THE RELATIVE CONTRIBUTION OF INPUT MODIFICATION, LEARNER AWARENESS, AND INDIVIDUAL DIFFERENCES TO SECOND LANGUAGE CONSTRUCTION LEARNING

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

SECOND LANGUAGE STUDIES

MAY 2014

By

Daniel O. Jackson

Dissertation Committee:

Richard Schmidt, Chairperson
John M. Norris
Luca Onnis
Lourdes Ortega
Peter Robinson
Amy J. Schafer

Keywords: Input, awareness, individual differences, second language learning, working memory, personality

ACKNOWLEDGEMENTS

I experienced a rare and ephemeral kind of happiness upon completing the final paragraph of this dissertation, which at times during the writing felt more like a raging confluence than a clear synthesis of ideas. Although the sensation of joyfulness that trickled into my consciousness then was utterly private, there were many special people behind it. I owe my satisfaction, as well as the present accomplishment, to them.

I am especially grateful to the members of my dissertation committee, whose expertise in second language acquisition and commitment to their students are remarkable. My chair, Dick Schmidt, has been a truly fantastic mentor, whose penetrating insight and enthusiasm for new ideas, I hope, are reflected in this dissertation. I am deeply thankful for his guidance. I was also extremely fortunate to have Peter Robinson join my committee. This dissertation is infused with Peter's ideas on SLA, culled from his profound writing or from beachside conversations he generously found time for. As expected, Lourdes Ortega has been a wonderful teacher and an unrivalled inspiration, whose keen advice on academic writing I will always remember. I also greatly benefitted from John Norris' coaching toward more transparent and meaningful quantitative reasoning. Luca Onnis helped substantially, from building the lab that provided a context for this study to offering thoughtful comments along the way. Luca, along with Amy Schafer, encouraged and supported the use of R software. I am further indebted to Amy for her outstanding service as university representative on the committee.

A special word of thanks is due to Min Young Cho, who assisted in coding data for this research. I must also thank those who developed and allowed the use of measures adopted in this study, including the Big Five Aspect Scales and the Automated Reading Span task. Many other scholars, as well, deserve recognition for their support. Professors Shuqiang Zhang, William O'Grady, and J. D. Brown lent me deep insights through their courses at UHM. I am also grateful to those who organized and attended related talks I gave at SLRF 2012 and CALICO 2013. Additionally, I gained a profound appreciation of Dick's contribution while working with the editors of *Noticing and second language acquisition: Studies in honor of Richard Schmidt*. Thanks Jim, Joara, and Sylvia!

The writing of this dissertation was supported by financial assistance from the Bilinski Dissertation Fellowship Fund and the *Language Learning* Dissertation Grant Program, for which I am grateful. Furthermore, I would like to thank the Linguistics Beyond the Classroom Program for assisting with participant recruitment, as well as the staff and volunteers in the Second Language Acquisition and Bilingualism Laboratory for enabling access to valuable resources. I also want to thank Emily Lee in the SLS Department, and the Center for Language and Technology staff, for clerical and technical help leading up to my dissertation defense meeting.

Many individuals provided moral support while I was working toward my doctoral degree and my sincere gratitude goes to them. I thank my fellow graduate students in the SLS Department, along with the editorial staff at *Language Learning & Technology*, and various program staff at UHM, for advice and encouragement. Lastly, I offer my greatest thanks to my family for their love and support. My parents, Dan and Lorraine, and my sister, Cara, have provided ongoing encouragement. Above all, Misa has been a constant source of happiness, from the beginning of this journey to its completion…*arigatou gozaimashita*.

ABSTRACT

This study adopts a constructionist approach to investigating three recent proposals grounded in cognitive perspectives on second language acquisition, which broaden the research agenda in the areas of input modification, learner awareness, and individual differences. Specifically, it was hypothesized that the following would support learning of artificial morphological constructions: (a) input modification in the form of partially repeated sequences, (b) awareness of precise form-meaning mappings, and (c) cognitive-affective learner differences. During the computerized training phase of the study, 90 English-dominant undergraduate participants were placed into three conditions: two experimental conditions exhibiting partial repetition similar to that found in caregiver discourse and a control condition exhibiting no such repetition. The testing phase utilized a picture-word matching task to assess participants' ability to judge correct versus incorrect constructions based on trained and untrained subtest items. Learner awareness was gauged through source attributions and a retrospective written questionnaire. Measures of learner differences included working memory tasks (reading span and dual 3-back) and personality scales (openness and intellect). A test-retest study with 20 of the participants showed these individual difference measures to be relatively stable. Results indicated that the study's hypotheses were largely supported. Learner awareness levels, reading span scores, and item type significantly predicted outcomes on the trained and untrained subtests of the picture-word matching task, based on a regression model including the random effects of items and participants. Also, a better fit to the data was achieved by using a fine-grained, rather than course-grained, system of coding awareness.

However, contrary to hypothesis, participants in the experimental conditions did not significantly outperform those in the control condition on the picture-word matching task. Another finding of this study was that participants reported use of both explicit (rule, memory) and implicit (guess, intuition) knowledge sources on the picture-word matching task. In conclusion, this study offers additional evidence concerning the roles of awareness and individual differences in second language learning, suggests ways to enhance the internal validity of research in this domain, and potentially informs future research on input modification as a pedagogic technique.

TABLE OF CONTENTS

Acknowledgements i

Abstract iii

List of Tables xii

List of Figures xiii

Chapter 1: Introduction 1

1.1 Cognitive perspectives on second language acquisition 1

1.2 Recent proposals concerning input, awareness, and ability 1

1.3 Research aims 4

1.4 Scope and potential contribution 6

1.5 The organization of this dissertation 7

Chapter 2: A rationale for variation sets in L2 learning 10

2.1 Introduction 10

2.2 Background to three experientialist approaches to input in SLA 11

2.2.1 Interactionist approaches 12

2.2.2 Cognitive linguistic approaches 16

2.2.3 Approaches based on general learning principles 19

2.3 Variation sets 21

2.3.1 Definition and overview 21

2.3.2 Additional distinctions and measurement 24

2.3.3 Research findings 26

2.3.3.1 Longitudinal studies 27

2.3.3.2 Experimental investigations 28

2.3.3.3 Corpus analyses 31

2.3.4 Summary of variation sets 33

2.4 Revisiting the three approaches: Multiple roles for variation sets? 33

2.4.1 Interactionist perspectives on variation sets 34

2.4.2 Cognitive linguistic perspectives on variation sets 41

2.4.3 General learning perspectives on variation sets 45

2.5 Potential implications 49

2.6 Summary 50

Chapter 3: Consciousness and L2 learning: Terminology, findings, and issues 51

3.1 Introduction 51

3.2 Terminology 51

3.2.1 Consciousness 52

3.2.2 Awareness 53

3.2.3 Attention 54

3.2.4 Noticing 56

3.2.5 Implicit and explicit learning and knowledge 58

3.2.6 Statistical learning 59

3.2.7 Additional terminology 61

3.3 Empirical findings concerning L2 learning and awareness 62

3.3.1 Early research and theoretical underpinnings 63

3.3.2 Comparison studies of implicit versus other conditions 64

3.3.3 Think-aloud protocols 68

3.3.4 Semantic implicit learning 69

3.3.5 Confidence ratings and source attributions 72

3.3.6 Interim summary 75

3.4 Issues in research on second language awareness 78

3.4.1 The explicit-implicit interface debate 78

3.4.2 The susceptibility of explicit versus implicit processes to IDs 81

3.4.3 Methodological issues in studying L2 awareness 84

3.4.4 Overview 87

3.5 Implications for the present study 90

Chapter 4: Working memory and L2 learning 95

4.1 Introduction 95

4.2 Current psychological perspectives on WM 95

4.2.1 Theoretical models of working memory 96

4.2.1.1 Baddeley's multicomponent model 98

4.2.1.2 Cowan's embedded-processes model 99

4.2.1.3 Engle and colleagues' resource-dependent inhibition model 99

4.2.1.4 Other perspectives on WM 100

4.2.2 General characteristics of working memory theories 101

4.2.2.1 Central executive as a key mechanism 101

4.2.2.2 Non-unitary nature of the system 102

4.2.2.3 Multiple determinants of capacity limits 103

4.2.3 WM and the lifespan 104

4.2.4 The neural correlates of WM 104

4.3 The measurement of WMC 105
4.3.1 Short-term measures 106
4.3.2 Complex measures 106
4.3.3 Dynamic measures 107
4.3.4 The criterion validity of WMC measures 107
4.3.5 Summary of WMC measures 109
4.4 WM and SLA: Conditions, processes, and outcomes 109
4.4.1 Learning conditions: Input delivery, modality, and modification 110
4.4.2 Cognitive processes: Noticing, understanding, and aptitude 114
4.4.3 Linguistic outcomes: Word meaning and morphology 117
4.5 Summary of implications for the current study 120
Chapter 5: Approaches to personality and L2 learning 123
5.1 Introduction 123
5.1.1 Eysenck's dimensions: P, E, N 123
5.1.2 The Myers-Briggs Type Inventory 124
5.1.3 The Big Five Model 126
5.2 Personality in L2 learning research 129
5.2.1 Personality traits and L2 outcomes 130
5.2.2 Potential covariates of personality in L2 learning 132
5.2.3 Research on other related learner variables 133
5.2.4 Issues in L2 research on personality 134
5.2.5 The current wave of personality research in L2 studies 135
5.3 Relationships between personality and cognitive factors 136

5.3.1 General intelligence 138
5.3.2 Fluid intelligence 138
5.3.3 Crystalized intelligence 139
5.3.4 Working memory 139
5.3.5 Speed 140
5.3.6 Summary 141
5.4 Personality and learning in experimental settings 141
5.5 Potential implications for future L2 studies 144
Chapter 6: Methodology 147
6.1 Introduction 147
6.2 Research questions and hypotheses 148
6.3 Pilot testing 151
6.4 The main study 152
6.4.1 Participant characteristics 152
6.4.2 Data collection 154
6.4.3 Materials and procedures 155
6.4.3.1 Artificial language 155
6.4.3.2 Training phase 161
6.4.3.3 Experimental conditions: Variation sets 166
6.4.3.4 Testing phase 169
6.4.3.5 Awareness measures 173
6.4.3.6 Individual difference measures 176
6.5 Analyses 180

Chapter 7: Results 182
7.1 Introduction 182
7.2 Preliminary results 182
7.2.1 Comparison across conditions on the attention-focusing tasks 182
7.2.2 Descriptive statistics and reliability for the picture-word matching task 183
7.2.3 Inter-rater reliability for coding of awareness categories 184
7.2.4 Reliability of the individual differences measures 186
7.3 Main results 188
7.3.1 Effects of the input variation conditions 188
7.3.2 Effects of awareness 192
7.3.3 Relationships between WMC and picture-word matching task scores 197
7.3.4 Relationships between personality, cognitive variables, and outcomes 203
7.3.5 Multiple predictors of scores on the picture-word matching task 206
7.3.6 Additional evidence from source attributions 211
Chapter 8: Discussion and conclusions 218
8.1 Introduction 218
8.2 Variation sets in adult L2 construction learning 218
8.3 Learner-generated awareness 221
8.4 Individual differences 224
8.4.1 The role of complex span and dynamic measures of WM 224
8.4.2 The influence of openness and intellect 227
8.5 Multiple predictors of L2 construction learning 229
8.6 Additional evidence from knowledge source attributions 233

8.7 Summary of major findings ... 235
8.8 Implications ... 237
8.9 Limitations ... 240
8.10 Future directions ... 241
APPENDIX A: Consent form ... 242
APPENDIX B: Images used during the experiment ... 244
APPENDIX C: Attention-focusing task performance ... 246
APPENDIX D: Picture-word matching task test items and statistics ... 248
APPENDIX E: Written questionnaire: Awareness, personality, and background ... 250
APPENDIX F: Histograms for the picture-word matching task ... 262
APPENDIX G: Histograms for individual difference measures ... 264
APPENDIX H: Inter-correlations between dependent and independent measures ... 268
APPENDIX I: Source attribution data ... 269
References ... 271

LIST OF TABLES

Table Page

2.1. Descriptive Features of Variation Sets 22

3.1. Overview of Studies on Implicit Second Language Learning (1986-2012) 77

3.2. Characterizations of L2 Learning in Terms of Cognitive Processes Invoked 91

4.1. Studies Reporting Significant Relationships between WMC and L2 Abilities 113

5.1. Domains and Facets in the Five-Factor Model of Personality 127

6.1. Demographic Characteristics of Participants ($N = 90$) 153

6.2. Lexical and Behavioral Characteristics of Nonwords in the Two Noun Classes 157

6.3. Inflectional Markers Presented in the Artificial Input 160

6.4. Partial Repetition in the Three Conditions and Percentage of Total Trials 168

6.5. Artificial Word Forms Used in the Experiment 172

7.1. Descriptive Statistics for the Picture-Word Matching Task 183

7.2. Descriptive Statistics for Individual Difference Measures 186

7.3. Summary of Reliability Indices (Individual Differences Measures) 188

7.4. Responses to the Trained and Untrained Subtests by Condition 190

7.5. Mixed-effect Regression Models Predicting Scores from Condition 191

7.6. Mixed-effect Models Predicting Scores from Awareness of Relevance 194

7.7. Mixed-effect Models Predicting Scores from Awareness of Mappings 196

7.8. Collinearity Diagnostics for Predictors Entered into the Models (VIFs) 207

7.9. Mixed-effect Regression Models Predicting Scores from Multiple Variables 210

LIST OF FIGURES

Figure Page

3.1. Measures adopted in L2 awareness studies ... 89

4.1. Three working memory model visualizations ... 97

6.1. A training trial from the experiment ... 162

6.2. A meaning-focused trial from the experiment ... 163

6.3. A form-focused trial from the experiment ... 164

6.4. A picture-word matching task trial from the experiment ... 170

6.5. Trial sequence in the dual 3-back task ... 179

6.6. Study design ... 180

7.1. Percentage of correct responses by condition ... 190

7.2. Mean scores grouped by awareness group: Course-grained coding ... 193

7.3. Mean scores grouped by awareness group: Fine-grained coding ... 195

7.4. Scatterplots: Picture-word matching (PWM) task scores and WM task scores ... 198

7.5. Scatterplots: PWM task scores and dual 3-back scores by condition ... 200

7.6. Scatterplots: PWM task scores and reading span task scores by condition ... 201

7.7. Scatterplots: PWM task scores and openness scores ... 203

7.8. Scatterplots: WM task scores and intellect scores ... 204

7.9. Scatterplot: Attention to form scores and intellect scores ... 205

7.10. Source attributions used by all participants ... 213

7.11. Source attributions used by number-high participants ... 215

7.12. Source attributions used by both-high participants ... 216

CHAPTER 1

INTRODUCTION

1.1 Cognitive perspectives on second language acquisition

Cognitive perspectives on second language acquisition (SLA) have expanded recently, as witnessed by numerous recent volumes (e.g., Bergsleithner, Frota, & Yoshioka, 2013; Dörnyei, 2009a; R. Ellis, Loewen, Elder, Erlam, Philp, & Reinders, 2009; Granena & Long, 2013; Rebuschat & Williams, 2012a; Robinson & N. Ellis, 2008; Sanz & Leow, 2011; Tyler, 2012), although earlier collections had already established a solid foundation for such work (e.g., Ellis, 1994; Schmidt, 1995a; Robinson, 2001a, 2002a). Some major current themes within this research domain include the adoption of novel theoretical approaches to language learning and use, particularly usage-based and cognitive linguistic accounts, the elaboration of psychological constructs, such as input, awareness, and ability (see below), the utilization of innovative research methodologies offering a window on the learner's mind, and the development of practical applications of this work in second language (L2) teaching and testing. Research along these lines is essential, as it holds promise for a better understanding of how second languages are learned. Against a backdrop of robust research activity, several recent proposals in need of further empirical investigation have emerged.

1.2 Recent proposals concerning input, awareness, and ability

First, work on adult statistical learning points to the conclusion that variation sets, or contiguous utterances consisting of partial repetition and conveying an approximately uniform intention, can facilitate the discovery of language structure (Onnis, Waterfall, & Edelman, 2008). This finding is important because it suggests practical applications in instructed SLA (Onnis, 2012). Indeed, an earlier body of work in second language pedagogy, reviewed in the following chapter, similarly advocated the use of input modifications based on partial repetition. Yet, as discussed later, recent cognitive approaches have made finer distinctions, and pursued a wider array of goals, based on the notion of variation sets, or linguistic units exhibiting partial self-repetition (Küntay & Slobin, 1996; Onnis, Waterfall, & Edelman, 2008; Waterfall, Sandbank, Onnis, & Edelman, 2010). Related definitions and examples will be given in the next chapter to illustrate applications of the construct of variation sets to the L2 literature.

Second, the longstanding issue of the role of awareness in SLA is often defined in terms of whether or not implicit learning, that is, learning in the absence of awareness, is possible (see, e.g., DeKeyser, 1994, 1995, 2003, 2009; N. Ellis, 1993, 1994, 2005, 2011; Hulstijn, 1989, 2005; Krashen, 1982; Robinson, 1994, 1995a, 1995b, 1996, 1997a, 1997b, 2003, 2005b, 2007a, 2010; Schmidt, 1990, 1994a, 1994b, 1995, 2001; Schmidt & Frota, 1986; Williams, 2005, 2009, 2013). At times, this question has been plagued by terminological confusion; most problematic when the conditions under which learning occurs, rather than the cognitive processes they invoke, are labeled as implicit, or when implicit *learning* is used as a surrogate term for implicit *knowledge*. These and other matters are reviewed in detail in Chapter 3. In short, recent proposals calling for methodological sophistication in research on learner awareness consider an assortment of techniques that may shed light on the issues (Leow, Johnson, & Záráte-Sandez, 2011; Leow & Hama, 2013; Rebuschat, 2013; Robinson, Mackey, Gass, & Schmidt, 2012).

Third, learners possess a wide range of individual abilities and traits, which, in general, influence adult second language (L2) learning (Dörnyei, 2005, 2009a). Relating this to awareness, as Schmidt has recently claimed, “individual differences are an important part of the story, and both inclinations and abilities affect who notices what” (2012, p. 44). Recent theoretical frameworks, despite varying in detail, aim to account for the complexity, interactivity, and dynamicity of factors influencing SLA as it unfolds across time and contexts, offering a revised view of individual differences (Dörnyei, 2009a, 2009b; Dörnyei & Skehan, 2003; Robinson, 2001b, 2002a, 2005a, 2007b). These situated accounts of learner abilities also offer exciting new possibilities for SLA research.

Shortcomings within each of these three areas are readily identifiable. For instance, the findings concerning variation sets noted above have not yet been extended to a wide range of linguistic structures, including those involving meaning. Also, while methods for gauging learner awareness are abundant, methodological triangulation in such studies is less common. Many empirical reports on individual differences (IDs) can be seen as one-sided for infrequently having considered how links between cognitive, affective, and conative variables all theoretically contribute to learning outcomes. These are a few of the gaps that the present study seeks to begin to address, although it will take a long-term research agenda to deal with them more comprehensively.

Another noteworthy feature of the present study is its assumption that the relative contribution of input variation, learner awareness, and individual differences is ideally suited to investigation within a constructionist approach. The idea of constructions as pairings of form and meaning that can be found at all levels of language has guided several programs of research within contemporary linguistic theory (see Croft & Cruse

2004; Bybee, 2010; Goldberg, 2006; Tomasello, 2003, for book-length overviews). For example, Croft (2001) described several types of constructions situated along the syntax-lexicon continuum, including complex but bound morphological constructions (e.g., [noun-s]). This study employed a miniature artificial language consisting of such morphological constructions to investigate adult L2 learning. Here, as in most discussions, the label construction is used synonymously with form-meaning mapping and form-function mapping. This approach aligns the present study with cognitive linguistic theories offering insight into how mental processes that are not exclusive to language nonetheless support its development (elaborated on in Chapter 2) as well as contemporary techniques in communicative language teaching, such as focus on form (e.g., Doughty & Williams, 1998), which take form-meaning mappings to be the target of L2 instruction. Thus, the view that L2 learners acquire constructions establishes theoretical coherence across the three distinct proposals examined here by situating results within an already well-articulated perspective on language.

1.3 Research aims

Based on the need for further research into the three proposals just outlined, this dissertation aimed to investigate the effects of variation sets in adult L2 learning, while addressing the need for further research on the kinds of linguistic constructions which may be learned without awareness, and also contributing to an understanding of the potential role of individual differences in such learning. To achieve these aims, 90 English-dominant participants from the undergraduate population at the University of Hawaiʻi at Mānoa were recruited to take part in a study carried out in the Second

Language Acquisition and Bilingualism Laboratory. The study employed a between-groups design involving three conditions administered by computer. These conditions presented learners with a miniature artificial language in which nonwords accompanying images of animals or objects were inflected for two values of the feature number (dual or plural) and two values of one additional feature (i.e., either animate or inanimate), yielding a total of four morphemes. The semantic components of number and animacy were thus conflated (see Talmy, 2000) on these morphemes (*-mi, -ku, -ba,* and *-zo*). Singular forms were bare. Type and token frequency remained constant across the three conditions. A control (scrambled) condition was employed which did not contain input variation, while the presence of variation sets in the two experimental conditions was manipulated based on previous studies (Onnis, Waterfall, & Edelman, 2008; Onnis, Edelman, & Waterfall, 2011). The difference between these two experimental conditions, which was unique to this study, was that they either repeated morphological markers and varied lexical stems (lexical variation) or repeated lexical stems and varied morphological markers (morpheme variation), across successive trials. During the training phase of the experiment, participants were instructed to read the words aloud and to look carefully at the pictures. In addition, all groups were required to respond to attention-focusing trials, which prompted them to type either: (a) the word appearing on the screen (copy trials), or (b) the English name for the image presented (meaning trials). These attention-focusing trials were intended to correspond to the distinction in L2 teaching and research between focus on forms and focus on meaning (Long & Robinson, 1998). During the testing phase, participants performed a picture-word matching task (see Jiang, 2012) to assess their ability to judge correct versus incorrect exemplars based on trained and untrained subtest items. This task also incorporated a trial-by-trial measure of awareness by asking participants to indicate the knowledge source to which they attributed their responses, from the following choices: guess, intuition, rule, or memory. The experiment was followed by retrospective questionnaire designed to tap participants' awareness (at the level of understanding) of the meanings indicated by each morpheme presented in training. ID measures included a set of questionnaire items pertaining to openness/intellect from the Big Five Aspect Scales (DeYoung, Quilty, Peterson, 2007), a dual 3-back task (e.g., Gray, Chabris, & Braver, 2003; Jaeggi, Buschkuehl, Perrig, & Meier, 2010), and an automated version of the reading span task (Unsworth, Heitz, Schrock, & Engle, 2005).

1.4 Scope and potential contribution

The present study seeks to contribute significantly to understanding adult SLA in several ways. First, it addresses the effects of variation sets in input presented to adult L2 learners. As argued later, this construct has potential applications that range from providing focused units of analysis for observational studies to offering techniques for enhancing L2 instructional design. Second, while researchers have concluded that implicit learning of certain form-meaning mappings appears to be possible (Schmidt, 2012; Williams, 2009), it remains to be seen whether learners can acquire generalizable knowledge of bound morphemes without awareness in the demanding, yet realistic, context of learning a novel lexicon. In this sense, this study broadens the methodological and linguistic scope of research on aware versus unaware L2 learning. Third, regarding IDs, a better understanding of how WMC and personality contribute to SLA may assist researchers seeking to explain how individual factors influence learning. Language teacher trainers may refer to such results when discussing the learner's role with teachers-in-training. L2 curriculum developers may also regard this knowledge as useful when considering how instruction can be specifically tailored to a range of learner types.

To fully appreciate how input modification might assist learners who show varying levels of awareness and ability, one must try to understand the relative effects of each. Moving toward multi-causal accounts of early L2 learning, it appears that various complex environmental, cognitive, and affective systems interact dynamically to facilitate learning (e.g., Beckner et al., 2009; Dörnyei, 2009a; Ellis & Larsen-Freeman, 2006). Fortunately, research tools to more appropriately consider these various effects, such as mixed-effects regression models, are now beginning to be used in SLA (e.g., de Zeeuw, Verhoeven, & Schreuder, 2012; Blom, Paradis, & Sorenson Duncan, 2012; Sonbul & Schmitt, 2013; for discussion, see Cunnings, 2012).

1.5 The organization of this dissertation

This dissertation has the following organization. The present chapter has given an overview of the territory, described the central focus of the study, indicated the research aims and how they were targeted, and stated the potential contribution to the L2 literature. The following chapters review the literature to provide a thorough background to the topics investigated herein.

Chapter 2 offers a rationale for examining the role of variation sets in L2 learning. The role of input is examined from interactionist, constructionist, and general learning principles, then, variation sets are defined in detail and discussed with reference to previous research, including longitudinal, corpus-based, and experimental studies. I then revisit the three approaches outlined at the start of the chapter and seek to identify ways in which a focus on variation sets might extend or expand each. Ultimately, I view variation sets as a multifaceted unit of analysis, independent of any given theory, but having potential implications for several approaches to understanding L2 input.

Chapter 3 explores the topic of L2 consciousness, starting with a review of terminology in this area, including coverage of the agreed upon (and not so agreed upon) meanings of consciousness, awareness, attention, and noticing in L2 research. I then discuss types of learning—including implicit, explicit, and statistical learning—distinguishing among these types. The chapter reviews a strand of L2 research that has concerned itself with carefully evaluating learner awareness, which is often thought to be necessary for adult L2 learning. It closes with a discussion of several issues in this area: the explicit-implicit interface debate, the interrelationship between IDs and awareness, and the measurement of conscious L2 processes.

In Chapter 4, the topic of WM is introduced and treated in depth. This chapter opens with a survey of current psychological perspectives on WM, including general characteristics, specific theoretical models, and measurement approaches. This sets the stage for a presentation of WM research in SLA, including the contexts, processes, and outcomes WM has been considered in relation to. Implications for the present study are discussed in closing.

Next, Chapter 5 is an introduction to personality, in psychology and in L2 research. There is a wealth of research on personality in the cognitive sciences, and this chapter draws on some of it, in order to identify personality variables that might profitably be examined in the context of the present study.

Based on this review of the literature, Chapter 6 begins by stating the research questions and hypotheses. Chapter 6 also provides details of the methodology of the present study. These details concern pilot testing, participant characteristics, data collection, materials and procedures, concluding with a sketch of the analyses used.

In Chapter 7, study results and analyses based on mixed-effect regression models are reported. Preliminary findings are stated, and then results for the following areas are given: (a) effects of input modification; (b) effects of awareness; (c) relationships between working memory and outcomes; (d) relationships between personality and outcomes; (e) overall predictors of learning, and lastly, (f) participants' use of implicit versus explicit source attributions.

Chapter 8 discusses the results and offers conclusions based on them. It revisits the research questions, restates the main findings, and comments on each of these in light of the study's hypotheses. Here, the results are also evaluated and elaborated on with respect to the literature reviewed. This chapter then summarizes the main findings, describes theoretical, methodological, and practical implications, and specifies future directions for research on the roles of input, awareness, and ability in L2 construction learning. Appendices and references appear last.

CHAPTER 2

A RATIONALE FOR VARIATION SETS IN ADULT L2 LEARNING

2.1 Introduction

This chapter provides an overview of the treatment employed in this study, in the form of a three-pronged rationale for using variation sets to enhance input provided to L2 learners. To begin, the chapter will present some historical background to the construct of input in SLA, in order to properly situate the notion of variation sets within this literature. There is room within three interrelated theoretical accounts of input for work on such units. The first of these is the well-established body of literature on learning second languages through interaction. The second is the growing literature that views first and second languages as comprised wholly of constructions, or form-meaning mappings. Finally, the literature on general learning principles, as applied to L2 learning and use, presents an explicit rationale. Then, variation sets (i.e., contiguous utterances consisting of partial repetition and expressing a roughly uniform intention) will be described, readers will be acquainted with relevant terminology, and the findings of recent research in the child language acquisition and artificial language learning literatures will be reviewed. Afterwards, the chapter will address problems of applying variation sets in adult L2 learning, by discussing their potential contribution to the three accounts (i.e., interactionist, constructionist, and general learning). Following that, some implications of these ideas for theory building, research methodology, and language pedagogy will be stated. The chapter will close with a summary of the case for variation sets in adult second language learning.

2.2 Background to three experientialist approaches to input in SLA[1]

The role of linguistic experience, while by no means the only factor underlying second language learning, has been given special consideration by theorists since the field's origin. Early scientific approaches to second language learning built, in part, on the observation that children typically learn the language of their environment to assert that, "all language learning occurs through experience" (Lado, 1964, p. 39). This fundamental assertion has been mirrored in numerous approaches to second language teaching and learning over the past several decades, during which it has also been repeatedly observed that the environment on its own does not suffice. Rather, it operates in tandem with psycholinguistic variables, as noted by Corder (1967):

> The simple fact of presenting a certain linguistic form to a learner in the classroom does not necessarily qualify it for the status of input, for the reason that input is 'what goes in' not what is *available* for going in, and we may reasonably suppose that it is the learner who controls this input, or more properly his [*sic*] intake. (p. 165)

In addition, an important paper by Wagner-Gough and Hatch (1975) emphasized the necessity of analyses of form and meaning in understanding the role of input data in L2 conversation, foreshadowing much of the research described later in this chapter. In the following sections, three major approaches focusing on L2 learners' experience within

[1] These three approaches roughly map on to two of the three orientations to L2 learning outlined in Norris and Ortega (2003), which included interactionist, generativist, and emergentist. For the time being, I leave aside the issue of variation sets within generativist approaches, because here it is the nature of input, rather than the innateness of language, that is of primary concern. The present chapter thus draws on three recent accounts of L2 learning, all of which share the assumption that learning derives from experience.

the linguistic environment are outlined in terms of their theoretical contributions, in order to provide some brief history relevant to this chapter.

2.2.1 Interactionist approaches

The interactionist approaches described here primarily concern the negotiation of meaning (i.e., repair moves aimed at resolving communication breakdowns) between interlocutors, which occurs in communicative exchanges involving L1 and L2 users, as well as in learner-learner interaction, both outside and inside of instructional settings. The goal of such interaction is to achieve mutual comprehension. This primary emphasis has grown to include studies on corrective feedback delivered during communicative interaction (e.g., recasts) and also communicatively oriented L2 instruction carried out via the use of technology (e.g., text-based computer-mediated communication). Related studies have also been done which focus on preemptively modifying discourse for the sake of learner comprehension (see discussion below). Interactionist approaches constitute a central area of theoretical development in SLA (e.g., Long, 1996; Pica, 1994; Swain, 1995) that has been fuelled by much observational and experimental research, conducted in classroom and laboratory settings (see Plonsky & Gass, 2011 for a recent review of 174 studies). A brief overview of this line of research follows.

Beginning in the 1980s, Krashen formulated his well-known input hypothesis, which argued that acquisition[2] occurs only when a learner understands language at a stage slightly beyond his or her current degree of competence ($i + 1$). This is made possible by

---

[2] Krashen's monitor model regarded *acquisition* as distinct from *learning* in that the former is similar to child language development, while the latter involves accruing knowledge *about* language.

assuming that “we use more than our linguistic competence to help us understand” (Krashen, 1982, p. 21). That is, contextual support facilitates understanding of language structure beyond what is already known. By focusing on additional extra-linguistic cues to meaning, unfamiliar input can be acquired.

Following this, Long (1983a) assumed that L2 learner comprehension is established by means of two kinds of input adjustments: linguistic and conversational. Larsen-Freeman and Long (1991, pp. 116-128) provided a helpful overview of both types of adjustments in speech directed to L2 learners. Linguistic adjustments include grammatical or ungrammatical foreigner talk (i.e., speech adjustments when addressing L2 users). These authors offered the following as an example of speech departing from the linguistic conventions ordinarily found in discourse:

(1) NS: …uh you’re from Kyoto, right?

NNS: Yeah.

Yeah. What does your father do in Kyoto?

(Larsen-Freeman & Long, 1991, p. 117)

In this example, the usage is marked because anaphoric reference, or *there* in place of *Kyoto*, would normally be preferred in the second question. In contrast to linguistic adjustments, which are based on the modification of phonological, morphological, syntactic, or semantic forms, conversational adjustments modify the interactional structure of communication as it occurs. As noted by Larsen-Freeman and Long, this may include, among other features, comprehension checks, confirmation checks, clarification requests, and repetition (including self-, other-, exact, semantic, complete, or partial repetition). The following example illustrates exact and semantic repetition:

(2) NS: Do you like California?

NNS: Huh?

Do you like Los Angeles?

Uhm…

Do you like California?

Oh! Yeah, I like it.

(Larsen-Freeman & Long, 1991, p. 120-121)

Long argued that such adjustments might serve a crucial function, in that, if adjustments facilitate comprehension, and comprehension facilitates acquisition (as Krashen claims), then it can be deduced that adjustments facilitate acquisition (1983a). In the same year, Long (1983b) further emphasized that input comprehensibility often appeared to be enhanced by *conversational* adjustments, describing a number of interactional devices used to achieve this end.

Although these premises led to evidence that conversational interaction (also referred to as negotiated interaction) benefits comprehension and, furthermore, can also subsequently foster grammatical development (e.g., Mackey, 1999), the question of whether 'what goes in' fosters acquisition remained elusive, hence the need to stipulate that, while input is necessary, it is insufficient to guarantee L2 development. This fact was clear from investigations documenting the restricted development of certain linguistic features despite of a wealth of input and interaction, such as Schmidt's (1983) longitudinal study of Wes, an L1 Japanese speaker residing in the United States. Furthermore, the results of Schmidt's diary study of his own learning of Portuguese yielded the observation that noticing a gap between i and $i + 1$ may benefit acquisition, but only in cases where noticing is a conscious process (Schmidt & Frota, 1986, p. 311). This substantially altered Krashen's input hypothesis, in which acquisition was posited to be an unconscious process. Ultimately, this departure from Krashen led to the hypothesis that noticing, or focal awareness, is necessary and sufficient for converting input to intake (Schmidt, 1990, 1993, 1994).

Most interactionist SLA researchers have adopted this viewpoint. For instance, Pica (1994) included attention to form as one of three learner-oriented conditions for SLA, alongside meaning comprehension and modified output. Long's (1996) subsequent reformulation of the interaction hypothesis claimed that negotiation leading to interactional modifications drives learning, "because it connects input, internal learner capacities, in particular selective attention, and output in productive ways" (p. 452). Shortly afterwards, Gass (1997) diagrammed the stages from input to intake in her model of second language acquisition, incorporating a range of learner internal factors believed to influence processing, such as language universals and prior linguistic knowledge. This concern for how learners subjectively experience the input—in psycholinguistic terms—has contributed to a broadening of the interactionist SLA agenda and has greatly expanded its empirical base, leading to more robust evidence for the second part of Long's syllogism, that comprehension benefits acquisition (see recent papers by Keck et al., 2006; Sauro & Smith, 2010; Plonsky & Gass, 2011, and volumes edited by R. Ellis, 1999; Mackey, 2007; Mackey & Polio, 2009; Robinson, 2011).

2.2.2 Cognitive linguistic approaches

Taking again the 1980s as a starting point, applied linguistics research has generated insights that are comparable to modern cognitive linguistic approaches. One example is Pawley and Syder's (1983) seminal paper on nativelike selection and nativelike fluency as linguistic capacities. Briefly, usage consists of countless form-meaning pairings that, although idiomatic, are not idioms in any strict sense (e.g., *it's a quarter to ten* rather than *it's three quarters past nine*). Furthermore, speech production is extremely demanding, yet shows a high success rate.[3] Pawley and Syder explained that possession of a large inventory of memorized units, reflecting experience within a given speech community, allows language users to solve the twin puzzles of selecting appropriate units and of generating fluent speech. In this account, as in much instructed SLA research throughout the 1990s (for review, see Doughty & Williams, 1998; Norris & Ortega, 2000), form-meaning mappings were the units to be acquired and usage was the primary means of doing so. Indeed, much literature on communicative language teaching appears to be tacitly in agreement with the tenets of usage-based linguistics. While varying in specific details, usage-based approaches to language tend to assume that: (a) language structure and use are related; (b) frequency drives learning; (c) processing (i.e., comprehension and production) constitutes linguistic ability, rather than the notion of competence (Chomsky, 1965); (d) processing further shapes acquisition in the individual; (e) language representations emerge from networks of neural activation; (f) usage data, such as that available from language corpora, grounds theory; (g) usage,

[3] For instance, Levelt (1989) notes that lexical retrieval (from a productive vocabulary of an estimated 30,000 words) is continuously and indefinitely maintained at a rate of two to five times per second during fluent speech (p. 199).

synchronic change, and diachronic change are intimately related; (h) linguistic and non-linguistic systems are interconnected; and (i) both linguistic and non-linguistic contextual factors influence linguistic systems (Kemmer & Barlow, 2000).

In this same vein, cognitive linguistics has developed as a robust area of contemporary linguistic inquiry. These approaches uniformly share the assumption that constructions, defined in the previous chapter as paired mappings of form and meaning, underlie language acquisition and use. It can further be noted that constructions occur at various levels of linguistic analysis, including, but not limited to, morphemes, words, complex words, idioms, argument structure, and syntactic structure (see Goldberg, 2006, for examples). Croft and Cruse (2004) and Goldberg (2006) offered detailed reviews of four major cognitive linguistic approaches, including Construction Grammar (e.g., Fillmore, 2003), Cognitive Grammar (e.g., Langacker, 2008), Radical Construction Grammar (e.g., Croft, 2001), and Cognitive Construction Grammar (e.g., Lakoff, 1987). Several key implications for SLA following from these approaches are that: (a) the construction is the basic unit of analysis, (b) form-meaning mappings may not be equivalent across languages (e.g., passive structures in various languages reveal unique meaning mappings which only partially overlap; see Croft, 2001), and (c) the concept of motivation helps to explain connections between structurally similar items that appear to be unrelated (e.g., using plural *–s* on singular forms such as *pants*, *shorts*, *stockings* but not on *skirt* is motivated by the bipartite structure of these referents; Goldberg, 2006; Langacker, 1987). The alternative presented, then, “is to look at linguistic competence, not in terms of possession of a formal grammar of semantically empty rules, but rather in terms of the mastery of a structured inventory of meaningful linguistic constructions" (Tomasello, 2003, p. 99).

Cognitive linguistics has recently influenced language pedagogy and SLA theory. Applications to teaching phrasal verbs (Dirven, 2001) and idioms (Kövecses, 2001), among other areas, appeared in a volume edited by Pütz, Niemeier, and Dirven (2001). Soon after, Achard and Niemeier (2004) assembled a variety of perspectives on L2 pedagogy situated in cognitive linguistic theories, including studies of motion events (Cadierno, 2004) and prepositions (Tyler & Evans, 2004). The volume edited by Robinson and Ellis (2008) brought together chapters by cognitive linguists and second language researchers working from such perspectives. Robinson and Ellis concluded by identifying numerous avenues for future research, which illustrate the nexus between cognitive linguistics and SLA. Moreover, Littlemore and Juchem-Grundmann (2010) commented that although interactionist and cognitive linguistics approaches are similar in their focus on learning through communicative use, what the latter contributes is, "a detailed description of the cognitive processes that are at work in language and thought enabling people to extract language knowledge from language use" (p. 1). Also recently, Tyler (2012) summarized many of these developments and reviewed evidence from effects-of-instruction studies grounded in cognitive linguistics. She grouped this research according to five key processes in cognitive linguistics: construal, conceptual metaphor, category formation, embodiment, and usage.

2.2.3 Approaches based on general learning principles

Slobin (1973) represents one early systematic attempt to derive, based on research into developmental psycholinguistics, a list of "universal operating principles which every child brings to bear on the problem of language acquisition" (p. 197). Examples of Slobin's principles include "pay attention to the ends of words" (p. 191) and "underlying semantic relations should be marked overtly and clearly" (p. 202). Using an artificial learning paradigm, MacWhinney (1983) found clear evidence for Slobin's principles in child, but not adult, language production. MacWhinney thus raised the question of whether such general learning principles may be overridden by other cognitive skills acquired in adulthood. For similar reasons, L2 researchers have often expanded on Slobin's principles in cognitive accounts of SLA (see, e.g., Andersen, 1989; Doughty, 2001 and also Peters, 1983, for elaboration in the case of L1 acquisition).

Essentially, two goals drive the search for general learning principles: descriptive adequacy and theoretical parsimony. A few recent examples follow. First, Tomasello (2003) pared down the learning mechanisms responsible for child language acquisition to two sets of skills: pattern-finding and intention-reading. In similar fashion, Bybee and McClelland (2005) put forth two principles, asserting that: (a) sensitivity to general and specific information influences knowledge representations, and (b) repeated practice leads to rapid performance that is not discrete, but integrated.[4] Lastly, Goldstein et al. (2010) used the acronym ACCESS to denote the following cognitive processes: Align Candidates, Compare, Evaluate Statistical/Social Significance, which they consider general principles for learning structure in real time. Each of these proposals described

---

[4] See Bybee (2010) for a longer list of domain-general processes supporting a view of language learning as the development of a complex system.

economical and sufficient principles, taking as their point of departure the computational level of analysis (Marr, 1982).

As before, ideas such as these have influenced perspectives on SLA. Indeed, a large constellation of approaches to adult second language learning has placed the explanatory burden on general principles. Among early proposals of this nature was Bley-Vroman's (1989, 1990, see also 2009) Fundamental Difference Hypothesis, according to which L1 knowledge and general problem solving shape adult L2 learning. N. Ellis (1998) noted similarities between functionalist, emergentist, and constructionist approaches, suggesting that, in each case, L2 outcomes arise from "simple learning processes applied, over extended periods of practice in the learner's lifespan, to the rich and complex problem-space of language evidence" (p. 641). Specifically, a wide range of basic processes that either facilitate or inhibit discovery, storage, and access in the case of SLA has been the focus of theoretical accounts in the past few decades. While space does not permit discussion of all terms and concepts, the following selective list gives a general impression of the kinds of processes considered relevant to initial L2 learning:

- Attention and noticing (Schmidt, 1990, 2001, 2012; Robinson, 1995b, 2003).
- Contingency learning (N. Ellis, 2006a), which includes statistical learning and associative learning.
- Learned attention (N. Ellis, 2006b), which covers: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning.

Other such processes, as well, have been proposed to account for the development of L2 fluency (e.g., automatization, see DeKeyser, 1997, 2001, 2007) and L2 proficiency (e.g., analysis/control, see Bialystok, 1994, 2011). What these processes have in common with each other, and with the aforementioned principles, is that they are held to be universal to all languages, learners, and contexts. Adding to these, in a recent chapter, Onnis (2012) drew on his research into statistical learning to propose and illustrate four general principles for adult L2 learning: (a) integrate information sources, (b) seek invariant structure, (c) reuse learning mechanisms, and (d) learn to predict (p. 204). Work to understand how these principles operate in SLA, and how they interact with input factors and learner variables, is ongoing.

2.3 Variation sets

This section describes variation sets, a feature of caregiver discourse that motivates the experimental treatments deployed in the current study. It will define and exemplify the construct, clarify distinctions made in the research, and review empirical findings from longitudinal, experimental, and corpus-based studies, closing with a summary. The goal of this section is to set the stage for a critical consideration of the role of these units in learning from L2 input. Then, in the following section, relationships between variation sets and the three theoretical approaches outlined above will be sketched.

2.3.1 Definition and overview

This section provides a definition of the concept of variation sets, including examples, based on work by Küntay and Slobin, who defined them, in the context of child-directed speech (CDS). To offer readers some idea of the scope of research related to the identification and measurement of this construct, Table 2.1 lists features used to describe variation sets, positioning formal dimensions alongside their functional aspects. Each of these features will be explained in detail throughout the following sections.

Table 2.1. *Descriptive Features of Variation Sets*

Formal features	Functional features
Lexical substitutions	Controlling listener (Directive)
Rephrasings	Providing information
Addition/Deletion of specific referential elements	Requesting information
Reordering	Mixed
Multiple utterances in one turn	
Utterances related to same event	
Close distance between paired utterances	
Similar AS across paired utterances	
Different syntactic forms	
Shared lexical content	
Length	
Frequency	
Proportion of entire corpus	
Simple (adjacent)	
Interleaved (non-adjacent)	
Diameter (Levenshtein distance)	
Prevalence	
Informativeness	
Statistical probability of alignment in corpus	
Shared lexical items (*n*-gram frequency)	
Types of grammatical constructions embedded	

To begin with, Küntay and Slobin (1996) gave the following English example (p. 267):

(3) Who did we see when we went out shopping today?

Who did we see?

Who did we see in the store?

Who did we see today?

When we went out shopping, who did we see?

Küntay and Slobin defined variation sets as series of utterances in child-directed speech that have a constant intention. They may exhibit several features, including lexical substitution or rephrasing, addition/deletion of specific reference, and reordering (each present in the above example). Analyzing data from Turkish child-directed speech, Küntay and Slobin (1996) illustrated the implications of linguistic forms occurring in variation sets. For instance, they explained how the ground-figure relationship between verb stems and suffixes becomes salient though morphological substitutions in the following example (p. 276):

(4) ***Çıkart-tı-m*** *benimkinin çekirdeğini.* [removes pit from fruit]

‘**I.removed** the pit from mine.’

