills). In future research aiming to continue this strand of research, using at least two measures of each type of L2 knowledge is recommended to strengthen construct validity and to closely match those measures in terms of task demands to ensure comparability.

Third, the very broad focus of this study in terms of the range of grammatical structures is both a blessing and a curse. While investigating only one to two target structures at a time arguably provides a more limited perspective in terms of what learners know at different stages of interlanguage development, investigating knowledge of ten structures did not allow for a desirably high number of exemplars per target structure in light of potential learner fatigue (i.e., 10 GR/10 UGR items for 10 target structures = 200 items). To increase the number of exemplars and confidence in the statistical results, researchers should analyze fewer target structures.

Alternatively, it would be wise in future research to divide the current linguistic structures into at least two groups based on their morphological, syntactic, and semantic properties. For example, linguistic structures could be grouped according to those requiring semantic processing (e.g., past tense aspectual distinction, the copular contrast to be, mood) versus those that do not (e.g., subject-verb, noun-determiner, noun-adjective agreement, constructions with gustar) or those that require morphosyntactic processing (e.g., present and
past subjunctive) versus morphological inflection only (e.g., person and gender agreement within the noun phrase). It would be interesting to compare the rate of development for each group as well as the potential interaction between learner IDs and properties of the linguistic structures.

The fourth limitation of this study is the analysis of oral latency data. While the data were screened for outlier RTs following standard data filtering and cleaning procedures in cognitive psychology and SLA (Jiang, 2011; Lachaud & Renaud, 2011; Leow, Grey, Marijuan & Moorman, submitted), it was not possible to filter RTs according to error response rates given the very low accuracy rates on this challenging task. Also, learner RTs were measured by hand using Praat due to a technical malfunction of a voice onset key. Thus, caution is warranted in interpreting all findings regarding oral reaction times. Future research should seek alternative, more reliable ways to measure learner RTs in oral production tasks and ensure through pilot testing that accuracy rates will not prevent analyzing latency for accurate responses only.

Relatedly, the latency results painted a somewhat different picture of the proficiency groups compared to the results of the placement test. Specifically, response speed was not significantly different between the intermediate and advanced groups whereas differences were more apparent in comparison to the lowest proficiency group. Thus, future research investigating L2 proficiency as an independent variable should consider including latency as a measure of proficiency in addition to course enrollment and standardized tests with the goal of establishing more robust proficiency assessment standards in experimental research (Tremblay, 2011).

Fifth, as with all research using learner self-reported data, caution is warranted in interpreting findings related to the quantified motivational constructs. The self-reported nature of L2 motivation questionnaires invites questionable veracity and impartiality (MacIntyre et al., 2009) as the attitudes, feelings, and effort reported by learners are not directly observable.
Sixth, while an effort was made to provide robust indices of all relationships and group differences by reporting power calculations and effect sizes (Larson-Hall, 2012), certain correlations with low power (< .80) run the risk of committing a Type I error, that is, "concluding that there is a difference between groups or a relationship between variables, when in fact there is not" (p. 465). One solution could be to set the alpha level below .05, to .01, and thus reduce the probability of committing a Type 1 error to 1%. Also, the sample size at time 3 was small (N = 33), which affects power and effect size, findings related to retention of L2 knowledge are tentative.

A seventh limitation is the amount of residual variance left unexplained in the combined regression models run with both cognitive and psychosocial variables. While an effort was made to incorporate a comprehensive battery of ID measures, other measures of cognitive abilities that focus on different functions of working memory capacity (e.g., inhibitory control, task-switching) and other affective variables such as personality should be studied in addition to those measured here with the goal of increasing the total amount of performance variance accounted for in the model.

Finally, the findings in this study are not generalizable to other native English speaking learner populations in different second or foreign language contexts. This is relevant especially in terms of L2 motivation which is highly dependent on the context in which learning takes place (Dörnyei, 1994) and in terms of L2 aptitude which may be differentially involved according to the target language being learned (e.g., Winke, 2005). Nevertheless, research questions guiding this study should be further explored with adult learners of different target languages in different contexts who are at different levels of proficiency in order to situate and build on the findings reported here.
APPENDIX A. Target Structures.

<table>
<thead>
<tr>
<th>Target structures ( (n = 10) )</th>
<th>Grammatical and *Ungrammatical Exemplars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Subject-verb agreement ( (\text{regular present tense}) )</td>
<td>Paula mira la televisión. *Luisa ves un oso.</td>
</tr>
<tr>
<td>2) Noun-adjective gender agreement</td>
<td>Manuel vive en una casa roja. *Ana tiene una falda negro</td>
</tr>
<tr>
<td>3) Noun-determiner agreement</td>
<td>Manuel corre afuera en el otoño. *Todos mis amigos viven en el estado de California.</td>
</tr>
<tr>
<td>4) Aspect</td>
<td>Ayer los niños jugaron hasta muy tarde. *Anoche los chicos estudiaban hasta las 3:00.</td>
</tr>
<tr>
<td>5) Ser-estar-Locatives</td>
<td>Los libros están en la biblioteca. *Mi cartera es en el apartamento.</td>
</tr>
<tr>
<td>6) Ser-estar-Events</td>
<td>La fiesta es el 28 de febrero. *El baile está a las 9:00.</td>
</tr>
<tr>
<td>7) Gustar</td>
<td>A los estudiantes no les gustan los exámenes. *A los niños les gusta los caramelos.</td>
</tr>
<tr>
<td>8) Present subjunctive-Nominal clauses</td>
<td>Mamá quiere que Paco estudie mucho. *Papá quiere que Elena estudie.</td>
</tr>
<tr>
<td>9) Imperfect subjunctive</td>
<td>Ana quería que su hijo mirara menos televisión. *El médico recomendó que mi papá trabaje menos.</td>
</tr>
<tr>
<td>10) Past hypothetical conditionals</td>
<td>Si Roberto hubiera estudiado, no habría sacado una mala nota. *Si tu hermano había sido médico, habría ganado mucho dinero.</td>
</tr>
</tbody>
</table>
APPENDIX B. Participant Biodata Questionnaire.

Welcome to the language background questionnaire portion of the Fall 2012 Spanish study.

This is a reminder that you have consented to participate in this research study about the development of Spanish as a foreign language. Please complete the following questionnaire so that we may learn about your previous experience with language learning. The questionnaire should take around 10 minutes of your time, and the information you provide is confidential. Please note that we need to collect your name and contact information to match your responses in this questionnaire to data from your other study sessions. However, only the researchers will have access to your information.