Sen de mi ***çıkart-ıcan***? [child nods]

‘Will.you.remove too?’

Çıkart *bakim* [child removes pit]

‘**Remove** (it), let’s see.’

Immh! Aferin yavrum! Sen de çekirdeğini ***çıkart-tın****.*

‘Mm-hm. Good for you! **You.removed** your pit too.’

İkimiz de çekirdeğini ***çıkart-tık****.*

‘Both of us **we.removed** the pit.’

Here, the interactional context yields a number of contrasts important to mastering verbal inflection. As Küntay and Slobin (2002) later summarized:

> Thus, cues for grammatical analysis are not encountered directly by the child, but come in a discourse-mediated package. It is highly plausible that the child is working on understanding the structure of interpersonal action. Simultaneously, as part and parcel of attempting to achieve social coordination, the child conducts the task of deciphering the linguistic structure. (p. 9)

Observations of children's own speech, as well, suggest that they may process language this way, as in the case of Nora's discovery of interchangeable constituents using the principle, "look for recurring parts in the formulas you know" (Wong Fillmore, 1979, p. 212) or the process described by Peters (1983) as fission, in which children segment previously heard material based on stored memories of previously extracted and recently heard speech.

Later, Küntay and Slobin (2002) expanded on their previous analyses by examining the main communicative functions of variation sets observed in Turkish motherese, which included controlling, providing information, and requesting information. As such, the variation set is a unit of analysis that incorporates functional and formal characteristics.

2.3.2 Additional distinctions and measurement

Waterfall (2006) acknowledged the dual nature of variation sets, in formulating contextual and linguistic criteria to identify them in her dissertation, which reported the first longitudinal study of their occurrence in English. To be counted as a variation set, she argued that a series of utterances must adhere to the following contextual criteria (Waterfall, 2006, p. 21):

1. There are multiple utterances in a single conversational turn.
2. These utterances relate to the same extra-linguistic event.
3. The first two utterances forming the set are within five utterances of each other.

In addition, the set of utterances must meet the following grammatical criteria:

4. The utterances have a similar argument structure.
5. They exhibit different syntactic forms.
6. They contain shared lexical content.

Waterfall's study, which looked at data from a carefully balanced sample of 12 mother-child dyads, employed several quantitative and qualitative measures of variation sets. These were based on (a) their frequency, length, and proportion in the discourse; (b) the four characteristics identified by Küntay and Slobin (1996): expansions, reductions, substitutions, and restructurings; and (c) a classification of the linguistic structures (e.g., verbs, nouns) that caregivers varied.

In corpus linguistics, researchers retrieve and analyze linguistic elements from language corpora (Gries, 2008). The corpus-based studies of variation sets described here appear to take as the object of analysis what Gries calls "collostructions" or lexico-grammatical co-occurrence patterns. Two such studies have taken variation sets as their focus, offering additional measures. First, a paper by Brodsky, Waterfall, and Edelman (2007) explored several measures related to the automated detection of variation sets in a corpus. Brodsky et al. distinguished between simple (adjacent) and interleaved (nonadjacent) variation sets, providing the following example of the latter:

(5) *Piggies / you want to read that? / Oh that is Piggies. / You want to read this one?*

They went on to illustrate how the Levenshtein (edit) distance can be used to evaluate the *diameter* of a variation set (in terms of the number of insertions, deletions, and substitutions). They also examined their *prevalence*, or the percentage of words occurring in variation sets, in several corpora. Lastly, they offered a method of calculating *informativeness*, noting that repetition and change must be in balance to achieve a high degree of informativeness. Second, a study by Waterfall, Sandbank, Onnis, and Edelman (2010) adopted the computational criteria established by Brodsky et al., defining variation sets as, "a contiguous sequence of utterances produced by a single speaker in a conversation and each successive pair of utterances has a lexical overlap of at least one element" (p. 687). As they pointed out, this expanded Waterfall's (2006) definition because it allows lexical content to change, as long as any element is repeated across consecutive utterances in the set. Waterfall et al. furthermore examined strictly adjacent variation sets and those with up to two interleaved utterances. They reported corpus analyses using measures of the proportion of utterances and word types in variation sets, the average length of the sets, the Levenshtein (edit) distance of paired utterances, and the probability that utterances in variation sets were aligned by chance.

2.3.3 Research findings

As noted above, methodological approaches to studying the role of variation sets in child and adult language acquisition include longitudinal studies, experimental investigations, and corpus analyses. The key findings from studies in each of these categories are summarized below.

2.3.3.1 Longitudinal studies

Three reports on variation sets in caregiver talk have generated longitudinal insights into their role in early language learning. First, Küntay and Slobin (1996) closely examined input from one mother addressing her daughter, who was being raised bilingually in Turkish and English, over seven months. At the time, Gül (a pseudonym) was between 20 and 27 months. The mother's discourse was analyzed to shed light on how Turkish children likely discover word classes based on distributional characteristics of the input. To summarize, in Turkish variation sets, word order is a weak cue to lexical class, but, overall, nouns are repeated less often and verbs show more morphological variation. Verbs also show less lexical diversity than nouns, and inflection is richer on verbs than nouns. Furthermore, the multiple frames around verbs (see Example 2) may facilitate the acquisition of word classes through communicative interaction. This study, while offering detailed insight into the behavior of word classes in variation sets, did not seek to answer questions about how mothers use them over time.

Second, a later study by Küntay and Slobin (2002) addressed this issue by analyzing a Turkish mother's speech to her daughter at the ages of 15 and 24 months. The study aimed to: (a) identify the communicative functions associated with variation sets, and (b) document changes in variation set use at different stages in child development. The research uncovered four types of communicative functions of variation sets (as noted earlier, these were controlling, providing information, requesting information, and mixed, which combined information and control types). The authors also reported changes in the functions of variation sets from 15 to 24 months (at these time points, 80 versus 36 instances were uncovered). At 15 months, information querying prevailed, however at 24 months, most variation sets were control-oriented. In sum, variation sets appear to be used more often with younger children and their communicative functions also seem related to the addressee's age, indicating a greater emphasis on eliciting and modeling language at earlier stages in development.

Third, Waterfall (2006) studied interactions in English between 12 mothers and their children, from a sample balanced according to child gender, birth order, and maternal education level. Here, the time points for the analyses ranged between 14 and 30 months. Building on the above literature, Waterfall compared mothers' use of linguistic items in variation sets to children's production of these same items. The results supported Küntay and Slobin's findings in that variation set use declined with age. The findings also revealed closer relationships between nouns, verbs, and syntactic features produced by the mother and those later used by the child, when items had occurred in variation sets. Waterfall noted, as well, that these structures were only a portion (10 to 20%) of the overall speech mothers provided to children (2006, p. 100-101).

2.3.3.2 Experimental investigations

Perhaps the earliest experimental study related to this topic is that of Morgan, Meier, and Newport (1989). These authors noted that *local* and *cross-sentential*[5] cues should facilitate the acquisition of structure. They demonstrated this in an experiment

[5] Cross-sentential cues, here, are those which are not overtly manifested in surface strings, result from an underlying rule system, and must be discovered by comparing input.

with one baseline and two experimental conditions, which used lexical substitution and rearrangement, respectively, to augment learning of a miniature artificial grammar instantiated by written (pre-segmented) words denoting shape stimuli as referents. During training, control group participants saw the baseline input repeated twice. Experimental group participants saw the baseline input and one of the two varied forms of input (substitution or rearrangement). The results, which showed better learning under conditions of input variability, demonstrated how, "such cues allow the learner to focus more efficiently on the linguistic domains—phrases—within which syntactic analyses are most optimally pursued" (p. 373). They also point out that such input features do occur in speech to children, citing corpus studies.

The findings of the child acquisition literature have inspired a more recent study to test whether adult learners may also benefit from the arrangement of input into variation sets. Onnis, Waterfall, and Edelman (2008) asserted that variation sets facilitate language induction through basic principles of alignment and comparison (Harris, 1946), operationalizing this notion as a 20% overlap in the words (Experiment 1) or phrases (Experiment 2) contained in adjacent sentence stimuli, contrasted with a control condition in which no overlapping material was presented. This percentage was based on child language corpora. The input was delivered via auditory presentation of an uninterrupted stream of phonemes encoding an artificial lexicon and structure. Performance on a two-alternative forced choice task showed that, in both experiments, participants in the variation set condition (20% overlap) outperformed those in the scrambled (control) condition. The authors concluded that variation sets can foster discovery of language structure via input in adult learning at the lexical and phrasal levels. The fact that learners in the experimental group did better even on words occurring outside of variation sets demonstrates that they could use local alignment to identify word boundaries only previously available globally. Furthermore, this study overcame an important limitation of Morgan et al.'s investigation in that partially repeated stimuli did not overlap temporally but were instead presented contiguously, as would be the case in naturally occurring spoken input.

At this point, it can be noted that previous studies do not appear to have examined interaction effects between variation sets and specific learner abilities, although this is clearly necessary for establishing their role in adult L2 learning. This issue is discussed in more detail later (see Section 2.4.3 and Research Question 3).

Lastly, a related study sought to test the application of the mechanisms facilitating the discovery of word and phrase boundaries in Onnis, Waterfall, and Edelman (2008) to another problem, namely the indeterminacy of reference, famously illustrated by Quine's *gavagai* example (1960). Building on research using a cross-situational learning paradigm (Yu & Smith, 2007), Onnis, Edelman, and Waterfall (2011) conducted three experiments to examine whether participants in conditions where input trials contained contiguous word-referent mappings (i.e., 80% of trials shared an exemplar with the previous trial) would outperform those in scrambled conditions (containing only 5% overlap). The first experiment, in which participants were exposed to nine word-referent pairs (using artificial stimuli), supported the hypothesis that contingency aids learning. The second experiment scrutinized whether this learning could be attributed to some mappings being presented earlier or later during training. It could not. The third experiment then increased the number of word-referent pairs to be learned from nine to

18. This time, there was no difference between the contiguous and scrambled conditions, though both performed above chance. The authors concluded that local cues of repetition and invariance seem important to learning, although global statistics also play a role, particularly when there are more items. One practical consequence is that locally aligned stimuli in experiments on L2 learning (even when these are presented at random) may influence results.

2.3.3.3 Corpus analyses

As mentioned earlier, two corpus-based studies have examined variation sets. Key findings from these studies are discussed here, in terms of the methods described in Section 2.3.2. Brodsky, Waterfall, and Edelman (2007) reported that their measure of the mean information value (or, informativeness) of variation sets correlated with children's vocabulary size in Waterfall's (2006) corpus. This implies an association between the proportion of repeated and new information in mothers' variation sets and children's vocabulary development. Next, Waterfall, Sandbank, Onnis, and Edelman (2010) examined the CHILDES database, with the overarching goal of building a computational framework for language learning meeting four requirements: (a) generativity, (b) use of real (corpus) data, (c) testability (e.g., by measures of recall and precision, which assess the proportion of unfamiliar strings a model accepts and the acceptability of a model's output by native speakers, respectively), and (d) psychological reality, or the correspondence of model assumptions to what is known about how the mind (and brain) construct languages. They then describe two grammar induction algorithms, ADIOS and

ConText, used to support their approach. Waterfall et al. reported findings related to three areas: statistical properties, *n*-gram frequency, and linguistic phenomena:

1. Regarding statistical properties, they found differences in the total percentage of utterances in variation sets, which were a function of the number of shared lexical items (i.e., *n*-grams). For instance, the percentage for variation sets with one word in common ranged from about 50% to 80% (for sets with a gap of zero versus two interposed utterances containing unrelated lexis, respectively). The percentage of utterances in variation sets for sets with two words in common, however, was between 20% and 45% (for the equivalent items).
2. Concerning the most frequent *n*-grams, the researchers noted that most bigrams and trigrams in variation sets occur at the beginning of an utterance. They also suggested that the bigrams (*did you*, *what you*, *are you*) and trigrams (*I don't know*, *what are you*, *you want to*) topping these lists are predictive of specific lexical classes.
3. Lastly, variation sets housed specific linguistic phenomena including dative constructions (found in roughly 12% of variation sets) and object complement constructions (found in roughly 27% of variation sets).

These results are important in that they indicate, in increasingly fine-grained detail, ways that caregivers may adjust the parameters associated with variation sets to potentially enhance their communicative and linguistic value.

2.3.4 Summary of variation sets

As a unit of analysis, variation sets offer the advantage of being definable in both formal and functional terms. Thus, they have been adopted by a range of studies in child language acquisition, cognitive psychology, and corpus linguistics. Studies to date have shown that mothers use them, in various ways, to map form and function in Turkish and English, their presence in input is related to children's production, and to adult's discovery of lexical and phrasal boundaries, and they exhibit important distributional tendencies in CDS, which may enable identification of lexical categories, and help children acquire grammatical features. Thus, it is reasonable to ask whether it may be possible to exploit variation sets to design more effective methods of L2 instruction. That is, by varying *some* intervening forms, portions of the input held constant should be easier to extract, either implicitly or explicitly (Onnis, 2012, p. 216). However, additional research is necessary before the construct can convincingly be adopted in second language studies. One precursor to this adoption is to look at the ways variation sets might be dealt with from the perspective of existing theoretical approaches, such as those outlined at the beginning of this chapter. I will turn to this in the next section.

2.4 Revisiting the three approaches: Multiple roles for variation sets?

This section revisits interactionist, constructionist, and general approaches to learning a language from input in order to critically evaluate the fit between these approaches and variation sets as a unit of analysis. The central problem, or puzzle (Popper, 1963; Jordan, 2004; Long, 2007) is that of whether variation sets might also hold promise for second language research, given the evidence for their substantial contribution to child learning. The focus here is mainly on theoretical issues, but, in determining whether this construct is relevant to the concerns of contemporary approaches to SLA, practical issues will also be raised. Because the research on adults so far has been conducted within artificial learning paradigms, issues concerning the frequency, and potential benefits, of variation sets need to be redefined for SLA. To this end, it may help to consider these issues through the prism of three theoretical approaches.

2.4.1 Interactionist perspectives on variation sets

The functions most commonly discussed in this literature include those related to the negotiation of *meaning* (usually concerned with confirmation checks, clarification requests, comprehension checks) and negotiation of *form* (e.g., recasts). Thus, the functional dimensions of interaction, which frequently concern breakdowns in two-way communication, are quite distinct from those found in the variation set literature. One assumption is that adult learners are already equipped with L1 knowledge of meanings and word classes—this is not the case for children.

A potential issue in applying the notion of variation sets is that they would likely constitute conversational adjustments in Long's (1983a) terms, but rather than arising from any kind of miscommunication, they seem to be based usually on a caregiver's perception of how to manage the interactional structure of discourse. In addition, the conversational adjustments investigated within the interactionist tradition are not typically multiple utterances in a single turn by a single speaker, as in the case of variation sets. A crucial distinction here is that conversational modification in interactionist SLA tends to initiate with the learner. For example, the trigger and resolution model as formulated by Varonis and Gass (1985) illustrated how speakers identify and resolve communication breakdowns based on a trigger (T), or problematic utterance, an indicator (I), or signal of the problem, and a response to the signal (R).[6] As an example, consider the following spoken exchange (which occurred by mobile phone, between two parties on opposite sides of the street):

(6) A: Can you across the street? (T)

B: Sorry, what? (I)

A: Can you cross the street? (R)

Importantly, this interactional sequence may provide implicit negative feedback to speaker A regarding the use of *across*. Feedback of this sort is not provided by variation sets. Thus, in L2 contexts, this may make them similar to grammatical foreigner talk, as they are typically preemptive modifications to discourse structure.

Surprisingly little research has been done on the use of various forms of repetition in instructed L2 settings. However, Chaudron (1988) reviewed several studies of teachers' use of self-repetition in L2 classroom settings, noting that it appeared to be used to enhance comprehension and improve retention.[7] He concluded that teachers use both exact and partial repetition, there is mixed evidence regarding whether repetition is more frequent in L1 or L2 classrooms, and studies controlling for the distance between the initial and repeated utterance may yield different results (pp. 84-85). These conclusions were mostly drawn from observational studies in classroom settings, so it is also

[6] A reaction to the response may also be considered as a final, optional move.
[7] However, whether L2 teachers' use of self-repetition is intentional or unintentional does not appear to have been closely examined.

important to consider studies adopting experimental designs, such as those discussed below.

In practice, variation sets in many ways resemble premodified input. This has been the basis for several L2 studies from an interactionist perspective. Four such studies are reviewed in the following paragraphs.

First, Pica, Doughty, and Young (1987) recruited 16 English language learners for an experimental study on the impact of interaction in listening comprehension. They employed two conditions: premodified and interactionally modified. These conditions differed in terms of the quantity, redundancy, and complexity of the input provided and whether or not interaction was encouraged. Participants in the study were asked to listen to a speaker give directions and then select and place objects on a board. To illustrate, those in the premodified condition heard the speaker say, "take the two mushrooms with the three yellow dots. Put the two *mushrooms* on the grass. Put the two *mushrooms* on the grass near the road" (p. 738, here underlining has been added to indicate repeated elements). In this condition, the speaker did not check on the listener's comprehension and no interaction was encouraged. Those in the interactionally modified condition heard less redundant input, but were then asked whether they understood and were encouraged to interact as much as necessary. Comprehension was assessed according to participants' success in selecting and placing the objects on the board. The study found comprehension scores around 70% in the premodified input condition versus nearly 90% for the interactionally modified condition. No control condition was included, however.

Second, Loschky (1994) reported an experimental study using premodified (n = 14), negotiated (n = 13), and control (n = 14) groups, to investigate the role of oral input modification in L2 Japanese learning. The study used a pretest/posttest design. The treatment comprised three listening tasks in which learners identified and numbered objects in still lifes, maps, or arrangements of shapes. For instance, during the treatment, learners in all groups heard the following baseline instructions (p. 312):

(7) Baseline sentence (heard by all groups)

Pen no migi ni monosashi ga arimasu.

Pen GEN right LOC ruler SUBJ exist

"To the right of the pen is a ruler."

In the premodified input condition learners were further exposed to elaborative modifications such as the as following (p. 312):

(8) Extra sentence read after the baseline sentence (premodified group only)

Sen o massugu kaku monosashi wa, pen no migi ni aru.

Line OBJ straight draw ruler TOP, pen GEN right LOC exist

"The ruler that draws straight lines is to the right of the pen."

In the negotiated group, the baseline input in Example 7 was augmented by repetitions, rephrasals, and explanations, during which the learners could ask questions. The study compared these three groups on measures of comprehension, vocabulary recognition, and sentence verification, involving locative constructions. Greater comprehension scores were found in the negotiated group. However, no group showed an advantage on the vocabulary recognition and sentence verification tests. Both experimental groups performed similarly to the control group.

Third, R. Ellis, Heimbach, Tanaka, and Yamazaki (1999) reported on two studies comparing premodified input to interactionally modified task input in the spoken modality. These authors developed three conditions (baseline, premodified, and interactionally modified) in each of which input varied according to quantity, redundancy, and complexity. An example from their listening task (p. 80) shows how partial or semantic repetition may occur during interactionally modified input (e.g., *sink is a place…it's a hole…sink is a hole*). This task engaged learners in identifying and placing objects in a kitchen using a printed sheet of images:

(9) T: We have an apple. And I'd like you to put the apple in the sink.

S: What is the sink?

T: <u>Sink is a place</u> to wash dishes. <u>It's a hole</u> where you wash dishes.

S: One more time please.

T: We have an apple. And I'd like you to put the apple in the sink.

S: What is sink?

T: <u>Sink is a hole</u> and you wash dishes in the hole.

Learners in the baseline and premodified input conditions were expected to listen, while those in the interaction condition were encouraged to listen and ask questions. In one of the studies these authors reported on, conducted with English learners at a Tokyo high school, the premodified and interaction groups outscored the baseline group on vocabulary translation posttests, though no differences were found between these two experimental groups. In another study, conducted at a high school in Saitama, the interaction group performed better than the premodified group on vocabulary, though data from the baseline group were not reported.

Fourth, and more recently, Robinson (2007c) incorporated premodified input into the design of a study based on the Cognition Hypothesis. Here, the input was a set of six task-relevant phrases based on two constructional types (e.g., *he is carrying a plank* versus *he walks tiredly*, p. 199). Speakers performed a storytelling task in three versions: simple, medium, and complex, which, being based on the picture arrangement subtest of the revised Wechsler Adult Intelligence Scale, varied according to intentional reasoning demands. As hypothesized, increasing the cognitive task demands led to more uptake, or use of premodified input. Specifically, there was a significant increase in instances of partial uptake across the three task versions. This result suggests that variation sets will be more readily incorporated into L2 learner speech under more complex task conditions.

These four reports, though not the only ones to incorporate the construct of premodified input in interactionist SLA research,[8] represent a solid line of investigation that might profit from insights yielded by the variation sets literature.

The construct of partial repetition with constant intention may therefore find a place within the interactionist approach. On one hand, indicated earlier, there are a multitude of features to consider in light of recent research on variation sets (see Table 2.1) that could be readily adopted as options for designing premodified input. It might be advantageous, for instance, to consider measuring independent variables such as informativeness in future studies. On the other, variation sets may also be considered to be of potential value in two-way interaction, particularly if their use is contingent upon learner utterances (see Example 9 above from R. Ellis et al., 1999). It is also possible to conceive of adaptations to providing corrective feedback that incorporate variation sets. The next paragraph provides an example.

---

[8] For example, another experimental study by Yano, Long, and Ross (1994) sheds light on the relative effectiveness of premodified written texts.

As Lyster (1998) noted, recasts and noncorrective repetition may be considered functionally equivalent in classroom discourse. This fact has led researchers to recommend intonationally modified recasts as a pedagogic technique (Doughty, 2001). For instance, Example 10a from a classroom study by Doughty and Varela (1998, p. 124), in which an L2 English user is reporting on a science experiment (10a), can be modified so that the target forms occur in a contiguous sequence of partial repetition (10b). Italics indicate targeted forms:

(10) a. Intonationally modified recast

Jose: I think that the worm will go under the soil.

T: I *think* that the worm *will* go under the soil?

Jose: (no response)

T: I *thought* that the worm *would* go under the soil.

Jose: I *thought* that the worm *would* go under the soil.

b. Recast with partial repetition

Jose: I think that the worm will go under the soil.

T: I *thought* that the worm *would* go under the soil.

T: I *thought* it *would* go under.

Jose: I thought it would go under.

The partial repetition in the third line of discourse in 10b (underscored text) might serve to winnow the learner's attention, focusing him on the corrected (and here, invariant) elements. Example 10b furthermore presents two alternatives to the learner's original utterance. Importantly, the recast with partial repetition is consistent with Doughty's proposal (2001, 2003, 2004) that the mechanism of cognitive comparison drives learning in form-focused interaction.[9]

Moreover, in computer-assisted learning environments, this strategy might be used to extend recent studies on the use of automated recasts (e.g., Sagarra, 2007, Sagarra & Abbuhl, 2013; Trovimovich et al., 2007). For instance, partial repetition may enhance recasts delivered on-screen and/or aurally in response to users' non-target-like spoken or keyed productions. As the number of options for feedback delivery increases, putting control of these settings in the hands of L2 learners will furthermore become increasingly important.

The lasting insight that conversation is selectively facilitative of L2 development (Sato, 1986) has led researchers to concede that its role in adult SLA is far from settled (Gass, Mackey, & Pica, 1998). While there appears to be no research adopting variation sets as a unit of analysis in interactionist SLA research, this does not mean that they hold no theoretical value for the approach. The suggestions above indicate gaps in our current understanding of L2 interaction, which a perspective incorporating a role for variation sets might usefully fill.

2.4.2 Cognitive linguistic perspectives on variation sets

It is easy to see how key processes identified by cognitive linguists, and claimed to have implications for SLA, including embodiment, construal, and categorization (Robinson & Ellis, 2008; Tyler, 2012) are related to the discourse phenomenon of variation sets. First, they appear to establish a joint attentional frame (Tomasello, 2006)

---

[9] Also highly relevant here are the formal constraints on noticing recasts identified by Philp (2003), which include length and the number of changes.

that functions mainly to request information or cooperation from young children. As Tomasello notes, linguistic symbols situated in this manner are not mere associations between form and meaning, but incorporate the additional dimension of socio-pragmatic intentions (2006, p. 28). Thus, the case for variation sets may rest strongly on understanding them as acts *embodied* in the physical and social environment. Furthermore, research by Childers and Tomasello (2001) examining “classic pivot constructions” (Braine, 1976, cited in Childers & Tomasello, 2001, p. 739) revealed that two and a-half-year-old children who heard verb constructions containing partial repetition (e.g., “Look! The cow’s pulling the car. See? He’s pulling it”, 2001, p. 741) were more likely to produce their own utterances incorporating a nonce verb, to which they were exposed later. A similar trend was reported for reception. Childers and Tomasello concluded that children need an anchor point, and some variation around it, to acquire constructional generalizations.

A related study on adult learners by Nakamura (2012) found that, contrary to expectation, input consistency (operationalized as use of proforms to instantiate constructions, e.g., *he there moopoed* versus *the rabbit the hat moopoed*) negatively affected adult L2 comprehension of novel constructions. This study was based on Casenhiser and Goldberg’s (2005) investigation of children’s learning of artificial constructions of appearance (e.g., *the rabbit the hat moopoed* was accompanied by a video of a rabbit appearing on a hat). Nakamura extended this study by developing a +/- input consistency condition, citing Childers and Tomasello. However, there was a crucial difference: Childers and Tomasello’s conditions presented children with either repeating nouns (“Look! The bird’s swinging the bathtub. See? The bird’s swinging the bathtub.”), or varying noun forms ("Look! The cow's pulling the car. See? He's pulling it.") (p. 741). In Nakamura's study, the + consistent conditions presented constructions containing only proforms (*she there moopo*, p. 28) whereas the - consistent conditions presented only concrete nouns (*the rabbit the hat moopo*, p. 26), without varying word classes. This was mainly for reasons related to the amount of input provided to participants. By the author's admission, the lack of variability within the constructions may explain the unexpected result for input consistency (Nakamura, 2012, p. 24).

Another feature related to cognitive linguistic accounts, is that varied elements offer alternative *construals* of a referent or action, as in the use of *push* versus *take* in this example of a mother talking to a child using a stroller to push dolls, from Waterfall et al. (2010, p. 685):

(11)	*You got to*	*push them*	*to school.*
		Push them.	
		Push them	*to school.*
		Take them	*to school.*
	You got to	*take them*	*to school.*

Here, the child is afforded a view of linguistic structure saturated with meaning and situated in usage. Furthermore, given its immediacy, the partial repetition in variation sets may offer evidence for generalization from retained instances, a hallmark of construction learning (Goldberg, 2006). Given the above example, it can be suggested that, upon encountering *push them home*, generalizing to *take them home* should be possible, because the variation set offers cues to the discovery of phrase structure *categories*

(Morgan et al., 1989). In sum, a variety of L2 mapping operations may be evoked when partially overlapping utterances are aligned in discourse.

Owing to the emphasis on usage, repetition in cognitive linguistics is often considered mainly in terms of relative frequency, or the probability of given types and/or tokens occurring in the input. Variation sets, however, have been claimed to play a role beyond frequency in child learning (Waterfall, 2006). It seems the effects of variation sets on adult SLA will depend on how and when they are used, but these matters may be hardly relevant to frequency-based accounts of SLA. Still, it is advantageous from a theoretical viewpoint to attempt to manipulate input frequency and input structure independently in order to understand their differential effects on adult learning (see Miyata, 2011). Having noted this concern, it is clear that cognitive linguistics has much to contribute to an understanding of how variation sets facilitate L2 learning.

As implied above, a major criticism from the perspective of cognitive, or usage-based, linguistics might be that variation sets are simply not widely available enough in language directed at adults to be granted any special status in these accounts of SLA. For the moment, this remains an empirical question, which could be addressed through detailed investigations of L2 classroom corpora on par with those described above in child-directed speech (Section 2.3.3.3). By definition, variation sets are mutable features of discourse, expressing syntagmatic and paradigmatic relations, and encoding semantic and pragmatic meanings while being situated in functional language use. This makes them prime candidates for cognitive linguistic analysis. As before, L2 studies are needed to investigate not only their effectiveness when present in artificial input, but their availability in formal and informal learning contexts. The notion of variation sets seems to align with many theoretical tenets within cognitive linguistics approaches (construal, categorization, and embodiment) and they have also been shown in first language acquisition to be replete with useful distributional information (Waterfall et al., 2010), but, once again, empirical gaps remain in L2 research.

2.4.3 General learning perspectives on variation sets

As mentioned earlier, research has, at times, focused on identifying a limited number of widely applicable principles by which languages may be learned. This is not to say that interactionist and cognitive linguistic approaches do not engage with valid learning principles. They do, albeit upon thorough consideration of the interactional context and linguistic richness encountered by the learner. The present accounts differ in that they take learning principles as their starting point.

Exemplary, related studies of this type include Robinson and Ha (1993), who, based on theories concerning the development of automaticity, offered insight into how repetition and variation within exemplars may influence processing in adult L2 learners. These researchers showed that repeating frames, with variation in the verb slot led to faster response times than when the surrounding frames were new and the verbs were repeated. Also closely related, Mintz (2002) drew on evidence of categorization in children to hypothesize that adults would form lexical categories based on distributional regularities in language, finding that learners could do so after just six minutes of exposure to an artificial language containing nonadjacent, static frames with variable items in medial position (see also Onnis, Christiansen, Chater, & Gómez, 2003). While these studies essentially dealt with the same phenomena, as mentioned earlier, Goldstein et al. (2010) and Onnis (2012) discussed the applicability of learning principles to the distributional qualities of input with explicit reference to variation sets.

First, Goldstein et al. (2010) consider variation sets to be one type of social interaction facilitative of learning, highlighting that such structures contain both behavioral and statistical cues (see their Figure 2, p. 253). They suggest that children learn structure by aligning candidates, comparing them, and evaluating their social and statistical significance. They propose that variation sets illustrate this claim, offering a detailed account of their prevalence and function in CDS. Sensitivity to the multiple cues available in such interactions, they argue, may help learners deal with the so-called poverty of the stimulus (p. 255). They also comment directly on applications to L2 learning. Regarding the use of partial repetition in meaningful teaching materials, they assert that, "intrinsic variation and social motivation should make structure more salient to the learner, stimulating alignment and comparison of adjacent sentences, without overtaxing working memory" (p. 256). They offer this as an outstanding question, presumably with implications for the L2 literature.

Later, Onnis (2012) also approached variation sets from the direction of general principles distilled from the literature on statistical learning, focusing exclusively on ways to diagnose learner needs and support learning in L2 research. Onnis presented and discussed a Spanish example of how partial repetition could serve to increase the salience of non-adjacent, invariant structures (p. 216). This contribution is a good start on understanding the implications of variation sets for grammatical induction. The stipulation that learners must seek invariance echoes Goldstein et al.'s arguments while also grounding them firmly in issues of L2 learning: "becoming sensitive to what changes versus what stays constant in the linguistic environment can highlight structural relations in language such as word boundaries, non-adjacent dependencies, syntactic phrases, and form-meaning mappings" (p. 222). This type of sensitivity might furthermore be related to learner differences in the area of priming. For instance, in the theory of L2 aptitude put forth by Robinson, "individual differences between learners in the extent to which prior exposure to a word or form "primes" and so speeds their subsequent recognition of it is proposed to be related to their success in incidental learning from speech or text" (p. 4; see Chapter 3 for further discussion of Robinson's aptitude theory).

Arguably, L2 learners may develop a sensitivity to the kind of variation in input which these authors suggest facilitates the discovery of structure. Yet, it seems clear that simply presenting such contrasts may not be sufficient. Other general principles including Bybee and McClelland's (2005) emphasis on repeated action as a route to changes in representation may thus also contribute to this argument. It is as yet unclear which L2 pedagogic tasks might be most amenable to the recommendations above, nor is it clear how much practice with them would be needed to bring about various forms of interlanguage development, from initial learning to restructuring. What is clear is that in spite of a principled foundation (Onnis, 2012) and encouraging initial results (Onnis, Waterfall, & Edelman, 2008), hardly any attempt has been made to test these ideas in adult L2 learning.

Viewed through the lens of general learning principles, variation sets hold promise as a mode of presentation for input to L2 learners. Still, an additional assumption widespread in not only the traditional SLA literature, but also in recent complex systems accounts—which specify that language emerges from the interaction of numerous factors, among them "preexisting cognitive abilities, processing idiosyncrasies, and limitations" (Beckner et al., 2009)—is that individual differences matter a great deal in adult learning. Particularly in light of recent work that establishes associations between, on the one hand, statistical learning abilities, and, on the other, L1 comprehension (Misyak & Christiansen, 2012) and L2 learning (Frost et al. 2013), it would be appropriate to consider how adult cognitive abilities may interact with L2 learning under variation set conditions. In this respect, Goldstein et al.'s claim that learning from variation sets reduces the burden on working memory is particularly intriguing and warrants investigation.

Yet another intriguing area of debate is the extent to which the statistical learning processes appealed to in general learning accounts overlap with implicit learning. Indeed, there may be some common ground between these two accounts of learning. Researchers such as Conway and Christiansen (2006) and Kidd (2012) use the cover term "implicit statistical learning". Others have sought to clarify precise nuances in the use of these terms to examine their (in)compatibility (e.g., Conway et al., 2010; Hamrick & Rebuschat, 2012; Misyak, Goldstein, & Christiansen, 2012; Perruchet & Pacton, 2006; Reber, 2011). Therefore, studies of variation sets that thread the needle between statistical and implicit accounts could contribute to a better understanding of the distinctions necessary to clearly formulate general learning principles. The gaps here approach bigger picture issues in adult L2 learning. In this respect, a clear advantage afforded by the literature on variation sets is the existence of computational frameworks building on tools such as ADIOS and ConText (see Waterfall, Sandbank, Onnis, & Edelman, 2010, discussed in Section 3.3.3), which, based on corpus data, provide empirical results concerning the relevance of proposed usage-based mechanisms to learning.

2.5. Potential implications

The potential implications of the construct of variation sets, as applied to adult language learning, can be classified as theoretical, methodological, and pedagogic. First, based on the above review, there is space to develop hypotheses about the role of partial repetition in second language work of various theoretical orientations. By this point, it should be obvious that as a unit of analysis, variation sets are, to some degree, independent of any given theory and thus might inform various researchers' perspectives. Second, methodologically, there may be implications for studies on implicit learning which report that the trial order of training input is "independently randomized for each participant" (Williams, 2005, p. 282). In this case, if some participants by chance see the input in variation sets, does this increase their likelihood of becoming aware? This could possibly explain some of the mixed results on studies investigating L2 awareness in experimental settings (related to this point, see Onnis, Edelman, & Waterfall's 2011 discussion of Yu & Smith, 2007). Last, as mentioned earlier, variation sets might foster innovation in terms of the delivery of recasts in classrooms, or in computer-assisted language learning (CALL). Regarding the latter, Chapelle (2009) has noted that interactionist and cognitive linguistic approaches have many implications for designing instructional strategies in CALL. Perhaps variation sets can be integrated into self-study tools, which might expand their value beyond established uses in CALL to include emergent technologies such as mobile devices (Levy & Stockwell, 2006). These are simply potential areas that could be served by a focus on how partial repetition may influence various processes and outcomes in L2 theory, research, and teaching.

2.6. Summary

In this chapter, I have dealt with the problem of assessing the potential value for SLA of a construct imported from child language acquisition research. I began by tracing the origins of three approaches to L2 input and carefully defining the construct under examination. Interactionist researchers are likely to consider variation sets as a distinct form of modified input, which can conceivably be applied to enhance learner comprehension, mainly within instructed settings. Cognitive linguists would analyze them in terms of how they fine-tune the learner's experience of language in ways that are statistical and embodied. Researchers whose interests lie with general principles view them as substantiating claims about how human cognitive abilities give rise to specific learning outcomes, based on which we might derive two complementary hypothesis regarding the role of variation sets: the automaticity hypothesis and the categorization hypothesis. Yet, these distinct approaches all converge on the explanation that it is social interaction and cognitive processes which underpin learning. Along these lines, this chapter has offered a coherent rationale for studying variation sets in adult L2 learning, but has also considered potential limitations of the construct. Later on in this dissertation, I make specific predictions about the role of variation sets and test them to provide evidence for or against the above views.

CHAPTER 3

CONSCIOUSNESS AND L2 LEARNING: TERMINOLOGY, FINDINGS, AND ISSUES[10]

3.1 Introduction

This chapter offers an overview of research on the role of consciousness in second language learning. To start with, it introduces key terminology from the literature on cognitive psychology, as defined within studies of second language learning, and used throughout this dissertation. Evidence from studies of the relationship between awareness and L2 learning is then reviewed, in order to identify linguistic features that may be learned without conscious awareness. Later, key ongoing debates in this area, which are in need of further investigation, are described. The chapter concludes by summarizing the implications for the present research.

3.2 Terminology

Within SLA research, labels abound for the great blooming, buzzing confusion described by James (1890). Those most relevant to the present study are defined here, followed by additional useful terminology. Since at least the early 90s, a number of full-length volumes and focused review papers have examined in detail phenomena related to L2 consciousness (e.g., Bergsleithner, Frota, & Yoshioka, 2013; DeKeyser, 2003, 2011; Hulstijn, 2003, 2005; N. Ellis, 1994; R. Ellis, Loewen, Elder, Erlam, Philp, & Reinders,

---

[10] A previous version of this chapter appeared in J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka, (Eds.). (2013). *Noticing and second language acquisition: Studies in honor of Richard Schmidt.* Honolulu: University of Hawai‘i, National Foreign Language Resource Center.

2009; Rebuschat & Williams, 2012a; Robinson, 2001a, 2002a, 2007a; Robinson, Mackey, Gass, & Schmidt, 2012; Sanz & Leow, 2011; Schmidt, 1990, 1994a, 1994b, 1995a, 1995b; Williams, 2009). Below, I present terms adopted throughout this literature.

3.2.1 Consciousness

In the field of cognition, there are several notable, yet distinct views of consciousness. Perhaps the one of the most influential in L2 research is that of Baars (1988, 1997), who brought together various metaphors for consciousness to develop a unified model called Global Workspace theory, which holds that consciousness is a "facility for *accessing, disseminating, and exchanging information,* and for *exercising global coordination and control*" (1997, p. 7, italics in original). As pointed out by Schmidt (1990, inter alia), however, consciousness is a vague term with many different senses, including awareness, attention, intention, knowledge, and control, each of which, when carefully defined, has great relevance to second language learning and teaching. Thus, terminological rigor is important to clarifying not only the practical implications of the role of consciousness, but also its significance for theories of first and second language acquisition.

Untangling these issues is complicated by the fact that the theoretical foundations of generative linguistics and SLA contain references to language operating "unconsciously and beyond the limits of awareness" (Chomsky, 1988, p. 157) or language acquisition as "a subconscious process" (Krashen, 1982, p. 10). As will be shown, though, the acquisition processes operating in L2 learning are more varied and distinctive. Consider, too, here, Block's (1995, 1996, 2007) definitions of phenomenal consciousness and access consciousness. The first refers to states within which we experience sensations, feelings, perceptions, mental images, desires, or emotions. Block controversially argues that these are separated from cognition, intention, and function (see 1995, p. 230). The second refers to states in which mental representations can inform reasoning, reporting, or action—representations which, though inextricably linked to language, may not be fully appreciated at times.

Lastly, one historic recent development is work on the physiological aspects of consciousness. Building on a number of detailed assumptions, this is framed in terms of the search for the neuronal correlates of consciousness (NCC), or the smallest coalition of neurons necessary for a given conscious percept (Koch, 2004). Koch has noted commonalities between his framework for the discovery of the NCC and Baars' Global Workspace theory, described above. In particular, Koch assumes that once neural activity exceeds a certain threshold, it activates a global network of neurons enabling access to working memory and other cognitive resources distributed across the brain (2004, p. 311, see also Dehaene & Changeux, 2004).

3.2.2 Awareness

Schmidt (1990) discussed three potentially dissociable levels of awareness: perception, noticing (focal awareness), and understanding. According to Schmidt, perception is the internal representation of external stimuli, yet is not always conscious. Noticing requires awareness at the level of subjective experience, which, under certain conditions, is available for verbal report. Understanding encompasses reflection, comprehension, problem solving, and other forms of thinking, including metacognition.

These last two levels, in particular, have wide-ranging implications for L2 research, and have been discussed at length in this literature. The distinction between noticing and understanding is of theoretical and practical importance, as it involves, in the former case, registering surface elements or instances, and in the latter case, recognizing abstract rules or patterns (Schmidt, 1993, 1995b). Of additional obvious relevance to cases of L2 learning after childhood is the role of metacognition, or more specifically, metalinguistic awareness, which has been defined and described in various ways (e.g., Bialystok, 2001; Jessner, 1999, 2006; Koda, 2007; Leow, 1997). There are several controversies surrounding the evidence for learning without awareness in SLA research and Section 3.4 below will describe some of these in more detail. One consequence of these debates is the increasingly fine-grained detail of accounts of L2 awareness (e.g., Leow, Johnson, & Záráte-Sandez, 2011; Robinson, Mackey, Gass, & Schmidt, 2012; Schmidt, 2012; Williams, 2009).

3.2.3 Attention

Tomlin and Villa (1994) presented several course-grained notions of attention, under which it can be conceived of as: (a) a limited-capacity system, (b) the process of selecting stimuli for further analysis, (c) effortful (in contrast to automatic) processing of information, and/or (d) control over information and action (p. 187; see also Robinson, 1995b, 2003). These authors also offered a fine-grained account of attention in terms of its key functions of alertness, orientation, and detection. Tomlin and Villa described the first of these as a readiness to process incoming stimuli, which can be manipulated. Orientation involves specifically directing attention toward some type of information and not another. This facilitates or inhibits subsequent detection, which is the registration of stimuli that makes it available for further processing. Contrary to Schmidt's noticing hypothesis (see Section 3.2.4), Tomlin and Villa contended that detection alone is all that is required for L2 learning, though awareness, along with alertness and orientation, plays a supportive role. Carr and Curran (1994) reached a similar conclusion in regard to the learning of structured sequences in SLA; it is primarily attention that is required for contextual coding, or storing relations among elements. Another distinction concerning attention in SLA is the difference between voluntary (controlled) and involuntary (passive) forms of attention (Schmidt, 2001).

Interestingly, researchers in visual cognition have recently marshaled evidence for dissociations between attention and awareness (Lamme, 2003, 2004; Dehaene et al., 2006; Koch & Tsuchiya, 2007; see Cohen et al., 2012a, 2012b; Tsuchiya, Block, Koch, 2012, for recent commentary). Logically, then, one could entertain the possibility of L2-related processing under any of the following conditions:

- +attention/+awareness
- +attention/-awareness
- -attention/+awareness
- -attention/-awareness

The proposal in the next section (on noticing) considers learning as the outcome of the first of these options.

The need to understand attention within SLA becomes clear in disputes over the difficulty of attending to form and meaning simultaneously (VanPatten, 1990, 1996; 2002a, 2004, for commentary, see DeKeyser, Salaberry, Robinson, & Harrington, 2002;

Doughty, 2004; Leow, Hsieh, & Moreno, 2008; Morgan-Short, Heil, Botero-Moriarty, & Ebert, 2012; VanPatten, 2002b; Wong, 2001). Controversy has arisen with respect to VanPatten's principles of input processing, particularly the assertion that L2 learners process the meaning of input before processing its form. Theoretical and empirical arguments have been mounted for and against this position. Briefly, alternative views of capacity limits lie at the heart of the debate, which has also been concerned with how the written modality may ease attentional demands and facilitate comprehension when learners pay attention to both form and content.

3.2.4 Noticing

Schmidt and Frota's (1986) study of Schmidt's learning of Brazilian Portuguese over a five-month period yielded the observation that acquisition occurs when learners notice the gap between their ability and features of the input. The noticing hypothesis states that noticing, which entails conscious registration of input, "is the necessary and sufficient condition for converting input to intake for learning" (Schmidt, 1994a, p. 17; see also 1990, 1993). It has furthermore been claimed that while noticing is necessary for L2 learning, understanding is facilitative but not required (e.g., Schmidt, 2012).

Building on this, Robinson (1995b, 2003) has presented an operational definition of noticing as detection plus activation in working memory, arguing that maintenance and elaborative rehearsal are implicated in the degree of awareness accompanying noticing. In brief, Robinson claims, based on his synthesis of work by Craik and Lockhart (1972), among others, that "[r]ehearsal processes can be of two kinds: *maintenance rehearsal*, requiring data-driven, instance-based processing, and *elaborative rehearsal*, requiring conceptually driven, schema-based processing" (2003, p. 652). According to Robinson, these processes are activated during noticing and influence the type of awareness it involves.