If you have questions about the study, you may contact: Ellen Serafini at (202) 285-4011 or esj23@hoyamail.georgetown.edu.

You may also contact the Georgetown University IRB Office at 202-687-6553 (8:30am to 5:00pm, Monday to Friday) if you have any questions about your rights as a research participant.

If you consent to participate, please click to continue this online portion. Thank you very much for your participation!

IRB #__________

LANGUAGE BACKGROUND QUESTIONNAIRE 2012

I. Background and demographic information

NAME (LAST, FIRST): _________________________ PARTICIPANT # ______________
SEX: M F DATE OF BIRTH: _______ EMAIL: _______@georgetown.edu

Please indicate which hand you write with: LEFT RIGHT
Do you have normal vision? YES NO
If NO, do you wear glasses or contacts? YES NO
Do you have any hearing problems that you are aware of? YES NO
If YES, please elaborate in the space below:
__________________________________________________________________________
_____________________________________________________

Please provide the following information:

Year of study: Freshman Sophomore Junior Senior
Area of specialization (if known): __________________________
Were you born in the United States?  YES  NO
If not, at what age did you begin residing in the US?  ________________

II. Spanish learning background

CURRENT SPANISH COURSE (FALL 2012): Please choose one.

SPAN: 004-Intro II  CURRENT SPANISH INSTRUCTOR: ________________
022-Intermediate II
104-Advanced II

Did you take the summer online placement exam or the fall campus placement exam prior to enrolling in the above course?  YES  NO
If so, did you enroll based on exam results?  YES  NO
If not, please explain: ________________

How many years have you studied Spanish in a school setting prior to enrolling in your current Spanish course? Please select the appropriate option and estimate according to any previous elementary school, middle school high school instruction and university level instruction.

For example, if you took 2 years of Spanish in high school and one semester in college prior to enrolling in your current course, you have studied Spanish for 2 1/2 years and the option "2.5 - 3 years" would be most appropriate.

0-1 year  1.5-2 years  2.5-3 years  3.5-4 years  4.5-5 years  5+ years
Other: ________________

Have you ever spent time abroad in a Spanish-speaking country?  YES  NO
If so, for how long were you there?
Weeks: ________________
Months: ________________
Years: ________________

What was the purpose of your trip?
Study Abroad  Travel  Other: ________________

III. Language learning background

This portion of the questionnaire asks about your native language and any previous language learning experience excluding Spanish. At the end of each set of questions, please click “Add _____ Language” (e.g., "Add Second Language") if you have not yet provided information on all of your languages. When you have answered questions for all of your languages, please click “Done.”

What is your native language (the language you first spoke)? If you speak a native language OTHER THAN English or if you speak a native language IN ADDITION TO English, please write it in the space below.
English
I speak a native language OTHER THAN English
I speak a native language IN ADDITION to English
Please specify your native language(s) here: __________________________________________

Native language (1): __________________

At what age did you begin learning this language? For example write "Birth" or another age such as "Five years" if you started learning it later on in childhood. ____________________________

Did you receive any formal instruction in this language? YES NO
If so, please estimate the amount of years you were formally instructed in this language. _________________

How would you rate your PRESENT proficiency in this language in each of the areas listed below? Please use the following scale:

6 = LIKE A NATIVE SPEAKER
5 = NEAR (ALMOST) NATIVE
4 = GOOD
3 = FUNCTIONAL
2 = POOR
1 = NONE OR ALMOST NONE

Speaking 1 2 3 4 5 6
Reading 1 2 3 4 5 6
Writing 1 2 3 4 5 6
Listening 1 2 3 4 5 6

How often do you CURRENTLY USE this language in each of the contexts listed below? Please use the following scale:

4 = ALWAYS 3 = FREQUENTLY 2 = RARELY 1 = NEVER

I speak it at work: 1 2 3 4
I speak it at home: 1 2 3 4
I read it at work: 1 2 3 4
I read it at home: 1 2 3 4
I write it at work: 1 2 3 4
I write it at home: 1 2 3 4
I listen to it at work: 1 2 3 4
I listen to it at home: 1 2 3 4
If you know another language that you also consider to be your native language, please click “Add Native Language (2)” and enter the requested information (i.e., age of exposure, formal instruction, present proficiency and frequency/context of use).

If you know another language that you learned after childhood (besides Spanish), please click “Add Second Language” and enter the requested information (i.e., age of exposure, formal instruction, present proficiency and frequency/context of use). You will have a chance to enter information for any other languages you know later on (i.e., "Add Third Language").

When you have completed all of the above information for all of the languages you know, click “DONE.”

**IV. Summary of language use and additional comments**

In this section we would like to ask you to estimate the percentage of time you spend using each of your languages at HOME/SCHOOL/WORK. (Note that percentages should add up to 100%. If they do not, please briefly explain why.)

Example:
1. English: 50%
2. Spanish: 30%
3. German: 20%

OR

1. English: 100%

Overall, during a typical week, approximately what percentage of the time do you spend using each of your languages at HOME/SCHOOL/WORK? (Note that percentages should add up to 100%. If they do not, please briefly explain why).

1. __________________________________________________________________________
2. __________________________________________________________________________
3. __________________________________________________________________________
4. __________________________________________________________________________

Comments: ____________________________________________________________

Please briefly describe any other language learning experience you have had that the previous questions have not sufficiently captured.

__________________________________________________________

Congratulations! You have completed the language background questionnaire. If you have questions about the study, you may contact Ellen Serafini at esj23@georgetown.edu or (202) 285-4011.
APPENDIX C. Classroom Teaching Style Questionnaire.

INSTRUCTOR QUESTIONNAIRE

LAST NAME: ___________________________ PARTICIPANT #__________
SEX: M  F (Circle one)

Thank you for filling out this brief questionnaire. The first part requests general background information in terms of your experience teaching Spanish and your self-perceived proficiency in Spanish and English. The second section aims to get a general idea of your classroom teaching style.

I. General background information

1) Which Spanish course are you currently teaching this Fall semester (2012)?
   SPAN: 004-Intro II
           022-Intermediate II
           104-Advanced II

2) Are you a graduate student? YES NO
3) If so, are you in a Master’s or Doctoral program? M.A./M.S. Ph.D.
   What year of study? _______________________
4) If you are not a graduate student, do you have a Master’s or Ph.D. degree? YES NO
   If so, please choose one.
   M.A./M.S. Ph.D.
   What is your area of study? _______________________

5) How many years have you taught Spanish at the university level? __________
6) Have you taught Spanish in other contexts? YES NO
   If so, please indicate which context(s): Primary Secondary

7) What is/are your native language(s)? _______________________________

Please rate your PRESENT proficiency in English and Spanish for each skill.