The noticing hypothesis has garnered much attention, along with, inevitably, criticisms, to which Schmidt (2012) has replied in detail. Frequently, the issues are methodological, as some have objected that: (a) diary studies are not appropriately fine-grained to allow inspection of attention (Tomlin & Villa, 1994) and (b) constructs within the hypothesis (e.g., the distinction between awareness at the level of noticing and at the level of understanding) have not been easily nor uniformly operationalized (Truscott & Sharwood Smith, 2011). Others have challenged the view of language learning implied by the hypothesis, stating that (c) certain linguistic features (words, sounds, orthography) may require noticing, while others (phonological, morphological, and syntactic rules) may not (Schachter, 1998), or that (d) models relying on environmental input as the sole impetus for L2 acquisition are insufficient, as this type of input is only indirectly related to the internal input (Carroll, 1999, 2006; see Gass, 1997, for a similar view that UG influences intake). Lastly, as noted above, the contribution of awareness to learning is debatable. Consequently, some take the position that (e) only attention (i.e., detection) is necessary; learning may proceed in the absence of awareness (Carr & Curran, 1994; Tomlin & Villa, 1994; Williams, 2005).

Research designed to assess noticing has employed a wide range of behavioral measures, which can be classified into concurrent measures, including think-aloud protocols (Alanen, 1995; Leow, 1997), underlining (Izumi & Bigelow, 2000), or note-taking (Izumi, 2002) and retrospective measures, including written questionnaires

(Robinson, 1995a), cued recall (Philp, 2003), and stimulated recall (Mackey et al., 2002). The use of eye-tracking methods to examine noticing (Godfroid, Housen, Boers, 2010; Smith, 2012) is another promising recent development. Studies of noticing have implications for practical areas such as classroom instruction and materials development, as well as theoretical areas, such as understanding the role of L2 learner-internal processes in instructed and non-instructed settings (see the chapters in Bergsleithner, Frota, & Yoshioka, 2013, for further commentary).

3.2.5 Implicit and explicit learning and knowledge

In the late 1960s, the psychologist Arthur Reber defined implicit learning as an unconscious process resulting in abstract knowledge (1967, 1989, 1993; see also Thorndike & Rock, 1934). In the area of L2 learning, implicit learning is hypothesized to occur when L2 learners acquire abstract phonological, morphological, syntactic, or other rules in the absence of awareness and without intention to do so (Schmidt 1990, 1994a, 1994b, 1995, 2001; for further discussion, see DeKeyser, 2003; Hulstijn, 2005; Williams, 2009). Explicit learning, in contrast, is accompanied by awareness and, often, though not always, intention. In either case, it is essential not to conflate learning and knowledge, as Schmidt (1994a) has noted:

> Implicit and explicit *learning* and implicit and explicit *knowledge* are related but distinct concepts that need to be separated. The first set refers to the processes of learning, the second to the end-products of learning (or sometimes to knowledge that is innate and not learned at all). (p. 20)

Thus, for learning to be implicit one must lack of awareness of the critical feature(s) at the time of learning. In support of this view, Leow and Hama (2013), highlighting the process/product dichotomy, suggested that researchers specify at which stage learners are claimed to be unaware: encoding or retrieval. This has far-reaching consequences, as some studies purporting to investigate implicit learning have, in fact, relied on measures that arguably assess implicit knowledge, not implicit learning. Regrettably, confusion over these matters has led some SLA scholars to conclude that, “there is no consensual definition of implicit learning” (R. Ellis, 2009, p. 7).

Another important point is that, often, research on implicit learning has targeted structural properties inherent in language. Examples include the use of artificial languages (see Onnis, 2013), which, in many studies, are not intended to convey meaning. In contrast to this, semantic implicit learning (Williams, 2013) depends on semantic distinctions potentially encodable in language (Bickerton, 2001; Talmy, 2000). A final consideration is that researchers often use the terms implicit/unconscious or explicit/conscious synonymously (e.g., Rebuschat, 2013).

3.2.6 Statistical learning

Aslin and Newport (2009) described statistical learning as learning: (a) based on frequency or probability; (b) through exposure alone (i.e., without feedback); and (c) in conditions where stimuli are presented rapidly, as in natural speech. The kinds of outcomes resulting from statistical learning include discovery of lexical categories via phoneme distributions (Christiansen, Onnis, & Hockema, 2009) and detection of non-adjacent dependencies via phonological characteristics (Onnis, Monaghan, Richmond, & Chater, 2005).

Due to the perceived overlap of implicit and statistical learning, recent reviews of implicit L2 learning have drawn on statistical learning research (Reber, 2011; Williams, 2009). At times, researchers have considered these processes similar enough not to warrant separate definitions. Several recent articles (Conway & Christiansen, 2006; Kidd, 2012, Yim & Rudoy, 2013) employed the cover term "implicit statistical learning" and Conway et al. (2010) went as far as to argue that the same phenomenon underlies both implicit and statistical learning. Others, such as Perruchet and Pacton (2006), have distinguished between implicit and statistical learning according to the precise nature of the operations involved. They illustrated how computations over the same data using transitional probabilities versus those resulting from chunk-strength and interference result in highly similar, but not identical, predictions (p. 236).

Before moving on, it should be pointed out that statistical learning originated in studies on infant learning (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and continues to be of great interest within this domain (Gómez & Gerken, 2000; Kuhl, 2004; Lany & Saffran, 2010, 2011). However, recent empirical studies have also examined statistical learning by adults (e.g., Frost et al., 2013; Hamrick & Rebuschat, 2012; Misyak & Christiansen, 2012; Onnis, 2012). Given the definition above, it is assumed here that certain distributional properties of input trigger specific internal cognitive mechanisms that can facilitate learning based on mere exposure in adults.

For example, it may be the case that statistical learning establishes task conditions which induce learners toward data-driven, rather than conceptually driven, processing.

The argument here is simply that more frequent, rapid exposure may shift attention toward surface features of the input, while simultaneously deterring schema-based processing, with substantial consequences for noticing and the use of working memory in adult L2 learning (see Sections 3.2.4 and Chapter 4).

3.2.7 Additional terminology

The abundance of terminology related to L2 learning makes exhaustive coverage nearly impossible. Several important constructs, however, include those related to knowledge and memory and those used to describe various degrees of instructed and uninstructed learning. These are discussed below.

To begin, a first set of constructs, pertaining to knowledge and memory, are interrelated, but distinct. Declarative and procedural knowledge are separate, in that the former encompasses information in verbalized form ("knowledge that") and the latter, behavioral routines ("knowledge how"), which may become automatized (DeKeyser, 1997, 1998, 2001, 2007, 2009). Furthermore, declarative knowledge consists of semantic and episodic memory. Tulving (2002) defines these as a system containing knowledge about the world (semantic memory) and an interrelated, but separate system enabling autonoetic awareness, or mental time travel (episodic memory). However, in addition to conscious knowledge of past events (explicit memory), memory may be unaccompanied by awareness (implicit memory), as noted by Daniel Schacter (1987). Implicit memory often refers to semantic memory without episodic memory, as when a subject shows priming effects (i.e., faster processing) for previously encountered stimuli, but has no recollection, or episodic memory, of having experienced the stimuli.

The remaining constructs are grounded in research on L2 instruction. Norris and Ortega (2000) defined explicit L2 instruction as involving rule explanation (deductive learning) or rule discovery (inductive learning), whereas implicit L2 instruction does not involve such attempts to target rules. The latter may include techniques such as input flood or recast, when these are not accompanied by any further guidance to attend to language features (see also DeKeyser, 1995, 2003). Antonyms for such explicit or implicit formal L2 instruction may include untutored and uninstructed learning.

Incidental learning, as used in the literature on L2 instruction, refers to learning something without the intention of doing so, which occurs as the by-product of activities primarily focused on communication or meaning (Hulstijn, 2001, 2003, 2005; Schmidt, 1994a). An example would be readers picking up the word *lomtick*, used to count toast slices in the novel *A Clockwork Orange*. Intentional learning, too, has a particular sense when in educational contexts it concerns L2 learning that is deliberate and has a clear goal. As Hulstijn (2003, 2005) notes, these terms are used differently in psychology, where incidental learning describes experiments in which participants are not told they will be tested afterwards. This is contrasted with intentional learning, as in experiments that do forewarn participants of testing (see for example, Dienes, Broadbent, & Berry, 1991, who found no difference in performance using incidental versus intentional instructions).

3.3 Empirical findings concerning L2 learning and awareness

This section tries to chart the development of empirical research on L2 learning and awareness and bring to light issues within it, according to evolving conceptual, theoretical, and methodological foundations. Is it possible to learn entire L2 systems (as Krashen many years ago suggested) or L2 subsystems (as others have more recently suggested) without awareness? In attempting to answer this question, the linguistic features that research has examined will also come into focus.

3.3.1 Early research and theoretical underpinnings

Three early studies established a conceptual base for research on implicit learning in SLA. First, Nation and McLaughlin (1986) used artificial grammars from Reber and Allen (1978) to test monolinguals, bilinguals, and multilinguals learning from implicit and explicit tasks. The two grammars used consisted of letter strings governed by a system of rules argued to be too complex to decode explicitly (Reber, 1993, p. 29). In the implicit task, which was presented first in the study, instructions to the participants were "intentionally nebulous; they were simply asked to pay close attention" (p. 46). Nation and McLaughlin found that the multilingual group outperformed the other two groups, scoring above 70% correct on grammaticality judgments of test strings in the implicit and explicit task conditions. They therefore suggested that experience plays a role in how learners process novel language stimuli (see also, McLaughlin & Nayak, 1989). Second, Hulstijn (1989) carried out two experiments with Dutch-speaking groups: L1 speakers studied a semi-artificial version of their L1, and L2 speakers learned subordinate clause structures in Dutch. Based on evidence from cued recall and timed copying tests, Hulstijn concluded that "attention to form at input encoding is a sufficient condition" for implicit learning of structural elements (p. 72). These elements included subordinate clause structures marked by word order, auxiliaries, and function words. Third, Ellis (1993) focused on soft mutation rules pertaining to initial consonants in Welsh (e.g., *Bangor* becomes *o Fangor*). Instructional conditions included random (considered "naturalistic" or implicit), rule plus instance, rule, and yoked random groups. Ellis reported that implicit learners' accurate judgments steadily increased for old and new well-formed examples, though these learners performed poorly on ungrammatical items and, in general, the rule groups performed better than the implicit group (p. 312).

In summary, these early studies suggest that implicit L2 learning may be: (a) sensitive to individual differences in L2 learning experience; (b) reliant on attention to form or *noticing*, the object of which is "elements of the surface structure of utterances in the input" (Schmidt, 2001, p. 5); and (c) accumulated slowly.

3.3.2 Comparison studies of implicit versus other conditions

Following these initial studies, an important contribution by DeKeyser (1995) was motivated by pedagogical concerns regarding the learnability of rules under varying exposure conditions. DeKeyser used an artificial language having categorical rules and prototypical patterns, hypothesizing that the former would be learned better under explicit conditions, whereas the latter would be learned better under implicit conditions. Participants in both conditions joined 20 training sessions (lasting 25 minutes each) plus a final testing session. During training, participants were exposed to sentences and accompanying pictures illustrating the target language. In the explicit condition, three of these sessions were preceded by grammar explanations, and two other sessions began with metalinguistic tests. To control for the added activities, the explicit group was divided into participants given the same time versus extra time to complete the training sessions. DeKeyser's first hypothesis was confirmed: explicit-deductive learners outperformed those in the implicit-inductive group on a test of categorical rules. Though there was some evidence for the second hypothesis (better learning of prototypical patterns by the implicit group), the difference was not statistically significant.

Likewise, Alanen (1995) was interested in the impact of providing explicit, metalinguistic information to L2 learners. In this study, thirty-six participants studied Finnish locative suffixes and consonant changes in two group sessions. There were four treatments differing in terms of whether participants received: (a) texts only, (b) enhanced texts with target forms italicized, (c) texts plus a page-length description of the target rules, (d) enhanced texts plus the rule description. The former two conditions were considered meaning-based (also called "implicit", p. 265), while the latter two were regarded as rule-based. During the group sessions, verbal reports were collected to assess whether participants noticed the target structures. Learners in more explicit, rule-based conditions performed better on a sentence completion test than those in more implicit, meaning-oriented conditions. Additionally, participants' comments on the structures correlated with performance. Alanen argued that these results supported the noticing hypothesis (Schmidt, 1990).

In several reports based on his 1994 doctoral dissertation, Robinson (1995a, 1996, 1997a) shed additional light on the issues of rule complexity, learner awareness, and instructional conditions. In his study, the implicit condition was compared to incidental, rule-search, and instructed conditions. Participants were English learners (*N* = 104) who responded to grammaticality judgment tasks assessing their learning of fronted adverbial clauses and pseudo-clefts of location. The groups each trained on 40 sentences, with procedures varying by condition. Those in the implicit group saw sentences and then indicated whether two words had appeared next to each other. Incidental learners answered a comprehension question about the sentences. Rule-search learners were prompted to figure out the rules governing the sentences and indicate when they knew them. Instructed learners were first taught the rules and then answered metalinguistic questions about the sentences. All but the rule-search learners received feedback after each trial.

A main effect was found for condition, with no interaction between rule complexity and condition. The highest scorers for both grammatical structures were the participants in the instructed condition. Awareness was measured using a post-experimental questionnaire asking participants whether they had noticed, searched for, and could verbalize rules. There were significant differences in the degree to which participants looked for rules across the four conditions, with fewer looking for rules in the implicit condition. Lastly, participants who could verbalize the rules scored higher than those who could not. As above, this suggests an advantage for participants who are aware of rules, irrespective of condition.

A subsequent study by Robinson (1997b) employed identical training conditions, but used artificial verbs based on English, to examine the question of whether learning is driven by memory for instances or rule abstraction. Sixty L1 Japanese speakers participated. The implicit, incidental, rule-search, and instructed groups again saw 40 sentences, here containing pseudo-verbs approximating English dative rules (e.g., “Peter pelled the English teacher an excellent paper,” p. 232). A grammaticality judgment task was used to assess accuracy and reaction times (RTs) for trained and new (grammatical and ungrammatical) items. Interestingly, accuracy and RTs for trained items were equivalent across the four study conditions. Still, instructed learners' responses to new, ungrammatical items were significantly more accurate, and there was also evidence for enhanced performance by instructed learners on new, grammatical items. Robinson concluded that, "rule-based knowledge…and implicit memory-based knowledge interact in decision-making" (p. 242).

Another comparison study by de Graaff (1997) examined learning in implicit versus explicit treatments. Both treatment groups completed a series of 10 computer-based lessons lasting about 1.5 hours each, on a version of Esperanto. These lessons incorporated a variety of activities, including meaning comprehension, translation, vocabulary practice, form-meaning mapping, and production. Both implicit and explicit participants received feedback after these activities and could redo problems, if incorrect. The difference was that, in the explicit condition, grammatical explanations appeared after several activities, whereas, in the implicit condition, these explanations were substituted by additional rehearsal. Thus, both groups received the same amount of input. Each group contained 27 learners. The targeted features were simple and complex morphological structures, including plural and imperative markers and placement rules for negatives and objects. A proficiency test with four sections was used to measure learning. Participants receiving explicit rule explanations scored better than those in the implicit (rehearsal) group. Unlike DeKeyser's (1995) study, though, the learning of specific rules and features did not appear to vary by condition.

A recurring theme in the research discussed up to this point has been the superior performance of learners characterized as having some degree of awareness, by means of either group assignment into explicit (i.e., instructed) conditions or self-reported awareness. Also, though there is some evidence to suggest that rule complexity may interact with implicit and explicit learning, generally, the advantage is in favor of explicit L2 learning. Implicit L2 learning may not extend beyond instances to which learners are exposed and may therefore be based on memory rather than rule induction (Robinson, 1997b). Finally, because most of these studies focused on the effects of assigning participants to different study conditions, they point to one obvious caveat. Namely, implicit instruction must be distinguished from implicit learning.

3.3.3 Think-aloud protocols

In contrast to these approaches, Leow (2000) was one of the earliest attempts in this literature to use think-aloud protocols to investigate awareness at the time of exposure.[11] Leow's study showed that aware learners scored higher than unaware learners on recognition and production tests after completing a crossword puzzle focusing them on the conjugated forms of stem-changing verbs in Spanish. Leow (2000) interpreted these results as contradicting claims that awareness and learning may be dissociated (Tomlin & Villa, 1994; Carr & Curran, 1994) and supporting claims that awareness at the level of noticing predicts learning (Schmidt, 1990, 1995; Robinson, 1995). Below, methodological issues will be addressed further.

[11] This work extended Leow (1997), adapting materials from that study such that some learners remained unaware of the target forms.

3.3.4 Semantic implicit learning

Another recent trend in the evolution of research on L2 awareness is the overt focus on the unaware learning of associations between form and meaning when participants are engaged in comprehending a novel L2 system (see Williams, 2013). Two experiments in this vein were reported by Williams (2004). In the first experiment, 37 participants learned an artificial language, based on Italian, in which nouns were coded for definiteness and animacy. Seven participants became aware of the role of animacy during training and were excluded from further analyses. The remaining participants scored 61% on a test of generalization items—this rose to 71% for participants whose L1 marked grammatical gender. Both figures were statistically above chance. In Williams' second experiment, 17 participants saw an artificial language consisting of four determiners marking nouns for both distance and animacy. Participants were taught that the markers *gi* and *ro* meant 'near' and *ul* and *ne* meant 'far.' They were not told that, in addition, *gi* and *ul* accompanied living referents, and *ro* and *ne* were used for non-living referents. This time, none of the participants became aware during training, scores on an initial test of generalization did not differ from chance, and little evidence of implicit learning of animacy was found.

Shortly afterwards, Williams (2005) achieved results more supportive of semantic implicit learning in two additional experiments building on Experiment 2 above. In contrast, the learning task in these experiments was simplified, the number of noun tokens was increased, and sentences (rather than phrases) were presented in training. The artificial determiner system was the same, however. In Experiment 1, 41 participants received instruction on the distance factor and then listened to and repeated sentences across 144 training trials. Eight of these participants reported awareness of animacy during testing. For unaware participants, performance was above chance on all test items. In Experiment 2, revised materials were used. This time, seven of 24 participants indicated awareness. The mean scores for the 17 unaware participants were above chance for both novel and trained items.

Later, Hama and Leow (2010) reported on an extension of Williams (2005), using concurrent verbal reports (i.e., think-aloud protocols) as a measure of awareness. They also increased the number of options on the multiple-choice test from two to four, employed a test of production, conducted training and testing phases in the spoken modality, and screened any participant who had a linguistics backgrounds. None of the 34 included participants became aware of the relevance of animacy, nor did they show learning. These results may be due to the verbal reasoning required of participants during the think-alouds, which may have blocked them from learning. The issue of whether such concurrent verbalization procedures are *reactive*, that is, the extent to which they increase task demands and alter mental processing, is a matter of some debate (Bowles, 2010; Leow & Bowles, 2005a, 2005b). A replication of Williams (2005) by Faretta-Stutenberg and Morgan-Short (2011) also found limited evidence of learning without awareness. Recently, however, another conceptual replication and extension of Williams' study argued, based on participants' source attributions, that some of the structural knowledge they had acquired remained unconscious (Rebuschat et al., 2013).

Leung and Williams (2011) investigated the possibility that implicit learning is limited to features which can be lexically derived (e.g., *gi* and *ul* are always associated with nouns denoting living things). They therefore set out to investigate whether unaware learning is possible when these determiners mark instead the thematic role of the noun. The sentence stimuli they constructed used *gi* and *ul* to signal agents and *ro* and *ne* to signal patients (with *gi* and *ro* here used for adults and *ul* and *ne* for children). Also, word order was flexible, so that it could not be used as a cue to thematic role. The procedure was based on the contextual cueing paradigm (Chun, 2000; Jiang & Chun, 2003). On each trial, participants first described a picture and then listened to a sentence containing the artificial thematic role markers. They next clicked a left or right mouse button to identify the agent from the picture and, lastly, reformulated the sentence in English. When participants unknowingly reached a series of trials in which the artificial system was violated, the form-meaning mappings in the input were reversed, although the task stayed the same. Five participants (of 25) displayed awareness in a subsequent debriefing. As for the remaining participants, the authors found a significant increase in reaction times on violation block items, when compared to the control block items. In light of these findings, Leung and Williams argued that implicit learning mechanisms enable detection of correlations between form and contextually-derived meanings based on depicted events.

Leung and Williams (2012) described two additional studies adapting the contextual cueing paradigm to implicit L2 learning research. As in their previous study, they examined the issue of exactly which meaning distinctions can be learned under implicit conditions. In Experiment 1, the artificial determiners from Williams (2005) encoded animacy and distance, whereas in Experiment 2 they encoded relative size and distance, based on the assumption that since relative size is not a grammaticalized concept, it should not interact as strongly with language processing as animacy (Leung &

Williams, 2012, p. 639). In Experiment 1, 13 of the 33 participants became aware. The violation effect for unaware participants was –68 ms, while for aware participants it was –232 ms. In both cases, these effects attained significance. In the second experiment, six out of 26 participants became aware. However, this time, neither group showed a significant violation effect. In their discussion, the researchers explained the different results of Experiments 1 and 2 in terms of the activation rate for animacy versus relative size.

Two more reports also speak to the issue of semantic implicit learning. First, Chen et al. (2011) used source attributions and verbal reports to demonstrate gains in unconscious knowledge of form-meaning pairs in a conceptual replication of Williams (2005) conducted with modified Chinese characters (Experiments 1 and 2), also showing that a linguistically-irrelevant feature was not learned this way (Experiment 3). Second, Guo et al. (2011) examined the development of implicit knowledge of semantic prosody (e.g., the typical usage of the verb *cause* with negative consequences) by Chinese learners of English by using pseudowords, this time employing confidence ratings and source attributions. Both studies expand the range of L2 features investigated in semantic implicit learning research with important implications.

3.3.5 Confidence ratings and source attributions

Although the idea of measuring awareness in SLA research is not new, recent studies have adopted innovative methods of assessing of implicit knowledge. For instance, within the cognitive science literature, two approaches to measuring implicit knowledge are confidence ratings and source attributions (for discussion, see Rebuschat, 2013, as well as Dienes, 2012; Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Scott, 2005). Paired with judgment tasks, these measures require participants to first judge test items, and then indicate either their confidence in, or the perceived source of, each judgment. Confidence ratings may be stated in terms of a percentage or scale descriptor such as *somewhat confident, very confident*, whereas source attributions are often stated in terms of options such as *rule knowledge*, *intuition*, *memory*, or *guessing*.

Rebuschat and Williams (2012b) reported a study on implicit knowledge of syntax employing these measures. The linguistic targets were several rules related to verb placement in German sentences. These rules were applied to sentences containing English lexical items, enabling the participants to access the meaning of the stimuli (e.g., "*Today bought John the newspaper in the supermarket,"* p. 9). Participants heard training items and indicated whether the items were semantically plausible. They were tested using a grammaticality judgment tasks accompanied by confidence ratings (Experiment 1) and source attributions (Experiment 2). Rebuschat and Williams then analyzed scores according to the participants' attested confidence and source identifications. With regard to the source attribution data, in Experiment 2, Rebuschat and Williams observed an accuracy rate of 59% when participants attributed their knowledge to intuition. Because this was significantly above chance, the authors suggest that accurate performance can be based on intuition and, further, interpret this as evidence of "unconscious structural knowledge" (2012b, p. 850).

A separate study by Tagarelli, Borges-Mota, and Rebuschat (2011) employed semi-artificial stimuli based on German verb placement rules, as in Rebuschat and Williams (2012b). Participants listened to 120 sentences and either judged their plausibility (incidental condition) or searched for underlying rules (rule-search condition). Once again, confidence ratings and source attributions followed the grammaticality judgment task. Both groups scored above chance on the grammaticality judgment task, though the explicit group scored better. Using source attributions, the study revealed that both groups scored above chance even when they claimed intuition as the basis for their judgments. Specifically, when attributing responses to intuition, learners in the incidental group scored 61%, and those in rule-search group scored 64%. Here again, the authors argued that these results constitute evidence of unconscious grammatical knowledge. It should be noted that these studies also employed a retrospective debriefing questionnaire. No participant in Rebuschat and Williams (2012b) could verbalize the rules using this measure and only three in the rule-search condition in Tagarelli, Borges-Mota, and Rebuschat (2011) could do so.

Other studies as well have employed source attributions and confidence ratings, providing further evidence upon which to assess the value of such methodological innovations. In one recent example, researchers combined verbal reports, confidence ratings, and an inclusion/exclusion task in order to assess L1 Cantonese speakers' awareness of Spanish word stress rules after training (Chan & Leung, 2012). One notable shortcoming of these approaches is that due to their retrospective nature, they can only possibly uncover implicit knowledge (see Section 3.2.5). Another recent study by Marsden, Williams, and Liu (2013) gathered source attribution data related to the initial learning of novel stems and suffixes, claiming that analyses revealed some implicit knowledge of both of these features.

At this stage, most commentary on the use source attributions to measure awareness has been addressed to studies of artificial learning (e.g., Norman & Price, 2010). Norman and Price raised at least five concerns with this methodology, including the following: (a) the application of source attributions in different tasks (e.g., AGL vs. SRT); (b) differing construct definitions of the categories (e.g., guessing, intuition, rule, memory) across different researchers and groups; (c) careful wording of category definitions within research studies; (d) interpretations of boundaries between categories (e.g. guessing vs. intuition, which are sometimes collapsed into one implicit category); and (e) individual or cultural variation in participants' use of such categories. Presently, the rapid increase of SLA studies adopting these measures warrants similar discussion regarding the validity and interpretation of subjective attribution judgments.

3.3.6 Interim summary

To recap, the null findings of Williams (2004, Experiment 2) as well as those of Hama and Leow (2010) demonstrate that L2 learning without awareness may be highly sensitive to experimental conditions and/or testing methods. Nevertheless, based on subsequent results, Williams has claimed that, "at least for some individuals, it is possible to learn form-meaning connections without awareness of what those connections are" (2005, p. 293). To the extent that they can be taken as support, the studies by Rebuschat and Williams (2012b) and Tagarelli, Borges-Mota, and Rebuschat (2011) showed evidence of implicit *knowledge*, as demonstrated by above chance judgments of syntactic rules (*not* form-meaning mappings), when participants claimed intuition as the source of their judgments. Finally, by using a reaction time methodology, Leung and Williams

(2011, 2012) showed effects in terms of behavior that reflects cognitive processes arguably beyond participants' conscious control (for discussion, see Norris & Ortega, 2003, p. 730). Accordingly, some researchers suggest that implicit learning of certain L2 constructions appears to be possible by some learners (Schmidt, 2012; Williams, 2009).

Table 3.1 indicates the L2 features targeted in the 20 reports reviewed in Section 3.3. Note that, because this table omits much important detail, it can, at best, be considered only an approximation of the research trends. As can be seen, various source languages have been drawn upon, and the kinds of L2 features investigated have mainly included morphology and syntax. Several studies also incorporated lexical learning. The evolving methodological context of implicit L2 learning research can be seen in this group of studies, in which the primary criteria for establishing a lack of awareness has been based on: (a) study designs assumed to involve unaware learning (research design), (b) conditions overtly labeled "implicit" (group assignment), or (c) use of subjective measures (see Section 3.4.3) to tap participants' awareness (e.g., think aloud, interview). Increasingly, studies use various combinations of subjective measures (see below for further discussion).

Table 3.1. *Overview of Studies on Implicit Second Language Learning (1986-2012)*

Study	Language	Linguistic feature(s)	Awareness criterion
Nation & McLaughlin (1986)	Artificial	Letter strings based on finite-state grammar	Research design
Hulstijn (1989)	Dutch	Semi-artificial Dutch & subordinate clauses	Research design
Ellis (1993)	Welsh	Consonant mutation rules	Research design
Alanen (1995)	Finnish	Locative suffixes & consonant changes	Research design
DeKeyser (1995)	Artificial	Plural, gender, & object suffixes	Group assignment
Robinson (1995a)	English	Fronted adverbials & pseudo clefts	Group assignment
Robinson (1997b)	Artificial	Verbs exhibiting dative alternation	Group assignment
de Graaff (1997)	Esperanto	Plurals, imperative, & negation	Group assignment
Leow (2000)	Spanish	Stem-changing verbs	Subjective measure
Williams (2004)	Artificial	Determiners for animacy	Subjective measure
Williams (2005)	Artificial	Determiners for animacy	Subjective measure
Hama & Leow (2010)	Artificial	Determiners for animacy	Subjective measure
Chen et al. (2011)	Chinese	Determiners for animacy & relative size	Subjective measure
Guo et al. (2011)	English	Semantic prosody using pseudo-words	Subjective measure
Faretta-Stutenberg & Morgan-Short (2011)	Artificial	Determiners for animacy	Subjective measure
Leung & Williams (2011)	Artificial	Determiners for thematic role	Subjective measure
Rebuschat & Williams (2012b)	German	German verb placement	Subjective measure
Tagarelli et al. (2011)	German	German verb placement	Subjective measure
Leung & Williams (2012)	Artificial	Determiners for animacy & relative size	Subjective measure
Chan & Leung (2012)	Spanish	Word stress rules	Subjective measure

3.4 Issues in research on second language awareness

Several outstanding issues presently characterize research on second language awareness. Below, three of these, which relate to notions of representation, learner differences, and L2 measurement, are discussed. Specifically, this section addresses the following questions:

- How might implicit and explicit knowledge interface?
- To what extent do implicit and explicit processes vary in their susceptibility to individual differences?
- Which measures have been used to study implicit L2 learning?

3.4.1 The explicit-implicit interface debate

Researchers have at times disputed not only the distinctiveness of implicit and explicit learning, but also interactions between these processes (Mathews et al., 1989). In SLA, this debate is usually framed in terms of noninterface, weak interface, or strong interface positions (N. Ellis, 1994; R. Ellis, 2009; Leow, 2007; Robinson, Mackey, Gass, & Schmidt, 2012). In their comprehensive review of the interface issue, Robinson et al. (2012) recounted arguments from the work of several researchers. Throughout, they offered an impartial summary of the debate, but concluded that, "conscious and unconscious processes operate in parallel" (p. 252), also stating that L2 learners gradually come to rely less on explicit knowledge and more on implicit knowledge as learning progresses. Four researchers' stances on the debate noted by Robinson et al., are briefly described in the following paragraphs.

First, Krashen's (1982) acquisition-learning distinction was an early formulation of the noninterface position. As noted earlier (Section 2.2.1), Krashen defined acquisition and learning as separate processes, the former of which concerns subconscious processes and the latter of which refers to learning about language. This model assumes that comprehensible input reaching brain areas responsible for language acquisition leads to acquired competence.

Second, more recently, Paradis (2004, 2009) has staunchly supported the noninterface position. He insists that there is evidence for neither a direct nor indirect connection between explicit metalinguistic knowledge and implicit linguistic competence. He also argues that any gradual substitution of explicit knowledge by implicit competence does not constitute an interface in the sense of two systems exchanging information across a boundary at a given point in time. Thus, according to Paradis' definition, while they may both contribute to L2 learning, implicit and explicit systems do not interact.

Third, N. Ellis (2005, 2011; Ellis & Larsen-Freeman, 2006) has defended the proposal that, while based on distinct neural architectures and knowledge stores, implicit and explicit learning do indeed interact during the conscious processing of an L2. Ellis has summarized his position with respect to the debate as follows:

> Krashen (1985) was correct to the extent that, as he termed it, acquisition and learning are different things; in psychological vernacular, explicit and implicit knowledge are distinct and dissociated; they involve different types of representation and are substantiated in separate parts of the brain…Paradis (1994) was correct in stating that explicit knowledge does not *become* implicit

> knowledge, nor can it be *converted* to it. Nevertheless, there is interaction. However unalike they are, these two types of knowledge interact. (2005, p. 307)

Ellis claims that "complex dynamic interactions of implicit and explicit knowledge" (p. 313) occur in consciousness, and that the extent of influence of metalinguistic knowledge on language processing justifies the assumption of an interface. This weak interface argument has other advocates, as well. In her discussion of explicit and implicit L2 knowledge, Roehr (2008) also adopted this position, claiming that:

> …the fine line between focused attention in the sense of stimulus detection and focused attention in the sense of noticing can be regarded as the threshold of conscious awareness, that is, the point of interface between implicit and explicit processes and representations. (p. 70)

Given the distinction between explicit and implicit knowledge that N. Ellis maintains, as well as Roehr's definitions of as explicit knowledge as categorical, stable, and context-independent and implicit knowledge as exemplar-based, flexible, and context-independent, consciousness is taken to be a workspace for mutual collaboration between two systems in L2 learning and performance, rather than a point of transmission.

Fourth, and last, DeKeyser, who has been associated with the strong interface position, has suggested that declarative knowledge may be converted to procedural knowledge that is "functionally equivalent" to implicit knowledge (2003, p. 329). However, in related work, Ullman (2005) distinguished between implicit/explicit knowledge and declarative/procedural memory, on the grounds that: (a) the former distinction relates to awareness and the latter to brain systems; (b) the mapping between implicit/explicit knowledge and declarative/procedural processes is not isomorphic; and

(c) declarative/procedural memory, in Ullman's model, interact with the L2 acquisition of lexis and grammar in specific ways, whereas no such predictions for implicit/explicit knowledge are available (pp. 160-161). In any event, a functional equivalence of the sort DeKeyser described does not seem to be compatible with the strong interface claim that implicit and explicit learning are closely intertwined.

3.4.2 The susceptibility of explicit versus implicit processes to IDs

The theoretical chasm between the implicit and the explicit runs deeper when one considers the role of individual differences in each type of learning. Briefly, researchers following a dual-process view argue for implicit and explicit L2 learning as distinct processes (e.g., Krashen, 1982; Reber, 2011). Others challenge this perspective by stipulating that L2 learning is a conscious process (Robinson, 1995, 2003; Schmidt, 1990, 2001, 2012). To a large extent, this debate echoes similarly conflicting assumptions among those in psychology who characterize these two systems in terms of their unique features (e.g., Reber, 1993; Dienes & Berry, 1997; Frankish & Evans, 2009), and those who remain skeptical of the existence of implicit learning (e.g., Brewer, 1974; Perruchet & Vinter, 2002; Shanks, 2005; 2010). Reber's (1993) prediction illustrates the consequences of these differing viewpoints for the study of learner factors in L2 learning: "we should expect to find fewer individual differences between people when implicit processes are in use than when explicit processes are" (p. 7). Has this prediction been upheld?

Elsewhere, Reber has stated that implicit learning should show less variability than explicit learning on measures such as intelligence (for discussion in SLA contexts, see Winter & Reber, 1994; Reber, 2011). Yet, when Robinson replicated the study upon which this claim is based (Reber, Walkenfeld, & Hernstadt, 1991), the results contradicted this claim. Robinson (2002b, 2005, 2010) reported on three experiments to investigate the generalizability of findings from the literature on artificial grammar learning to L2 populations. Experiment 1 replicated Reber et al. (1991). Reber and colleagues had assumed that, because implicit learning systems phylogenetically pre-date conscious functioning, they would show less variance than explicit systems and correlate less strongly with standard measures of cognitive ability (Reber, 1993; Reber & Allen, 2000). They demonstrated both of these outcomes, offering higher correlations between IQ and an explicit series-solution task (r = .69) than an implicit artificial grammar learning (AGL) task (r = .25) as supporting evidence (N = 20). The replication by Robinson, using an experimental group of 37 participants, also found less variance in scores on the implicit AGL task than on the explicit problem-solving task. In this same experiment, though, Robinson unexpectedly found a negative correlation between implicit learning and scores on an IQ test, which he attributed to high-IQ participants' unsuccessful attempts to process the letter strings analytically (but see also Gebauer & Mackintosh, 2007).

To add to this, there is a robust literature investigating the role of individual difference variables in implicit learning (see Jackson, 2013). Briefly, these variables have included: language learning experience (Nation & McLaughlin, 1986; Williams, 2005), working memory (Andrade & Baddeley, 2011; Bo, Jennett, and Seidler, 2011; Tagarelli et al., 2011), personality (Kaufman et al., 2010; Norman, Price, & Duff, 2006; Norman, Price, Duff, & Mentzoni, 2007; Pretz, Totz, and Kaufman, 2010; Woolhouse & Bayne, 2000), learning strategies (Nayak, Hansen, Krueger, & McLaughlin, 1987, cited in McLaughlin & Nayak, 1989), motivational states (Eitam, Hassin, & Schul, 2008; Eitam, Schul, & Hassin, 2009) and thinking styles (Kiyokawa, Dienes, Tanaka, Yamada, & Crowe, 2012). In general, these studies demonstrated positive associations between implicit learning and individual differences. Although the size of these effects is varied and often small, they show that the role of individual differences in learning without awareness remains an area of considerable interest.

How might the various individual differences in the preceding paragraph have mattered in the implicit learning experiments these studies described? Considering the cognitive processes outlined at the beginning of this chapter, one possibility is that attention is shaped by individual differences. Implicit learning relies on attention (Hsiao & Reber, 1998; Shanks, 2003), thus it would not be surprising if attentional capacity (see Chapter 4) mediated such learning. A second possibility is that learning involves both attention and awareness at the level of noticing, and that, as Schmidt (2012) has claimed, noticing is susceptible to experiential and cognitive factors. In this case, it stands to reason that implicit knowledge gained through a process of unconscious generalization across noticed instances would show the influence of individual differences. A third, related possibility is that learners possess complexes of abilities that are invoked by particular tasks, and that these underlie the relationship between learning outcomes and ability measures. This is the gist of Robinson's aptitude complex/ability differentiation model, discussed in the next paragraph.

In summary, Robinson (1995a, 1997a) offered a complementary hypothesis to the fundamental difference hypothesis, or the idea that L2 learning in adulthood differs from in childhood, by drawing not on universal grammar, but rather, on existing language knowledge and general problem-solving mechanisms (Bley-Vroman, 1989, 1990, 2009). Robinson's proposal, the fundamental similarity hypothesis, states that adult L2 learning is fundamentally similar across learning conditions because it relies on conscious processing in response to task demands (Robinson, 1997a, p. 85). Based on this, Robinson (2001b, 2002c, 2005a, 2007b) has advanced two further hypotheses. First, following Snow (1994), conscious processing engages distinct patterns of cognitive abilities, as called for by the setting and task. The aptitude complex hypothesis considers the hierarchal structure of such abilities for instructed SLA. Thus, an aptitude cluster related to learning from recasts would hypothetically be comprised of (a) abilities required to notice gaps (i.e., perceptual speed, pattern recognition) and (b) memory for contingent utterances (i.e., phonological working memory capacity and speed of working memory). Second, following Deary, Egan, Gibson, Austin, Brand, and Kellaghan (1996), during the transition from childhood to adulthood general intelligence decomposes into distinct sets of abilities. The ability differentiation hypothesis suggests that L2 learning abilities undergo a similar age-related change, resulting in more differentiated aptitude complexes among older learners which instruction must try to cater to in order for learning to be successful.

3.4.3 Methodological issues in studying L2 awareness

In cognitive psychology, two standards have been proposed to evaluate awareness tests (Shanks & St. John, 1994). The first of these is the information criterion, or the requirement for measures to elicit information responsible for learning. As Shanks and St. John stated:

> Before concluding that subjects are unaware of the information that they have learned and that it is influencing their behavior, it must be possible to establish that the information the experimenter is looking for in the awareness test is indeed the information responsible for performance changes. (1994, p. 373)

One potential obstacle to meeting the information criterion in artificial grammar learning is the abundance of theoretical accounts regarding knowledge representation. Pothos (2007), in a review of nine such accounts (rules, microrules, fragments, the chunking hypothesis, specific similarity, generalized context model similarity, single recurrent networks, PARSER, and fluency), noted that rules, similarity, and associative learning constitute distinct approaches to the form of underlying knowledge. The second standard is the sensitivity criterion, which requires that tests exhaustively measure knowledge pertaining to learning. These two guidelines can help evaluate measures used in L2 research to assess awareness, which are described next (for recent discussion, see also Leow, Johnson, & Záráte-Sandez, 2011; Rebuschat, 2013; Robinson, Mackey, Gass, & Schmidt, 2012).

Looking back at the studies reviewed in Section 3.3 above, the most popular approach to measuring awareness, has been to conduct participant interviews. Additional measures of awareness included think-aloud protocols, written questionnaires, and, more recently, confidence ratings and source attributions. These measures are here discussed in detail.

Interviews. As noted, many implicit L2 learning studies have employed retrospective interviews to investigate participants' awareness of rules (DeKeyser, 1995; de Graff, 1997; Williams, 2004, 2005; Leung & Williams, 2011, 2012). For example, Leung and Williams (2012) described a standardized interview in which participants were asked: (a) what they thought the study was about; (b) whether they noticed anything odd (presumably, the violation block); and (c) what feelings they had about the morphemes used.

Think-aloud protocols. Building on the classic methodology laid out by Ericsson and Simon (1993) for the study of cognition through verbalization, three studies on implicit learning used think-aloud procedures, each in different ways (Alanen, 1995; Hama & Leow, 2010; Leow, 2000). In the SLA field, Bowles (2010) reported a meta-analysis of 14 unique studies employing verbalization. One of her recommendations is that studies using verbalization also include a control (silent) group in order to gauge reactivity. Unfortunately, none of the studies reviewed above did this.

Written questionnaires. According to Brown (2001), questionnaires can tap knowledge about language learning processes (p. 31). Illustrative of this approach, Robinson has used written questionnaires as awareness measures in at least two studies (1994, 1997b). In both studies, items written in participants' L1 asked whether they had noticed rules, had been looking for them, and could describe them. After these were coded, they complemented subsequent analyses of the data.

Confidence ratings and source attributions. As already pointed out, two studies by Rebuschat and colleagues have utilized confidence ratings and source attributions to assess the extent of participants' awareness. Rebuschat and Williams (2012b) employed confidence ratings according to two criteria proposed by Dienes, Altmann, Kwan, and Goode (1995, p. 1325). The *guessing criterion* is met when participants who believe they are guessing respond accurately. The *zero correlation criterion* is met when participants' confidence in their correct decisions is as high as it is for their incorrect decisions. These two criteria are restricted to participants' judgment knowledge. Briefly stated, *judgment knowledge* (e.g., providing a response on a grammaticality judgment task) describes the knowledge constructed when participants decide whether a test item follows the same rules as training items. Basically, this means knowing that a particular exemplar is either correct or incorrect. In contrast, the underlying knowledge such decisions are based on, or knowledge of *why* something is correct or incorrect, is called *structural knowledge* (Dienes & Scott, 2005; Dienes, 2012). To determine the extent of access to structural knowledge, then, source attributions, which ask participants whether their responses are based on intuition, memory, rules, or guesses, were also used. Although Rebuschat and Williams (2012b) found no evidence of unconscious knowledge by the guessing and zero order criteria, they reported that participants scored above chance when attributing their knowledge to intuition (as did Tagarelli, Borges-Mota, & Rebuschat, 2011).

3.4.4 Overview

Put simply, establishing implicit cognition involves finding "a dissociation between performance and awareness" (Tunney & Shanks, 2003, p. 1060). The forgoing discussion illustrates ways in which such dissociations can be constrained in order to support claims of implicit knowledge. The information and sensitivity criteria may be thought of as principles guiding the design of studies on awareness. The interpretation of results benefits from specifying, in detail, which construct a given test is assumed to rely on (e.g., implicit memory, structural knowledge, or judgment knowledge). On a broader scale, a related issue is the classification of measures into *objective* and *subjective tests.*

In the context of measuring awareness, objective tests are defined as those requiring participants to generate responses based upon factual knowledge they have acquired (e.g., two-alternative forced choice tasks). On the other hand, subjective tests are those that elicit subjective responses based on participants' feelings, intuitions, or self-assessed mental states (e.g., confidence ratings) (see Tunney & Shanks, 1993; Shanks, 2005; Robinson, Mackey, Gass, & Schmidt, 2012; Ziori & Dienes, 2006). As Ziori and Dienes (2006) further explain, objective tests can indicate either conscious or unconscious knowledge when participants succeed. Because objective measures are largely indifferent to the reasons behind success or failure, subjective tests are necessary to determine whether participants' knowledge is conscious or otherwise. To summarize, Figure 3.1 classifies examples of measures used in L2 learning research on awareness according to this distinction, also incorporating another distinction deemed important in the L2 literature: whether measures are concurrent or retrospective to the learning phase of experiments.

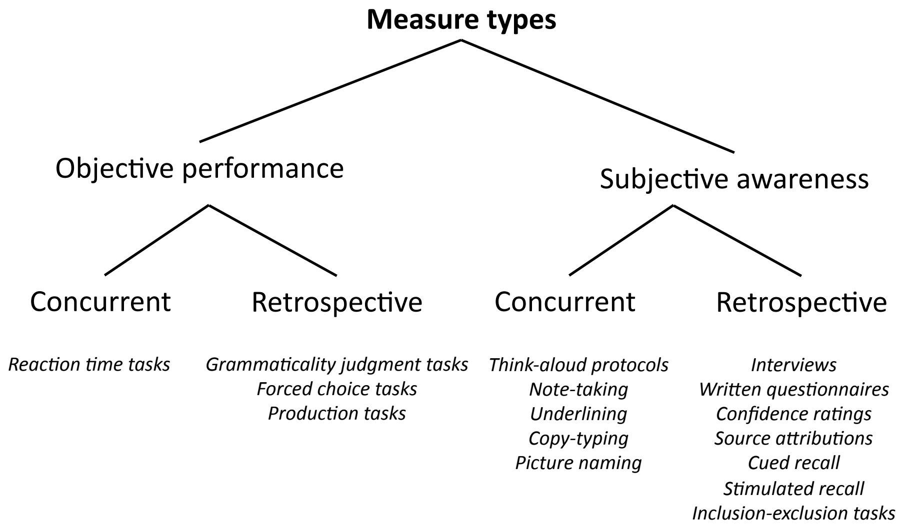

Figure 3.1. Measures adopted in L2 awareness studies.

What is essential about the objective/subjective distinction is that comparing results from different measures can yield triangulated interpretations about the nature of participants' underlying knowledge. Therefore, one imminent direction for future research on implicit learning is to continue to carefully design research that assesses learner knowledge by looking at its varied subjective dimensions. In doing so, it is also important to consider how the information and sensitivity criteria may guide choices among the measures listed in Figure 3.1. First, careful matching of objective and subjective tests may help to ensure that information gathered regarding performance and awareness, in theory, relates to a common underlying source of knowledge. Second, while it may be impossible to exhaustively measure awareness, utilizing multiple subjective measures appears to be one way of addressing the sensitivity criterion, since awareness encompasses many levels, including noticing, understanding, and metalinguistic awareness. Last, but not least, interpretations need to be based on a clear acknowledgement of the timing of measures—concurrent or retrospective—and an understanding of the validity concerns arising from the use of each.

3.5 Implications for the present study

Regarding terminology (Section 3.2), in the remainder of this dissertation, I will generally use the labels untutored or uninstructed learning to refer to the learning my participants demonstrated. Alternate, roughly equivalent expressions include learning outside formal instruction, learning in non-educational settings, and learning in experimental settings. As above, explicit learning is learning that is attended and aware. It can be either intentional or incidental. However, because incidental learning has two senses, both related to this study, when this term is used, I will specify whether this refers to having learned deliberately (sense 1) or having not been apprised of an upcoming test (sense 2). Following those who maintain the process/product distinction, I restrict my usage of implicit learning to the possibility of learning without awareness *at the time of encoding*. I will also use implicit knowledge, where appropriate, to describe accurate performance in the absence of awareness *at the retrieval stage*. This is consistent with the assumption that learning may occur as a process of unconscious generalization based on noticed instances. Importantly, awareness may be partial, as detailed later in this study.

Table 3.2 offers a guide to the meaning of these types of L2 learning, cross-referenced against the cognitive processes they invoke, when used strictly to refer to such processes, rather than instructional or experimental methods. The binary classification used for the sake of convenience here is not to deny that all of these processes are, in fact, continuous. Typically, one experiences a fluctuating level of attention, awareness, or intention in accordance with the ebb and flow of daily life.

Table 3.2. *Characterizations of L2 Learning in Terms of Cognitive Processes Invoked*

	Attention	Awareness	Intention
Explicit learning			
Intentional learning	+	+	+
Incidental learning	+	+	-
Implicit learning	+	-	-

In light of the review in Section 3.3, this dissertation seeks to contribute further evidence to assess the problem of whether semantically motivated morphological features can be learned without awareness. As can be seen from the earlier Table 3.1, studies have targeted bound morphology, including number (e.g., DeKeyser, 1995; de Graaff, 1997), but evidence of learning without awareness in these reports was limited. More recently, studies have demonstrated that participants show unaware learning of animacy, if simultaneously taught to attend to distance (Chen et al., 2011; Williams, 2005). Like previous studies, the present study targeted the acquisition of morphological markers indicating two semantic features (e.g., number and animacy). However, the investigation detailed here involved: (a) bound morphology; (b) concurrent learning of a set of concrete nouns; (c) no explicit instruction to focus on any aspect of the morphology. Thus, this study extends earlier work in terms of target forms and learning conditions. In particular, because neither feature is taught, the study permits examination of naturalistic learning under varying conditions of awareness. Because a high level of performance in such experiments requires attention to form and meaning, the stimuli in this experiment were presented in written and visual form.

Revisiting the key issues mentioned in Section 3.4, in the present study, first, concerning the interface debate, it is assumed that implicit and explicit knowledge are mutually, if not equally, facilitative of L2 learning and performance. I accept the weak interface argument in this respect. Although it is not entirely clear how the exchange of data between implicit and explicit systems could be conducted, particularly if their representational formats differ, perhaps the most likely candidate for modeling such exchanges in SLA resides in connectionist approaches (e.g., Ellis & Schmidt, 1997, 1998; see discussion in MacWhinney, 1997). Moreover, if the intent behind the weak interface position is merely to acknowledge a potential contribution to L2 learning of both types of knowledge, then it may indeed shed new light on adult SLA.

Secondly, on the susceptibility of different kinds of learning to individual differences, it seems advantageous to assume that any influence of IDs, no matter how small, is important. The findings of research on implicit learning and intelligence are mixed, recent studies have detected an influence of various IDs on such learning, and several explanations can be invoked for these findings. Because it clearly acknowledges the hierarchal structure of learner abilities and their multifaceted influence in specific settings, Robinson's aptitude complex hypothesis has potential to contribute to our understanding of this area. The present study similarly assumed that combinations of multiple factors, including awareness, working memory, and personality, influence learners' processing of novel L2 input, and their subsequent performance, and, furthermore, that input conditions may shape the degree to which these factors come into play. Thus, it is consequently argued that any performance based on implicit knowledge will not be unaffected by IDs, though this influence may also vary by input condition.

The third issue, of research methodology, basically concerns the internal validity of investigations of L2 awareness. This is dealt with in greater detail when discussing the study's methodology in Chapter 6. For the time being, I will consider two points following from the discussion above. To begin, the type of knowledge this study is concerned with is the comprehension of L2 form-meaning mappings. So, it is advisable to employ an objective measure that tests precisely this information. For this reason, a picture-word matching task (see Jiang, 2012) was chosen. Moreover, because the artificial morphological system in question could be described verbally (i.e., in terms of the general relevance of features or specific form-meaning mappings), an offline questionnaire would constitute an appropriately sensitive subjective measure of awareness. Any dissociation between performance on the picture-word matching task and the offline questionnaire could be considered as evidence for implicit knowledge. In these ways, the study attempted to address the information and sensitivity criteria.

To end with a concise summary, the methods employed by this study relate to the constructs described throughout this chapter in the following way. During the learning phase, learners' *attention* was focused by having them perform two online tasks: (a) word copying and (b) picture naming. Their *judgment knowledge* was elicited using the picture-word matching task (i.e., they were asked to classify each form-meaning pairing as a match or non-match). Their *structural knowledge* was evaluated by eliciting source attributions (i.e., guessing, intuition, rule, or memory) for each item. This knowledge type was also assessed by examining whether or not they could verbally report *awareness* of the semantic features associated with each morphological marker on a written questionnaire. The questionnaire also yielded data on their *intention* to learn.

CHAPTER 4

WORKING MEMORY AND L2 LEARNING

4.1 Introduction

Given this dissertation's focus on individual differences in SLA, this chapter focuses on one variable of substantial interest in recent studies, working memory. Because the notion of WM was originally postulated, and continues to be developed, outside the field of SLA, the chapter will first broadly review a portion of the voluminous literature on WM in cognitive psychology, especially that concerning construct definition and measurement. These two alone are major areas of interest in psychology. Yet, there is also abundant research and theory exploring the connection between WM and L2 learning. Within this body of research, work directly related to L2 contexts, processes and outcomes will be discussed. Lastly, the implications of WM research for the present study, which focuses on novel L2 learning (here, the learning of morphological constructions not previously encountered by the learner) will be summarized.

4.2 Current psychological perspectives on WM

This section first introduces theoretical frameworks articulated by various researchers. It then specifies characteristics that, in general, define the construct of WM. Later, changes in WM systems across the lifespan are discussed and, finally, the neural architecture underpinning WM is briefly addressed. Thus, this section covers theoretical, developmental, and biological perspectives on WM.

4.2.1 Theoretical models of working memory

Three models widely discussed in the WM literature, and, to a good extent, in L2 research, are presented in this section. Each having evolved over the primary author's career, these models have supported a large body of empirical work and generated ongoing theoretical discussion. After each model is introduced separately, other noteworthy perspectives on WM are mentioned. The three models reviewed here are depicted together in Figure 4.1.

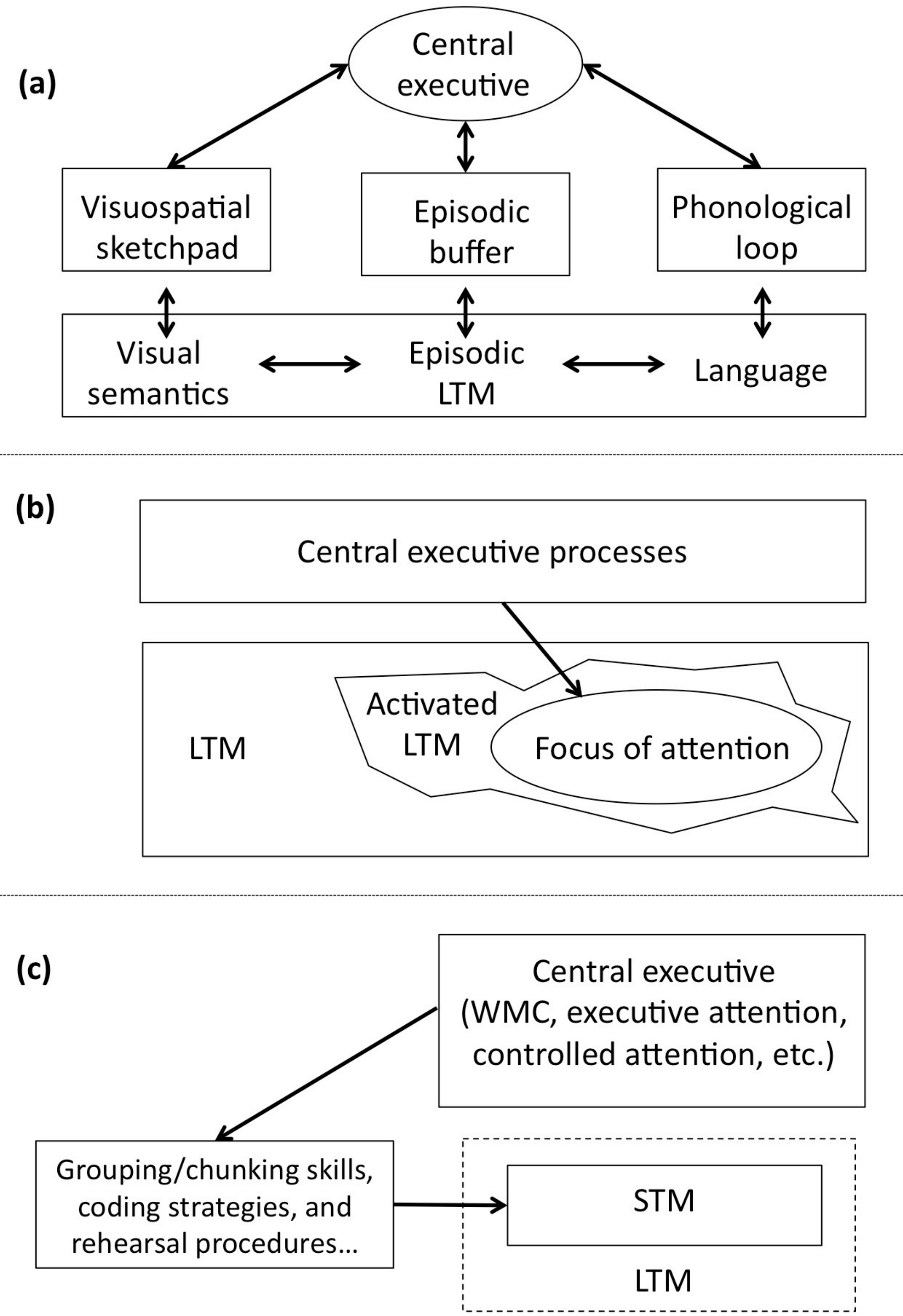

Figure 4.1. Three WM model visualizations, adapted from primary sources: (a) Baddeley's multicomponent model (Baddeley, 2000), (b) Cowan's embedded-processes model (Ricker, AuBuchon, & Cowan, 2010); and (c) Engle's measurement model of the WM system (Kane, Conway, Hambrick, & Engle, 2007). LTM = long-term memory; STM = short-term memory; WMC = working memory capacity.

4.2.1.1 Baddeley's multicomponent model

The foremost name associated with the theoretical development of WM is that of Alan Baddeley. Based on a series of experiments, Baddeley and Hitch (1974) were the first to elaborate on WM as a workspace of limited capacity to handle storage and manipulation of information. Their proposal extended the notion of short-term memory, which refers to storage only, in order to highlight the WM system's functional role (Baddeley, 2007). The specific details of the model have evolved over time, from the original three-component model, which included a central executive and two subsystems, the phonological loop and the visuospatial sketchpad, to the addition of a fourth component, the episodic buffer (Baddeley, 2000).

Presently, the multicomponent model consists of a central executive assumed to carry out four processes: (a) focusing, (b) dividing, and (c) switching attention, as well as (d) linking WM and long-term memory. As above, this executive component connects to three subcomponents. The phonological loop and the visuospatial sketchpad allow storage and rehearsal of information in STM that is phonological and articulatory, or visual and spatial, respectively. The latest component, the episodic buffer, serves to bind information from various sources into episodes. Recent accounts of the model retain this structure, while adding detail related to the information processed by each component (Baddeley, 2012; see also Baddeley, 1986; Baddeley & Logie, 1999; Gathercole & Baddeley, 1993).

4.2.1.2 Cowan's embedded-processes model

By comparison with the multicomponent model, one difference in Cowan's model of information processing (1988, 1995, 2005) is the emphasis placed on the intersection between memory and attention. In order to understand the mutual influence of these two areas on each other, this model distinguishes between two phases of memory activation: an initial, brief sensory store and a longer, short-term store, which is the activated subset of long-term memory, rather than a separate component. A central executive guides attention toward either external stimuli, or internal, long-term memories. Memory activation is limited in time, with its two phases unfolding on distinct timescales, while attentional focus is limited in capacity to about four chunks in normal adults (Cowan, 2000). Together, these features have been described as an embedded processes model, in which mechanisms of memory activation, attention, and long-term memory jointly comprise working memory (Cowan, 1999). Accordingly, this model, more than those that describe WM as control over executive attention, emphasizes the interaction of various processes at a general level (Cowan, 2005, 2008).

4.2.1.3 Engle and colleagues' resource-dependent inhibition model

Research by Engle and associates has contributed the theoretical insight that inhibition, or the ability to disregard irrelevant information during tasks[12], is also crucial to working memory (Conway & Engle, 1994; Engle, Conway, Tuholski, & Shisler, 1995). In an overview of this approach, Engle, Kane, and Tuholski (1999) illustrated the connections between the central executive, long- and short-term memory, and procedures

[12] For example, the picture-word matching task in the present study required participants to disregard certain cues on wrong items (see Section 6.4.3.4).

for maintaining activation (see Fig. 1), arguing that, “individual differences on measures of working memory capacity primarily reflect differences in capability for controlled processing” (p. 104). This model has also shed light on the relationships between WMC, general fluid intelligence, and executive attention. A review by Hambrick, Kane, and Engle (2005) describes support from studies linking WM tasks and attentional control tasks, as well as other studies which link WM and fluid intelligence. Given its emphasis on domain-general processes of maintenance and retrieval in the face of interference, this model has also come to be labeled, “an executive attention theory of WMC” (Kane, Conway, Hambrick, & Engle, 2007, p. 22; see also Engle, 2002).

4.2.1.4 Other perspectives on WM

Many other views on WM have been developed, three of which are briefly reviewed here. First, Ericsson’s model posited no fixed capacity limits, instead focusing on how expert domain knowledge supports WM during familiar tasks, such as bicycling or reading, through the mechanisms of long-term working memory and short-term working memory (Ericsson & Kintsch, 1995; Ericsson & Delaney, 1999). Second, Jonides and colleagues (2008) made a substantial contribution by targeting the (a) structure, (b) processes, and (c) failures of short-term memory, putting forth a layered model specifying cognitive, neural, and task events as they unfold in short-term memory. Third, Oberauer (2009) spelled out requirements for the WM system, theorized that it consists of separate declarative and procedural systems, and related its architecture to dual-process theories by positing analytic and associative processing modes. These are only a few, recent alternate perspectives on WM. For more extensive systematic reviews and comparison of multiple theories, readers are referred to the volumes edited by Miyake and Shah (1999) and Conway, Jarrold, Kane, Miyake, and Towse (2008).

4.2.2 General characteristics of working memory theories

This section describes general characteristics of WM theories, in order to expand on the definition provided in Chapter 1: WM is fundamentally a system of temporary storage and attentional control. Several general features of WM appear across the main three models reviewed above. There are varying degrees of consensus about each of these characteristics.

4.2.2.1 Central executive as a key mechanism

In each of the aforementioned models posited by Baddeley, Cowan, and Engle, the central executive seems to personify WM in spite of any differences in emphasis arising amongst these theorists. Indeed, some regard the label working *memory* as a misnomer, because many theories stress the aspects of attention and control, rather than memory storage (Kintsch, Healy, Hegarty, Pennington, & Salthouse, 1999, p. 438). Baddeley (2007, 2012) has lamented the fact that the central executive was under-theorized in his earlier work, even though he recently elaborated on this component, by delineating attentional processes belonging to it (i.e., focusing, dividing, switching attention). From the beginning, Cowan's model, too, has included a central executive to direct attention and control processing (1988, 1995, 2005). Based on this, Cowan (2005) raises important, unanswered questions about the relationship between attentional control and attentional capacity (see Section 4.2.2.3). In Engle's work, attention takes center stage, particularly for the reason that differences in WMC are attributed to differences in the control of attention. Here, the central executive absorbs a range of processes: WMC, executive attention, controlled attention, and so on. Furthermore, each author gives precedence to the central executive by placing it above all other components in visual representations (see Fig. 4.1).

4.2.2.2 Non-unitary nature of the system

The three models align with slightly different perspectives on the division of WM into domain-specific components. Just as there are domain-general aspects of WM, such as attention, that are shared across theories, so, too, are there domain-specific components. In other words, researchers have tended to accept a non-unitary view of WM (Miyake & Shah, 1999). This is obviously true in the case of the multicomponent model, which from its early days has suggested roles for visual and auditory subsystems (Baddeley & Hitch, 1974). Cowan (1999) remarks that his model appears more unitary than the multicomponent model. He states that activated memory, at the core of the model, may take the form of any modality, arguing that, “different codes may be processed according to the same principles” (p. 79). Similarly, the group led by Engle regarded WMC as domain free (Engle, Kane, & Tuholski, 1999), supporting this assertion with evidence that separate WM span tasks performed in verbal and spatial modalities correlated more strongly with each other than did verbal and spatial short-term memory tasks. Based on this, they concluded that WMC is primarily domain general and secondarily domain-specific (Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004).

4.2.2.3 Multiple determinants of capacity limits

Miller (1956) identified limits on memory and attention on the order of seven plus or minus two, noting in conclusion that we can expand this bottleneck by organizing input into chunks. WM research provides insight into such limits, though there is little in way of consensus. First, Baddeley (2007) reviewed three distinct hypotheses: the speed hypothesis, the resource pool hypothesis, and the inhibition hypothesis. In sum, according to Baddeley, processing speed and storage impose separate limits, performance may be limited by storage capacity or executive capacity, and the role of inhibition in determining capacity limits appears unresolved. Next, Cowan (2000, 2005) has built a case that it is the focus of attention within active memory which is limited. Rather than accepting Miller's magical number, he proposes that adults can process a fixed limit from three to five chunks. Cowan concedes that this is not an agreed-upon hypothesis—he has summarized ten contrasting viewpoints on the issue of capacity limits (2000). Lastly, Engle's view is oriented toward attentional control. As he put it, "although individuals possessing different WM capacities will show differences in the number of items stored in a variety of memory tasks, this is a result of differing ability to maintain and inhibit information, particularly in the face of distraction and interference" (2002, p. 21). In sum, of the three, Baddeley seems the most sympathetic toward the notion of capacity limits in both storage and processing, while others appear to agree slightly more with Craik and Lockhart's (1990) suggestion, that "limited capacity turns out not to be an invariant structural property of a memory system, but a limitation imposed by processing" (p. 103).

4.2.3 WM and the lifespan

The WMC of children, young adults, and older adults varies. This is invaluable for understanding SLA, particularly considering Newport's (1990) 'less is more' hypothesis, which attributed variation in language learning ability across the lifespan to "differences between adults and children in the way linguistic input is perceived and stored" (p. 23). Newport pointed out that children's more limited STM capacity might give them an advantage over adults in learning morphology. Gathercole and Baddeley (1993) documented some of the increases in the operating efficiency of WM from about four years of age. These authors explained that while the phonological loop, central executive, and visuospatial sketchpad are present in young children, the functioning of each system changes dramatically throughout early childhood (see Hitch, 2006, for a recent review drawing similar conclusions). In addition to these early changes, there is clear evidence for a gradual decline in WMC from the 20's onward. Park and Payer (2006) attribute the age-related cognitive decline associated with WM to processing speed in their review, noting, as well, the contrast with verbal abilities, which steadily increase with age. Clearly, SLA research seeking to understand the role of cognitive abilities among learners of various ages must not ignore these findings.

4.2.4 The neural correlates of WM

A spate of research in the early 1990s helped to identify neural regions activated during various WM tasks. Specifically, activation within the prefrontal cortex was found for verbal tasks (Petrides et al., 1993), for spatial tasks (Jonides et al., 1993), and for concurrent semantic and spatial tasks (D'Esposito et al., 1995). Building on this, a further area of investigation has been the possibility of a left/right hemispheric dissociation for verbal versus nonverbal WM tasks, as well as a dorsal/ventral dissociation for spatial versus nonspatial tasks (e.g., Nystrom et al., 2000). A meta-analytic review of 60 studies by Wager and Smith (2003) yielded some evidence for these dissociations. At a course-grained level, it is clear that the neural activity evoked by WM tasks is distributed throughout the brain. However, further basic research is needed in order to develop an account of WM performance grounded in cognitive neuroscience.

In one relevant study, these basic findings provided a foundation for research on IDs in WMC. Osaka and Osaka (2007) explored the distinct patterns of distributed activation in high- and low-span participants. Using functional magnetic resonance imaging, their investigations have encompassed the dorsolateral prefrontal cortex, the anterior cingulate cortex, and the superior parietal lobule, which they associate with maintaining attention, monitoring conflict, and focusing attention, respectively. They present evidence to support the hypothesis that connectivity between these regions accounts for IDs in performance on WM tasks, such as reading or listening span.

4.3 The measurement of WMC

Having reviewed theoretical, developmental and biological perspectives, it is now appropriate to consider the measurement of WMC. As described previously, WM is not a unitary construct, so it is not surprising to find a vast quantity of measures designed to tap its various components. For example, in their meta-analysis, Ackerman, Beier, and Boyle (2005) listed 59 different WM tests (alongside a similarly large number of STM tests).

For the purpose of this review, I discuss measures of STM and WM (also distinguishing between complex and dynamic WM measures). In the models above, STM generally refers to the storage component of WM. Therefore such measures are relevant to the present chapter. This brief overview will follow the organization used by Juffs and Harrington (2011), but will also incorporate the complex/dynamic distinction made by Conway et al. (2005).

4.3.1 Short-term measures

Measures of STM include simple span measures and phonological STM. Among simple span measures (Jacobs, 1887, cited in Kane et al., 2007), there are tests such as digit span and word span. These measures require examinees to read or listen to sets of items and recall them in serial order. The participants' L1 or L2 may be used. A widely used measure of phonological STM is the nonword repetition task (e.g., Gathercole & Baddeley, 1990; Service, 1992). In this case, participants are asked to recall pseudowords, which are typically presented aurally, either individually or in a list. Because the form of the items can be manipulated, potential confounds due to prior language knowledge are controlled. For instance, Service (1992) employed two lists: one consisting of English-sounding pseudowords (e.g., *disajoinance*) and the other, Finnish-sounding ones (e.g., *hainuksia*).

4.3.2 Complex measures

Perhaps the most popular approach to assessing WMC is through the use of complex span measures. These measures consist of instruments like the reading span test

(Daneman & Carpenter, 1980). During this test, participants read a set of sentences and then recall the last word of each sentence (e.g., given the sentence *The taxi turned up Michigan Avenue where they had a clear view of the lake,* participants recall *lake,* Daneman & Carpenter, p. 453). The number of items in a set varies from, for example, two to six sentences. Span measures have been developed within a wide range of knowledge domains; some variations include operation, counting, rotation, symmetry, and navigation span (for details of these, see Kane et al., 2004).

4.3.3 Dynamic measures

Dynamic measures of WMC require participants to "monitor a continuous stream of stimuli, often of uncertain length, and to respond according to only a subset of the stimuli presented" (Conway et al., 2005, p. 779). Examples include the running span and *n*-back tasks. In the *n*-back, participants monitor stimuli presented at a predetermined rate and respond if the current item is identical to the one *n* items before it, according to visual, auditory, or even olfactory characteristics of the stimuli (see Owen et al., 2005 for review).

4.3.4 The criterion validity of WMC measures

Individuals who perform highly on tests of WMC often show high levels of performance on other measures of cognitive ability, including, for instance, fluid intelligence and language comprehension.[13] First, intelligence tests have incorporated memory span measures (e.g., the Wechsler Adult Intelligence Scale includes a WM

[13] The relationship between WMC and personality will be discussed in Chapter 5.

subtest). But more to the point, recent research on intelligence (*g)* has been informed by Cattell's (1943) distinction between fluid intelligence (*Gf*) and crystallized intelligence (*Gc*). Therefore one possibility entertained by researchers is that WM underlies *Gf*, or a person's capacity, from birth, for general problem solving (see Ackerman, Beier, & Boyle, 2005*;* Conway, Getz, Macnamara, & Engel de Abreu, 2011; Hambrick, Kane, & Engle, 2005). Though the exact nature of this relationship remains a mystery, Pretz and Sternberg (2005) noted that WMC explains roughly 35% of the variance in *Gf.* In reviewing these findings, they further remark that measures tapping executive attention appear to underlie the relationship between WM and fluid intelligence.

Second, concerning language learning and processing, in their classic paper, Daneman and Carpenter (1980) appealed to working memory (WM) as an explanation for individual differences in reading comprehension. A later meta-analysis by Daneman and Merikle (1996) reported average weighted effect sizes for verbal WM measures and specific language comprehension tasks. Importantly, the correlation was higher when these measures involved storage *and* processing ($r = 0.52$), rather than storage alone ($r = 0.40$) and confidence intervals for these two figures did not overlap. Although the role of WM in linguistic processing has been fiercely debated (see Waters & Caplan, 1996; MacDonald & Christiansen, 2002), a study by Roberts and Gibson (2002) demonstrated that several measures (*n*-back and a composite measure of reading, math, and category span) accounted for over 50% of the variance in performance on a task requiring participants to listen to clauses of varying length.

4.3.5 Summary of WMC measures

The differences between the three classes of measures can be understood in terms of task demands. Simple span measures involve only recall of stored material (STM), whereas complex measures invoke both recall and processing (WM). Dynamic measures also do this, but, in the case of *n*-back, require encoding, storage, rehearsal, matching, ordering, inhibition, and response processes (Jonides et al., 1997, p. 471) in a speeded task. To the extent that these specifications map on to relevant processes in language learning and comprehension, each type of measure is clearly important for understanding how cognitive abilities relate to SLA.

4.4 WM and SLA: Conditions, processes, and outcomes

In this section, I present a review of studies on WM and SLA. Indeed, there is an abundance of research on this topic, much of which has been reviewed elsewhere (see N. Ellis, 2001, 2005; Juffs & Harrington, 2011; Robinson, 2003; Wood Bowden, Sanz, & Stafford, 2007; Watanabe & Bergsleithner, 2006; Wen, 2012, Williams, 2012). So, rather than attempt to comprehensively account for this large subfield of SLA research, this section looks at studies addressing the relationship between WM and (a) learning conditions, (b) other L2 cognitive processes, and (c) linguistic outcomes. Within these three fundamental areas of interest, additional distinctions are made, and findings relevant to the role of WM in novel L2 learning are summarized.

4.4.1 Learning conditions: Input delivery, modality, and modification

First, regarding input delivery, Erlam (2005) reported a classroom study based on learning French in three treatment groups: deductive instruction, inductive instruction, and structured input. Each group contained roughly 20 students. In this case, WM ability, as measured using a test of phonological capacity, correlated significantly with immediate and delayed tests of written production in the structured input group ($r = .49$ and .57, respectively). In the other groups, correlations between WM and outcomes were mostly positive, though not significant. Though it is difficult to draw conclusions based on the large number of correlations run in this study (besides written production, WM was paired with 8 other outcome measures per group), it seems likely that such outcomes may depend to some degree on the specific learner-internal processes invoked by instructional type. Thus, carefully controlled laboratory studies may also offer clear insight into this issue.

For example, a laboratory study by Tagarelli, Borges-Mota, and Rebuschat (2011) further shows that what learners are expected to do with the input matters. These authors reported an artificial language learning experiment employing incidental and rule-search conditions. In this study, two working memory tasks (letter-number ordering and operation span) were administered. Both groups scored above chance, though the rule search significantly outperformed the incidental group. The rule-search group ($n = 31$) showed a significant correlation of $r = 0.47$ between the grammaticality judgment task and the letter-number ordering task. No such correlations, however, were found for the incidental group. Previously, Robinson (1995) reported similar results using the paired associates subtest from the MLAT as a measure of memory. In that study, significant correlations were found for rule-search and instructed groups, but not for an incidental group. Based on the admittedly different studies reviewed in this section, it appears that WM may interact with how input is delivered, but more work on this topic would surely be a welcome addition to the literature.

Next, a separate issue is that of input modality. Considering the bifurcation between phonological and visuospatial subsystems in the multicomponent model, one might ask whether WM effects have been found for aural versus written input. This is an important question, though the matter is slightly more complicated, as several studies show. The following brief review covers studies showing effects of WMC on performance in aural and written tasks, which, interestingly, do not appear to be dependent on the modality of the WM measures used.

Robinson (2005b, see also 2002b) conducted a laboratory study on the incidental learning of Samoan by 37 Japanese college students in which WMC was assessed using an L1 reading span task (RST). In this case, WM significantly correlated with scores on a listening GJT, $r = .42$. Furthermore, Kormos and Sáfár's (2008) study of Hungarian EFL learners used two verbal WM tasks: nonword repetition and backward digit span task (DST). Based on a sample of 45 beginning learners, the latter task showed a significant, moderate correlation with listening ability, $r = .37$.

The processing of written input also appears to be influenced by WMC. One of the earliest studies in this area, carried out by Harrington and Sawyer (1992), showed that L2 reading span correlated significantly with L2 reading comprehension in a group of 34 Japanese learners of English. Taken together, these results point to WM effects in the written modality.[14] A recent study by Alptekin and Erçetin (2011) extended these results. Based on results from 62 Turkish EFL learners, they found that L2 reading span predicted L2 reading comprehension on items requiring readers to make inferences. Lastly, other studies have adopted measures of WM in both L1 and L2, including Geva and Ryan's (1993) investigation of L1 English-speaking children learning L2 Hebrew. These authors reasoned that the use of both languages would provide a more controlled design. A model fit to the data included Hebrew opposites span and word span among significant predictors of L2 reading, as measured by a cloze test. These measures were based on listening to words and supplying antonyms (Hebrew working memory opposites, or HWM-O) and listening to words and repeating them (word span).

Thus, associations with WM clearly do arise in both listening and reading tasks. Table 4.1 furthermore reveals that significant relationships have been reported irrespective of the formats of the WM and L2 test. That is, studies using WM tests based on the written modality have found relationships with L2 outcomes in the aural modality and vice versa. This is not to deny that, as argued by Wen (2012), it may be preferable to employ measures relying on the same modality at times. There is mixed evidence as to whether effects occur across modalities. Consider, for instance, Taguchi's (2008) null finding. In that study, no significant correlation was found between L1 reading span and an L2 listening measure. The point is simply that, like the decision to adopt L1 versus L2 WM tasks, various rationales can be presented for pairing WM tests with L2 measures.

---

[14] This study also revealed the language-independent nature of WM measures. Digit, word, and reading span in the participants' L1 and L2 were significantly correlated. Similarly, Osaka and Osaka (1992) reported strong cross-language correlations using English and Japanese versions of the reading span.

Table 4.1. *Four Studies Reporting Significant Relationships between WMC and L2 Abilities*

		L2 measure	
		Listening	Reading
WM test	Spoken	DST × Listening test (Kormos & Sáfár, 2008)	HWM-O × Cloze test (Geva & Ryan, 1993)
	Written	RST × Listening GJT (Robinson, 2005b)	RST × Inference test (Alptekin & Erçetin, 2011)

Lastly, concerning input modification, it has been suggested that WM may be among the ID variables related to learners' use of modified input, based on a seminal investigation by Mackey, Philp, Egi, Fujii, and Tatsumi (2002). Their study found a marginally significant relationship between a composite measure of WMC and noticing of recasts during a stimulated recall task. Also, there was stronger evidence of sustained language development in the high rather than low WM group, although this was based on a comparison of only seven learners.

Other recent studies contribute to understanding the role of WM in learning from face-to-face L2 interaction. An investigation by Mackey, Adams, Stafford, and Winke (2010) and a small-scale study of older L2 English learners ($N = 9$) by Mackey and Sachs (2012) both suggest that WM may influence L2 outcomes in environments where input is partially reformulated for the sake of elucidating form-meaning connections. Another study by Révész (2012) also reported results showing that the performance of learners in a recast treatment group on written and oral measures of L2 learning was related to L1 reading span and oral WM tests. Lastly, Goo (2012) reported that reading and operation span tests predicted learning from recasts, but did not predict learning from metalinguistic feedback. His explanation for this finding is that learning from metalinguistic feedback may demand less cognitive control than recasts.

Studies employing computer-delivered input modification have yielded additional, though somewhat conflicting results. Sagarra (2007) reported that L1 reading span scores predicted ability to learn from computer-delivered recasts in her study, whereas Trofimovich, Ammar, and Gatbonton (2007) found that attention control, but not L1 letter-number sequencing, corresponded to learning in their study. Following up on this, Sagarra and Abbuhl (2013) linked reading span scores to the performance of groups receiving computerized recasts in the oral, but not written, modality. These authors argue that the inconsistent findings of Trofimovich et al. may be due to the timing of outcome measures, positing a delayed effect for WM on language performance.

4.4.2 Cognitive processes: Noticing, understanding, and aptitude

To begin, WM is closely linked to noticing, or the subjective awareness of surface features of the input (Schmidt, 1990, 1995, 2001, 2012). Schmidt has specified that, "noticing is related to rehearsal within working memory and the transfer of information to long-term memory, to intake, and to item learning" (1993, p. 213). Robinson has expounded on this relationship by characterizing noticing as detection plus rehearsal and awareness in WM (2003, see also 1995b). As noted in Chapter 3, Robinson assumes that rehearsal can be of two types: maintenance and elaborative rehearsal (Craik & Lockhart, 1972, 1990). These correspond to data-driven versus conceptually driven learning processes in long-term memory.

To date, few studies have operationalized noticing such that links between WM and the learning mechanisms described above can be established. One challenge lies with finding the appropriate methodological tools to observe noticing behaviors. Retrospective measures of noticing have been used for this purpose. The above-cited paper by Mackey et al. (2002) utilized stimulated recall protocols (Gass & Mackey, 2000) to demonstrate that those with higher composite scores on nonword recall, L1 listening span, and L2 listening span displayed more instances of noticing related to target forms during videotaped interactions.

Next, at higher levels of awareness, such as metalinguistic understanding, the role of WM is less clear. Here, the bulk of the evidence suggests that any such relationship is highly sensitive to the measures adopted. An early study by Ellis and Sinclair (1996) tested explicit rule learning of novel Welsh constructions, finding that participants who repeated input aloud during the learning phase outperformed those whose WM function was suppressed by having them recite an unrelated number sequence. Thus, the quality of input processing in WM may also be crucial when learning is based on awareness at the level of understanding.

However, three other studies point to the conclusion that metalinguistic awareness may involve processes beyond WMC. Robinson (2002b), who found no association between a reading span task and an unspeeded GJT, noted that these measures do not exhibit process continuity (p. 256). Bell (2009) showed that inductive language learning ability, but not reading span, predicted metalinguistic awareness of French gender rules. Lastly, Roehr and Gánem-Gutiérrez (2009) also found no significant relationship between span tests and untimed measures of metalinguistic knowledge. They also suggested this result was due to the incompatibility of measures.

The last psychological construct to be considered in this section is aptitude. The idea that WM is closely related to language aptitude has received support from cognitive psychologists (Miyaki & Friedman, 1998), aptitude test developers (Stansfield & Winke, 2008), and SLA researchers (Skehan, 2002). Yet, there is quite a range of positions on their precise relationship. Miyake and Friedman's (1998) WM-as-language-aptitude proposal suggests that WM may be the primary construct behind aptitude for language learning. Others hypothesize that WM plays a recurring role in various aptitude constructs corresponding to stages of L2 processing such as segmentation, noticing, and pattern identification (Dörnyei & Skehan, 2003). Still others see WM as one of a number of cognitive resources underlying aptitude complexes which apply in a given task and setting, such as learning from recasts (Robinson, 2001b). Thus, as the traditional notion of aptitude derived from the use of the MLAT—which included phonetic coding ability, grammatical sensitivity, memory abilities, and inductive language learning ability (Carroll, 1990)—gives way to newer aptitude theories, the status of WM within them varies considerably.

Two recent studies embody these theoretical tensions. Sáfár and Kormos (2008) established significant, moderate correlations between the Hungarian version of the MLAT and a backward digit span task in a study of IDs among EFL learners. Based on a regression analysis including both as independent measures, they argued, "working memory is a better predictor of language learning success than the traditional construct of language aptitude" (p. 129). Building on Carroll's work, Doughty et al. (2010) conducted a study on the reliability of the High-level Language Aptitude Battery (Hi-Lab). In a factor analysis of the test battery, WMC was one of several factors, which also included tolerance of ambiguity, task switching, perceptual acuity, rote memory, and speech perception in noise. As such, WM seems unlikely to supersede the construct of aptitude entirely, though it stands out as a significant component in newer conceptualizations of language aptitude (see also Granena & Long, 2013).

4.4.3 Linguistic outcomes: Word meaning and morphology

As already implied, numerous L2 outcomes have been investigated in relation to the potential role of working memory or short-term memory. Included among these are the comprehension and production of written and spoken language, as well as the online processing of written sentence structure (e.g., Alptekin & Erçetin, 2011; Erlam, 2005; Juffs, 2004; Kormos & Sáfár, 2005; Weissheimer & Mota, 2009; see Juffs & Harrington, 2011 and Williams, 2012 for two recent reviews). In this section, I will narrow down the review in order to focus on the specific outcomes addressed in the current study. Given the artificial L2 constructions used in this study, the most relevant outcomes are vocabulary and morphology. In particular, I focus on studies in which participants learned word meanings and bound morphology with no prior knowledge of these target forms.

Several studies have offered empirical evidence linking WM and the acquisition of word meanings. For example, Atkins and Baddeley (1998) used a training method in which learning trials presenting 32 adult participants with English-Finnish sentence pairs were interspersed with test trials requiring them to translate from English to Finnish.

Verbal WM span (but not visuo-spatial span) was significantly related to their rate of learning, in that, as span scores rose, the number of test trials until participants provided correct translations fell, *r* = -.52. Next, Speciale, Ellis, and Bywater (2004) reported two studies on the acquisition of foreign language vocabulary by undergraduates. In these studies, correlational analyses showed that nonword repetition was significantly related to oral production of English-to-German translation equivalents (Study 1, *N* = 38) and to lexical comprehension on a Spanish course examination (Study 2, *N* = 44). Also, a recent laboratory training study by Martin and Ellis (2012) reported correlations between a composite measure of artificial vocabulary learning and three independent measures: nonword repetition, nonword recognition, and listening span, all of which showed significant correlations (ranging from *r* = .34 to .40), based on 40 participants. In sum, working memory capacity clearly supports the ability to translate to and from a new lexicon (for reviews, see Baddeley, Gathercole, Papagno, 1998; Ellis, 1996).

Studies have also explored the role of WMC in the development of morphological knowledge. A study by Williams and Lovatt (2003) targeted the role of phonological STM (PSTM) in acquiring determiner-noun agreement rules in novel input. In the first experiment, involving Italian stimuli, no significant relationship between nonword repetition and generalization was found based on a partial correlation controlling for language background. However, in their second experiment, Williams and Lovatt used artificial stimuli to reduce the influence of language background with 21 participants. This time, the mediating influence of language background on PSTM disappeared, and PSTM and language background contributed independently to generalization (see Williams, 2012 for further discussion). In the last test cycle, this correlation reached *r*

= .53, which was significant. Memory for morpheme combinations (among trained items) was also a significant predictor of generalization (*r* = .64). Here, free morphemes encoded grammatical gender, definiteness, and number, in contrast to other studies on bound forms (see below).

Brooks and Kempe have conducted a series of investigations to assess the contribution of IDs to the learning of bound morphology in Russian, particularly nominal inflections marking gender and case. An overview of their work suggests that it is not only WM, but measures of intelligence, similarly involving control of attention, such as Cattell's Culture Fair Intelligence Test, that predict outcomes. For instance, they have argued that "a capacity for memorizing unfamiliar items along with an ability to effectively allocate attention to the analysis of distributional characteristics of the input is crucial for learning complex inflectional paradigms" (Kempe & Brooks, 2008, p. 742). These researchers have demonstrated how learning is shaped by the interaction of these cognitive factors and (a) type variation in the input (Brooks, Kempe, & Sionev, 2006), (b) linguistic characteristics of the input (Kempe, Brooks, & Kharkhurin, 2010), and (c) metalinguistic awareness emerging from training (Brooks & Kempe, 2013). Specifically regarding WM, Brooks, Kempe and Sionev (2006) reported significant relationships between reading span and learning of a case marking paradigm on old items (*r* = .69) and on new items (*r* = .53) with 20 participants, when the target vocabulary size was 24 tokens, but not in conditions where vocabulary size was only 6 or 12 items. Subsequent regression analyses revealed that, in this condition, reading span independently predicted scores alongside intelligence for the old items, but not the new ones.

In the studies reviewed in this section, one explanation often turned to is the critical mass hypothesis, which assumes that morphosyntactic development is driven by vocabulary size (Marchman & Bates, 1994). Kempe, Brooks, and Sionev (2006) found that although increasing the number of vocabulary items improved generalization, there was significant variance in learners' ability to benefit from this condition related to IDs. Kempe and Brooks (2008) asserted that nonverbal intelligence predicted both vocabulary and grammar learning—thus, IDs account for learning in substantial ways. Martin and Ellis' (2012) study also demonstrated that while vocabulary accounted for the most variance in knowledge of inflectional morphology, WM also made a further independent contribution. Hence, even among researchers who place lexical and grammatical knowledge along a continuum, it appears that it is not only vocabulary growth, but also IDs, that impact the learning of inflectional morphology.

4.5 Summary of implications for the current study

At present, few, if any, studies have investigated possible links between measures of WMC and the picture-word matching task employed in this study to gauge participants' knowledge of novel L2 morphological constructions. To my knowledge, no study has attempted to do so using both complex span and dynamic measures of WM. Thus, one contribution of the current study is that it is designed to evaluate the relationship between picture-word matching task performance and WM using two distinct measures. As noted earlier, both of these measure types may offer insight into the nature of processes underlying L2 learning.

Despite the large number of studies supportive of a connection between WMC and L2 learning, it should be stated clearly that WM is most likely selectively associated with learning. In Section 2.4.3, I reviewed Goldstein et al.'s (2010) prediction regarding WM and variation sets. This study was designed to test their prediction. WM scores should be positively related to learning regardless of input condition, but the association will be stronger when no partial repetition is available to support learners' processing and learning. This hypothesis is consistent with results reviewed in Section 4.4.1, in which it was noted that the influence of WM appears to vary to a certain extent on input delivery and input modification.

This study also aims to broaden the investigation of WMC in SLA research by examining its potential relationship with affective variables. Research indicating links between personality and WMC is described in the next chapter (Section 5.3.4). For the present, suffice it to say that, in order to develop a fuller account of adult SLA, it is important not only to understand relationships between WMC and criterion measures of L2 learning, such as those detailed in this chapter, but also to explore interrelations between WMC and other ID variables.

Based on the preceding review, statistical learning from variation sets (Chapter 2), awareness of form-meaning mappings in the input (Chapter 3), and WM (this chapter) should all underpin the learning of novel morphological features. The present study considers each of these potential explanations. Because the training phrase of this study presented both verbal and visual stimuli, two measures of WMC were employed: one which relies on processing, storing, and retrieving verbal information (the complex reading span task) and the other which necessitates encoding, matching, and responding to visual stimuli (the dual 3-back task). Reconsidering the non-unitary nature of the WM system, it is expected that both of these measures may be related to picture-word matching task scores.