6 = LIKE A NATIVE SPEAKER
5 = NEAR (ALMOST) NATIVE
4 = GOOD
3 = FUNCTIONAL
2 = POOR
1 = NONE OR ALMOST NONE

English:

Speaking  1  2  3  4  5  6
II. Teaching style

In this section, please indicate your answers to the questions that follow by clicking the answer, or filling in the blank where appropriate, that best approximates your teaching style based on what one might observe on a typical day in your Spanish classroom. Try to give the first answer that comes to mind as we are most interested in your gut reaction to these questions.

During a typical day in your Spanish classroom......

1) Please rate the frequency with which you employ the following types of tasks and activities on using the following scale from 1 to 4.

4 = ALWAYS  3 = FREQUENTLY  2 = RARELY  1 = NEVER

a. Exercises or discussions based on content _____
b. Exercises or discussions based on grammar _____
c. Listening activities _____
d. Speaking activities _____
e. Reading activities _____
f. Writing activities _____
g. Individual learner tasks _____
h. Paired or small group work _____
i. Teacher-led discussions _____
j. Other: ______________________  _____

2) Overall, how often do typically correct student errors during classroom activities? Please click one.

4 = ALWAYS  3 = FREQUENTLY  2 = RARELY  1 = NEVER

3) Please indicate the degree to which you agree with the following statements regarding the types of feedback you typically provide your students during oral classroom activities using the following scale of 1 to 5.

1 = STRONGLY AGREE
2 = AGREE
3 = NOT SURE
4 = DISAGREE
5 = STRONGLY DISAGREE

a. I ask my students to clarify in response to their errors. _____
b. I use facial gestures to indicate they have made an error. _____
c. I repeat the student's error in a questioning tone. _____
d. I overtly correct my students' errors while they are speaking. _____
e. I use student errors as a way to discuss grammar rules. _____
f. I correct more errors related to grammar than to meaning. _____
g. I correct more errors related to meaning than to grammar. _____
h. I demand my students self-correct. _____
i. If they do not respond quickly, I usually correct their errors. _____
j. I never correct my students' errors. _____
k. Other: ________________ _____

4) Please provide an overall estimate of the percentage of time you typically speak and the percentage of time your students speak during a typical class period. For example, the following indicates that the instructor generally speaks 40% of the time while her students typically speak 60% of the time.

Me: 40%
My students: 60%

Me: _____
My students: _____
APPENDIX D. Outside Exposure/Use Questionnaire.

I. **End of the semester**

Please click the appropriate answer or provide the requested information below.

i. Outside of class, around how many hours a week did you spend studying for and preparing assignments in your Spanish course?
   a. 0-3 hrs   b. 4-6 hrs   c. 7-9 hrs   d. 10-12 hrs

ii. In addition to time spent working on class assignments, around how many hours a week did you spend using Spanish on your own?
   a. 0-3 hrs   b. 4-6 hrs   c. 7-9 hrs   d. 10-12 hrs

iii. Please indicate the extent to which you did each of the following activities on your own volition outside of class time this fall semester by clicking the appropriate answer for each activity.

   *I watched TV and/or movies in Spanish.*
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

   *I listened to music in Spanish.*
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

   *I spoke with native Spanish-speaking friends and/or colleagues.*
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

   *I interacted with native Spanish-speaking friends online.*
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

   *I read newspapers or magazine articles in Spanish.*
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

   *I received one-on-one tutoring sessions.*
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

   **Other:** __________________________________________
   ALWAYS   FREQUENTLY   SOMETIMES   NEVER

II. **After holiday break**

Please click the appropriate answer or provide the requested information below.

i. After the end of the Fall 2012 semester and before the start of the Spring 2013 semester, did you study Spanish on your own? ______
If so, around how many hours a week?

a. 0-3 hrs 

b. 4-6 hrs 

c. 7-9 hrs 

d. 10-12 hrs

ii. Did you speak or listen to Spanish at all during the holiday break? ______

If so, around how many hours a week did you spend using Spanish on your own?

a. 0-3 hrs 

b. 4-6 hrs 

c. 7-9 hrs 

d. 10-12 hrs

iii. Over the holiday break, please indicate the extent to which you did each of the following activities on your own volition outside of class time this fall semester by clicking the appropriate answer for each activity.

*I watched TV and/or movies in Spanish.*

ALWAYS FREQUENTLY SOMETIMES NEVER

*I listened to music in Spanish.*

ALWAYS FREQUENTLY SOMETIMES NEVER

*I spoke with native Spanish-speaking friends and/or colleagues.*

ALWAYS FREQUENTLY SOMETIMES NEVER

*I interacted with native Spanish-speaking friends online.*

ALWAYS FREQUENTLY SOMETIMES NEVER

*I read newspapers or magazine articles in Spanish.*

ALWAYS FREQUENTLY SOMETIMES NEVER

*I received one-on-one tutoring sessions.*

ALWAYS FREQUENTLY SOMETIMES NEVER

Other: ______________________________

ALWAYS FREQUENTLY SOMETIMES NEVER
APPENDIX E. Elicited Oral Imitation (OI) Task Script.

Slide 1:

**Welcome to the Oral Imitation task.**

In this task, you will hear a series of sentences in Spanish.

Before you hear each sentence, you will see a slide with an asterisk *. After you hear each sentence you will be prompted to immediately say SI or NO to indicate whether you agree with the **content**, or **meaning**, of the statement. Give your first "gut" response.

Then, you must **repeat** the sentence out loud in **correct Spanish** as quickly as possible. (Sometimes this will mean repeating the sentence exactly as you heard it, and sometimes this could mean changing some part of the sentence.) If you do not remember the entire sentence, repeat as much as you can.

Finally press the **space bar** to advance to the next sentence.

Press any key to continue.

Slide 2:

The following is a sample series in this task:

1) You **see**: *

2) You **hear**: *La casa blanca está en DC.*

3) Immediately respond to **meaning**: Sí (o No)

4) Immediately **repeat** aloud in correct Spanish: *La casa blanca está en DC.*

5) Press **space bar** to advance to the next sentence.