CHAPTER 5

APPROACHES TO PERSONALITY AND L2 LEARNING

5.1 Introduction

Although the relationship between personality and L2 learning is little understood, and may even be reciprocal, this chapter attempts to advance the notion that personality variables may influence the noticing and learning of L2 features. To open this discussion, I will first review three major approaches from within psychology. Eysenck, Myers and Briggs, and McCrae and Costa have all had a substantial impact on the L2 literature and their contributions are reviewed in this section.

Later in this chapter, I will move on to a brief review of the L2 literature on personality (Section 5.2), examine personality from the perspective of research on human cognitive abilities (Section 5.3), review research on artificial learning and personality (Section 5.4), and, finally, weigh the implications of these literatures for research on adult second language learning (Section 5.5).

5.1.1 Eysenck's dimensions: P, E, N

Eysenck (1947, 1952; Eysenck & Eysenck, 1985) was concerned with developing a scientific taxonomy of human behavior. An early definition of personality put forth by Eysenck stated that:

> Personality is the sum-total of the actual or potential behaviour-patterns of the organism, as determined by heredity and environment; it originates and develops through the functional interaction of the four main sectors into which these

behaviour-patterns are organized: the cognitive sector (intelligence), the conative sector (character), the affective sector (temperament) and the somatic sector (constitution). (1947, p. 25)

Seeking to confirm the hypothetical structure of personality, Eysenck’s factor analytic research isolated two dimensions of personality, labeled neuroticism and extraversion-introversion. Later writings added psychoticism as a third dimension of personality (Eysenck, 1952).

The labels *type, concept, or dimension,* are contrasted with *trait* in this work. The former terms refer to three superordinate descriptors that arise from the inter-correlations among subordinate traits, or sets of correlated behavioral patterns. Thus, Eysenck and Eysenck (1985) defined the type concepts of psychoticism, neuroticism, and extraversion in terms of the traits corresponding to each. Psychoticism entails aggressive, cold, egocentric, impersonal, impulsive, antisocial, unempathetic, creative, and tough-minded traits. Extraversion is based on sociable, lively, active, assertive, sensation-seeking, carefree, dominant, surgent, and venturesome traits. Lastly, neuroticism consists of traits including anxious, depressed, guilt feelings, low self-esteem, tense, irrational, shy, moody, and emotional (pp. 14-15). These three types (P, E, N) are considered to be normally distributed and universal to all cultures.

5.1.2 The Myers-Briggs Type Inventory

According to an updated report originally authored by Isabel Briggs Myers (1998), the most widely used self-report measure of personality is the Myers-Briggs Type Indicator® (MBTI). The instrument was developed in accordance with Jung’s view of psychological type. This theory distinguished between extraversion–introversion, sensing–intuition, thinking–feeling, and perceiving–judging. Taken together, these lead to 16 MBTI types, which are derived from binary preferences for the aforementioned categories. An example would be ISTJ, or people who are introverted, sensing, thinking, and judging. The test developers characterize this type as responsible, loyal, committed, and preferring to work either independently, or in clearly defined team roles. Other characteristics of this type are that they share their insights only with close companions and may not immediately appreciate others' needs, although once they do, they are supportive of them. An obvious issue with the MBTI is the dichotomization of personality variables. It is unclear how those who are somewhere in between introverted and extraverted, or those who prefer both sensing and intuition, would be represented. Regardless of this, applications far and wide have been found—for example, in job placement, counseling, and education. Baseline type distributions have been published for a number of occupational sectors, including the arts, business, education, health, government, and science and technology, as well as for students (see McCaulley, 1990 for discussion).

The MBTI has been challenged on theoretical and psychometric grounds, but particularly owing to the treatment of personality variables as dichotomous preferences rather than continuous traits. For these reasons, investigators have frequently sought to compare the MBTI with the NEO-PI (described below). Namely, McCrae and Costa (1989) found no support for the 16-type classification asserted by the MBTI. Their results indicated overlap between the four constructs represented by the MBTI and those comprising the Big Five, except for neuroticism. Furnham (1996) replicated this finding, as neuroticism was neither strongly nor consistently related to MBTI variables in his comparison of the two measures (see also Furnham, Moutafi, & Crump, 2003).

5.1.3 The Big Five Model

Taxonomic description was, once again, the goal of the Five-Factor Model (or, the Big Five Model), the development of which is documented by McCrae and Costa (2003). However, this model also has explanatory aims. That is, McCrae and Costa (2003) define personality traits at two levels. At the level of the phenotype, traits are "dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings, and actions" (p. 25), but they can also be understood, at the level of the genotype, as "endogenous Basic Tendencies that give rise to consistent patterns of thoughts, feelings, and actions" (pp. 204-205). In the penultimate chapter of their book, McCrae and Costa (2003) make the controversial claim that personality traits originate solely from biological bases. As such, the model attempts to facilitate both observation and prediction.

As noted by McCrae and Costa (2003), the most widely used measure of the Big Five Model is the NEO Personality Inventory. The current version, NEO-PI-R, assesses five factors, or domains, within which there are six facets measured by subscales. Its methods of administration include self-report or observer ratings. Table 5.1 indicates the domains and facets included in the original and revised versions of the instrument. McCrae and Costa (2003) define each variable in terms of its entailments at the facet level (pp. 47-51) and also provide descriptions of high and low scorers (e.g., a person who scores low on neuroticism is calm and relaxed, whereas a high scorer is thin-skinned and irritable).

Table 5.1. *Domains and Facets in the Five-Factor Model of Personality*

Domain	Facets
Original NEO-PI	
Neuroticism	Anxiety
	Angry Hostility
	Depression
	Self-Consciousness
	Impulsiveness
	Vulnerability
Extraversion	Warmth
	Gregariousness
	Assertiveness
	Activity
	Excitement Seeking
	Positive Emotions
Openness	Fantasy
	Aesthetics
	Feelings
	Actions
	Ideas
	Values
Additions to NEO-PI-R	
Agreeableness	Trust
	Straightforwardness
	Altruism
	Compliance
	Modesty
	Tender-Mindedness
Conscientiousness	Competence
	Order
	Dutifulness
	Achievement Striving
	Self-Discipline
	Deliberation

Numerous validation studies have been done on the Big Five Model (e.g., McCrae & Costa, 1987). As mentioned previously, in the study by McCrae and Costa (1989), it was demonstrated that each of the four MBTI scales converged with one of the five factors measured by the NEO-PI. Another study tested the generalizability of the model by administering translations of the NEO-PI in German, Portuguese, Hebrew, Chinese, Korean, and Japanese to speakers of these languages, and comparing results with those of an American sample (McCrae and Costa, 1997). Factor loadings for the instrument were similar across these groups, leading the authors to assert the universality of the Big Five Model.

These results notwithstanding, McCrae and Costa themselves suggested that while its five domains have been substantiated, the empirical basis of the facets within the model is weaker:

> Researchers coming from many different traditions have agreed on the importance of the five factors because these broad dimensions recur in most personality measures in one form or another. The finer distinctions within the domains, however, are more arbitrary. For example, some theorists would combine Warmth and Gregariousness into a Sociability facet or break Anxiety down into Tension and Apprehension facets. There is certainly no reason why each domain should have exactly six facets. (2003, p. 47)

One can also infer from this statement that the two-tiered structure of the model (with domains and facets) is not entirely agreed upon, either. In fact, DeYoung, Quilty, and Peterson (2007) found that each domain could be broken down into two aspects composed of multiple facets. Based on this finding, they developed a more concise measure of the Big Five Model called the Big Five Aspect Scales (BFAS). In their approach, the aspect scale descriptors corresponding to each Big Five domain are: neuroticism (volatility/withdrawal), extraversion (enthusiasm/assertiveness), openness (intellect/openness), agreeableness (compassion/politeness), and conscientiousness (industriousness/orderliness).

5.2 Personality in L2 learning research

Some early views on the role of personality in L2 learning are traceable to the strong emphasis on learning from the linguistic environment outlined in Chapter 2 (see Dewaele & Furnham, 1999, pp. 519-520, for discussion). For instance, Seliger (1977) studied High Input Generators (HIGs) versus Low Input Generators (LIGs). Input generation refers to the degree to which learners are active or passive in classroom settings, and also the extent to which they seek practice opportunities outside of the classroom. This distinction seems driven by extraversion, which plays a major role in all three of the approaches to personality described in Section 5.1. Interestingly, Seliger found that, although HIGs did interact more in the classroom, and also had outside contact with their target language, they did not score higher on a cloze test. Another early study placing emphasis on the role of learner and the environment, and one that has shaped much thinking on the issues, is Naiman, Fröhlich, Stern, and Todesco's (1978) report on the good language learner.

Naiman et al. (1978) assumed a model of language learning comprised of six factors: context, learner, teaching, environment, learning, and outcome. Based on this, they carried out two studies: one involving interviews with adult language learners of a wide array of languages and another classroom study with learners of French from several different school-age populations. The latter study used tests of French achievement and elicited imitation as criterion measures and measures of cognitive style, personality, student attitudes, and observed classroom behaviors as predictors. Results indicated that tolerance of ambiguity could positively impact classroom L2 learning, though this was not the case for all grade levels, and other tests of personality were not significantly related to outcomes.

In their conclusions, Naiman et al. pointed out the "need for other and better measures of personality factors thought to be related to language learning" (1978, p. 68). Ultimately, these researchers drew attention to the role of the learner and prompted commentary (e.g., Rubin, 1975, see also Griffiths, 2008) as well as further research into the influence of personality, among many other learner variables, on L2 learning. The remainder of this discussion will examine personality and (a) linguistic outcomes, (b) covariates, (c) related constructs, (d) research issues, and (e) current studies.

5.2.1 Personality traits and L2 outcomes

Although what follows is by no means a comprehensive review, L2 outcomes investigated in relation to personality factors have broadly included proficiency, achievement, and placement. First, regarding proficiency, Genesee and Hamayan (1980) conducted a study with L1 English-speaking children in a French immersion setting, finding that a factor consisting of enthusiasm, assertiveness, and self-reliance did not predict literacy skills, listening comprehension, oral production, or reading comprehension. Ehrman and Oxford (1995) studied U.S. Department of State employees and, using the MBTI, reported that intuitive types outperformed others on proficiency measures and on aptitude tests, though aptitude was a better predictor overall of proficiency. Extraversion-introversion, thinking-feeling, and judging-perceiving were unrelated to language training outcomes. On the contrary, Carrell, Prince, and Astika (1996) found that extraversion-introversion *was* related to language scores in a study of Indonesian EFL learners. They found that introverts scored significantly better than extraverts on vocabulary, though this proved to be the case for no other personality type-proficiency test combination.

Besides proficiency scores, personality has been examined in relation to outcomes such as oral interviews, achievement (i.e., course grades) and language program placement in other studies. Busch (1982) conducted a widely cited study in which the introversion-extraversion scores of Japanese college EFL learners were compared to their performance on both proficiency and oral interview tests. She found a significant relation only for introversion and better pronunciation during the interviews, which she ascribed to carefulness, or a lack of impulsivity. Onwuegbuzie, Bailey, and Daley (2000) found that cognitive, affective, personality, and demographic variables predicted the course grades of university foreign language students in the U.S. In this study, cooperativeness was linked to underachievement. The authors attribute this somewhat puzzling finding to classroom dynamics. As a remedy, they suggest that instructors make greater use of group work, role-play, and other instructional methods likely to appeal to cooperative types. Next, Brown, Robson, and Rosenkjar (2001) analyzed factors predicting placement in an intensive English program at a university in Japan. They found that three personality variables—Inferiority Feelings, Nervousness, and Thinking Extraversion[15]—were the most reliable predictors of high (relative to middle and low) placement.

The handful of studies reviewed here help to locate personality in the context of second and foreign language teaching. Any conclusions, however, must be tempered by the fact that cognitive variables were superior predictors to personality variables in the studies led by Ehrman and by Onwugbuzie. There were also mixed results for some personality traits (e.g., extraversion-introversion). As well, the impact of situational press is an important consideration in personality studies conducted in educational settings and the use of global L2 measures in such studies is another potential issue. These and other factors have forced some to the pessimistic conclusion that, "there is really very little evidence that personality systematically makes any difference" (Bialystok & Hakuta, 1994, p. 142).

5.2.2 Potential covariates of personality in L2 learning

As noted in the previous section, personality is often considered a lesser influence than other ID variables in L2 studies. While that section offered commentary on potential links between personality and outcomes, this section explores variables that may covary with personality in the L2 literature. These include age, gender, and major. Regarding age, personality may be in flux during the early years of life, but even more clearly, the self-report measures typically employed to assess personality are less reliable when used with children. This might account for the null results obtained by Genesee and Hamayan (1980), who administered the Early School Personality Questionnaire to children in their

[15] These are three of the twelve traits assessed by the Yatabe-Guilford Personality Inventory.

study. The mean age of these children was about 6 years old. Also, many researchers note gender differences in personality type. Carrell, Prince, and Astika (1996) confirmed such differences in an EFL population. Women in their study were more likely to be classified as Introverted-Sensing-Feeling-Judging than men. Lastly, a study by Moody (1988) using the MBTI examined differences in personality type across college majors. Moody reported that students enrolled in language majors at the University of Hawai‘i were more likely to be intuitive than sensing types, and also showed a greater tendency for thinking over feeling, when compared to a larger general sample of college students.

5.2.3 Research on other related learner variables

Three other personality-related variables encountered in L2 research deserve mention here. First, as noted by Dörnyei (2005, 2009a), *anxiety* is a multifaceted construct, because it can be construed as either: (a) beneficial or harmful to L2 performance and (b) a state or trait phenomenon. Language anxiety may influence academic, cognitive, social, and personal dimensions of L2 learning (MacIntyre, 2002). As an example of its state characterization, MacIntyre and Gardner (1994) conducted a study in which they aroused anxiety by exposing participants to a video camera, leading to lower scores on a vocabulary task. Second, *creativity* can be understood as related to personality, but has been investigated somewhat separately from it in L2 research. Albert and Kormos (2004) associate creativity with openness in the Big Five model and report findings relating three facets of creativity (originality, flexibility, and creative fluency) to narrative task performance (see also Albert, 2011). Third, personality is one of the foundations, along with intergroup climate, of MacIntrye, Dörnyei, Clément, and Noels'

(1998) model of *willingness to communicate*. This is a situated model of a person's readiness to use an L2 encompassing multiple contextual and individual variables that has prompted recent studies in second and foreign language settings (e.g., MacIntrye & Legatto, 2011; Yashima, 2002).

5.2.4 Issues in L2 research on personality

The previous literature on personality raises two central issues that are acknowledged by researchers. The first issue is the lack of a prevailing theoretical model of personality on which to base L2 studies. As described at the start of this chapter, there are differences in the models put forth by Eysenck, Myers and Briggs, and Costa and McCrae. Ortega (2009) reviewed these three approaches and noted that these authors emphasized temperament, cognitive style, and personality, respectively (p. 194). Furthermore, each model contains conceptual gaps. This inconsistency makes cross-study comparisons a serious challenge. However, as indicated above, the Big Five model is well established in psychology and different authors have recently expressed the view that this model may represent a way forward for SLA studies (Dörnyei, 2005; Ortega, 2009).

The second issue concerns measurement incongruity. That is, the measurement process differs across the constructs of personality and learning. The nature of these distinct assessment models may thus obscure the interpretation of any relationship between the two constructs (see Norris & Ortega, 2003; Messick, 1981; and also Section 5.3 below). This issue explains various problems noted in reviews of the L2 literature on personality. For example, Dewaele and Furnham (1999) argued that the oral or written modality of outcome measures is crucial to understanding the influence of extraversion. Dörnyei, too, has suggested that researchers must carefully conceptualize the "personality–achievement contingencies" (2005, p. 30) inherent in their work.

One exemplary study that tackles both of these issues is Verhoeven and Vermeer (2002). Rather than focusing on proficiency, Verhoeven and Vermeer targeted three dimensions of communicative competence in their research: organizational, pragmatic, and strategic competence. These components of language ability were assessed using a test battery, role-play tasks, and rating scales. Also, instead of depending on children's self-assessment, a teacher-scored instrument was used to assess their personalities according to the Big Five model. In light of McCrae and Costa's (1997) results on the universality of this model, it is noteworthy that the authors found no difference in the personality scores of the L1 and L2 children in their study—these children came from Dutch and Turkish or Moroccan backgrounds, respectively. Results for the L2 learners were as follows: organizational competence was highly correlated with conscientiousness and openness, pragmatic competence was moderately correlated with openness, and strategic competence was highly correlated with openness and extraversion. This study stands out for having avoided potential pitfalls in the selection of predictor and criterion measures and for having yielded results that can be considered comparable to psychological research on the Big Five.

5.2.5 The current wave of personality research in L2 studies

Three studies from the current decade indicate the extent to which L2 research has managed to (a) establish conceptual clarity with regard to the personality literature and

(b) adopt increasingly sophisticated measurement practices. To begin, Ockey (2011) is a refinement in that it focused on assertiveness (a facet of extraversion) and self-consciousness (a facet of neuroticism). This study also used a group oral task as the outcome measure. Results indicated that assertiveness predicted ratings of pronunciation, fluency, grammar, vocabulary, and communication skills. Next, Ramírez-Esparza, Harris, Hellerman, Richard, Kuhl, and Reeder (2012) coded the personality behaviors (i.e., extraversion-introversion) of 20 adult ESL learners classified as either high-education or low-education. Using this methodology, the researchers found that low-education learners displayed more introverted behaviors while high-education learners showed more extraverted behaviors. Furthermore, introversion was negatively related to literacy test scores, whereas extraversion positively related to such scores. Lastly, Ożańska-Ponikwia and Dewaele (2012) looked at the influence of personality on immigrants' L2 use and self-perceived proficiency. Linear regression analyses showed openness to be a significant predictor of both outcome variables. It is of note that these studies generally did not investigate solely proficiency test scores nor did they rely exclusively on self-report measures of personality.

5.3 Relationships between personality and cognitive factors

On a broad scale, the influences of affective and motivational processes combine with cognitive abilities in learning and performance. Brief examples of how affective variables impinge on performance include the detrimental effect of anxiety on test performance and the beneficial impact of positive moods on tasks requiring creativity and problem solving (Corno et al., 2002). Research dealing with relations among the three categories of affect, cognition, and conation, however, is somewhat scarce. As Ackerman (2013, p. 63-63) notes, there are several barriers to understanding such triadic relations. The first is the fact that personality is constrained by environmental or situational press, or "strong cultural norms or other situational constraints on the individual, which in turn tend to diminish the individual's range of expressed behaviors" (Ackerman, 2013, p. 63). The second is that, since personality traits never exclusively predict an individual's response, their role is more evident in aggregate behaviors. A third point, related to this, is that measurement practices differ according to the constructs under investigation. Maximal performance is the focus of measures of cognitive abilities while typical behavior is the focus of measures of personality. Nonetheless, while instruments differ substantially, at the level of theory, these constructs do come into close contact (see Ackerman, 2013; Ackerman & Heggestad, 1997; DeYoung, 2011).

In fact, there are empirically grounded approaches that situate personality alongside cognition. One manifestation of this is Ackerman's PPIK theory of intelligence (1996, 1999), which derives its name from four components: intelligence as process, personality, interests, and intelligence as knowledge. This theory presupposes that distinct relationships hold between personality and two types of intelligence. These are intelligence as process, positively yet weakly related most to personality factors, except neuroticism and anxiety, and intelligence as knowledge (or, content), related to openness and typical intellectual engagement. These two types supplant the more familiar distinction between fluid and crystallized ability (Cattell, 1943).[16] The latter, more commonly invoked labels will be used in this section. The organization of this section

[16] Others have chosen to replace Cattell's descriptors with *non-verbal* versus *verbal intelligence* (DeYoung, 2011).

loosely follows Carroll's (1993) hierarchy of human cognitive abilities, beginning at the top with general intelligence, then moving to the next, lower stratum, at which fluid intelligence, crystallized intelligence, memory, and speed are located.

5.3.1 General intelligence

Most of the studies on personality and cognitive ability reviewed in the extensive meta-analysis by Ackerman and Heggestad (1997) addressed general intelligence (along with fluid and crystallized intelligence). To give a bird's eye view of this project, Ackerman and Heggestad identified ten categories of ability tests, along with 19 personality variables, in the literature they reviewed. This yielded 190 correlation coefficients from 135 included studies containing nearly 65,000 participants. Ackerman and Heggestad reported significant positive or negative correlations between general intelligence and 12 distinct personality traits (see their Table 1). Of these, the strongest relationships were found for openness (r = .33) and test anxiety (r = -.33).

5.3.2 Fluid intelligence

A recent study by Soubelet and Salthouse (2011) sought to examine relationships between cognitive variables and the Big Five personality traits across the lifespan. In a sample of 2,317 adult participants, these researchers found that that only openness was significantly related to fluid intelligence (as measured by a battery of several tests) in each of the three age groups examined (i.e., 18–39, 40–59, and 60–96 years old). This finding extends previous work because Ackerman and Heggestad's (1997) meta-analysis included only two studies (N = 205) addressing connection between fluid intelligence and openness. Based on these earlier studies, the relationship was not significant. DeYoung, Peterson, and Higgins (2005) also reported significant correlations between fluid intelligence and openness/intellect.

5.3.3 Crystalized intelligence

The findings are interesting for crystallized intelligence. In their meta-analysis, Ackerman and Heggestad (1997) reported a significant correlation (r = .35) between crystallized intelligence and typical intellectual engagement, a personality measure they associated with the construct of intelligence-as-typical performance. Big Five variables significantly linked to crystallized intelligence included both extraversion (r = .11) and openness (r = .30). The study by Soubelet and Salthouse (2011) also found significant, positive relationships for their three age groups between crystallized intelligence and openness. They suggested that openness corresponds to both processing and acquiring information (hence its relationships with both *Gf* and *Gc*). In addition, extraversion was significantly but, here, negatively correlated with crystallized intelligence. Soubelet and Salthouse noted that this finding is supported by recent work.

5.3.4 Working memory

Given the purpose of this study, this section focuses on working memory capacity, although broader notions of memory have often been applied in understanding its relationship with personality. The review by DeYoung (2011) asserted that openness, or particularly the aspect of it labeled intellect by work on the BFAS (DeYoung, Quilty, & Peterson, 2007), is positively related to working memory. DeYoung, Shamosh, Green, Braver, and Gray (2009) conducted a study using a 3-back task that supported this conclusion. Kaufman, DeYoung, Gray, Jiménz, Brown, and Mackintosh (2010), who used an operation span task, also found support. Counter to these findings, other studies using personality measures have reported no direct link between dynamic (Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012) or complex span (Unsworth, Miller, Lakey, Young, Meeks, Campbell, & Goodie, 2009) tests of working memory. Insofar as WMC is related to intelligence (see Section 4.3.4), its relationship with openness can be expected to be positive.

5.3.5 Speed

Perceptual speed measures and extraversion showed a significant, though small, positive correlation in Ackerman and Heggestad (1997). An updated meta-analysis by Wolf and Ackerman (2005) confirmed this small effect, but also noted that trends in the use of measures and in sample age may lead to conflicting patterns of results over time. Intriguingly, Soubelet and Salthouse (2011) did not confirm the speed-extraversion relation, but instead found that speed and openness were positively and significantly related for all age groups. DeYoung (2011) reasoned that any correlations between speed and extraversion may be an artifact of the positive relationship between extraversion and openness/intellect (p. 723). Also, in Soubelet and Salthouse (2011), speed was related positively to conscientiousness, but only for those between ages 40 and 59 years old.

5.3.6 Summary

It is useful to keep in mind that intelligence, memory, and speed, are all closely related variables. Studies often show they are positively correlated (e.g., Carroll, 1993). Furthermore, as DeYoung (2011, p. 719) writes, intelligence is "relatively narrow when compared to traits like the Big Five that represent very broad regularities in personality", so perhaps it should not be surprising to find specific personality traits or aspects that correspond to it. Nevertheless, it does seem striking that of the many personality variables, a positive relationship between openness and cognitive abilities surfaces time and again. Chamorro-Premuzic and Furnham (2005) offer five possible scenarios underlying this correlation. First, more open individuals may gravitate toward activities that enhance intelligence. Second, reversing the cause-effect relationship, preexisting intelligence may give rise to openness. Third, individuals with more intelligence may engage (openly) in activities to further their intellectual competence. Fourth, survey measures of openness may reflect self-report of intelligence. Fifth, related to the latter, intelligent people may discern and provide socially desirable responses to openness scales (Chamorro-Premuzic & Furnham, 2005, pp. 60-61). Awaiting a more precise explanation, most researchers seem to agree that the robust connection between openness and intelligence highlights the benefit in seeking to redress the traditional separation of cognition and personality in the intelligence literature.

5.4 Personality and learning in experimental settings

The literature on artificial learning has generated evidence directly bearing on the central thesis of this chapter—that personality variables can influence the processing of novel input. In this work, performance on SRT and AGL tasks was compared to participants' scores on personality measures. Importantly, these findings support and extend the results discussed in the last section by revealing that links between cognitive and affective variables can also be found in experimental learning contexts.

First, Norman and her colleagues conducted two studies comparing performance on the NEO-PI to several outcomes related to an SRT task, yielding conflicting findings. Norman, Price, and Duff (2006) had participants learn a sequence of circles determined by a second-order conditional rule (see Reed & Johnson, 1994). In one of their two conditions, in which the response-stimulus interval[17] was zero, positive correlations were found between openness to feelings, on the one hand, and RTs, ratings, and inclusion task scores on the other. The researchers noted that this result might explain previous inconsistent findings and put forward the idea that openness to feelings may be a marker of ability to examine the contents of one's consciousness. However, a later study failed to replicate this finding. Norman, Price, Duff, and Mentzoni (2007) used two second-order conditional rules implemented probabilistically. The sequence was instantiated by colored shapes. In contrast to the previous study, openness to feelings was not related to learning. The authors thus speculated that this relationship may depend on the quality of participants' consciousness, and that the null results could be due to changes in the stimuli used.

A study by Kaufman, DeYoung, Gray, Jimenéz, Brown, and Mackintosh (2010) also addressed this issue, employing a wide range of personality measures, including the MBTI, NEO-PI, and BFAS, in addition to several cognitive tasks. The learning task was

---

[17] RSI is a factor that has been shown to influence the degree of implicit versus explicit learning in laboratory studies (Destrebecqz & Cleeremans, 2001).

a probabilistic SRT task with sequences encoding two second-order conditional rules (sequence A occurred with a probability of .85 and sequence B occurred with a probability of .15). Learning showed significant correlations with BFAS openness, NEO aesthetics, and MBTI intuition. These relationships were even stronger when latent variables representing personality were used. The authors construe implicit learning as an ability measured by the SRT task and claim that meaningful individual differences reside within this ability. To illustrate, they present a model linking Openness to implicit learning and Intellect to working memory. The authors' remarks on Openness are especially relevant here:

> The engagement with the perceptual world that characterizes Openness may be facilitated by implicit learning [ability]. Of course, it is also possible that those high in Openness are better at implicit learning because they have a wider focus of attention. (Kaufman et al., 2010, p. 336)

Though the direction of causality remains uncertain, overall, this study is an impressive demonstration of how cognitive and affective variables are conjoined, and it has direct implications for understanding how personality may be related to implicit sequence learning, which has been demonstrated to be an ability with potential relevance to L2 aptitude by Granena (2012).

In addition to these findings, two other recent studies have found associations between personality traits and the learning of letter strings based on artificial grammars. Pretz, Totz, and Kaufman (2010) used the Rational–Experiential Inventory (Pacini & Epstein, 1999) as a cognitive style measure. These investigators found that rational ability (i.e., "a high level of ability to think logically and analytically", Pacini & Epstein, 1999, p.

974) was correlated significantly and positively with scores on the AGL task, as well as an SRT task. Xie, Gao, and King (2013) employed the Thinking Styles Inventory as a measure of participants' tendencies to handle tasks in certain ways. They had participants perform the same task under two conditions. Learners in the implicit group were told to memorize strings and those in the explicit group were asked to discover rules. Here, legislative style, or, "a preference for creating rules and autonomy" (p. 270) was significantly related to learning in the explicit, but not the implicit condition. This is precisely what might be expected because the instruction itself urged participants to look for rules. No significant correlations occurred with implicit scores. In sum, the influence of learners' personal styles may be found only under certain conditions of learning novel stimuli. Clearly, more such work is needed to fully understand the personality–learning link.

5.5 Potential implications for future L2 studies

Section 5.4 no doubt raises questions about the extent to which findings regarding artificial learning and personality generalize to SLA—inside or outside of educational settings. For one thing, situational press, or norms and other factors constraining the expression of a range of behaviors, is inherent to laboratory settings. Thus, the personality effect might even be greater outside the lab. The brief training phases typical of such studies implies that, in SLA, any such effects might accrue over longer periods of time, and this is what is shown by L2 research on personality (see Section 5.2). Also, the learning targets in experimental studies bear only limited resemblance to SLA, as the strings and sequences of natural language come bundled in phonological, semantic, and pragmatic packages which the learner must also unravel. The correlational design of most of these studies furthermore poses challenges to interpretation. Nevertheless, the observation that meaningful differences in novel learning of language-like input are associated not only with cognitive factors, such as working memory, but also personality factors, should concern L2 researchers. Among other relevant issues, these studies raise the question of whether personality is related to noticing in early SLA.

Considering the research trends discussed in this chapter, in general, scholars seem amenable to a reappraisal of the role of personality in SLA. In keeping with the popularity, within psychology, of the Big Five Model, and the trend toward the use of specific measures of language outcomes in applied linguistics, Dörnyei concluded that, "the application of the Big Five model in L2 studies is likely to shed new light on the relationship between personality and language learning, particularly if elaborate language measures are employed as criterion variables" (2005, p. 29). This sentiment is matched by Ortega, whose own review of personality concluded with the comment that, "openness to experience…emerges as a fruitful site for future personality research in SLA" (2009, p. 196). Naturally, a focus on specific variables, as in the case of some L2 studies reviewed herein, should lead to more precise hypotheses and clearer interpretations within this area of study. In addition to the implications for research, as a final, theoretical note, Dörnyei (2009a) viewed such new directions in individual differences research as having a dual advantage. Namely, they acknowledge both the reconciliation of the trilogy of mind (cognition, motivation, and affect) in psychology and the movement toward dynamic systems theory in SLA. He proposed that cognitive, motivational, and affective systems interact dynamically, while IDs also function as more or less stable attractor states. The next chapter presents hypotheses related to the potential role of personality in this study.

CHAPTER 6

METHODOLOGY

6.1 Introduction

The present research brings together methodological traditions which have, in the past, been considered very distinct. More than a half a century ago, Cronbach (1957; see also Cronbach & Snow, 1977) described the historical divide between experimental and correlational psychology, as a prelude to arguing for their unification. The strengths of experimental methods, which target causal description, and correlational methods, which identify associations between variables, can be coupled. A major advantage to this coupling is the possibility of investigating interactions. One reason for doing so is that, “we can expect some attributes of persons have strong interactions with treatment variables” (Cronbach, 1957, p. 680). In the context of the present dissertation, do learner attributes interact strongly with treatments based on the statistical learning literature?

In summary, this study employed a randomized design comparing three treatments based on exposure to artificial language input, followed by a battery of measures probing learner-related factors. Regarding the study design, the effects of the treatment were observed by means of a single posttest consisting of two subtests—one comprised of instances presented during treatment and the other comprised of novel exemplars based on the same system.

Because it included three treatment conditions as well as several measures of cognitive and affective differences (described below), the design allowed for the analysis of both experimental and correlational results, as well as interactions between treatment and learner variables. This design was deemed suitable for investigating how learners generalize beyond the input they receive, for understanding the potential roles that learner abilities play in generalization, and, lastly, for seeking evidence of interactions between input properties and learner variables. Each of these is an important aim in cognitive approaches to second language acquisition.

6.2 Research questions and hypotheses

The following research questions and hypotheses guided the study. Each of the six questions below was motivated by recent research into input modification, learner awareness, and individual differences in SLA.

Research Question 1: What are the effects of the input conditions on picture-word matching test scores?

- Hypothesis 1: Trained and untrained scores will be higher in the variation set conditions than in the scrambled condition (details below). This hypothesis assumes that the effect of input variation observed by Onnis et al. (2008) when using artificial language stimuli with no semantic component can be extended to the learning of form-meaning connections. The findings of Research Question 1 will contribute to the establishment of a solid empirical base that seeks to assess the potential value of statistical learning in adult L2 populations.

Research Question 2: What are the effects of awareness on picture-word matching test scores?

- Hypothesis 2a: Participants who exhibit awareness (on the post-experiment written questionnaire) of the relevance of the two features of number and animacy to the artificial language to which they have been exposed will show a higher proportion of correct responses on the picture-word matching task than those who are not aware of the relevance of those features, for both instances encountered during the learning phase and novel instances.
- Hypothesis 2b: Participants who exhibit accurate awareness of the specific form-meaning mappings (in their responses to the written questionnaire) will show a higher proportion of correct responses on the picture-word matching task than those who cannot accurately report those mappings, for both trained and novel exemplars of these constructions.
- Hypothesis 2c: Awareness (accurate report) of specific form-meanings will lead to better performance than awareness of the significance of semantic features alone, measured in terms of the proportion of correct responses to the picture-word matching task.

The procedures used to assess each participant's level of awareness are described in detail in Section 6.4.3.5.

Research Question 3: Is WMC (as measured by complex span and dynamic measures) positively related to picture-word matching task scores?

- Hypothesis 3a: For participants in all three conditions generally, the relationships between complex span and dynamic measures of WMC and picture-word matching task scores will be positive. This acknowledges the important role of WMC in learning novel lexis and morphology in a second language (e.g., Martin & Ellis, 2012; Brooks & Kempe, 2013).

- Hypothesis 3b: In the scrambled condition, the positive association between complex span and dynamic measures of WMC and picture-word matching task outcomes will be stronger than in the variation set conditions. That is, there will be an interaction between condition and WMC. This hypothesis assumes that the variation set condition will lessen the demand on WMC because the variation set conditions prompt alignment and comparison of the input to a greater degree than the scrambled condition. This tests the hypothesis put forth by Goldstein et al. (2010), previously discussed in Section 2.4.3, in the context of adult L2 learning.

Research Question 4: How are openness and intellect related to measures of performance and cognitive abilities used in the experiment?

- Hypothesis 4a: Openness will be positively related to any learning found in the study. This seeks to extend evidence in Kaufman et al. (2010), which demonstrated that openness was related to implicit sequence learning, to the learning of form-meaning mappings.
- Hypothesis 4b: Intellect will be positively associated with WMC and attention to form. This hypothesis seeks to provide empirical evidence for affective–cognitive relationships posited by DeYoung, Quilty, Peterson (2007) and DeYoung et al., (2009) within the context of L2 construction learning.

Research Question 5: Which of the experimental and observational variables in this study predict scores on the trained and untrained subtests? Do interactions between these variables predict outcomes?

- Hypothesis 5: Input variation, learner awareness, and individual differences will each contribute to L2 construction learning as measured by the picture-word matching task. This hypothesis accords with recent views that L2 learning ability, and therefore outcomes, emerge from multiple, dynamic processes specific to given tasks and situations (e.g., Beckner et al., 2009; Dörnyei, 2009a; Robinson, 2007b).

Research Question 6: What do additional measures of awareness (i.e., source attributions) indicate about participants' learning of the artificial constructions? Specifically, do participants use both explicit and implicit knowledge sources during picture-word matching task performance?

- Hypothesis 6: Participants will report use of both explicit (rule, memory) and implicit (guessing, intuition) knowledge sources for their responses to the picture-word matching task. This hypothesis acknowledges claims that learners use both implicit and explicit knowledge when processing their L2 (N. Ellis, 1994, 2005, 2011; R. Ellis, 2009; Rebuschat, 2013).

6.3 Pilot testing

An initial pilot experiment was conducted in spring term 2012 and then followed up on in subsequent pilot studies during fall term 2012. Several issues arose during piloting. Each of these matters is dealt with in the description of the procedures later on in this chapter, in the sections noted in parentheses below. Briefly, modifications to the study based on the pilot research concerned: (a) fine-tuning picture stimuli so that participants were able to easily identify meanings associated with them by the experimenter (Section 6.4.3.1); (b) guaranteeing that noun classes were not inadvertently associated with statistical cues (e.g., word-initial or word-final letters) (Section 6.4.3.1); (c) revisions based on item analyses for the picture-word matching task (Section 6.4.3.4); (d) technical adjustments to experimental scripts (Section 6.4.3.2); and (e) streamlining of the awareness measures used across the experiment (Section 6.4.3.5).

6.4 The main study

In this section, details are given for the main study, conducted after the pilot testing period, including participant characteristics, data collection, and materials and procedures. Later, I also briefly summarize the analyses employed in this study.

6.4.1 Participant characteristics

The population for this study consisted of undergraduate students enrolled at the University of Hawaiʻi at Mānoa whose dominant language, as judged by a self-assessment, was English. A total of 93 participants joined the study. Data from three participants were excluded due to ineligibility or apparent failure to comply with instructions. Table 6.1 provides demographic characteristics for the sample of included participants.

Table 6.1. *Demographic Characteristics of Participants (*N = 90*)*

Characteristic	*n*	%
Gender		
Female	64	71.11
Male	26	28.89
Undergraduate year		
1st year	27	30.00
2nd year	16	17.78
3rd year	29	32.22
4th year	17	18.89
Other (transfer)	1	1.11
College or school		
Arts & Sciences		
Arts & Humanities	4	4.44
Languages, Linguistics & Literature	16	17.78
Natural Sciences	12	13.33
Social Sciences	16	17.78
Other Arts & Sciences	2	2.22
Business	9	10.00
Education	7	7.78
Engineering	3	3.33
Tropical Agriculture & Human Resources	2	2.22
Architecture	1	1.11
Hawaiian Knowledge	1	1.11
Nursing & Dental Hygiene	6	6.67
Ocean & Earth Science & Technology	1	1.11
Travel Industry Management	3	3.33
Double major	4	4.44
Undeclared	3	3.33

Note. See http://manoa.hawaii.edu/academics/ for details.

At the time of the study, participants' ages ranged from 17 to 36 (M = 21.20; SD = 3.10). The participants were also asked to indicate their L1, which was frequently, but not always, English (72 participants, or 80% of the total). First languages other than English included (as given by the participants): Cantonese (2), Chinese (3), Filipino (1), Ilokano (2), Polish (1), Spanish (3), Tagalog (1), Tibetan (1), and Vietnamese (2). In addition, one participant simultaneously acquired English and Japanese as first languages, while another learned both English and Cantonese as first languages. Participants also reported having knowledge of between zero and six languages in addition to English ($M = 2.24$; $SD = 1.27$).

6.4.2 Data collection

The study was publicized by means of on-campus flyers, electronic messages to departmental electronic mailing lists and the university calendar website, and announcements during teaching faculty meetings. Additionally, the study was included in a list of research projects open to students whose instructors offered course credit through the Linguistics Beyond the Classroom program. All participants contacted the researcher by email to arrange a meeting time. They then visited the Second Language Acquisition and Bilingualism Laboratory to take part in the research.

Prior to joining, participants signed a consent statement (See Appendix A) for the study, which had been approved as exempt by the University of Hawaiʻi Committee on Human Studies. Participants were assigned to conditions based on the order in which they joined the study, alternating between the conditions. That is, the assignment forced equal sample sizes (Shadish, Cook, & Campbell, 2008). Because each participant joined at his or her convenience, this assignment was based on chance. During the study, data were collected at computer workstations in the lab. Upon completion, participants received as compensation either: (a) ten US dollars ($n = 50$) or (b) course credit, through the Linguistics Beyond the Classroom program ($n = 40$). No significant difference was found between the scores of participants on the trained subtest of the picture-word matching task, $t(88) = -0.73$, $p = 0.47$. This was also the case for the untrained subtest, $t(88) =$ -

0.34, $p = 0.73$. To protect anonymity, participants were assigned a number to enter for each of the tasks below; full names were not stored with participant data.

Given that this study aimed to investigate individual differences, a target sample size of approximately 100 participants was established. Rules of thumb regarding the sample size necessary to conduct multiple regression analyses, such as used here, are often based on the ratio of participants to independent variables. A baseline number of participants may also be recommended. For example, in their discussion of multiple regression analyses, Tabachnick and Fidell (2007) suggested a minimum of 50 participants, plus eight times the number of independent variables (p. 123). Thus, the minimum number required to investigate the overall effects with five independent variables would be 90 (i.e., 50 + 8(5) = 50 + 40 = 90).

6.4.3 Materials and procedures

The materials and procedures described next were used in this study. The learning and testing phases of the experiment were followed by assessments of awareness and individual differences. All participants completed these tasks in the same order.

6.4.3.1 Artificial language

Strong rationales for using artificial languages in second language research have been offered by many researchers (Ellis & Schmidt, 1998; Hulstijn, 1997; Yang & Givón, 1997). Here, first and foremost, it was necessary to use an artificial language because the population is multilingual[18], thus there is a need to carefully control for previous exposure and language knowledge given the research questions above. Also considering the research questions, it is important to control linguistic factors that may influence learning, including lexical, morphological, and cross-linguistic factors (Jarvis & Pavlenko, 2008). Nonetheless, there are also limitations of artificial languages, among which is the fact that they may not generalize to language learning outside of the laboratory. For this reason, an artificial lexicon incorporating semantically-motivated morphological distinctions, which, though arguably rare in the languages predominantly learned and used today, occur in at least one documented language, was used.

The learning phase of the study exposed participants to a miniature artificial language comprised of 12 nonwords accompanied by images. The process of designing artificial stimuli began by selecting words from a nonword database, grouping them into two noun classes (animate or inanimate), and assigning them meanings. Nonwords were drawn from the English Lexicon Project (ELP) database (Balota et al., 2007). To reduce the possibility of effects due to word characteristics, nonwords were selected from the database and assigned to noun classes on the basis of several criteria. First, all words were four characters in length. Second, syllable structure was either CCVC or CVC (both were used with nearly equal frequency across the two word classes). Third, 66.7% of all initial letters and 83.3% of final letters co-occurred across the two word classes, so that these cues would be unlikely to assist in categorization (see Onnis & Christiansen, 2008).

[18] Several independent sources attest to this description. For instance, 70% of the total student population at UHM is from Hawaiʻi (Mānoa Institutional Research, 2010), where slightly more than 25% of the population speaks a language other than English (LOTE) at home (Shin & Kominski, 2010). In addition, during the past decade, there were 80 different languages other than English taught here (MLA Historical Enrollment Data, 2002-2009).

Fourth, lexical characteristics generated by the ELP database, including the number of orthographic neighbors (i.e., the number of words that result from replacing one letter while holding others constant) were compared across the two sets of words (for evidence of the effects of orthographic neighborhood on recognition, see, e.g., van Heuven, Dijkstra, & Grainger, 1998). Fifth, data from a lexical decision task, also reported in the ELP database, were examined. This included the average reaction time (RT) and proportion of accurate responses for each nonword. Based on independent *t*-test results, no significant differences in means were found in terms of these lexical or behavioral characteristics (see Table 6.2). Additional criteria for matching the nonwords to English meanings included letter correspondences (e.g., *haun* was not chosen for horse).

Table 6.2. *Lexical and Behavioral Characteristics of Nonwords in the Two Noun Classes*

	Animate		Inanimate		
Characteristic	*M*	*SD*	*M*	*SD*	Diff.
Ortho_N	2.67	1.03	3.17	1.60	-0.50
RT	719.80	43.62	713.98	60.17	5.82
Accuracy	.95	.08	.96	.05	-0.02

Note. Ortho_N = orthographic neighbors; RT = reaction time; Accuracy = proportion of accurate responses.

The target feature in this study was an inflectional paradigm comprised of features marking animacy and number. These features had the values animate and inanimate, and singular, dual, and plural. As in many languages, the bare form was used to denote singular. Four affixes marked dual and plural. Words were shown in these bare or inflected forms throughout the learning phase. The use of different morphology to mark number across different noun classes is an attested feature in the world's languages (Corbett, 2000). For example, as in the artificial system here, bare nouns denote singular forms while stems are inflected for dual and plural depending on the noun class to which they belong in Resígaro, a language of the Arawakan family (Allin, 1976).

There were three reasons for designing the number marking system this way. First, the semantic features are recoverable from context. Thus, they do not pose the same challenges, to optimal design or to optimal L2 learning, as uninterpretable features, or purely formal markers of syntactic dependencies (see Chomsky, 2000; Slabakova, 2009). Second, the choice is consistent with Leung and Williams' (2011) argument, based on Talmy (2000) and Bickerton (2001), that implicit learning may be more likely in cases of form-meaning mappings which are known to be encoded in language. Because the inflectional morphology here is attested, there should be no question as to its learnability. Third, it is consistent with a view of language acquisition that seeks to uncover learning principles applicable to a diverse range of form and meaning mappings. As Evans and Levinson comment on linguistic diversity, "the vast difference in morphological complexity is mirrored by differences in grammatical organization right through to the deepest levels of how meaning is organized" (2009, p. 434). The use of bound morphology to denote number and noun class has yet to be investigated in research on artificial L2 learning.