Let's do a couple practice examples. Press any key to continue.

Slide 3:

Let's practice.

(Hears) *Los Estados Unidos es un país pequeña.*

*  

(Judges meaning) Sí o No

(Repeats aloud & Presses space bar)
If you have any questions, please ask the researcher now. If you are ready to begin the task, press the space bar to advance to the first sentence.
Welcome to the untimed grammaticality judgment task!

In this task, you will see a series of sentences that are either grammatical or ungrammatical in Spanish.

For example, the sentence

*Yo voy a la biblioteca todos los días.*

is grammatical whereas

*Yo comes pizza los viernes.*

is not a correct sentence in Spanish.

Your job in this task is the following:

1) **First**, click "Grammatical" or "Ungrammatical" based on your judgment of the sentence.

2) **Next**, choose "Feel", "Guess", or "Rule" to indicate on what basis you made your judgment. "Feel" indicates that you used your intuition, or it "seemed" right or wrong. "Guess" means you weren't certain and guessed. "Rule" means you used a Spanish grammar rule to make your decision.

3) **Finally**, if you chose "Rule" to explain an ungrammatical sentence, you will be asked to simply describe it in English.

There is no time limit to this task. Good luck!
PART III. SPELLING CLUES

Each item below has a group of words. The word at the top of the group is not spelled in the usual way. Instead, it is spelled approximately as it is pronounced. Your task is to recognize the disguised word from the spelling. In order to show that you recognize the disguised word, look for one of the five words beneath it that corresponds most nearly in meaning with the disguised word. When you find this word or phrase, make a mark in the appropriate space on your answer sheet (SIDE B, Part III, Spelling Clues). DO NOT WRITE IN THIS BOOKLET.

Here are some sample items:

S. luv
   A. carry
   B. exist
   C. affection
   D. wash
   E. spy
   luv is a disguised spelling of love, which corresponds most nearly in meaning to affection, so C has been marked in Example S on the answer sheet.

T. ernst
   A. shelter
   B. sincere
   C. slanted
   D. free
   E. impatient
   ernst is a disguised spelling of earnest, which corresponds most nearly in meaning to sincere, so B has been marked in Example T on the answer sheet.

NOW GO RIGHT AHEAD WITH THE TEST. WORK RAPIDLY!
PART IV. WORDS IN SENTENCES

This is a test of your ability to understand the function of words and phrases in sentences.

Look at the following sample item:

V. LONDON is the capital of England.

He liked to go fishing in Maine.

V. A B C D E

In the first sentence, which we will call the key sentence, LONDON is printed in capital letters. Which word in the second sentence does the same thing in that sentence as LONDON does in the key sentence? The right answer is the word "he," because the key sentence is about "London," and the second sentence is about "he." Therefore, answer space A has been marked for Example V on your answer sheet (SIDE B, PART IV, Words in Sentences).

Here is another sample item:

W. Mary is cutting the APPLE.

My brother John is beating his dog with a big stick.

W. A B C D E

In the key sentence, APPLE is the name of the thing which is being cut; in the second sentence, dog is the thing which is being beaten. Therefore, answer space D has been marked for Example W on your answer sheet.

Here is one for you to try. Mark your answer in Example X on your answer sheet.

X. MONEY is his only object.

Not so many years ago, most farming was done by hand.

X. A B C D E

The right answer is farming; it performs the same function in the second sentence as MONEY does in the key sentence. Therefore you should have marked space D.

Here is still another one, except that this time the choices are found in four sentences instead of only one. Be sure to look over all choices and choose the one which functions in its sentence in the way which most closely resembles that of the capitalized word in the key sentence. Mark your answer in Example Y on your answer sheet.

Y. There was much TALK about a rebellion.

Where is John?

Y. A

There is no doubt about it.

B C

There lay the dead horse.

D

There I found my answer.

E

The right answer to sample question Y is doubt, space B.

When the examiner gives the signal turn the page and start the test. Remember, always look over all the choices to find the one which functions most nearly like the word or phrase in the key sentence.

In answering the test, use the separate answer sheet (SIDE B, PART IV, Words in Sentences). Try to answer every item; if you are not certain of the answer, give your best guess.

DO NOT TURN THE PAGE UNTIL THE SIGNAL IS GIVEN.
PART V. PAIRED ASSOCIATES

Instructions. Your task is to MEMORIZE the Kurdish-English vocabulary below. Wait for the signal, then you will be given two minutes to study the vocabulary printed below. At the end of the two minutes the examiner will give you the signal to start filling in the blanks in the lower half of the PRACTICE EXERCISE SHEET which you have placed over page 11, to the right of this page. You are allowed to look back at the vocabulary on this page when you are filling in the blanks on the practice exercise sheet. After filling in the blanks, continue studying if there is still time.

Vocabulary (Memorize for 2 minutes)

<table>
<thead>
<tr>
<th>Kurdish</th>
<th>English</th>
<th>Kurdish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>hij</td>
<td>draw</td>
<td>kete</td>
<td>camel</td>
</tr>
<tr>
<td>naq</td>
<td>that</td>
<td>chie</td>
<td>few</td>
</tr>
<tr>
<td>sidqu</td>
<td>news</td>
<td>yong</td>
<td>hawk</td>
</tr>
<tr>
<td>nente</td>
<td>lady</td>
<td>hui</td>
<td>fall</td>
</tr>
<tr>
<td>ja</td>
<td>day</td>
<td>xozo</td>
<td>easy</td>
</tr>
<tr>
<td>ngoz</td>
<td>dark</td>
<td>mep</td>
<td>on</td>
</tr>
<tr>
<td>tsep</td>
<td>enter</td>
<td>lah</td>
<td>wolf</td>
</tr>
<tr>
<td>lohong</td>
<td>ask</td>
<td>wener</td>
<td>book</td>
</tr>
<tr>
<td>mupa</td>
<td>anger</td>
<td>mi</td>
<td>touch</td>
</tr>
<tr>
<td>nung</td>
<td>frog</td>
<td>jate</td>
<td>sun</td>
</tr>
<tr>
<td>chomco</td>
<td>body</td>
<td>e</td>
<td>bowl</td>
</tr>
<tr>
<td>roo</td>
<td>art</td>
<td>hon</td>
<td>cold</td>
</tr>
</tbody>
</table>

Instructions. In the next part of the test, you will be given the Kurdish words and 5 choices in English. Following is a sample item.

hij
A. frog
B. fall
C. cold
D. draw
E. book

Since hij means draw you would place a black mark in space D, as shown in Example Z on your answer sheet.

These questions are to be done from memory.

DO NOT LOOK BACK AT THIS PAGE!
DO NOT TURN PAGE UNTIL TOLD TO DO SO!
APPENDIX H. Operation Span Task.

Instruction slides:

Slide 1:
In this experiment you will try to memorize letters you see on the screen. However, to make it more difficult, you will have to do a second task between the presentation of each letter. Specifically, you will have to solve simple math problems.
Before we begin, we’ll practice solving math problems out loud.
For example, if you saw the following:

\[ IS \ 4 + 1 \ ? \]

You would say, “Is 4 plus 1 equal to one? No.” You would say “No” because \( 4 + 1 = 5 \), not 1.
The math problems in this task will be a little more challenging. Let’s practice.

Participant is now presented with 5 practice math problems.

Slide 2:
Now in the actual task, you will see a letter appear at the end of the equation. Your job is to read the equation OUT LOUD, then verify if the answer provided is correct or not by saying “yes” or “no,” and then immediately read the letter that follows the equation OUT LOUD.
For example, if you saw the following:

\[ IS \ (2 \times 1) + 1 = 3 \ ? \ D \]

You would say, “Is 2 times 1 plus 1 equal to 3?” then you would say “yes,” because \( (2 \times 1) + 1 \) DOES equal 3, and then you would immediately say “D.”

Slide 3:
After you say the letter aloud, you will see a new equation and letter appear on the screen, for example:

\[ IS \ (3 \times 1) - 1 = 5 \ ? \ N \]

Here, you would say, out loud: “Is 3 times 1 minus 1 equal to 5?...no...N.” Here you would say “no” because \( (3 \times 1) - 1 = 2 \), not 5.
Slide 4:
After some number of these equations and letters, you will see three question marks appear in the center of the screen like this:

???

This is your cue to write down all the letters that you saw in that set, in the same order you saw them in.

So, for this example, you would write “D” in the first blank on your answer sheet, and “N” in the second blank.

Slide 6:
Let’s begin with some practice.

If you can’t think of all the letters, please leave a blank space for any letters you can’t remember. There is no penalty for guessing.

Do you have any questions?

To Begin the Practice Trials press ‘1.’
To Repeat Instructions press ‘2.’

Participant is now presented with three practice sets of stimulus items.

Slide 8:
Now the real task is going to work just like practice, but there will be a different number of equations and letters in each set. Sometimes there will be just 2 in a set, like in practice, but other times there will be 3, 4, or 5 in a set. The order of these sets is randomized, so you won’t know how big a set is until you are finished with it.

Remember, if you can’t remember a letter, please leave a blank space for it. There is no penalty for guessing.

To Begin the Actual Task, press ‘1.’
To repeat Practice, press ‘2.’

Participant is now presented with 12 sets of 2 – 5 stimulus items.

Slide 9:
Congratulations! You have completed the Operation Span Task. Please wait for instructions from the experimenter.
APPENDIX I. Digit Span Task.

*Instructions:* You will hear increasingly longer sets of numbers and your task is to repeat the numbers aloud in the same order you heard them.

*Practice for Digit-Span Task:*

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5-3</td>
<td>2-4</td>
<td>7-3-8</td>
<td>4-1-3</td>
<td></td>
</tr>
</tbody>
</table>

*Items for Digit-Span Task:*

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2)</td>
<td>4-2</td>
<td>5-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3)</td>
<td>1-8-3</td>
<td>9-4-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4)</td>
<td>7-6-9-1</td>
<td>2-8-4-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5)</td>
<td>3-5-1-8-2</td>
<td>6-9-2-4-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6)</td>
<td>8-3-5-6-2-1</td>
<td>4-8-2-5-1-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7)</td>
<td>1-5-2-8-9-4-3</td>
<td>2-6-3-9-4-1-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8)</td>
<td>6-8-3-2-5-1-9-4</td>
<td>5-7-2-9-4-3-1-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9)</td>
<td>8-2-4-1-7-3-9-6-5</td>
<td>3-9-7-2-5-8-4-1-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX J. Digit Symbol Coding Task (DSCT).

Instructions: Use the key to copy the symbols that correspond to the digits in the empty boxes. You will be given two minutes to complete as much of the grid as you can.
### DSCT Answer Grid

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2 | 1 | 3 | 7 | 2 | 4 | 8 | 2 | 1 | 3 | 2 | 1 | 4 | 2 | 3 | 5 | 2 | 3 | 1 | 4 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 5 | 6 | 3 | 1 | 4 | 1 | 5 | 4 | 2 | 7 | 6 | 3 | 5 | 7 | 2 | 8 | 5 | 4 | 6 | 3 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 7 | 2 | 8 | 1 | 9 | 5 | 8 | 4 | 7 | 3 | 6 | 2 | 5 | 1 | 9 | 2 | 8 | 3 | 7 | 4 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 6 | 5 | 9 | 4 | 8 | 3 | 7 | 2 | 6 | 1 | 5 | 4 | 6 | 3 | 7 | 9 | 2 | 8 | 1 | 7 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 9 | 4 | 6 | 8 | 5 | 9 | 7 | 1 | 8 | 5 | 2 | 9 | 4 | 8 | 6 | 3 | 7 | 9 | 8 | 6 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2 | 7 | 3 | 6 | 5 | 1 | 9 | 8 | 4 | 5 | 7 | 3 | 1 | 4 | 8 | 7 | 9 | 1 | 4 | 5 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 7 | 1 | 8 | 2 | 9 | 3 | 6 | 7 | 2 | 8 | 5 | 2 | 3 | 1 | 4 | 8 | 4 | 2 | 7 | 6 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

347
APPENDIX K. Attitudes and Motivational Test Battery (AMTB) (adapted from Yanguas, 2007).

a. General Motivation Questionnaire

Instructions: Please respond to the following statements by circling one of the following options: a) Strongly disagree, b) Moderately disagree, c) Slightly disagree, d) Neutral, e) Slightly agree, f) Moderately agree, or g) Strongly agree.