Morphological inflections were selected with several criteria in mind. To begin, in their meta-analysis, Goldschneider and DeKeyser (2001) noted the influence of the following characteristics of morphemes on acquisition order: (a) perceptual salience, (b) semantic complexity, (c) morphophonological regularity, (d) syntactic category, and (e) frequency. Perceptual salience refers to the number of phones, syllabicity, and sonority of a morpheme. The four affixes in this study were similar in terms of number of phones (e.g., *ba)*, syllabicity (1 syllable), and sonority (each ended in a vowel). In terms of semantic complexity, as indicated above, each morpheme expressed two meanings (number and animacy). They were morphophonologically regular in that homophones did not occur among the four items. All belonged to the same category of bound morphemes and occurred with the same frequency (as described below). Distinct consonants and vowels were selected to ensure that associations between form and meaning were arbitrary (Saussure, 1972).

Next, as pointed out by Luk and Shirai (2009), the natural order studies, upon which Goldschneider and DeKeyser's analysis was based, tended to neglect the role of L1. So, two additional characteristics from the early literature (Weinreich 1953/1963, cited in Luk & Shirai, 2009) can also be considered: congruence and phonetic similarity to L1 English. In part, the artificial grammar here is similar to English number marking, which also uses bare singular forms and suffixes to mark plural. However, there is a lack of congruence in that English does not inflect nouns for dual, nor does it require different regular plural endings based on nominal categories. It is therefore not clear whether transfer based on congruence can be expected, although it can be assumed that English-dominant participants (most of whom listed English as their L1) would notice distinctions in number marking more readily than the animacy marking. Furthermore, no phonetic similarity exists between these inflections and those used to mark number in English (see Table 6.3).

Table 6.3. *Inflectional Markers Presented in the Artificial Input*

	Singular	Dual	Plural
Animate	-Ø	*-mi*	*-ku*
Inanimate	-Ø	*-ba*	*-zo*

Lastly, Gor (2010) presented a list of factors to consider in the processing of L2 morphology, raising the issues of decomposability (i.e., the extent to which stem plus morpheme patterns are identifiable), L1 transfer, L2 proficiency, input properties, type of exposure, and individual differences (see p. 15). Though several of these points have just been addressed, it is worth noting the fact that inflected and uninflected stems were presented in written format throughout the experiment, most likely increasing the likelihood of participants segmenting the forms.

The picture stimuli used in the experiment were digital images freely available as Microsoft Office Clip Art, or were purchased through the company istockphoto. They were converted to .pict files using Microsoft PowerPoint and presented on a white background during the experiment. The same image was used in duplicate to represent the dual form of words. The size and position of these two images varied. In the case of plural, half of the picture stimuli used to represent each noun class showed three images, while the other half showed more than three. Again, these images were resized and repositioned so that they would appear not to be identical. Care was taken to ensure that other salient characteristics besides animacy and number would not serve as cues. For example, objects of various sizes, shapes, and materials were used, and animals from distinct taxonomic classes, orders, and families were shown (see Appendix B).

6.4.3.2 Training phase

PsyScope X software (Version B 57) was used to administer the training and testing phases of this study. Regarding the presentation of materials, the screen position and timing were consistent across the three conditions. That is, on each training trial, the picture stimuli, which were 250 by 250 pixels in size, appeared at the center of a white screen. The target words were presented simultaneously in black, using `Courier` font, to the left of the pictures. There were six blocks consisting of 24 training trials, for a total of 144 trials. Blocks consisted of the 24 trained word forms shown below, each repeated only once during the block. As such, all trained items repeated six times during the learning phase. Note that this involved repeating stems twice per block, once in one form and once in another, different form. On each trial the word and picture were displayed for 2500 ms. A 30 second rest-break occurred after each block.

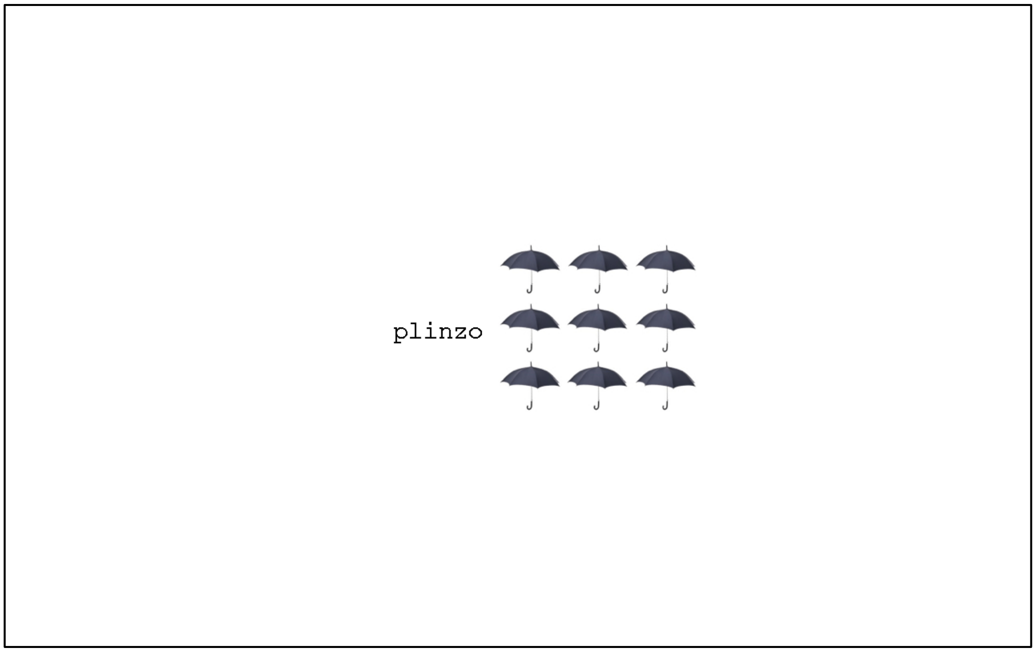

Figure 6.1. A training trial from the experiment.

All three conditions also contained attention-focusing trials, which were interspersed within the training trials. There were two types of attention-focusing trials, referred to here as *meaning* trials and *copy* trials. These two types alternated during training.

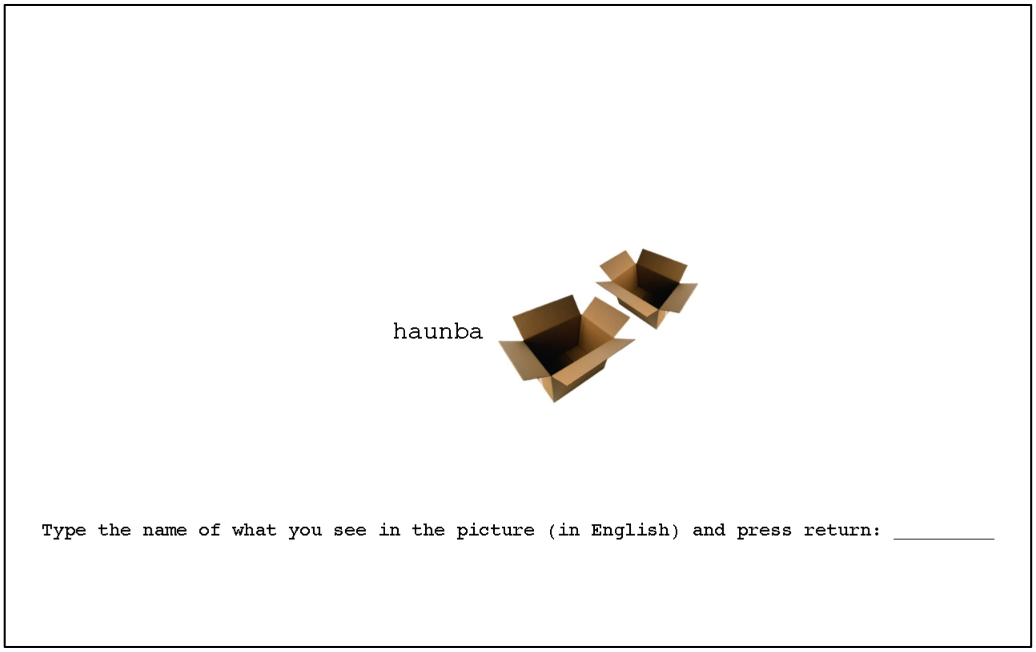

Figure 6.2. A meaning-focused trial from the experiment.

On meaning trials, participants saw a word and picture and were prompted to name the picture in English and press return.

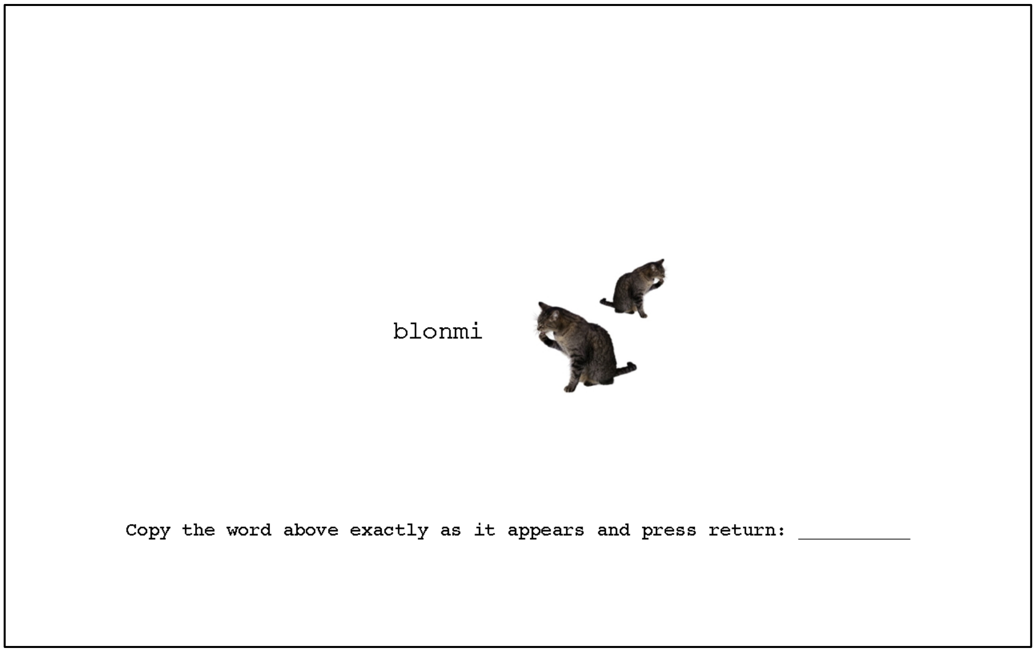

Figure 6.3. A form-focused trial from the experiment.

On copy trials, participants again saw a word and picture, but were instead asked to copy the word exactly as it appeared. These trials were speeded, so that the total duration of the training phase could be controlled for all participants. Participants were informed at the beginning of the experiment that they would have exactly 10 seconds to enter their responses to these trials. Regarding the attention-focusing trials, the same word-picture pairs were presented in the same order in all conditions. Thus, performance on them was comparable across conditions. At the beginning of the experiment, participants saw a set of instructions informing them to read the words aloud quietly and to look at the pictures. They were instructed to rest their hands on the keyboard as they would occasionally be asked to respond to the words and images by typing. The training trials commenced thereafter.

The word-copying task was used as an online measure of participants' attention to form, as it can be said to invoke alertness (participants were told about the task, and it required a speeded response), orientation (instructions were to type the words exactly as they appeared) and detection, or registration, of the written stimuli (evident in the responses). Items were scored as follows: (a) correct, unedited responses were taken as evidence of pristine detection plus rehearsal and received a score of four; (b) correct, edited responses were regarded as correct registration plus self-correction based on monitoring and received a score of three; (c) incorrect, edited responses did not constitute accurate detection but nevertheless showed some monitoring; these received a score of two; (d) incorrect, unedited responses indicated neither accurate detection nor evidence of additional monitoring and thus received a score of one; (e) failures to respond or follow the instructions were scored as zero. This coding system was inspired by the one used by Sauro and Smith (2010) to analyze written responses in L2 computer-mediated communication. Table C1 indicates these results.

The meaning trials provided a measure of the extent to which participants gave meaningful labels to the images used in the experiment. Because the same image, shown in different quantities, consistently appeared with only one stem, it can be assumed that word-referent associations established by the participants are revealed on these trials. Responses to these trials were examined to ensure that participants clearly understood the implied meaning of each word-picture pairing. The meaning trial data also indicated the extent to which alternate construals for the images shown could occur. For example, different participants labeled the same picture as *fish* or *goldfish*, however the range of responses to this task was highly restricted, and showed a preference for basic level terms

(Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1974). Table C2 provides further details of the meaning trials.

These data, however useful, reveal only awareness of form (in the case of the copy trials) or awareness of meaning (in the meaning trials). They do not lend direct insight into participants' awareness of form-meaning connections. Section 6.4.3.5 below describes procedures for assessing participants' awareness of the form-meaning mappings.

6.4.3.3 Experimental conditions: Variation sets

The three conditions were identical apart from the presentation (or, list order) of the paired words and images. As mentioned earlier, the experiment was designed to test the effects of variation set structure on learning outcomes (Onnis, Waterfall, & Edelman, 2008; Onnis, Edelman, Waterfall, 2011) using meaningful stimuli. Three lists were constructed, one of which constituted a baseline pseudo-random condition (i.e., the scrambled condition) and the other two of which attempted to exploit the statistical learning principle of highlighting invariance (Onnis, 2012).

It is important to understand the similarities and differences in the presentation of input tokens in the three conditions. First, based on the review of the literature in Chapter 2, variation sets were operationalized as input sequences exhibiting partial repetition of a meaningful linguistic element across (at least) two successive trials. In the present study, this element could be either morphological or lexical. Because it is not known which, if either, of these options may be more effective in facilitating adult L2 learning of novel constructions, two experimental conditions were employed. In the lexical variation set condition, suffixes repeated (i.e., were carried over from the previous trial) on 33% of the training trials, while no repetition of lexical stems occurred. In the morphological variation set condition, lexical stems repeated on 33% of the training trials, while no repetition of suffixes occurred. Because the constructions as a whole had a different meaning on each trial, the exact same images were never repeated across successive trials. As such, the goal of presenting these constructions in a context of *partial* repetition was not to merely reinforce their surface features, but to also enable access to the semantic components they encoded. The one-third percentage was thought to be sufficient based on previous research. Lastly, in the scrambled (i.e., pseudo-random or control) condition, neither stems nor suffixes repeated across consecutive trials. Thus, no repetition whatsoever occurred in the scrambled condition; across successive trials only variation occurred. Table 6.4 summarizes the conditions used in this study (the abbreviated condition names given in the table will also be used throughout this dissertation). To illustrate further, the first six trials of each condition displayed the following words (underlined word pairs occurred in variation sets). The lexical variation set condition began with: *blon, drosba, tremba, plin, blonmi, doirmi*, the morphological variation set condition began with: *fralba, fralzo, seor, seorzo, blonmi, teunku*, and the scrambled condition began with: *blon, tremba, flisku, dros, blonmi, fralzo*. In each list, the fifth word appeared in a copy trial condition. As noted previously, items used in the copy and meaning trials were identical across the conditions.

Table 6.4. *Partial Repetition in the Three Conditions and Percentage of Total Trials*

Condition	Stem repetition	Suffix repetition
Lexical varset	0 (0%)	48 (33%)
Morpheme varset	48 (33%)	0 (0%)
Scrambled	0 (0%)	0 (0%)

Data collection for the main study began in spring 2013, during which time participants were randomly placed into either the morpheme or scrambled conditions. At the end of this term, preliminary analyses were conducted, based on the mean difference and standard deviation for 60 participants. The mean difference in scores on the untrained picture-word matching task scores was 3.33% in favor of the morpheme group, and the standard deviation was 14.41. No significant difference was found based on a two-sample *t*-test. Next, a power analysis was performed using the R function power.t.test. Assuming that power is .80 (i.e., that the probability of a Type II error is .20), this analysis revealed that approximately 231 participants would be needed in each group in order to demonstrate a significant advantage for the morpheme group. Following Clark-Carter's (2003) guidelines for effect sizes and *t*-tests, the estimated effect size for the contrast between the morpheme group and the scrambled group was $d = 0.23$.

Rather than continuing to employ only these two conditions, it seemed potentially informative to add an additional, third experimental condition. Hence, the lexical condition was developed in order to explore another operationalization of the construct of variation sets within the same experiment. This condition was also based on the literature reviewed in Chapter 2, which notes a number of factors that can be taken into account in defining variation sets. Building on the existing materials, the only change was to construct a new list order for this condition (as described above). Then, during Fall 2013, participants who joined the experiment were placed into the lexical condition. Uniform recruitment procedures were maintained across these two semesters of data collection and the same software, lab equipment, and protocol were used for the entire duration of the study.

6.4.3.4 Testing phase

A picture-word matching task was used to assess participants' knowledge of the form-meaning mappings presented in the exposure phase. This task has been employed in several first language studies (Laine, 1999; Michelon & Zachs, 2003; Saalbach & Imai, 2007; see Jiang, 2012 for review). As noted by Jiang, the task "involves the recognition of both words and pictures and the activation of related concepts or meaning" (p. 161). It was adapted for the present study as follows.

The testing phase took place immediately after training. Instructions for this part of the experiment, as well as two practice items for each subtest, were presented. A message on screen informed participants that they would see words and pictures and be asked to decide whether these words and pictures matched. They were told that half of the items would match and half would not. Each test trial began with the presentation of a fixation cross (500 ms), and then, simultaneously, a word and picture, positioned at the center of the screen. Importantly, the stem and the picture displayed on each trial always matched; however, the match between the bound morphological inflection and picture was either congruent or incongruent with the exposure phase input. The question "is the word a good choice to name the picture?" appeared below these stimuli and participants were instructed to respond by pressing a 'yes' or 'no' label on the keyboard.

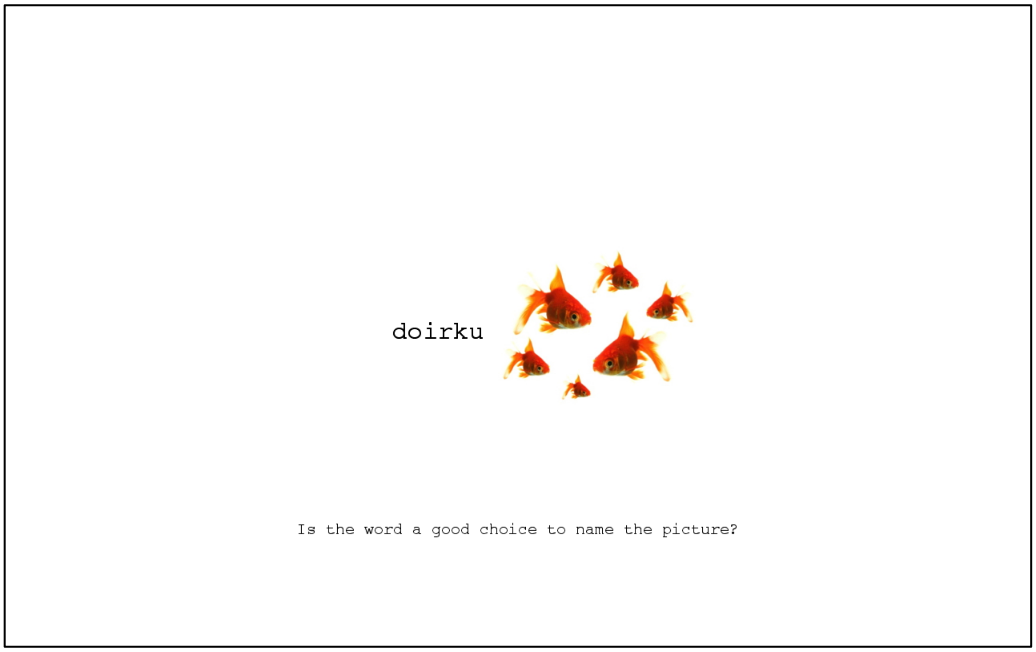

Figure 6.4. A picture-word matching task trial from the experiment.

There were two subtests: a trained subtest and an untrained subtest. The first of these consisted of items seen during the exposure phrase, which directly preceded this subtest. The second subtest tested items for which the inflected form was absent from the exposure phase. Each subtest included 12 items. Six showed correct mappings and six others showed incorrect mappings. Of these latter incorrect mappings, three items showed a word carrying the correct inflection for the noun class of the depicted item (animate or inanimate), but indicating the *wrong* number and three others showed an appropriate number marker given the picture (singular, dual, or plural), but marked *incorrectly* for the feature animacy[19].

Note that some learning based on co-occurrence patterns is possible. That is, it could be surmised that stems taking –mi also take –ku, and that stems taking –ba also take –zo. To reduce the possibility of this type of learning facilitating test performance, no doubly wrong items were used. That is, on incorrect items, a single cue signaled the error, while another cue acted as a lure. All incorrect number items were based on stem plus suffix combinations which were permissible (lure), but which were semantically incongruous in the context of the image shown (signal). Wrong animacy items showed the appropriate number of items for the given suffix (lure), but the stem and suffix combination was not permissible in form (signal). In addition, all untrained incorrect animacy items used stems that did not appear with both suffixes during training (i.e., *blon*, *teun*, *trem*).

During this part of the experiment, the same random item order was used for all participants. The following list (Table 6.5) indicates the training and test items.

[19] Due to an error, this was only the case for the untrained test. On the trained test, there were, again six correct and six incorrect items, though they consisted of five incorrect number items and one incorrect animacy item.

Table 6.5. *Artificial Word Forms Used in the Experiment*

Noun class	Meaning	Singular	Dual	Plural
Animate	fish	doir	**doirmi**	**(doirku)**
Animate	horse	flis	**(flismi)**	**flisku**
Animate	cat	**blon**	blonmi	**(blonku)**
Animate	bird	**(huil)**	huilmi	**huilku**
Animate	monkey	**(cref)**	**crefmi**	crefku
Animate	dog	**teun**	**(teunmi)**	teunku
Inanimate	chair	seor	**(seorba)**	**seorzo**
Inanimate	phone	dros	**drosba**	**(droszo)**
Inanimate	key	**trem**	tremba	**(tremzo)**
Inanimate	book	**(fral)**	fralba	**fralzo**
Inanimate	box	**(haun)**	**haunba**	haunzo
Inanimate	umbrella	**plin**	**(plinba)**	plinzo

Note. Items in parentheses were withheld from training. Items in boldface were tested, with trained and withheld items appearing on separate subtests.

Test items used on the trained and untrained tests, along with individual item statistics, appear in Tables D1 and D2 (see Appendix D). The reliability for the picture-word matching task is reported later (Section 7.2.2).

After each picture-word matching task trial, participants were asked to indicate the basis of their response by choosing one of the following options: guessing, intuition, rule, or memory. This procedure is described in greater detail in the next section. Upon the conclusion of the testing phase, participants saw a message indicating that this part of the experiment had ended.

6.4.3.5 Awareness measures

Though researchers have described the construct of awareness as slippery (Leow, Johnson, & Zárate-Sández, 2011), recent methodological advances seem to hold promise for its interpretation in L2 research. L2 researchers have turned increasingly to cognitive psychology for advice regarding the internal validity of such studies. The present study incorporated measures of awareness at the level of subjective experience (i.e., source attributions) and at the level of understanding form-meaning mappings (i.e., a written questionnaire). This approach allows for triangulation of data sources in order to more firmly grasp the nature of consciousness during L2 learning.

First, based on recent discussion of the role of subjective measures in L2 learning research, an additional feature during the testing phase was the incorporation of trial-by-trial measures of participants' judgment knowledge, or the basis for their yes/no responses to the word-picture matching task. Several recent studies focusing on awareness in adult language learning have incorporated this methodology (Rebuschat, 2008; Rebuschat & Williams, 2012b; see Rebuschat, 2013, for discussion).

After responding to each item on the picture-word matching task, participants were asked on the next screen to attribute the source of their response to one of four choices. The following options and descriptors were used:

1. Guessing - You do not know the answer and made a guess.
2. Intuition - You feel that your answer is correct but you do not know why.
3. Memory - You based your answer on memory.
4. Rule use - You based your answer on a rule that you can state.

Because word-picture matching has rarely, if ever, been used in combination with trial-by-trial measures of source attributions in second language research, this constitutes an unique feature of the methodology.

Second, a more conventional approach to assessing awareness in L2 research is through retrospective measures of verbal report. These may rely on structured interviews, or written instruments, as in the case of questionnaires. Two limitations to these approaches should be noted, related to the aforementioned issue of internal validity. First, there is the issue of veridicality (Leow & Bowles, 2005). That is, regardless of their intentions, participants may forget or misrepresent information underlying their performance. However, Bowles (2010) observes that ensuring only a short delay between performance and verbal report measures can minimize this risk. Second, it is also possible that the act of reporting ones' cognitive processes itself may induce awareness which was not present during performance. Nevertheless, verbal reports, in oral or written format, have been used to probe awareness in numerous L2 studies (e.g., DeKeyser, 1995; de Graff, 1997; Leung & Williams, 2011, 2012; Robinson, 1994, 1997b; Williams, 2004, 2005).

For the reason mentioned above, the post-experiment questionnaire used in this study was administered immediately upon completion of the word-picture matching task. Participants were required to type answers to six open-ended questions delivered through a web browser (Appendix E). The form and sequence of the questions followed the logic of previous studies (e.g., Leung & Williams, 2012). The first item asked participants what they thought the experiment was about. Next, they were asked to report what they had noticed about the words and pictures. They were then asked to state any relationships between the word endings and the pictures they had become aware of. Following this, they were prompted with the word endings, and asked to indicate the meaning of each, providing as much information as possible, even if guessing. The remaining items inquired into participants' intention to learn, strategy use, and motivation for joining the study. No time limit was set for answering any of these questions.

The questionnaire data were coded according to the following categories. These categories reflected two, interrelated levels of awareness important to this study. First, responses were coded according to whether or not participants showed awareness of the relevance of each semantic feature (animacy and number). For example, participants who noted that, "Words without a suffix refer to singularity. There are different suffixes for plurality than for multiplicity. There are also a different set of suffixes for animals than for inanimate objects" or similar were classified as being aware of the relevance of both features. Second, if the participant was aware of the relevance of a feature, the coding further probed whether he or she was accurate or not in describing the forms used to mark this feature. Comments such as, "-mi is when there are 2 animals -ba is 2 objects -ku is 3 or more animals -zo is 3 or more objects" reveal this more specific type of awareness. This example gives an accurate report of the mappings for both features. The written questionnaire was designed to investigate awareness of relevance and accuracy.

The following procedures were used in coding the data. For each coding category, definitions along with examples were written up based on the preliminary data collection. Using this coding protocol, two raters then independently coded questionnaire responses randomly selected from 10 participants, marking each as either aware or unaware of each of the two features. Raters first read participants' responses to each of the questionnaire items and judged whether they showed awareness of the relevance of the features number and animacy, coding individuals as ‘yes’ or ‘no’ for each of the two features. Next, in the case of participants who showed awareness of the relevance of a given feature, responses were additionally judged according to the participants’ degree of accuracy in describing the form-meaning mappings. Accuracy was coded as ‘high’ or ‘low’. After coding the data from 10 participants, the two raters then compared and discussed their use of the coding system to facilitate subsequent rating.

Then, to assess the reliability of the coding scheme, the two raters separately coded responses from 20 additional participants, again selected at random. Based on results from these 30 participants (one third of the data), simple percentage agreement and Cohen’s kappa were computed to gauge the level of agreement between raters. In cases of disagreement, participants’ responses were carefully reviewed and coding decisions revised, as needed. Results in terms of inter-rater reliability are given in the next chapter (Section 7.2.3).

6.4.3.6 Individual difference measures

Independent variable measures included in this study were: openness, intellect, dynamic working memory, and working memory span. Upon completing the awareness questionnaire, participants responded to 20 items from the Openness and Intellect scales of the Big Five Aspect Scales (DeYoung, Quilty, & Peterson, 2007). These items are shown in Appendix E. Each scale contained ten 5-point Likert scale items. Responses were recorded via a web browser and downloaded later for scoring. Scale scores were computed from the average of these ten items, after adjusting for reverse scored items.

The internal reliability of these scales in the present sample was assessed using Cronbach's alpha and their test-retest reliability was also estimated (Section 7.2.4).

In addition, participants filled out a background questionnaire immediately after the personality survey was administered, as part of the same web-based survey. The section above on participant characteristics (Section 6.4.1) contains a summary of these data. Appendix E lists questionnaire items from the participant background questionnaire.

This study used two tests of working memory capacity. The rationale for choosing these two measures was to examine the role of WM in L2 learning from the perspective of measures regarded by researchers to be distinct in terms of the mental processes invoked. There is good evidence to support the argument that WM is domain-general (Hambrick, Kane, & Engle, 2005). For instance, span measures employing rather different verbal versus visuospatial stimuli correlate strongly (e.g., Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). Nonetheless, there is also a distinction between complex span and dynamic WM tasks, which correlate less strongly to span tests than various span measures do with each other (Kane, Conway, Miura, & Colflesh, 2007; for review, see Conway, Getz, Macnamara, Engel de Abreu, 2011). As described by Kane et al. (2007) complex span measures involve serial recall of target stimuli from memory, whereas dynamic tasks such as *n*-back require speeded recognition of stimuli needing to be distinguished from similar targets (p. 621). Considering that both sets of processes—serial recall and speeded recognition—may, in theory, aid considerably in processing novel linguistic input, this study employed one measure of each type.

A dual 3-back task was designed to test WM relying on speeded recognition of visuospatial information. To create stimuli, the researcher followed guidelines to control for the familiarity (and, hence, verbalizability) of the stimuli (Attneave & Arnoult, 1956). Shapes were presented in one of four quadrants on the screen for 500 ms each, separated by a 2500 ms inter-stimulus interval. Participants were informed that they should press the spacebar whenever an item matched the one appearing three before it, according to both shape and position (see Figure 6.5). 48 trials were used, with 16 (33.3%) targets, and 6 (12.5%) lures. Lures presented an item of either the same shape *or* position at two items back. In this task, participants were instructed to respond only to targets. These task specifications drew on previous studies (Gray, Chabris, & Braver, 2003, Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Perrig, & Meier, 2010, Jonides et al., 1997, Nystrom, Braver, Sabb, Delgado, Noll, & Cohen, 2000; Owen, McMillan, Laird, & Bullmore, 2005). Following Haatveit et al. (2010), in order to assess participants' ability to accurately discriminate target from control trials, *d'* was computed by subtracting z-scores for the proportion of correct responses (hit rate) minus the proportion of incorrect responses (false alarm rate).

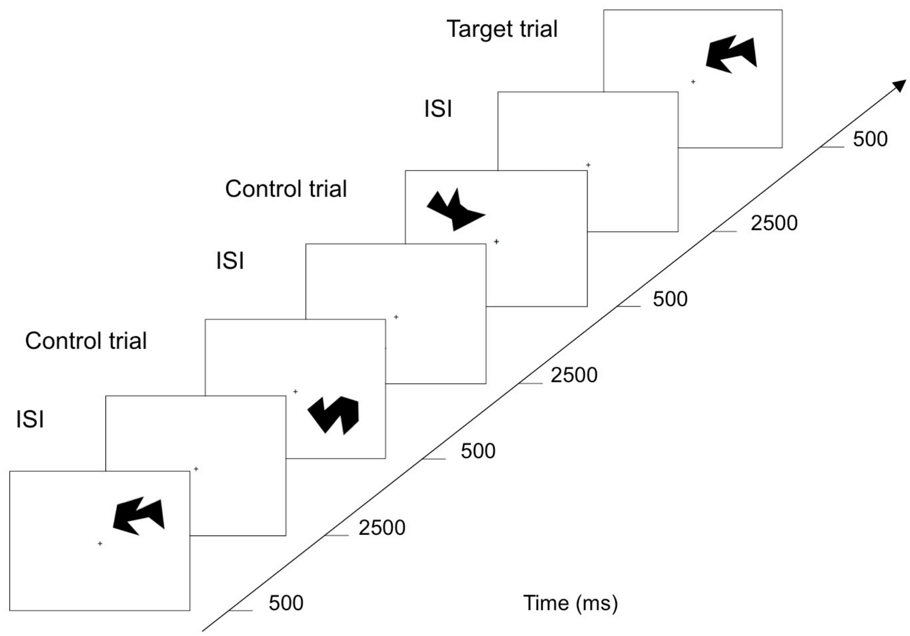

Figure 6.5. Trial sequence in the dual 3-back task.

An automated version of the reading span task (Unsworth, Heitz, Schrock & Engle, 2005) was used to measure participants' WM. Participants were asked to recall a series of alphabet letters appearing after sentences that they read and judged for plausibility. For example, the sentence "every now and then I catch myself swimming blankly at the wall" would be judged as implausible. The task employed 15 items, each consisting of a set of three to seven sentences. Scoring was carried out by the partial-credit, unit-scoring procedure recommended in Conway et al. (2005). That is, the mean proportion of correctly recalled letters within each item was taken as a measure of WM capacity. This task and the dual *n*-back task described below both included a practice session prior to the scored trials. E-Prime 2.0 software was used to administer the automated reading span task. Unsworth, Heitz, Schrock, and Engle (2005) reported reliability and other information for this measure.

For both WM tests, reliability in the present sample was assessed using Cronbach's alpha and also by test-retest correlations. These are reported in the following chapter under preliminary analyses. Figure 6.6 provides an overview of the key variables investigated and how they were measured.

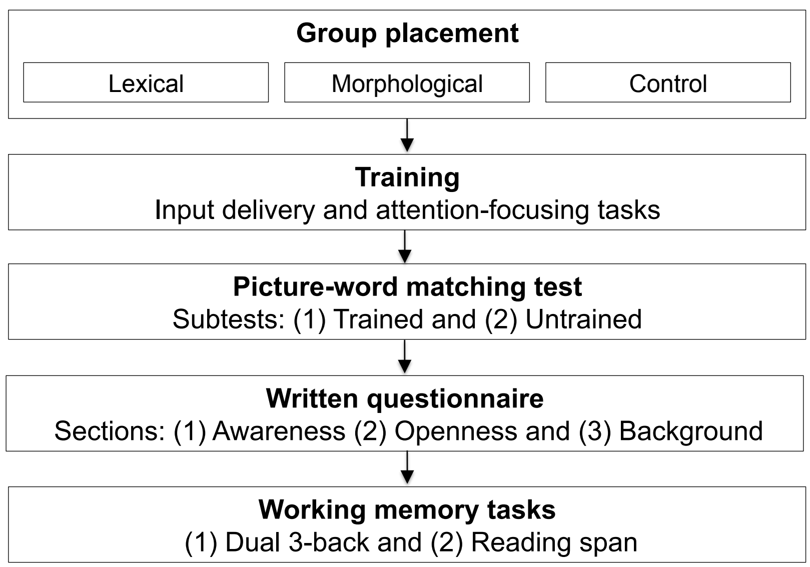

Figure 6.6. Study design.

6.5 Analyses

The general thrust of the analyses was to explore the data set in order to select models accounting for picture-word matching task scores based on relevant categorical and continuous predictors. To answer Research Question 1, binomial regression models with condition as a fixed-effect and subject and item as random-effects were used (e.g., Baayen, 2008; Johnson, 2008). To answer Research Question 2, awareness was added to the model. Research Questions 3 and 4 were addressed mainly using Pearson product-moment correlations. Based on the outcomes of these analyses, individual difference variables and their interactions were modeled, to seek an answer to Research Question 5. Lastly, to answer Research Question 6, the source attribution data were examined according to key features of the study design.

Specifically, mixed-effects binomial regression models were chosen because the criterion variable in the study (picture-word matching task scores) had only two possible outcomes (correct or incorrect). These models furthermore distinguish between effects based on repeatable levels (fixed effects) and effects sampled from a larger population (random effects). Such analyses assume that predictors are uncorrelated; thus the data were checked for collinearity (see Section 7.3.5). Additional correlational analyses were used to explore potential associations among study variables. These involved checking for normality. When considering these analyses, it should be noted that the use of repeated inferential tests as well as the possibility of measurement error should inform readers' interpretations of the significance of the findings.

CHAPTER 7

RESULTS

7.1 Introduction

In this chapter, results and analyses based on the study are reported. To begin, preliminary results are given. Later, analyses for each of the six research questions are presented in separate sections. The analyses reported in this chapter were conducted using R version 3.0.1.

7.2 Preliminary results

This section reports preliminary results from the two attention-focusing tasks (employed as control variables), the picture-word matching task, which was the dependent (criterion) variable measure, and each of the independent (predictor) variable measures. The predictor variables included, in the order in which participants experienced them: (a) the written awareness questionnaire, (b) the openness and intellect scales from the BFAS, (c) the dual 3-back task, and (d) the automated reading span task.

7.2.1 Comparison across conditions on the attention-focusing tasks

Results for word-copying and picture-naming trials, which participants completed during the training phase, are given in the appendix (Tables C1 and C2, respectively). Based on the scoring procedures described in Chapter 6, the word-copying task scores increased slightly after the first trial. Performance on this task across the three conditions was roughly similar, though the average scores of participants in the lexical variation set condition were slightly greater than in the other conditions. In the picture-naming task, as well, accurate performance increased after the first trial. Table C2 also indicates that the words used most frequently by participants to label the pictures were consistently identical to the meanings assigned to the pictures by the experimenter (e.g., *dog*, *box*).

7.2.2 Descriptive statistics and internal reliability for the picture-word matching task

Descriptive statistics for the picture-word matching task appear in Table 7.1. As can be seen, scores were substantially higher on the trained subtest, and more widely dispersed on the untrained subtest. The scores for both subtests appeared to be normally distributed (see also Appendix F for histograms).

Table 7.1. *Descriptive Statistics for the Picture-Word Matching Task*

	Trained	Untrained
Mean	81.20	61.85
Median	83.33	62.50
Skewness	-0.42	-0.16
Kurtosis	-0.60	-0.60
Min	50.00	25.00
Max	100.00	100.00
Range	51.00	76.00
SD	12.59	15.90

Cronbach's alpha was used to compute the internal reliability of the picture-word matching task. First, separate analyses were performed for the scores from the trained and untrained subtests, each of which contained 12 items. Because one item on the trained subtest had an item facility of one (i.e., all participants answered correctly), this item was removed for these analyses. The alpha values for the trained and untrained subtests were .48 and .66, respectively. For the entire test, or the two subtests combined, the alpha was .69. The noticeably lower value for the trained subtest raises concerns about its unidimensionality. Appendix D contains item statistics based on each subtest.

7.2.3 Inter-rater reliability for coding of awareness categories

As mentioned earlier, one third of the participants' responses to the written questionnaire were read and coded by two raters. Agreement was perfect for the yes/no coding of awareness of relevance of animacy and number. Agreement was also substantial for the coding of high versus low accuracy in describing the form-meaning mappings. Here, 96.67% (kappa = 0.95) agreement was obtained for the accuracy of animacy mappings and the accuracy of number mappings. The researcher coded the remaining data independently.

Considering first the course-grained analysis based on awareness of the relevance of the features, there were very few participants who did not report awareness of either feature (3, or 3.33%). Most participants reported being aware of the relevance of number, but did not comment on animacy (61, or 67.78%). The remaining participants reported that both number and animacy were relevant to the experiment (26, or 28.89%). No participant reported the relevance of animacy without also commenting on number.

The more fine-grained analysis of awareness of form-meaning mappings yielded the following results. The three participants who did not report the relevance of either feature also reported none of the mappings correctly. This group was thus labeled “none”. Of participants aware of the relevance of only number, 52 (57.78% of the total) gave reports of number mappings that were not accurate. For example, when asked to indicate the meanings of the word endings -ba, -mi, -ku, and -zo, one of participant in this group replied, “I think these endings may refer to how many items you’re referring to. Like mi might be two, ba might be 3 and so forth. I didn’t pay attention to this rule during the experiment, I only realized now that this might be the meaning of each ending”. In the following analyses, this group was given the label “number-low”. Another nine (10.00%) participants correctly reported the number mappings but without noting anywhere the animacy distinction (e.g., “ba and mi mean two ku and zo mean more than two”). These participants were categorized as “number-high”.

Among those aware of the relevance of both number and animacy, three participants, or 3.33%, gave reports that were inaccurate or lacked detail in terms of number and animacy mappings (e.g., “I think that the mi, and ku endings were added to words that were items or things, whiles the ba, and zo endings were added to animals”). They were classified as “both-low”. Only one (1.11%) person provided a report that was accurate in terms of number but not accurate in terms of animacy (“if I’m not mistaken, -ba is 2 organic objects, mi is 2 non-organic objects, -ku is 3 or more organic objects, and -zo is 3 or more non-organic objects”). Two individuals, or 2.22% of the sample, gave reports that were incorrect in terms of number but correct in terms of animacy (e.g., “mi-singular case for animate objects? ba-singular case for inanimate objects? ku-plural case for animate objects? zo-plural case for inanimate objects?”). The three participants who showed high accuracy in reporting some mappings but low accuracy in reporting others were classified under the label “both-mixed”. Lastly, 20 participants, or 22.22% of the total, reported the right mappings for both number and animacy (e.g., “-mi is plural for two animate objects -zo is plural for three and more inanimate objects -ba is plural for two inanimate objects -ku is plural for three or more animate objects”). This group was labeled “both-high”.

7.2.4 Reliability of the individual differences measures

Descriptive statistics for the individual differences measures are provided in Table 7.2. (see Appendix G for histograms). It can be seen from the table that, although the values for skewedness are generally around zero, the reading span test scores were negatively skewed. This means that care must be taken in interpreting findings based on the reading span measure. In the remainder of this section, procedures for calculating the reliability of these measures are reported, along with their estimates.

Table 7.2. *Descriptive Statistics for Individual Difference Measures*

	Openness	Intellect	Dual 3-back	RSPAN
Mean	3.69	3.33	1.06	0.79
Median	3.70	3.30	1.02	0.82
Skewness	-0.36	0.18	0.04	-1.22
Kurtosis	0.04	-0.49	0.30	1.48
Min	1.70	1.90	-0.71	0.29
Max	4.90	4.60	2.83	1.00
Range	4.20	3.70	4.54	1.71
SD	0.63	0.60	0.66	0.14

For each measure, two reliability estimates are given: Cronbach's alpha, which estimates internal reliability, and Pearson's *r*, which was based on test-retest scores. Test-retest scores came from a sample of 20 participants from the main study who agreed to retake these measures, for an additional compensation of $10. The average time between test administrations for these participants, which ranged from four to 39 days, was approximately 16 days.

First, two separate sets of reliability analyses were conducted for the openness and intellect scales from the BFAS used in this study. These results are reported in Table 7.3. As can be seen, for both scales, the internal reliability coefficient was adequate. In addition, high correlations were found among test-retest scores, indicating the stability of these measures of personality.

Next, previous reports of the reliability of the *n*-back task (Kane, Conway, Miura, & Colflesh, 2007) were consulted to calculate the internal reliability of the dual 3-back task. The analysis presented here was based on scores for target items only ($k = 16$). Two subscales were constructed using the scores from each of four blocks. The estimated reliability coefficient for these scales was 0.68, which is comparable to values reported by Kane et al. (2007). Furthermore, a significant relationship was found between participants' scores on two separate administrations of this task (see Table 7.3).

Lastly, the procedure for computing the internal reliability of the automated reading span task followed guidelines in Unsworth, Heitz, Schrock, and Engle (2005, p. 501). Specifically, participants' scores for the 15 items were combined into three subscores containing the first, second, and third presentation for each set size (corresponding to sets of 3, 4, 5, 6, and 7 letters). The Cronbach's alpha obtained from these three subscores was 0.81, suggesting that the internal reliability of the measure was good. The retest scores for the reading span task also showed a significant correlation.

Table 7.3 summarizes the results of this section, along with significance tests for correlations based on the test-retest study. Because positive correlations between separate administrations of the same measures were expected, one-tailed tests were used to compute *p*-values.

Table 7.3. *Summary of Reliability Indices (Individual Differences Measures)*

		Internal consistency (N = 90)	Test-retest reliability (N = 20)
Measure	*k*	Cronbach's alpha	Pearson's *r*
Openness	10	.77	.83***
Intellect	10	.77	.90***
Dual 3-back	16	.68	.44*
RSPAN	15	.81	.58**

Note. * $p < .05$ **, $p < .01$, *** $p < .001$

7.3 Main results

The main results of the study, which inform Research Questions 1 through 6, are provided in this section.

7.3.1 Effects of the input variation conditions

Moving on to research question one, this section examines the effects of input conditions (lexical variation set, morphological variation set, and scrambled) on outcomes (the picture-word matching subtests). To provide an overview, proportions of accurate responses in each condition are presented first. Thereafter, mixed-effects logistic regression models are introduced. These analyses, which model the variance attributable to items and participants as random effects, are based on correct versus incorrect responses to each item (i.e., not aggregated scores). Separate models were constructed for the trained and untrained subtests. The method used throughout this chapter began with null models based on only the random effects. Then, various predictors were added incrementally. Because it was part of the design of the experiment, condition was added first to the models. Here, and elsewhere in this chapter, summary statistics for models are reported and comparable models are examined for significantly better fit using the χ^2 statistic. The reporting conventions adopted here follow recent work published in the journal *Language Learning* (e.g., de Zeeuw, Verhoeven, & Schreuder, 2012; Blom, Paradis, & Sorenson Duncan, 2012; Sonbul & Schmitt, 2013).