1) I really enjoy improving my Spanish.
   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

2) I tend to approach my Spanish homework in a random and unplanned manner.
   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

3) I want to improve my Spanish so much that it really becomes like English to me.
   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

4) It worries me that other students in my class seem to speak Spanish better than I do.
   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

5) I enjoy meeting and listening to people who speak other languages.
   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

6) Studying and improving my Spanish is important for me because it will allow me to better understand the Spanish-speaking culture in the U.S.
   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

   a. strongly disagree b. moderately disagree c. slightly disagree d. neutral e. slightly agree f. moderately agree g. strongly agree

8) I feel anxious if someone I don’t know asks me something in Spanish.
9) I often wish I could read newspapers and magazines in other languages than English.

10) When called upon to use my Spanish, I feel very much at ease.

11) Studying and improving my Spanish is important for me because it will allow me to feel more confident when meeting and conversing with native Spanish speakers.

12) Improving my Spanish can be important to me because I think it will be someday useful in getting a good job.

13) I really have no interest in languages.

14) I find the study of Spanish very boring.

15) The more I get to know members of the Spanish-speaking community in the U.S., the more I want to be fluent in Spanish.

16) Because of the importance of Spanish in the U.S., I think that all U.S. schools should teach Spanish.
17) I would feel comfortable speaking Spanish in an informal gathering where both English and Spanish speaking people were present.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

18) Improving my Spanish is important for me because it will increase my ability to influence others.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

19) I keep up to date with Spanish by working on it almost everyday.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

20) I really work hard to improve my Spanish.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

21) I never feel quite sure of myself when I am speaking in our Spanish class.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

22) Speaking Spanish bothers me.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

23) Spanish speakers in the U.S. are very sociable, warm-hearted and creative people.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

24) Improving my Spanish is important because it will make me appear more cultured.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
   disagree     disagree     disagree     agree     agree     agree

25) I make a point of trying to understand all the Spanish I see and hear.
   a. strongly   b. moderately   c. slightly   d. neutral   e. slightly   f. moderately   g. strongly
disagree disagree disagree agree agree agree

26) Spanish is really great.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

27) I get nervous and confused when I am speaking in my Spanish class.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

28) I have a tendency to give up when our Spanish instructor goes off on a tangent.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

29) I would feel calm and sure of myself if I had to order anything in Spanish.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

30) I can’t be bothered trying to understand the more complex aspects of Spanish.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

31) I don’t usually get anxious when I have to respond to a question in my Spanish class.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

32) It embarrasses me to volunteer answers in our Spanish class.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

33) I love improving my Spanish.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
disagree disagree disagree agree agree agree

34) By promoting Spanish, the Spanish-speaking community in the U.S. has shown that they deserve less, not more, consideration from U.S. citizens.

a. strongly b. moderately c. slightly d. neutral e. slightly f. moderately g. strongly
35) I wish I was as fluent in Spanish as in English.

36) If I stayed in another country, I would make a great effort to learn the language even though I could get along in English or Spanish.

37) I find I am losing any desire I ever had to improve my Spanish.

38) I don’t bother checking my corrected assignments in my Spanish courses.

39) I haven’t any great wish to learn any more Spanish than I already know.

40) Studying languages is not a pleasant experience.

41) The more I learn about the Spanish-speaking community in the U.S., the less I like them.

42) I feel quite relaxed to speak to anyone in Spanish.

43) Knowing English and living in the U.S., it is not important for members of the Spanish-speaking community to learn other languages.
a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
44) I hate Spanish.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
45) Spanish fervor is a real threat to U.S. unity.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
46) When I finish this course, I will give up the study of Spanish entirely because I am not interested in it.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
47) I don’t understand why other students feel nervous about using Spanish in class.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
48) I wish I could speak languages other than English perfectly.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
49) Improving my Spanish is important because it will give me the edge in competing with others.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
50) Improving my Spanish is a waste of time.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
    disagree  disagree  disagree  agree  agree  agree
51) I am sometimes afraid the other students will laugh at me when I speak in Spanish.
52) When I have a problem understanding something we are learning in Spanish class, I always ask the instructor for help.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

53) If it were up to me, I would spend all my time trying to improve my Spanish.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

54) I feel confident when asked to participate in my Spanish class.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

55) If the U.S. should lose the Spanish-speaking community, it would be a great loss.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

56) I wish I had started studying Spanish formally at an early age.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

57) I plan to improve my Spanish as much as I can.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

58) I sometimes daydream about dropping the Spanish class.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

59) I don’t pay too much attention to the feedback I receive in my Spanish class.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly

60) I feel uncomfortable speaking Spanish under any circumstances.

a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
61) I do not get anxious when I am asked for information in my Spanish class.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

62) I would like to improve my Spanish as much as possible.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

63) I would really like to learn other languages.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

64) Most members of the Spanish-speaking community are so friendly and easy to get along with that the U.S. is fortunate to have them.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

65) Students who claim they get nervous in Spanish class are just making excuses.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

66) I would rather see a film dubbed in English than see it in the original foreign language version.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

67) I would like to know more members of the Spanish-speaking community in the U.S.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

68) I would get nervous if I had to speak Spanish to someone in a store.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree

69) Improving my Spanish isn’t really an important goal in my life.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly 
   disagree  disagree  disagree  agree  agree  agree
70) I would rather spend my time on courses other than Spanish.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

71) It bothers me to speak Spanish on the telephone.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

72) Studying and improving my Spanish is important for me because it will allow me to gain more friends more easily among members of the Spanish-speaking community in the U.S.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

73) To be honest, I really have little desire to improve my Spanish.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

74) It doesn't bother me at all to use my Spanish.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

75) Most other languages besides English sound crude and harsh.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

76) Spanish speakers in the U.S. should try to maintain their cultural identity.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

77) Studying and improving my Spanish is important for me because it will allow me to participate more freely in the activities of the Spanish-speaking community in the U.S.
   a. strongly  b. moderately  c. slightly  d. neutral  e. slightly  f. moderately  g. strongly
   disagree  disagree  disagree  agree  agree  agree

78) When I am studying Spanish, I ignore distractions and stick to the job at hand.
b. Situational motivation: Evaluation of Spanish course and Spanish teacher

Instructions: The purpose of this section of the questionnaire is to determine your ideas and impressions about your Spanish course and Spanish teacher this semester which we refer to here as concepts. Your answers in this section will remain anonymous. In answering this section, you will be asked to rate these concepts on a number of scales. On the following pages, there is a concept given at the top of the page, and below that a group of scales. You are asked to rate each concept on each of the scales. Below is an example of how to use the scales.

If the word at either end of the scale very strongly describes your ideas and impressions about the concept at the top of the page, you would place your X as shown below:

Friendly _X___: _____: _____: _____: _____: _____: _____ Unfriendly

If the word at either end of the scale somewhat describes your ideas and impressions about the concept at the top of the page (but not strongly so), you would place your X as follows:

Dangerous _____: __X___: _____: _____: _____: _____: _____ Safe

If the word at either end of the scale only slightly describes your ideas and impressions about the concept at the top of the page, you would place your X as follows:

Fast _____: _____: __X___: _____: _____: _____: _____ Slow

If neither word at either end of the scale is related to your ideas and impressions about the concept, your X would be placed as follows:

Fast _____: _____: _____: __X___: _____: _____: _____ Slow

And so on… There are no right or wrong answers and again, all answers are anonymous. We want you to indicate on your own ideas and impressions. If you have any questions, please ask them now. In answering this part of the questionnaire, work quickly and don’t stop to think about each scale. It is your immediate impressions that we are most interested in. Thank you.

My Spanish Teacher

Efficient ____: ____: ____: ____: ____: ____: ____ Inefficient

Insensitive ____: ____: ____: ____: ____: ____: ____ Sensitive

Cheerful ____: ____: ____: ____: ____: ____: ____ Cheerless

Competent ____: ____: ____: ____: ____: ____: ____ Incompetent
Insincere: ______: ______: ______: ______: ______: ______: ______ Sincere

Unapproachable: ______: ______: ______: ______: ______: ______: ______ Approachable

Pleasant: ______: ______: ______: ______: ______: ______: ______ Unpleasant

Trusting: ______: ______: ______: ______: ______: ______: ______ Suspicious

Incapable: ______: ______: ______: ______: ______: ______: ______ Capable

Tedious: ______: ______: ______: ______: ______: ______: ______ Fascinating

Friendly: ______: ______: ______: ______: ______: ______: ______ Unfriendly

Exciting: ______: ______: ______: ______: ______: ______: ______ Dull

Organized: ______: ______: ______: ______: ______: ______: ______ Disorganized

Unreliable: ______: ______: ______: ______: ______: ______: ______ Reliable

Unimaginative: ______: ______: ______: ______: ______: ______: ______ Imaginative

Impatient: ______: ______: ______: ______: ______: ______: ______ Patient

Polite: ______: ______: ______: ______: ______: ______: ______ Impolite

Colorful: ______: ______: ______: ______: ______: ______: ______ Colorless

Unintelligent: ______: ______: ______: ______: ______: ______: ______ Intelligent

Good: ______: ______: ______: ______: ______: ______: ______ Bad

Industrious: ______: ______: ______: ______: ______: ______: ______ Unindustrious

Boring: ______: ______: ______: ______: ______: ______: ______ Interesting

Dependable: ______: ______: ______: ______: ______: ______: ______ Undependable

Disinterested: ______: ______: ______: ______: ______: ______: ______ Interested

Inconsiderate: ______: ______: ______: ______: ______: ______: ______ Considerate

My Spanish Course

Meaningful: ______: ______: ______: ______: ______: ______: ______ Meaningless
Enjoyable: _____: _____: _____: _____: _____: _____: _____ Unenjoyable
Monotonous: _____: _____: _____: _____: _____: _____: _____ Absorbing
Effortless: _____: _____: _____: _____: _____: _____: _____ Hard
Awful: _____: _____: _____: _____: _____: _____: _____ Nice
Interesting: _____: _____: _____: _____: _____: _____: _____ Boring
Good: _____: _____: _____: _____: _____: _____: _____ Bad
Simple: _____: _____: _____: _____: _____: _____: _____ Complicated
Disagreeable: _____: _____: _____: _____: _____: _____: _____ Agreeable
Fascinating: _____: _____: _____: _____: _____: _____: _____ Tedious
Worthless: _____: _____: _____: _____: _____: _____: _____ Valuable
Necessary: _____: _____: _____: _____: _____: _____: _____ Unnecessary
Useless: _____: _____: _____: _____: _____: _____: _____ Useful
Elementary: _____: _____: _____: _____: _____: _____: _____ Complex
Pleasurable: _____: _____: _____: _____: _____: _____: _____ Painful
Educational: _____: _____: _____: _____: _____: _____: _____ Noneducational
Unrewarding: _____: _____: _____: _____: _____: _____: _____ Rewarding
Difficult: _____: _____: _____: _____: _____: _____: _____ Easy
Satisfying: _____: _____: _____: _____: _____: _____: _____ Unsatisfying
Unimportant: _____: _____: _____: _____: _____: _____: _____ Important
Pleasant: _____: _____: _____: _____: _____: _____: _____ Unpleasant
Exciting: _____: _____: _____: _____: _____: _____: _____ Dull
Clear: _____: _____: _____: _____: _____: _____: _____ Confusing
Colorful: _____: _____: _____: _____: _____: _____: _____ Colorless
APPENDIX L. Possible Selves Questionnaire (adapted from Dörnyei, 2010).

*In this survey, I would like you to tell me how much you agree or disagree with the following statements by simply circling a number from 1 to 7. Please do not leave out any items.*

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Moderately disagree</th>
<th>Slightly disagree</th>
<th>Neutral</th>
<th>Slightly agree</th>
<th>Moderately agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

1. I can imagine myself living abroad and having a discussion in Spanish
2. I can imagine myself studying in a university where all my courses are taught in Spanish
3. Whenever I think of my future career, I imagine myself using Spanish
4. I can imagine a situation where I am speaking Spanish with foreigners
5. I can imagine myself speaking Spanish with international friends or colleagues
6. I can imagine myself living abroad and using Spanish effectively for communicating with the locals
7. I can imagine myself speaking Spanish as if I were a native speaker
8. I can imagine people mistaking me for a native speaker of Spanish
9. I imagine myself as someone who is able to speak Spanish
10. The things I want to do in my future require me to use Spanish
11. I study Spanish because close friends of mine think it is important
12. Learning Spanish is necessary because people surrounding me expect me to do so.
13. I consider learning to speak Spanish with a near native accent important because the people I respect think that I should do it
14. Studying Spanish is important to me in order to gain the approval of my peers, teachers, family, boss, etc.
15. Studying Spanish is important to me because an educated person is
supposed to be able to speak another language.