This section presents descriptive statistics based on the proportion of correct and incorrect responses on each subtest for participants in the three study conditions. These responses were dummy coded using 1 for correct and 0 for incorrect. Recall that there were 30 participants in each condition and 12 items per subtest. Thus, for each subtest, the total number of responses per condition was 360. In Table 7.4, these responses are broken down according to whether they were correct and incorrect and percentages are noted.

Table 7.4. *Responses to the Trained and Untrained Subtests by Condition*

	Correct	%	Incorrect	%
Trained subtest				
Lexical varset	293	81.39	67	18.61
Morpheme varset	299	83.06	61	16.94
Scrambled	285	79.17	75	20.83
Untrained subtest				
Lexical varset	228	63.33	132	36.67
Morpheme varset	224	62.22	136	37.78
Scrambled	216	60.00	144	40.00

Figure 7.1 conveys the same results based on the percentage of items correct.

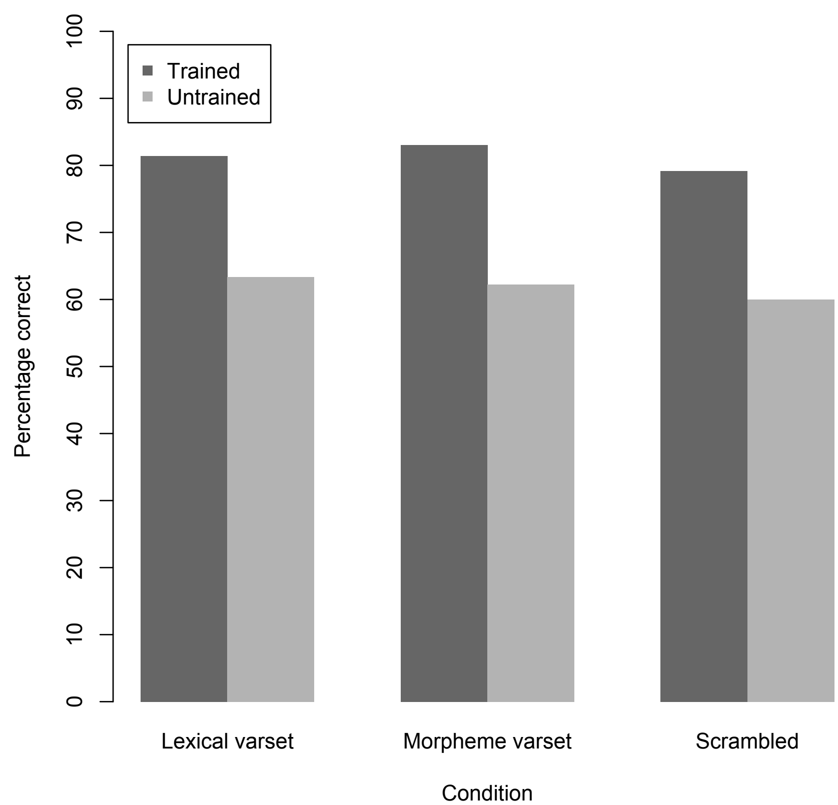

Figure 7.1. Percentage of correct responses in each study condition by subtest.

Language researchers have recently recommended using mixed binomial logistic regression models in the case of binary dependent variables (Baayen, 2008; Jaeger, 2008; Johnson, 2008). In line with these recommendations, the regression analyses that follow are based directly on each participant's yes/no responses to the subtest items, rather than the aggregated means of their test scores. These models incorporate both fixed and random effects. First, null models, or, those with only random effects for the intercepts of item and participants, were computed. Specifically, these two factors were chosen because, were the study to be repeated, one could reasonably expect differences in subtest items, which are based on randomly sampled combinations of stem-suffix-picture, and individuals, whose participation was arbitrary (see Baayen, 2008). Then, to examine whether it affects scores over the random effects, condition was added to the model as a fixed effect. As indicated by the model output (Table 7.5), the scrambled condition was selected as the reference level. This, in turn, allows for consideration of Hypothesis 1, which stated that trained and untrained scores would be higher in the variation set conditions than in the scrambled condition. Results for condition are presented in Table 7.5.

Table 7.5. *Mixed-effect Regression Models Predicting Scores from Condition*

	Trained subtest				Untrained subtest			
Predictor	*B*	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
(Intercept)	1.81	0.42	4.35	0.00	0.47	0.25	1.89	0.06
Lexical varset	0.16	0.25	0.66	0.51	0.16	0.19	0.82	0.41
Morpheme varset	0.30	0.25	1.24	0.22	0.11	0.19	0.57	0.57

As the estimates of the coefficient for each subtest show, the effects (in log-odds) for the variation set conditions were positive in comparison to the scrambled condition, but no significant main effects were found.

Next, likelihood ratio tests were computed to compare the null models to those incorporating the effects of condition. The χ^2 statistic was used to assess whether adding condition as a predictor led to a significantly better fit to the data. These latter models did not lead to a significant improvement in fit in comparison to the null models, for either the trained ($\chi^2(2) = 1.49$, $p = 0.48$) or untrained ($\chi^2(2) = 0.70$, $p = 0.71$) data.

7.3.2 Effects of awareness

The findings in this section address research question two, which concerned the effects of awareness on picture-word matching test scores. First, in order to visualize the data, participants' individual mean scores are presented according to the levels of awareness they showed on the questionnaire. Given the hypotheses that (a) participants aware of the relevance of both semantic features will outperform those who do not report both features, and (b) participants aware of the specific form-meaning mappings will also outperform those do not report these mappings, two logistic regression models were computed. In the first, the group codings were based on awareness of relevance, and in the second, the coding was based on awareness of mappings. These models were then compared and, finally, summary statistics were prepared for the best-fitting models. The model comparisons in this section thus further serve to evaluate whether a more fine-grained coding of awareness may help model its effects more clearly.

Beginning with the course-grained view of awareness, in order to visually examine the data, scores for each participant on the two subtests were averaged and used to generate the plots shown in Figure 7.2.

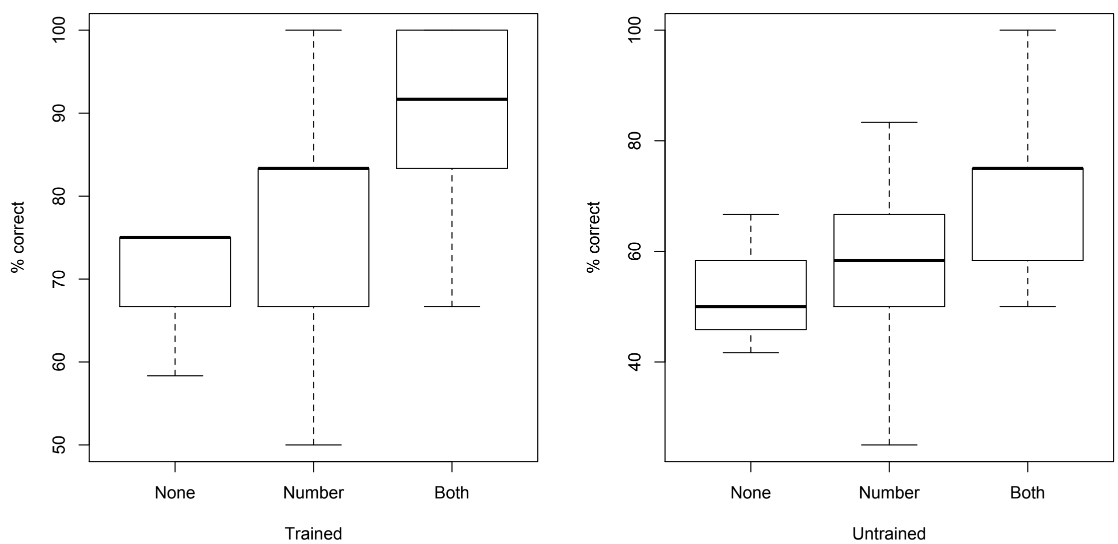

Figure 7.2. Aggregated mean scores for participants on trained and untrained subtests grouped by awareness according to course-grained coding procedures.

Note that the horizontal axis of each plot is comprised of the coding categories described in Section 7.2.3. The plots above used only three levels of coding (based on awareness of the relevance of the two semantic distinctions), whereas those appearing later (Figure 7.3) were based on low or high awareness of the mappings of these distinctions on to suffix forms. As shown in Figure 7.2, scores steadily increased depending on participants' awareness level.

Based on these results, mixed-effects models incorporating awareness were run in order to evaluate its usefulness as a predictor, in addition to the random effects for item and participant. Because condition did not yield significant results, it was not included in these models.[20]

Table 7.6 gives the output from the models employing the course-grained awareness categories (i.e., awareness of relevance).

Table 7.6. *Mixed-effect Regression Models Predicting Scores from Awareness of Relevance*

	Trained subtest				Untrained subtest			
Predictor	*B*	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
(Intercept)	1.14	0.58	1.95	0.05	0.13	0.45	0.29	0.77
Number	0.56	0.46	1.23	0.22	0.26	0.41	0.63	0.53
Both	1.60	0.49	3.24	0.00	0.90	0.42	2.13	0.03

As indicated, the scores of participants in the group aware of the relevance of both number and animacy showed significant increases over the reference group, in which participants did not report awareness of either feature. In this latter group, the proportion of correct responses was 69.44% for the trained subtest and 52.78% for the untrained subtest. The group aware of number only replied correctly on 78.14% of trained items and on 58.33% of untrained items. Lastly, the group aware of both produced 89.74% and 71.15% correct responses on the trained and untrained tests, respectively. The models reported in Table 7.6 showed a better fit than the null models, for the trained data ($\chi^2(2) = 22.37$, $p = 0.00$) and also for the untrained data ($\chi^2(2) = 14.53$, $p = 0.00$).

[20] It can also be noted that awareness and condition were unrelated. A chi-square test using the course-grained coding indicated that the observed frequencies for awareness and condition were independent ($\chi^2(4) = 2.57$, $p = 0.63$).

Next, models employing the fine-grained awareness categories (i.e., awareness of mappings) were constructed. These models incorporated a larger number of categories than the course-grained models. In addition, analyses indicated that a slightly better fit to the data could be obtained by using a larger reference group, composed of participants whose written questionnaire responses had been coded as none and number-low.[21] Participants in these groups were similar in that they were unaware of animacy and reported no specific form-meaning mappings.

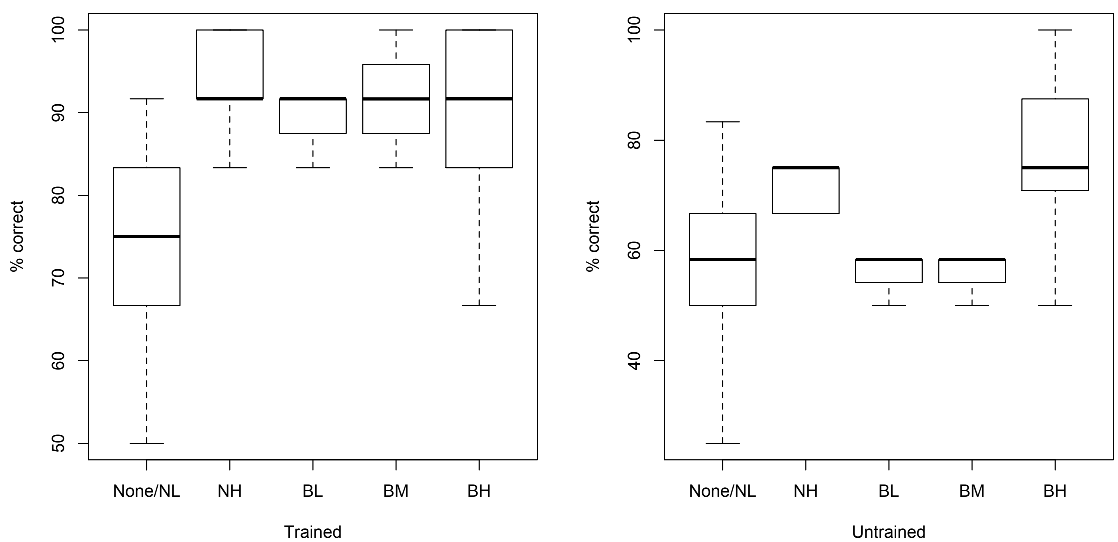

Figure 7.3. Aggregated mean scores for participants on trained and untrained subtests grouped by awareness according to fine-grained coding procedures (NL = Number-low; NH = Number-high; BL = Both-low; BM = Both-mixed; BH = Both-high).

[21] Fine-grained models assigning participants from both none and number-low ($n = 55$) groups to the reference group had lower AICs than those using only none ($n = 3$) as the reference group. The AIC, or Akaike Information Criterion, is used to judge model fit. A lower value for AIC indicates better fit.

To illustrate the picture-word matching task score differences in these groups, Figure 7.3 shows the mean scores of participants based on their awareness of both relevance and specific form-meaning mappings. Compared with Figure 7.2, these plots reveal some exceptions to the steady increase in scores based on participants' awareness of relevance, particularly for those who knew that both number and animacy were relevant, but did not report these mappings accurately (i.e., those classified as both-low or both-mixed).

The output from the models employing the fine-grained awareness categories (i.e., awareness of mappings) is provided in Table 7.7.

Table 7.7. *Mixed-effect Regression Models Predicting Scores from Awareness of Mappings*

	Trained subtest				Untrained subtest			
Predictor	*B*	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
(Intercept)	1.48	0.39	3.76	0.00	0.26	0.22	1.17	0.24
Number-high	1.76	0.42	4.22	0.00	0.81	0.24	3.36	0.00
Both-low	1.12	0.57	1.98	0.05	-0.01	0.36	-0.03	0.98
Both-mixed	1.47	0.64	2.31	0.02	-0.01	0.36	-0.03	0.98
Both-high	1.20	0.24	4.93	0.00	1.01	0.18	5.67	0.00

As the table indicates, the scores of those in several groups showed significant increases over the reference group, in which participants indicated no more than partial awareness. In the reference group, which combined none and number-low participants, the proportion of correct responses was 75.15% (trained subtest) and 55.78% (untrained subtest). It is useful to compare these values with those of the groups listed in Table 7.7.

In each group, for the trained and untrained subtests, respectively, these proportions were as follows: number-high (93.52% and 72.22%), both-low (88.89% and 55.56%), both-mixed (91.67% and 55.56%), and both-high (89.58% and 75.83%).

On the trained subtest, the largest effect was found for the number-high group and on the untrained subtest, the largest effect was found for the both-high group. To gain additional perspective on these results, confidence intervals (CIs) were estimated based on the coefficients for the significant predictor groups. For the trained test scores, Table 7.5 shows clear, significant effects for number-high, 95% CI [0.94, 2.58], both-mixed [0.22, 2.71], and both-high [0.72, 1.68]. For the untrained test scores, on the other hand, there were significant effects for number-high [0.34, 1.27] and both-high [0.66, 1.36].

The fine-grained models proved a better fit than the null models (trained, $\chi^2(4) = 45.57$, $p = 0.00$; untrained, $\chi^2(4) = 35.49$, $p = 0.00$). In addition, the fine-grained awareness coding categories also provided a significantly better fit than the course-grained categories did, for the trained data ($\chi^2(2) = 23.21$, $p = 0.00$), as well as the untrained data ($\chi^2(2) = 20.97$, $p = 0.00$).

7.3.3 Relationships between WMC and picture-word matching task scores

Based on Research Question 3, the relationship between scores on the picture-word matching task and the two measures of WMC (dual 3-back and reading span) was considered next. First, the data were examined to determine whether the assumptions underlying these analyses were met. The scores for the dual 3-back were normally distributed. However, it should be pointed out that scores on the reading span were unevenly distributed. Again, because the reading span scores were negatively skewed (-

1.22), care must be taken when interpreting the following results. Appendix G provides histograms for these measures.

Next, the data were plotted to inspect relationships between scores on both WM tasks and the trained and untrained subtests of the picture-word matching task.

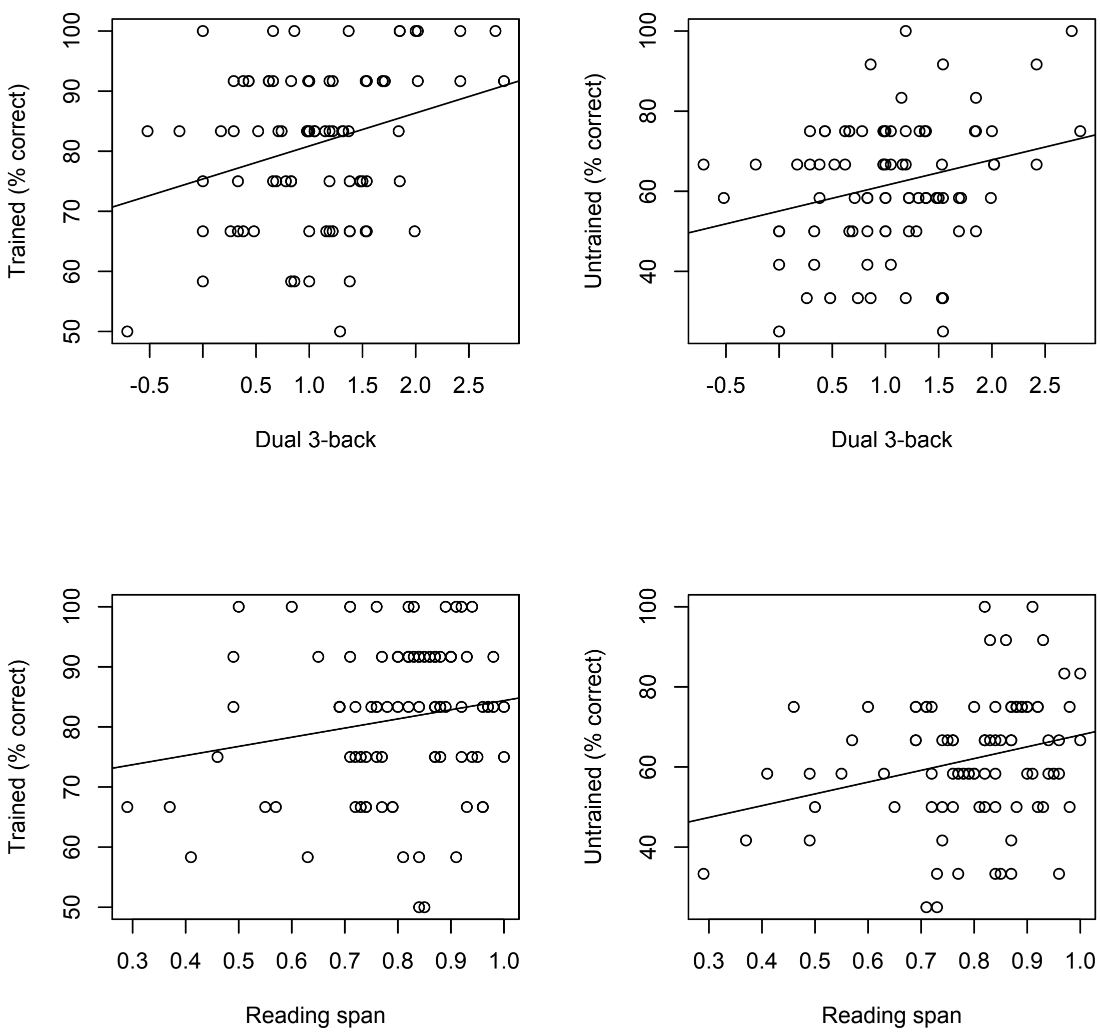

Figure 7.4. Scatterplots indicating the relationship between picture-word matching task scores and WM task scores.

The plots shown in Figure 7.4 include regression lines. As shown in the figure, there were positive, linear relationships between WM and picture-word matching task scores. The correlations between these variables were also computed. Hypothesis 3a assumed that complex span and dynamic measures of WMC would be positively related to picture-word matching task scores. Thus, critical values of the Pearson correlation coefficient were based on a one-tailed test. For the dual 3-back task[22], the correlations were significant in the case of the trained ($r(88) = 0.29$, $p = 0.00$) and untrained ($r(88) = 0.27$, $p = 0.01$) subtests. It is important to note here that low reliability, as found in the case of the dual 3-back task, especially, should be kept in mind as a factor that may reduce the magnitude of a correlation. For the reading span task and the trained subtest, the p-value was not below the criterion level of 0.05, though, as expected, the correlation between these scores was positive ($r(88) = 0.17$, $p = 0.05$). Reading span and untrained subtest scores were weakly, yet significantly, correlated ($r(88) = 0.27$, $p = 0.01$). These results suggest that, at best, the amount of shared variance between WMC and picture-word matching task scores was between 7 and 8 percent.

In addition to these analyses combining the data from all three conditions, further analyses were carried out to investigate the potential influence of WMC on outcomes under each separate condition. As a reminder, Hypothesis 3b argued that in the scrambled condition, the positive associations between complex span and dynamic measures of WMC and outcomes would be stronger than in the variation set conditions. The following plots illustrate the data separately for the lexical variation set, morphological variation set,

[22] Results for the two WM tests are presented in the order in which participants took them, with the dual 3-back first. When discussing these tasks elsewhere, the order may be reversed to reflect the greater popularity of complex span measures in the literature.

and scrambled conditions, first for the dual 3-back task (Figure 7.5) and then for the reading span task (Figure 7.6).

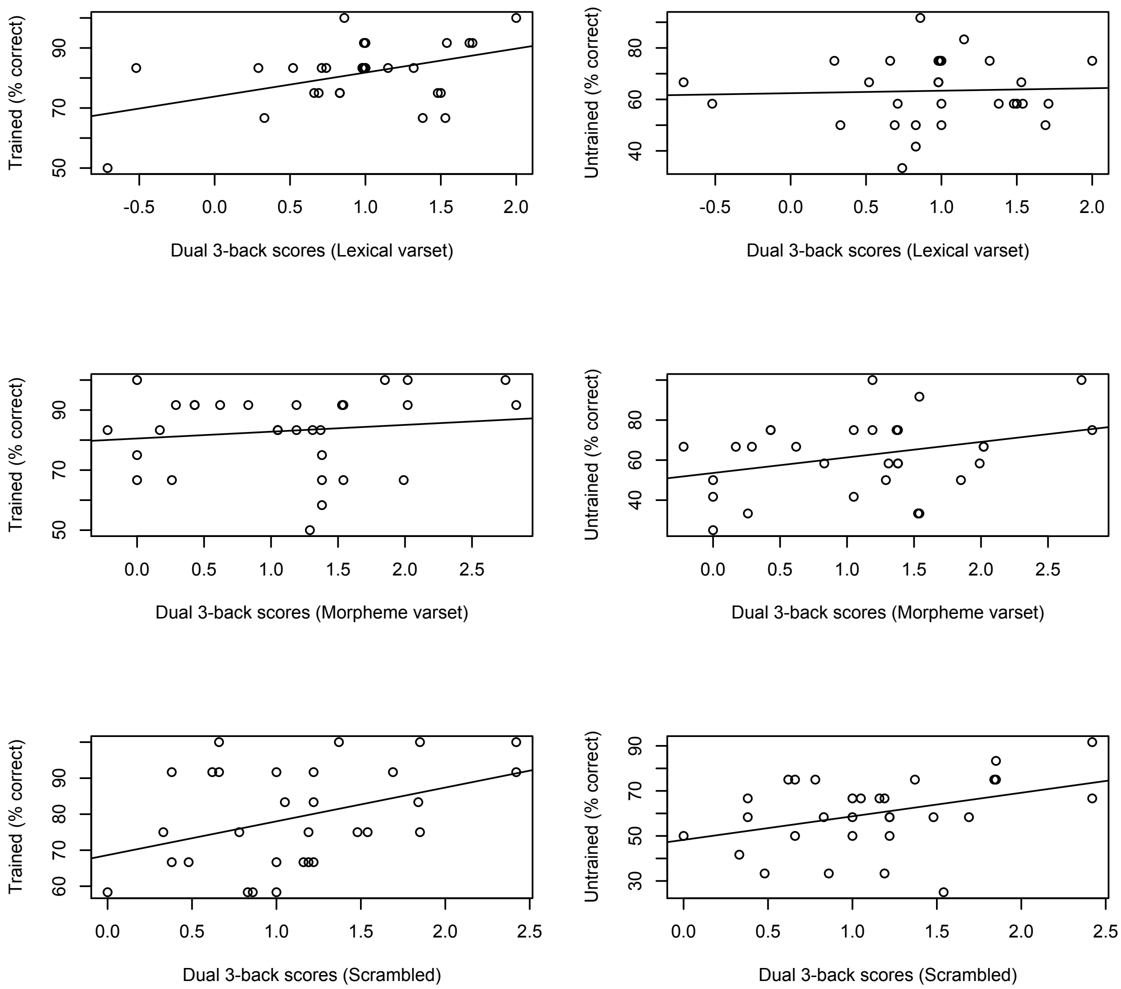

Figure 7.5. Scatterplots indicating the relationship between picture-word matching task scores and dual 3-back task scores by condition.

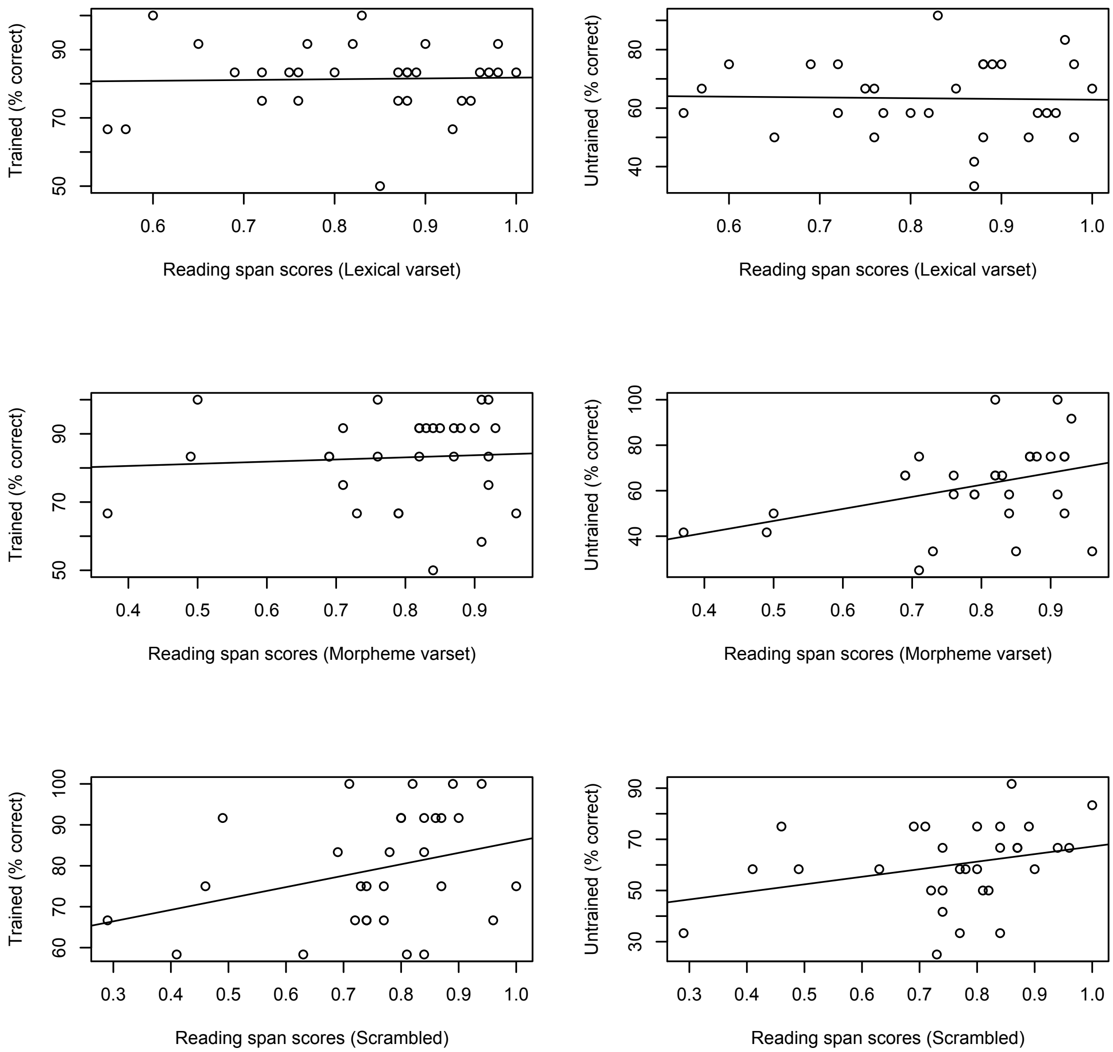

Figure 7.6. Scatterplots indicating the relationship between picture-word matching task scores and reading span task scores by condition.

To further evaluate Hypothesis 3b, correlation coefficients for WM and subtest scores were calculated for each condition. Below, these are reported first for the dual 3-back task and second for the reading span task.

Regarding the dual 3-back, the findings for the lexical variation set group differed across the trained ($r(28) = 0.45$, $p = 0.01$) and untrained ($r(28) = 0.04$, $p = 0.82$) subtests. For the morpheme variation group, the results were not significant for either the trained ($r(28) = 0.14$, $p = 0.47$) or untrained ($r(28) = 0.33$, $p = 0.07$) subtests. Lastly, for the scrambled group, significant correlations emerged both for the trained ($r(28) = 0.40$, $p = 0.03$) and untrained subtests ($r(28) = 0.39$, $p = 0.03$). Bearing in mind that no differences were found in learning outcomes across these conditions (see Section 7.3), the results from the dual 3-back task indicate some support for the notion that variation set conditions may lessen the burden on WM.

Next, considering the results from the reading span task, in the lexical variation set group, the correlations were consistently low between this ID measure and the trained ($r(28) = 0.03$, $p = 0.88$) as well as the untrained subtests ($r(28) = -0.03$, $p = 0.90$). In the morpheme variation group, a significant correlation was not found for the trained subtest ($r(28) = 0.07$, $p = 0.73$), but was for the untrained subtest ($r(28) = 0.39$, $p = 0.03$). Finally, for the scrambled group, the relationship was positive, but nonetheless did not reach significance for either the trained ($r(28) = 0.32$, $p = 0.08$) or untrained subtests ($r(28) = 0.30$, $p = 0.10$). Here, the evidence for Hypothesis 3b is merely suggestive. While it is intriguing that in the scrambled group the correlation coefficients between reading span and both subtest scores were in the range $r = .3$, these findings were not significant.

As a final, important caveat, it should be noted that the number of statistical comparisons used to evaluate Hypothesis 3b increases the likelihood of correlations coefficients attaining significance by chance alone. As a further point of reference, Appendix H contains inter-correlations for all dependent and independent variables.

7.3.4 Relationships between personality, cognitive variables, and outcomes

Research Question 4 was included to explore possible links between personality (i.e., openness, intellect) and measures of (a) learning outcomes and (b) cognitive ability. As noted earlier, scores on the written survey scales used to measure these ID variables were normally distributed. Figures G2 and G3 in the appendix show the histograms.

The data were visualized using scatterplots fitted with a regression line for each pair of variables. These visuals appear in Figure 7.7, Figure 7.8, and Figure 7.9. These are discussed in turn below.

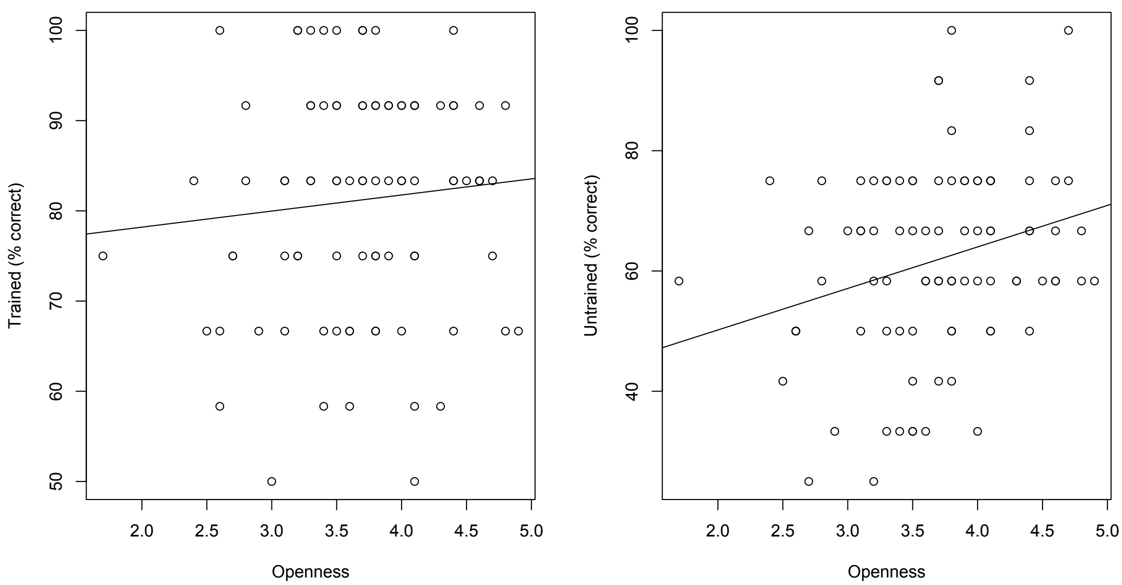

Figure 7.7. Scatterplots indicating the relationship between picture-word matching task scores and openness scores.

The data for subtest scores and openness were widely dispersed. There was a pattern of rising scores on the untrained subtest as openness increased. Of the only two participants who attained a perfect score on the untrained test, both were above the mean score for openness.

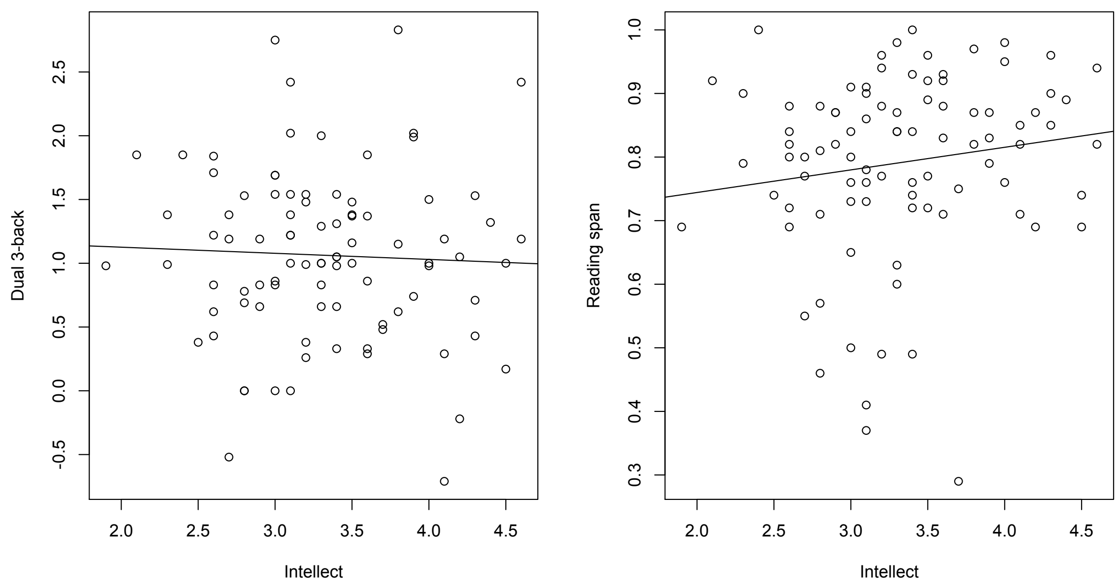

Figure 7.8. Scatterplots indicating the relationship between WM task scores and intellect scores.

When intellect was plotted against WM scores, the data were dispersed. The trend was slightly negative for the dual 3-back and slightly positive for the reading span task. Thus no consistent pattern emerged for intellect and WM.

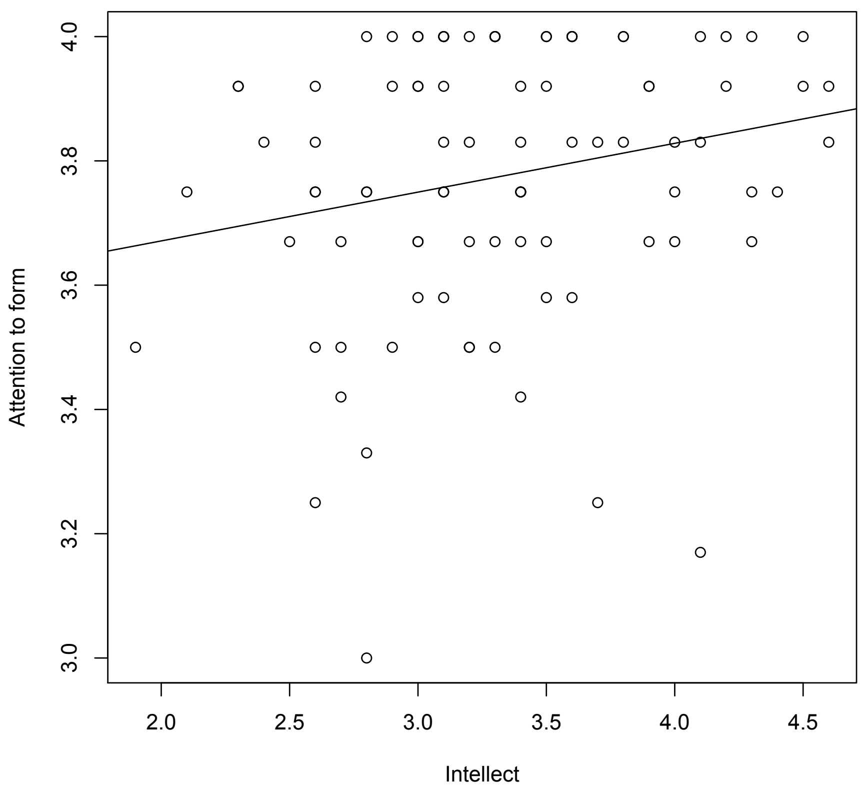

Figure 7.9. Scatterplots indicating the relationship between attention to form task scores and intellect scores.

Scores for attention to form (i.e., the word-copying task) and intellect clustered at the top of the graph, due to the uneven distribution of scores for attention to form. A small positive trend appeared in the data. There were, however, a few outliers whose scores were above the mean for intellect but who nonetheless scored low on attention to form.

Based on the hypotheses outlined in the previous chapter, tests of the correlations between the above variables were carried out. Hypothesis 4a assumed that openness would be positively related to any learning in the study, that is, the scores on the picture-word matching subtests. Statistical support was not found for the trained subtest ($r(88) = 0.09$, $p = 0.41$). However, a significant, positive relationship was found between openness and untrained subtest scores ($r(88) = 0.27$, $p = 0.00$). Hypothesis 4b asserted that intellect would be positively associated with WM scores and with attention to form. No evidence for a relationship between intellect and WM can be claimed for either the dual 3-back task ($r(88) = -0.04$, $p = 0.69$) or the reading span task ($r(88) = 0.15$, $p = 0.08$). On the contrary, intellect was significantly related to attention to form ($r(88) = 0.21$, $p = 0.04$), although, as noted above, the scores for attention to form were negatively skewed.

7.3.5 Multiple predictors of scores on the picture-word matching task

This section seeks to answer research question five, which considers the influence of multiple factors on L2 construction learning. The preceding sections have already shed some light on this: condition did not appear to predict outcomes, though awareness did. Furthermore, while the random effect of items was noted, the effects of specific item types (correct, incorrect animacy, incorrect number) can also be modeled. Also, WM and personality were, to varying degrees, positively associated with mean scores on the picture-word matching test. So, in this section, these variables will be added to the previously described models based on awareness (see Table 7.7). The aim is to find out whether these other experimental and individual factors can account for any additional variance.

One important assumption in regression analyses is that predictors in a model should not be overly correlated. Thus, to begin with, the collinearity of the predictor variables was assessed as follows. First, generalized regression models were fitted with accuracy as the outcome and awareness, item type (coded as correct, incorrect number, and incorrect animacy), openness, dual 3-back, and reading span as predictors. Second, the variance inflation factors (VIF) for these models were extracted. These values, along with averages, are shown in Table 7.8.

Table 7.8. *Collinearity Diagnostics for Predictors Entered into the Models (VIFs)*

Predictor	Trained model	Untrained model
Awareness	1.15	1.20
Dual 3-back	1.12	1.15
Item type	1.02	1.02
Openness	1.09	1.07
Reading span	1.15	1.15
Average	1.11	1.12

It is often noted that no value of VIF should exceed 10, nor should the average VIF be considerably greater than 1. Given that collinearity appears unlikely in the present case, these predictors were added incrementally to the above models, and the likelihood ratio test was again used to compare models to previous ones.

First, considering the target constructions and test design, item type was added to the models. Subtest items containing anomalous constructions (i.e., incorrect items, to which participants should have responded ‘no’), showed clear, negative effects, which were stronger when the items displayed mismatches involving animacy markers. That is, scores decreased (relative to correct items) on items showing the wrong number marker and decreased more when items showed the wrong animacy marker. With item type included, the model fit was significantly improved. Compared to the models in which the only fixed effect was awareness, the likelihood ratio test result was significant for the trained ($\chi^2(2) = 16.09$, $p = 0.00$) and untrained ($\chi^2(2) = 12.74$, $p = 0.00$) data.

Next, based on theory and research, scores on the reading span measure were added to the models. The fit of these models was also better than when only awareness and item type were included as fixed effects. For the trained data, this result was only marginally significant ($\chi^2(1) = 3.68$, $p = 0.05$). However, for the untrained data, the result was clearly significant ($\chi^2(1) = 10.34$, $p = 0.00$). Next, dual 3-back scores were added, but this variable did not lead to any significant improvement in fit for either trained ($\chi^2(1) = 1.16$, $p = 0.28$) or untrained ($\chi^2(1) = 1.60$, $p = 0.21$) outcomes. Thus, dual 3-back was removed. Then, openness was added as a predictor to the models, however, this led to no significant improvement in model fit for trained ($\chi^2(1) = 0.17$, $p = 0.68$) or untrained ($\chi^2(1) = 2.36$, $p = 0.13$) scores.

Finally, based on hypotheses 3b, interactions between WM and condition were added to these models to examine their potential effects. To test for interactions, reading span was removed from the models and they were run with fixed effects for item type, awareness, and the interaction between condition and WM (for both WM tests). These models were then compared to those modeling the effects of awareness, item type, and reading span independently. Unfortunately, this did not improve model fit, when the condition by reading span interaction was entered (trained: $\chi^2(4) = 3.50$, $p = 0.48$; untrained: $\chi^2(4) = 2.70$, $p = 0.61$) or when the condition by dual 3-back interaction was entered (trained: $\chi^2(4) = 2.50$, $p = 0.64$; untrained: $\chi^2(4) = 0.00$, $p = 1.00$). Thus, no interactions were included in the final models. Summaries of the best-fitting models, which include item type, awareness and reading span as predictors, are shown in Table 7.9.

Table 7.9. *Mixed-effect Regression Models Predicting Scores from Multiple Variables*

	Trained subtest				Untrained subtest			
Predictor	*B*	*SE*	*Z*	*p*	*B*	*SE*	*Z*	*p*
(Intercept)	1.53	0.53	2.92	0.00	-0.45	0.42	-1.07	0.28
Item type: Incorrect animacy	-2.84	0.60	-4.77	0.00	-1.49	0.31	-4.76	0.00
Item type: Incorrect number	-1.85	0.36	-5.10	0.00	-0.56	0.31	-1.78	0.08
Awareness: Number-high	1.70	0.42	4.05	0.00	0.72	0.24	2.97	0.00
Awareness: Both-low	1.21	0.57	2.13	0.03	0.11	0.37	0.31	0.76
Awareness: Both-mixed	1.43	0.64	2.24	0.03	-0.07	0.36	-0.20	0.84
Awareness: Both-high	1.24	0.25	5.05	0.00	1.07	0.18	5.91	0.00
Reading span	1.13	0.58	1.94	0.05	1.54	0.48	3.21	0.00

As above, the CIs for the coefficients in both models were estimated. These are reported here for all significant predictors shown in Table 7.9. For the model fitted to trained test scores, significant effects were found for item type, including incorrect animacy, 95% CI [-4.01, -1.67] and incorrect number items [-2.56, -1.14], and for awareness, or those assigned to the categories number-high [0.88, 2.52], both-low [0.10, 2.33], both-mixed [0.18, 2.68], and both-high [0.76, 1.73]. For the model fitted to untrained test scores, significant effects were again found for item type, but only for incorrect animacy [-2.10, -0.87], for awareness, here limited to those classified as number-high [0.24, 1.19] and both-high [0.71, 1.42], and for reading span [0.60, 2.48].

Before concluding, a measure of the goodness of fit for these final models is presented. For logistic regression models, fit can be evaluated using the *C* index of concordance (Baayen, 2008; Gries, 2009). Gries indicates that a model performs well at classifying outcomes when the value of *C* is near or above 0.8. For the models in Table 7.9, the values for the *C* index were .81 and .73, for the trained and untrained data, respectively.