16. It will have a negative impact on my life if I don’t learn Spanish.

17. My parents believe that I must study Spanish to be an educated person.

18. I have to study Spanish because if I do not study it, I think my parents will be disappointed in me.

19. Studying Spanish is important to me because other people will respect me more if I have a knowledge of Spanish.

20. I consider learning Spanish important because the people I respect think that I should do it.

21. If I fail to learn Spanish I’ll be letting other people down.

22. I would like to take more Spanish courses at my university in the future.

23. If my Spanish teacher would give the class an optional assignment, I would certainly volunteer to do it.

24. I would like to study Spanish even if I were not required to do so.

25. I would like to spend lots of time studying Spanish.

26. I would like to concentrate on studying Spanish more than any other topic.

27. I am working hard at learning Spanish.

28. I am prepared to expend a lot of effort in learning Spanish.

29. I think that I am doing my best to learn Spanish.

30. Compared to my classmates, I think I study Spanish relatively hard.
APPENDIX M. Scoring procedure: Oral Imitation Task.

Elicited Oral Imitation Task Scoring Procedure: Accuracy

**Scoring A** (following Bowles and R. Ellis): Form based

0 Incorrect: Target structure is inaccurately repeated or not repeated

1 Correct: Correctly imitates sentence with obligatory target structure

**Scoring B** (adapted from Ortega and colleagues): Form and meaning based

0 Incorrect target structure; repeats little to no part of the sentence

1 Incorrect target structure; repeats a quarter to half of the sentence

2 Incorrect target structure; repeats over half of the sentence

3 Correct target structure; repeats up to half of the sentence

4 Correct target structure; repeats over half of the sentence

**Examples**

<table>
<thead>
<tr>
<th>Sentence Stimulus</th>
<th>Sentence Repetition</th>
<th>Score A</th>
<th>Score B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Muchos estudiantes tomaron clases de español el año pasado.</strong> (Cierto)</td>
<td>‘Muchos estudiantes tomar el clase de español la año pasado’ (Beg-3)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>(Aspect-GR)</td>
<td>‘Muchos estudiantes tomar las clases de español el año pasado’ (Int-3)</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Los exámenes finales están en febrero.</strong> (Falso)</td>
<td>‘Los exámenes finales están en febrero’ (Adv-5)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>(Ser/Estar-Events-UGR)</td>
<td>‘Los exámenes finales son en febrero’ (Int-1)</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Sarah Palin quería que muchos demócratas votaran en las últimas elecciones.</strong> (Falso)</td>
<td>‘Sarah Palin, no, um…’ (Int-5)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(Imperfect Subj-GR)</td>
<td>‘Sarah Palin quieren que muchas valen ultimas elecciones’ (Beg-4)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>‘Sarah Palin quiere que muchos demócratas voten en las últimas elecciones’ (Adv-2)</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
APPENDIX N. Project IRB Approval.

Georgetown University Institutional Review Board

Date: February 9, 2012

To: Ellen Johnson, PhD Student
1848 Wyoming Avenue NW, #103
Washington, DC 20009
esj25@georgetown.edu

From: David Blanco
Institutional Review Board

Title: Exploring the relationship between cognitive and affective individual differences and the long-term development of implicit/explicit second language (L2) knowledge in learners of varying proficiency levels

IRB#: 2012-027

Annual Approval Date: January 13, 2012
Expiration Date: January 12, 2013

Action: Approved as submitted
Expedited Initial Review
C-1 signed 12/20/11
C-3 signed 12/20/11 (category 7a, 7b)
ICF for learners
ICF for instructors
Cognitive ID measures
Johnson: CV, SSDF, Human subjects protection training

Your above-referenced protocol and consent forms were approved through expedited review by Dr. Heidi Li Feldman, the IRB Chair or a designee, on February 8, 2012.

This is to inform you that you may commence your project. Approval for this study is through January 12, 2013.

When consenting participants, please be sure to use only the most recent version of the informed consent form that is stamped by the IRB with the current approval and expiration dates.

This study will automatically become inactive when its approval expires on January 12, 2013 unless a continuing review submission for the study is received and approved by the IRB before that date. The IRB requires that you submit an application for annual renewal at the end of each approval period and/or at study completion. It is the PI’s responsibility to submit the application for annual renewal and the appropriate IRB forms at least one month before the expiration date.

Please remember to:
1. Seek and obtain prior approval for any modifications to the approved protocol. All recruitment materials or incentives for participation must be submitted for IRB review and approval prior to their use.
Bibliography


Psychological Corporation.


University Press.


IBM SPSS. New York: Routledge.


Conditions, processes, and knowledge in SLA and bilingualism (pp. 101-113). Washington, DC: Georgetown University Press.


Sagarra, N. & Abbuhl, R. (Forthcoming, 2013). Optimizing the noticing of face-to-face recasts via computer-delivered feedback: Evidence that oral input enhancement and working memory help L2 learning. _Modern Language Journal_.


VanPatten, B. (2010). Some verbs are more perfect than others: Why learners have difficulty with *ser* and *estar* and what it means for instruction. *Hispania, 93,* 29-38.


The textbooks and accompanying workbooks used at the introductory II level (SPAN 004) are *Vistazos: Un curso breve* (3rd ed.) (VanPatten, Lee, & Ballman, 2010), *¡Avance!* (2nd ed.) (Bretz, Dvorak, Kirshner, Bransdorfer, & Kihvet, 2008) at the intermediate II level (SPAN 022) and *Repase y escriba* (6th ed.) (Dominicis & Reynolds, 2010) at the advanced II level (SPAN 104).

The proficiency level of participants in Roehr & Gánem-Gutiérrez (2009) is not explicitly stated but they are likely at an upper-intermediate level given that they reported having taken four years previous L2 instruction on average.

All three composite ID measures of L2 Aptitude (MLAT-Comp) and L2 Motivation (AMTB-Comp, Selves-Comp) also significantly predicted performance on the test of implicit knowledge, but it was more of interest to ascertain the impact of individual ID components; thus, only components were entered into the combined Level-2 GCA.

As found for the test of implicit knowledge, all three composite ID measures of L2 Aptitude (MLAT-Comp) and L2 Motivation (AMTB-Comp, Selves-Comp) also significantly predicted performance on the test of explicit knowledge, but again only individual ID components were entered into the combined Level-2 GCA.

In addition to proficiency level, previous language learning experience seems to be an uncontrolled variable in R. Ellis' studies. Fifteen of the participants had studied a foreign language in addition to English, including eleven who had studied it for more than 2 years (2009, p. 41), which could have influenced performance on the measures of implicit and explicit knowledge of the target language, English.