7.3.6 Additional evidence from source attributions

In this section, Research Question 6, regarding the role of knowledge source attributions in describing L2 construction learning, is considered. Initially, the overall pattern of responses within the dataset was examined. This gives a broad perspective on the use of each attribution and its potential relevance to any learning that may have occurred during the study.

Then, to focus the analysis, source attributions from the two groups whose performance was significantly better than the unaware group on the untrained subtest (the number-high and both-high groups) were analyzed in detail. There were two reasons for this. First, based on their picture-word matching task scores, these groups arguably learned the most (see Figure 7.3). So, if implicit and explicit knowledge sources coexist in L2 processing, it should be observable within these groups. Second, these participants uniformly reported precise mappings—a clear indicator of explicit knowledge. Thus, the source attributions associated with their responses are of particular interest because Hypothesis 6 assumes that both implicit and explicit knowledge may interact during performance. Thus, this approach considers the use of attributions generally, as well as their use by specific groups, insofar as both of these approaches are potentially valuable for developing this study's final interpretation.

The data were graphed to show the number and distribution of responses for each attribution category, as well as the accuracy of picture-word matching task responses associated with them. These graphs were divided by subtest and item type. The vertical axis for each graph is based on the total number of responses per item type. The bars give raw counts of correct and incorrect responses side-by-side, nested within each item type. This approach allows success rates for each attribution type to be seen easily.

First, these analyses were conducted for the entire sample of 90 participants. They were then repeated based on the data from only those in the number-high group ($n = 9$) and the both-high group ($n = 20$). Figures 7.10, 7.11, and 7.12 present these results. These data appear in table format in Appendix I, which reports source attribution use and accuracy by subtest and item for all awareness groups.

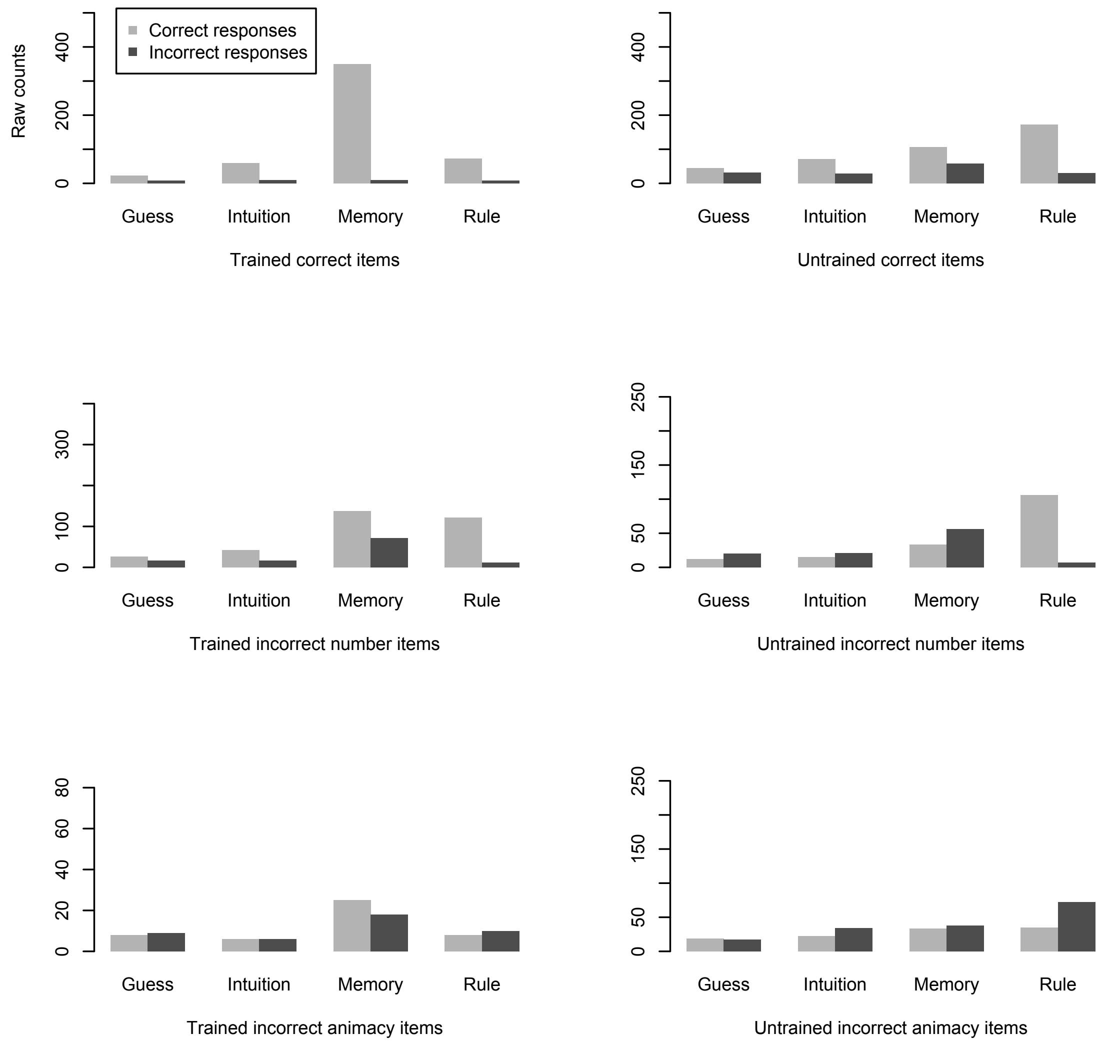

Figure 7.10. Source attributions used by all participants (*N* = 90) for each subtest and item type.

The overall results suggest that, for trained items, memory is a powerful knowledge source, particularly for correct items (Figure 7.10). However, memory can also lead participants astray, as seen in the number of incorrect responses when it was used to judge infelicitous items. To perform well on those items, participants should have ignored certain cues (presumably available in memory) and focused on other cues to identify the source of the error. Namely, wrong number items were all permissible forms, but did not match the number shown in the picture. Conversely, wrong animacy items were formally incorrect, but their attached suffixes indicated the number of items shown in the picture. On untrained items, rule knowledge came into play more than other attributions. Given the categorical nature of the target construction, this is somewhat unsurprising. On untrained incorrect items, the importance of rule knowledge stands out even more: non-rule attributions almost invariably resulted in more incorrect than correct responses. However, even rule knowledge was not helpful for judging untrained incorrect animacy items. These overall results also suggest a limited role for intuition. As can be seen, intuition was more helpful than harmful only when used to respond to three item types: trained correct items, untrained correct items, and trained incorrect number items.

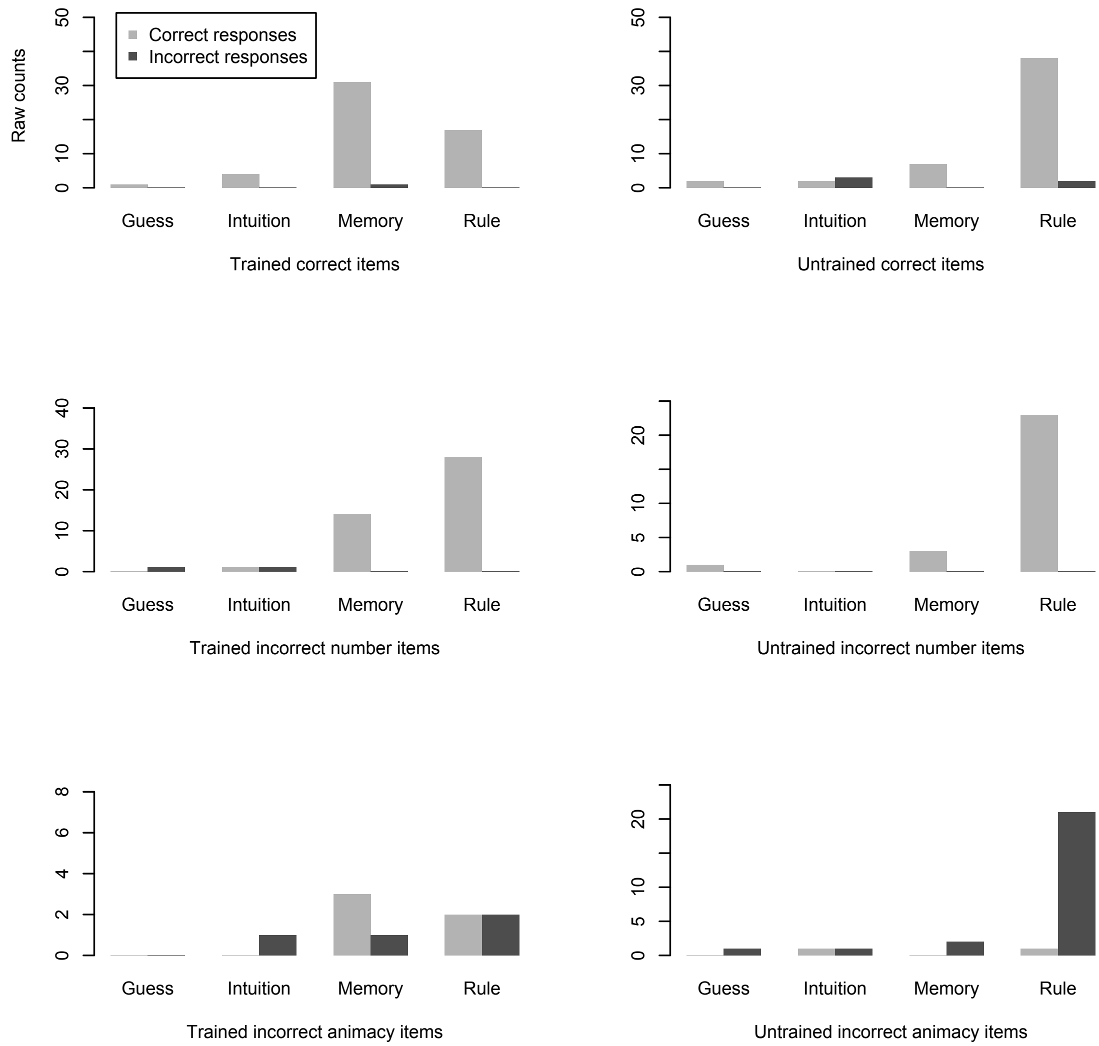

Figure 7.11. Source attributions used by number-high participants (*n* = 9) for each subtest and item type.

Turning to the results for number-high participants, memory was again favored on trained correct items (Figure 7.11). Also, rule knowledge predominated on trained incorrect number items, and on all untrained item types. This served these participants well, until they were faced with incorrect animacy items. Then, applying a rule based on number only had dire consequences. Intuition and guess attributions had a lesser role for these participants.

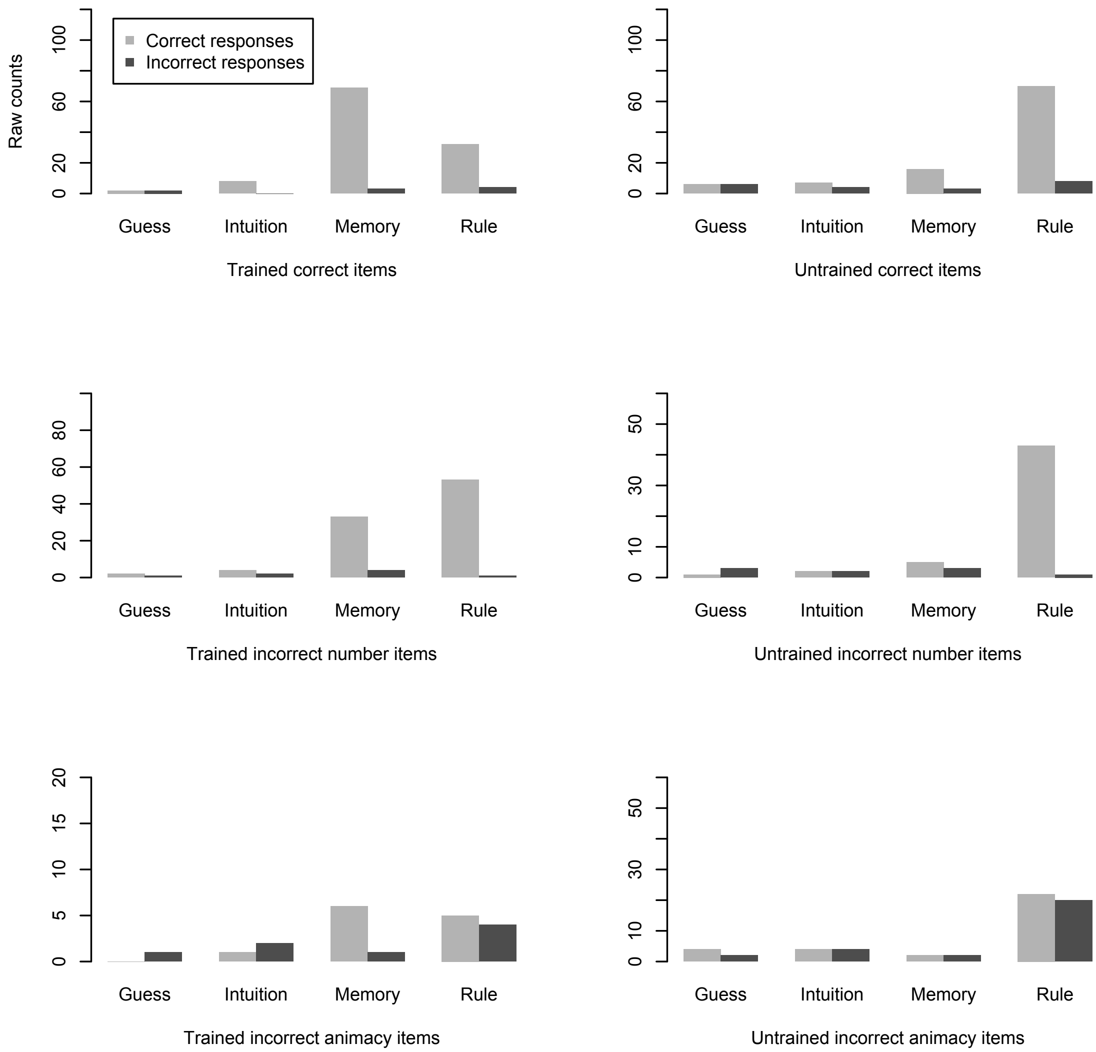

Figure 7.12. Source attributions used by both-high participants (*n* = 20) for each subtest and item type.

Lastly, considering the both-high group, participants alternated between memory and rule attributions on the trained subtest (Figure 7.12). Both of these attribution categories facilitated their performance. On the untrained subtest, rule knowledge was deployed more often than any other category. Rule attributions were consistently associated with a greater proportion of correct responses on all item types, regardless of subtest. For these participants, use of intuition was infrequent and sometimes unsuccessful.

CHAPTER 8

DISCUSSION AND CONCLUSIONS

8.1 Introduction

This closing chapter begins with three sections based on several potential factors in L2 construction learning considered within this dissertation: (1) partial repetition in the input, (2) learner-generated awareness, and (3) individual differences in WM and personality. Within each of these sections, I will revisit the research questions and hypotheses, summarize the outcomes, interpret and explain major findings, and seek to contextualize results with respect to the literature reviewed previously. The relative contribution of these variables is then discussed. Afterwards, this chapter presents a summary of the main findings of the study before elaborating on theoretical, methodological, and practical implications related to this area of research. The final section seeks to articulate future directions for the study of L2 construction learning based on an understanding of the roles of input variation, learner awareness, and individual differences.

8.2 Variation sets in adult L2 construction learning

Research Question 1: What are the effects of the input conditions on picture-word matching test scores?

- Hypothesis 1: Trained and untrained scores will be higher in the variation set conditions than in the scrambled condition.

Based on the operational definitions used in this study (which included morphological and lexical forms of input variation), no advantage was found for participants in the variation set conditions. While it is true that participants in these conditions performed slightly better than those in the scrambled condition, a model fit to the data revealed no main effects for condition as a predictor of the proportion of correct responses on either the trained or untrained subtests. Furthermore, this model did not account for the data any better than a null model containing only item and participant as random effects. Thus, no substantial support can be claimed for Hypothesis 1.

These results suggest that the findings of Onnis et al. (2008) showing better performance in variation set conditions cannot readily be compared to the learning of form-meaning connections in the present study. Several possible explanations for these divergent findings are related to differences in the two studies. First, the former study examined the learning of word and phrase boundaries from artificial speech streams, while the present study looked at acquiring semantically-motivated morphological units. This was in order to examine whether variation sets could facilitate the learning of L2 constructions, which is distinct from learning structural characteristics. In particular, constructional generalizations require learners to go beyond surface features. Although it was assumed that arranging the input into variation sets might help them do so, it was found that participants' awareness on the post-experiment questionnaire was unrelated to input condition. Had awareness been dependent on condition, then a clear link between the mechanisms underlying performance in the Onnis et al. study and the learning of L2 constructions in the present study would have emerged. Second, the addition of picture cues added a rich source of information about the learning targets, such that discerning potentially relevant semantic features based on them may have competed with analyzing the structural aspects of the written word forms. Indeed, this process may be regarded as at least as important, if not more, to generalizing from the input provided. Related to this, it is also worth noting that the written forms were often segmented (that is, stems and suffixed forms both appeared in the trained input, e.g., *doir* and *doirmi*), so that identifying these boundaries, presumably a substantial part of the learning in the Onnis et al. study, would have been easier here. Third, each of the conditions in the present study required that all participants pay attention to the word forms and the meanings of the images. This was successfully achieved through the word-copying and picture-naming trials. As noted, performance on these trials was very similar across conditions. Thus, it is perhaps not surprising to find that there was little difference on the picture-word matching task for participants in the three conditions. In summary, in the present experiment: (a) the nature of the linguistic target, (b) the presentation of multiple cues, and (c) the manipulation of attention appear to have led to similar performances, regardless of how the input was presented. Note that, on top of this, as an experimental control, the type and token frequency of the input was identical in all conditions.

In spite of this, there was a small trend toward better performance in the variation set conditions on both trained and untrained subtests. This non-significant trend is consistent with the idea that formal and functional cues within variation sets are valuable to the discovery of language structure (e.g., Küntay & Slobin, 2002; Onnis et al., 2008; Waterfall et al., 2010). However, any advantages shown over the scrambled condition were very small (about 2-3%) and, in the present case, it furthermore remains unclear whether holding both stems and picture cues constant while varying the suffix cues (lexical variation) works better than holding only the suffix cues constant (morphological variation) in adult L2 construction learning. Any differences between these two conditions were extremely minor and neither one led to consistently better performance than the other across the trained and untrained subtests.

8.3 Learner-generated awareness

Research Question 2: What are the effects of awareness on picture-word matching test scores?

- Hypothesis 2a: Participants who exhibit awareness (on the retrospective written questionnaire) of the relevance of the two features of number and animacy to the artificial language to which they have been exposed will show a higher proportion of correct responses on the picture-word matching task than those who are not aware of the relevance of those features, for both instances encountered during the learning phase and novel instances.
- Hypothesis 2b: Participants who exhibit accurate awareness of the specific form-meaning mappings (in their responses to the written questionnaire) will show a higher proportion of correct responses on the picture-word matching task than those who cannot accurately report those mappings, for both trained and novel exemplars of these constructions.
- Hypothesis 2c: Awareness (accurate report) of specific form-meanings will lead to better performance than awareness of the significance of semantic features alone, measured in terms of the proportion of correct responses to the picture-word matching task.

This study considered two approaches to describing participants' awareness. The first approach, based on awareness of only the relevance of semantic features, showed that those aware of both animacy and number performed significantly better on the outcome measures than those who were unaware of these features. Thus, Hypothesis 2a was supported. The second approach, based on a more detailed examination of awareness of the precise mappings of form and meaning encountered during training, indicated that awareness of relevance alone may not be sufficient for generalizing to previously unseen instances: here, the effects were significant only for participants in the number-high and both-high categories. Thus, Hypothesis 2b was also supported. The better fit of this model to the data further substantiates the coding of awareness based on reporting of form-meaning mappings. Another key observation was that, among participants uniformly aware of the relevance, of only number, or of both number and animacy, those with higher awareness of the precise form-meaning mappings outperformed those with less complete awareness of these mappings on the picture-word matching task (see Figure 7.3). Therefore, Hypothesis 2c was confirmed. Methodologically, these results demonstrate that learner awareness is not inscrutable, if carefully and thoroughly assessed.

These findings agree with those of previous studies on the learning of artificial form-meaning mappings by adult learners. The literature has yielded evidence for more robust learning of bound morphology under conditions of awareness, in the case of locative suffixes (Alanen, 1995), plural, gender, and object markers (DeKeyser, 1995), and imperative forms (de Graaff, 1997). This study demonstrated the contribution of learner-generated awareness by focusing on suffix markers encoding number and animacy.

Relatedly, the study was also broadly consistent with the aim of developing increasingly detailed accounts of L2 awareness (Leow, Johnson, & Záráte-Sandez, 2011; Rebuschat, 2013; Robinson, Mackey, Gass, & Schmidt, 2012; Schmidt, 2012; Williams, 2009). Past research has at times classified learners as either aware or unaware based on concurrent or retrospective reports (Leow, 2000; Leung & Williams, 2011; 2012; Williams, 2005). The present study extended these approaches by applying a fine-grained coding system, which attained high inter-rater reliability, to retrospective data.

These results (see Table 7.7) showed that the number-high group ($n = 9$) and the both-high group ($n = 20$) each performed significantly better than the reference group ($n = 55$) on the untrained (generalization) subtest. Also, the both-high group had a higher coefficient (and a slightly narrower CI) than the number-high group. As a reminder, this model included random effects for item type and participant and was based on participants' yes/no responses to the picture-word matching task. The mean percentage scores of these two groups on the untrained subtest were 72.22% and 75.83%, respectively. These scores indicate a large percentage gain over the unaware group, in which the untrained mean score was 55.78%.

The pattern that emerged across the two models—one for scores on trained items and the other for scores on untrained items—is interesting, as well. For trained items, which participants were exposed to, performance in the both-low (88.89%) and both-mixed (91.66%) groups was significantly better than the reference group (75.15%). However, when these participants had to judge novel exemplars using the same grammatical system, this advantage disappeared. Only the groups who reported accurate mappings performed significantly better on the untrained subtest. Knowledge of the relevance of a semantic feature may not be adequate for generalizable learning to take place.

8.4 Individual differences

The present section discusses the results for the individual difference variables examined in this study: working memory and personality. First, the research questions and hypotheses will be restated. Then, the results for the measures used, and their relationship to L2 construction learning, will be summarized. Lastly, key findings will be examined and discussed within the context of previous investigations. This section sets the stage for discussion of the results for Research Question 5.

8.4.1 The role of complex span and dynamic measures of WM

Research Question 3: Is WMC (as measured by complex span and dynamic measures) positively related to picture-word matching task scores?

- Hypothesis 3a: For all participants in all three conditions generally, the relationships between complex span and dynamic measures of WMC and picture-word matching task scores will be positive.

Hypothesis 3a was mostly supported. The relationship between the complex reading span task and picture-word matching task scores was not significant for the trained subtest (though it was positive, as expected), whereas, for the untrained subtest, a significant correlation was found (r = .27). The complex reading span measure was assumed to involve processing, storing, and retrieving verbal information. This study showed an association between these abilities and participants' judgments of previously unseen instances of artificial L2 constructions morphologically encoding number and animacy on concrete nouns. This result adds to previous research linking span measures to outcomes when learners were exposed to novel L2 input (e.g., Atkins & Baddeley, 1998; Brooks, Kempe, & Sionev, 2006; Martin & Ellis, 2012; Robinson, 2005b) by extending this finding to the picture-word matching task, which involves attention to both form and meaning.

The results are even clearer for the dynamic measure of WMC. The dual 3-back correlated significantly with both the trained (r = .29) and untrained (r = .27) subtests. This is a key finding because few previous studies have investigated associations between measures requiring participants to rapidly encode, match, and respond to stimuli and L2 outcomes (however, see Jackson, in press). In concluding their review, Juffs and Harrington (2011) called for L2 research that seeks to explicate the relationship between WMC and attention, in terms of "processes that involve maintenance of immediate attention and resistance to distraction and that are sensitive at the millisecond level" (p. 160). The dual 3-back may be one way to do this, because it requires constant attention, invokes both facilitation (on target trials) and inhibition (on lure trials), and taps these abilities within a brief time window. One possible interpretation for the results here is that the online pattern recognition ability invoked by the dual 3-back task supports the learning of novel grammatical features under uninstructed conditions (see Dörnyei & Skehan, 2003; Skehan, 2002).

- Hypothesis 3b: In the scrambled condition, the positive association between complex span and dynamic measures of WMC and picture-word matching task outcomes will be stronger than in the variation set conditions.

This dissertation provides some weak empirical support for Hypothesis 3b. Generally, the positive correlations between WM and picture-word matching task scores in the scrambled group were higher than in the variation set groups. Although not strong, these correlations were consistently in the range of r = .3 to .4. On the other hand, the correlations between WM and scores on the outcome measures in the variation set groups were mostly near zero. There were, however, some exceptions to this. In fact, the highest correlations for each of the two WM measures were not found in the scrambled group, but in either the lexical or morphological variation set groups.

Hypothesis 3b assumed that partially repeated sequences of input would reduce cognitive load by "stimulating alignment and comparison of adjacent sentences, without overtaxing working memory capacity" (Goldstein et al., 2010, p. 256). It is important to note that this premise was based on the observation that input variation appears to guide learning in young children, whose WMC is known to be underdeveloped. Here, it was proposed that variation sets might facilitate statistical learning processes in adults by highlighting the internal structure of linguistic forms while also repeating information germane to their meanings. Though small, the consistent effect for WM in the scrambled conditions appears to suggest that, in the absence of partial repetition (or other types of supportive interventions) learning uniformly shows the influence of WM. In contrast, when partial repetition is available, the role of WM may be less pervasive, but learning nonetheless remains susceptible to its influence. This explains the less consistent pattern of results in the variation set conditions. At a more general level, this account implies that, to some extent, WM is selectively facilitative in SLA.

8.4.2 The influence of openness and intellect

Research Question 4: How are openness and intellect related to measures of performance and cognitive abilities used in the experiment?

- Hypothesis 4a: Openness will be positively related to any learning found in the study.

This study found a significant positive relationship between openness and scores on the untrained subtest of learning, but no correlation between openness and trained scores. These findings offer partial confirmation for Hypothesis 4a, but pose a challenge as far as their interpretation is concerned. It should be recalled that including openness in the regression models did not significantly improve their fit. Also, considering the correlational nature of this finding, it is not possible to assert that untrained scores depended on openness, rather than the opposite. For example, Ożańska-Ponikwia and Dewaele (2012) showed self-perceived proficiency to be related to openness, so it could be that participants who (rightly) imagined they had performed well on the harder, untrained subtest subsequently gave themselves higher scores on the openness measure. However, the reliability of the openness measure, which, along with intellect, was most stable of the four ID variables in this study, speaks against this interpretation. Still, it is difficult to grasp the precise role of openness in this study, given the numerous possibilities that may underlie the relationship between openness and cognitive abilities (Chamorro-Premuzic & Furnham, 2005). Perhaps it could be speculated that Hypothesis 4a was confirmed only for untrained scores because openness indicates an ability to reflect on one's consciousness (Norman, Price, & Duff, 2006) and conscious awareness guided performance, especially on this subtest. Yet, without further research, this would be mere conjecture.

- Hypothesis 4b: Intellect will be positively associated with WMC and attention to form.

Mixed support was found for Hypothesis 4b. Measures of WM and attention in this study included the dual 3-back task, the reading span task, and the copy-typing task (during the training phase). Of these three measures, intellect showed a positive relationship with the reading span and copy-typing tasks, which reached significance in the case of the latter. No association was found between the dual 3-back and intellect. The literature reviewed earlier parallels these findings in offering conflicting evidence for a link between WM and intellect. It was noted in Chapter 5 that two studies did support this view, when WM was measured using a 3-back task (DeYoung, Shamosh, Green, Braver, & Gray, 2009) and an operation span task (Kaufman, DeYoung, Gray, Jimenéz, Brown, and Mackintosh, 2010). However, counterevidence was also cited from studies using these same types of measures: other studies found no association between personality and complex span (Unsworth, Miller, Lakey, Young, Meeks, Campbell, & Goodie, 2009) or dynamic (Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012) WM tasks. In light of these sparse and conflicting results, it may be problematic to assume that a small handful of measures, rather than a battery of related, but distinct cognitive tasks, will clearly reveal such connections.

Relating this issue to language performance, the significant, positive relationship between intellect and attention to form (as measured by copy-typing) calls for further investigation. As discussed earlier (Section 8.2), scores on the attention to form measure were consistently high, but it can further be noted that variation in this measure was positively related to intellect. If replicated, this finding could help to establish empirical support for recent claims that certain learner traits may affect noticing (Schmidt, 2012). Although the evidence presented here is not entirely straightforward, the assumption that meaningful relationships may hold between cognitive and affective factors, as these relate to SLA, is grounded in contemporary views of psychology. Adding personality to the agenda in future studies on novel L2 construction learning is one way to extend this perspective in L2 research.

8.5 Multiple predictors of L2 construction learning

Research Question 5: Which of the experimental and observational variables in this study predict scores on the trained and untrained subtests? Do interactions between these variables predict outcomes?

- Hypothesis 5: Input variation, learner awareness, and individual differences will each contribute to L2 construction learning as measured by the picture-word matching task.

In a nutshell, item type, awareness and WM significantly predicted outcomes on the picture-word matching task used in this experiment. Thus, Hypothesis 5 received partial support. As already noted, input variation, as operationalized here, did not influence learning outcomes. Nonetheless, the inclusion of item type and WM accounted for learning better than awareness alone. Building on the results described in Section 8.3, the effects associated with item type and reading span were significant. This means that certain item types and higher WM scores affected picture-word matching task scores on the trained and untrained subtests. Specifically, in the case of the untrained test, looking at the mean scores of all participants on the correct (72.78%) versus the incorrect number (61.48%) and incorrect animacy (40.37%) item types reveals substantial differences. Importantly, these effects were independent of those found for awareness, indicating that L2 learning is influenced by the precise nature of the target construction, as well as by individual differences, at varying levels of awareness.

Specifically, it appears to have been difficult for participants to reject anomalous constructions. This finding should be interpreted cautiously, however, because the number of incorrect item types was small and was only perfectly balanced on the untrained subtest. Yet, on the trained subtest, both incorrect item types led to significantly reduced scores, and the effect was stronger for items showing wrong animacy markers. However, on the untrained subtest, which required generalization to previously unseen word-picture pairs, scores were significantly lower only for incorrect animacy items. These findings are consistent with the results from the written questionnaire, as fewer participants commented on animacy. This pattern of results furthermore resonates with accounts of learning L2 morphology acknowledging the important role of transfer (e.g., Luk & Shirai, 2009). It is perhaps not surprising that these English-dominant participants found it easier to classify erroneous instances when number rather than animacy was miscoded. In the artificial language used in this study, the marking of number at the end of nouns was similar to English, even if the values it took (including dual) were not.

As described in Section 8.3, learner-generated awareness may have triggered the development of semantic categories leading to generalizable knowledge, which also accounted substantially for picture-word matching task outcomes. Based on this, it can be claimed that it is not awareness of the relevance of semantic features, but awareness of the precise constructional mappings involved in expressing those features, that prompts generalization. Comparing the results for both subtests, a contrast can be seen in that, on the trained subtest, learners labeled as number-high, both-low, both-mixed, and both-high significantly outperformed those who did not report awareness of the relevance of either feature, whereas on the untrained test, it was only the number-high and both-high groups who did significantly better than the reference group (see Table 7.7). Learners can and do learn more about L2 constructions than what they experience in the input they receive, but, here, this ability to generalize seems largely comprised of explicit knowledge of form-meaning mappings (Schmidt, 1990, 2001, 2012; Robinson, 1995, 2003; for related accounts, see Bergsleithner, Frota, & Yoshioka, 2013; Robinson & Ellis, 2008; Robinson, Mackey, Gass, & Schmidt, 2012).

Another important point is that the effects of awareness coexist alongside the influence of individual differences. Previous research (e.g., N. Ellis, 2001, 2005; Juffs & Harrington, 2011; Robinson, 2003; Williams, 2012) has suggested that WM is facilitative of L2 learning. How might it have related to outcomes in the present case? The reading span measure gauged participants' ability to process, store, and retrieve verbal information in terms of span size, or the number of letters they could recall after carrying out the task of reading and judging the semantic plausibility of a series of sentences. As such, the processing demands of this WM task are roughly similar to those posed by the picture-word matching task: participants were shown novel constructions based on those they had previously attended to and were required to judge whether they were semantically acceptable. Controlled attention and semantic processing based on verbal information underlie both tasks.

Intriguingly, the role of WM was not quite the same across the two subtests used in this study. It should be noted that the coefficient for reading span was low, and only narrowly significant, in the model based on the trained data. On the contrary, reading span had the largest coefficient in the model based on the untrained data and it was clearly a significant predictor. It might be safe to assume that participants used episodic memory to judge trained instances, thus lessening the burden on WM. The untrained test version, however, was more demanding, as it required participants to judge previously unseen instances based on the same underlying system as correct or incorrect. In that case, directing attention to the relevant aspects of the items and controlling explicit processing based on rule knowledge (see Section 8.6) facilitated better performance.

Theorists have proposed that IDs in SLA are multi-componential, hierarchal, and process-dependent (Dörnyei, 2009a, 2009b; Dörnyei & Skehan, 2003; Robinson, 2001b, 2002c, 2005a, 2007b; Skehan, 2002). This study adds empirical support to studies addressing the contribution of multiple predictors to early L2 learning (e.g., Brooks & Kempe, 2013). It has demonstrated that the constructs of input, awareness, and IDs in themselves are multi-faceted and amenable to various operational definitions. These variables subsume so many distinct, interacting processes that even when constrained to learning a simple artificial language over a brief exposure period, their system dynamics turn out to be remarkably complex. For instance, despite the aforementioned pattern of correlations, regression modeling offered no evidence for an interaction between condition and working memory. This, however, could be due to the size of the lexicon learners encountered. Here, it consisted of only 12 items. Brooks, Kempe, and Sionev (2006) reported significant correlations between WM and novel grammar learning only when the lexicon contained 24 items (not 6 or 12 items). Furthermore, it seems likely that WM effects are sensitive to task and condition (see Section 8.4.1). Thus, the role of such interactions should remain a key component of the IDs research agenda.

8.6 Additional evidence from knowledge source attributions

Research Question 6: What do additional measures of awareness (i.e., source attributions) indicate about participants' learning of the artificial constructions? Specifically, do participants use both explicit and implicit knowledge sources during picture-word matching task performance?

- Hypothesis 6: Participants will report use of both explicit (rule, memory) and implicit (guessing, intuition) knowledge sources for their responses to the picture-word matching task.

Overall, participants reported making use of both implicit and explicit knowledge sources. Consequently, Hypothesis 6 was supported. The general pattern of results leaned toward explicit sources: memory and rule knowledge were used more than guessing or intuition. Crucially, the use of knowledge source attributions appeared to be sensitive to factors such as subtest (trained vs. untrained), item type (correct vs. incorrect), and participant awareness as later reported on the written questionnaire.

Participants reported using explicit knowledge on 78% of trained subtest responses and 69% of untrained subtest responses. Memory was used more for trained items, while rule was used more often for untrained items. Especially in the groups aware of mappings, rule knowledge was applied when incorrect items were judged, either successfully or unsuccessfully. In addition to use, another facet of knowledge attributions this study examined was the accuracy of responses associated with different sources. As might be expected, given the categorical rules comprising the learning object used here, rule attributions accompanied more accurate performance on trained and untrained subtests and on correct and incorrect items, but participants also performed well on trained items when using memory. These explicit knowledge sources showed the highest accuracy overall.

Participants reported implicit knowledge as the source of 22% of their trained subtest responses and 31% of their untrained subtest responses. They were more often right than wrong when they used intuition to respond to correct items (both subtests) and incorrect number items on the trained subtest. Despite this finding, this study appears to support the conventional wisdom that intuition is precarious, particularly as a basis for accurate grammatical performance. Relative to the entire sample, use of the intuition attribution seemed to decline, and intuition-based performance seemed to weaken, for participants aware of constructional mappings in this study. One straightforward way of explaining this finding might be that, as adult learners begin to develop hypotheses about L2 grammar, they gradually relinquish the use of implicit knowledge as a guide to performance.

These findings build on the longstanding assumption that learners use both implicit and explicit knowledge during L2 processing (N. Ellis, 1994, 2005, 2011; R. Ellis, 2009; Rebuschat, 2013). The present study indicated that this is indeed a plausible assumption. It also revealed possible constraints related to how distinct knowledge sources are deployed during the early stages of adult language learning. Specifically, learners reported using certain sources of knowledge depending on whether they had seen test items before or not, and also whether the items were correct or not.

The evidence from the source attributions also validates the coding of awareness based on the written questionnaire. That is, the proportion of rule attributions tended to increase as participants reported in greater detail their awareness of these grammatical constructions. Inversely, use of intuition declined as participants reported more awareness (see Appendix I). The groups classified as number-high and both-high consistently had the highest proportion of rule attributions on both subtests. And these participants also scored highest on the untrained version of the picture-word matching task. This triangulation among knowledge source attributions, picture-word matching task responses, and retrospective written reports contributes to the internal validity of this study.

8.7 Summary of major findings

The main findings of this dissertation study, which concern the roles of input variation, learner awareness, and individual differences, are summarized here:

1. Neither of the variation set conditions lead to a clear advantage over the scrambled condition on the picture-word matching task. A model fit to the data revealed no main effects for condition as a predictor of the proportion of correct responses on either trained or untrained picture-word matching subtests.

2. Using a course-grained coding system (based on awareness of the relevance of semantic features), participants aware of both animacy and number performed significantly better on the picture-word matching task than those who were unaware of both of these features on the trained and untrained subtests.
3. Using a fine-grained coding system (based on awareness of specific form-meaning mappings), the results on the picture-word matching task differed by subtest: participants in the number-high, both-low, both-mixed, and both-high groups performed significantly better than those who were unaware on the trained subtest, while only those in the number-high and both-high groups performed significantly better than those who were unaware on the untrained subtest.
4. Participants with awareness of the precise form-meaning mappings outperformed those with awareness of only the relevance of semantic features on the picture-word matching task.
5. Scores on the complex reading span task and picture-word matching task were positively and significantly correlated (on the untrained subtest).
6. Scores on the dual 3-back task were correlated positively and significantly with both the trained and untrained subtests of the picture-word matching task.
7. Scores on the openness measure were correlated positively and significantly with the untrained picture-word matching subtest.
8. Scores on the intellect measure were correlated positively and significantly with the copy-typing task.

9. Item type, awareness and working memory significantly predicted outcomes on the trained and untrained subtests of the picture-word matching task, based on a regression model including the random effects of items and participants.
10. Participants reported using both explicit (rule, memory) and implicit (guess, intuition) knowledge sources on both picture-word matching task subtests.

8.8 Implications

The implications of this study for theory, research, and practice are briefly described here. First, a question that motivated this study concerns how learners achieve infinite ability to comprehend an L2 based on finite means. Participants in this study learned constructional generalizations based on very brief exposure to a small lexicon presented in a meaningful context. If constructions are defined as pairings of form and meaning, and awareness is required for learning meaning in an L2 (see Ellis, 2005), then it follows logically that awareness is required for learning L2 constructions.

Relevant to this study, it has been argued that when two semantic features are conflated onto one morphological form, then attending to only one meaning component (e.g., distance) is adequate for the learning of the other component (e.g., animacy) (Williams, 2005; see also Leung & Williams, 2011, 2012; Rebuschat et al., 2013; Marsden et al., 2013). In many studies reaching this conclusion the novel target morphemes occurred with familiar lexical items, or they involved a more explicit training regime than in the present study. The findings of this dissertation suggest that, when both the conflated morphological suffixes and the lexical stems are new, attending separately to referent meaning and surface form (as participants did during the attention-focusing trials) is insufficient for generalizable learning of form-meaning mappings. In the absence of a focus on any specific form, acquiring these constructions depended on learner-generated awareness of both animacy and number. Moreover, the role of awareness in the learning process was clearer when a fine-grained system, based on awareness of form-meaning mappings, was applied to the data. Therefore, this study may offer some evidence to rein in the hypothesis that two form-meaning relations can be had at the awareness cost of one.

This study also shows clearly that individual differences play a role in L2 construction learning under uninstructed conditions. In particular, the results for complex span and dynamic WM tasks, which were relatively stable, were positively and significantly associated with trained learning outcomes (although the strength of this association seemed to vary based on the form of input modification employed). As mentioned previously, although these effects may be small, it can be argued that no amount of variation in L2 learning ability is negligible. Considering global trends toward bilingual and multilingual education, these differences are too important to ignore. As such, these results are useful for the development of theoretical approaches, but also may help illustrate the role of learner abilities such as WM in language teacher training, and help inform the design of L2 instructional treatments that afford better access for all learners. Furthermore, this study offered an expanded view of personality in SLA, as well as evidence for its relationship to performance and learning.

As for research implications, the results of this study suggest that, unlike implicit L2 learning, with its elusive and controversial nature, implicit L2 knowledge can be a fruitful construct for investigation. The use of source attributions is a powerful tool for exploring the nature of L2 knowledge, as recently argued by Rebuschat (2013). It can support internal validity if used to triangulate among methods. But this study does also raise some questions. The accessibility of implicit knowledge, the reactivity of source attributions, and the issue of the timing of measures of awareness are each areas in which the SLA field needs to refocus (for recent reviews, see Leow, Johnson, & Záráte-Sandez, 2011; Robinson, Mackey, Gass, & Schmidt, 2012). There have been concerns raised about the use of source attributions in cognitive psychology (Norman & Price, 2010). Presumably, it will not be long before similar questions surface in the SLA literature.

Another key methodological feature of this study is the use of mixed-effect regression models to examine the role of multiple factors in artificial L2 learning. Language studies adopting this approach have recently increased in popularity, owing to the widespread availability of software and related resources for conducting such analyses. Yet, this timing also coincides with greater scrutiny regarding practices surrounding the conventional use of statistics in L2 studies and a theoretical shift in SLA toward increasingly multi-causal accounts of L2 learning. So, the data analysis options leveraged in the present study seem likely to become mainstream in the next few years.

Lastly, concerning practice, although it was originally hoped that this study could serve as a proof-of-concept for how variation sets might be employed as an effective technique for facilitating the learning of form-meaning mappings in natural languages, the results were not so promising in this regard. However, it should be clear by now that variation sets can be defined and implemented in many ways (see Chapter 2 for suggestions). In some way, perhaps this study can inform future effects-of-instruction type studies inspired by the variation sets literature. As a coda to the present attempt at exploiting input variation, it can at least be noted that the effects for the variation set groups were in the expected direction, even if small and non-significant.

8.9 Limitations

Several limitations of this study include its reliability, scope, and statistical methods. First, it was noted earlier that the reliability of trained picture-word matching subtest was low. This may be due to the relatively small number of items used on the test. It may also suggest that the measure does not tap a single underlying construct. However, because similar results were obtained for both the trained and untrained subtests in this study, it is not entirely clear how these issues with the trained subtest affected results. The reliability of the dual 3-back WM task and its potential impact on the correlational findings was also noted (see Section 7.3.3). Second, regarding the scope of the study, using artificial stimuli inevitably limits generalizability, despite the advantages gained by using novel input. Future research should address the issues dealt with here using other construction types. Related to this, while ID measures were selected based on a thorough review of the literature, it was not feasible to examine the influence of prior language learning. Although the artificial language was designed to avoid confounds due to prior knowledge, it would nonetheless be interesting to see how learner variables such as years of language study, number of languages learned, knowledge of specific languages, and foreign language proficiency might impact performance when learning a novel construction. Third, regarding the procedures, it is clearly problematic to rest one's case on multiple statistical comparisons, as were used to address Research Questions 3 and 4. Still, these questions were more exploratory in nature and the same data were subjected to more rigorous testing to answer Research Question 5. A final point is that, in spite of the benefits of mixed-effects modeling (these models test fixed and random effects, do not rely on aggregated mean scores, and enable comparison of relative effects), there are also drawbacks in that models often do not fit other data well. This also limits the generalizability of the study's findings.

8.10 Future directions

More research is needed to relate the diverse range of phenomena considered within experientialist SLA approaches to views of the learner embracing cognitive, affective, and conative differences. With this goal in mind, studies on the role of learner-generated awareness in diverse populations offer one promising direction. As noted already, in addition to WM and personality, incorporating predictors based on prior language learning would be desirable. Another avenue to explore is the range of linguistic outcomes, including reaction times, testable using the picture-word matching task, as well as ways to enhance its reliability. A focused research agenda dealing with these issues might also foster the development of theoretical frameworks, methodological tools, and practical procedures relevant to understanding the statistical and embodied aspects of L2 construction learning. For the time being, researchers are certain to improve on the methods employed in this study to examine these and many other important issues in second language learning.

APPENDIX A

Consent form

AGREEMENT TO PARTICIPATE IN RESEARCH

Project title: Individual differences in second language learning

Investigator: Daniel Jackson, Ph.D. student
Faculty supervisor: Dr. Richard Schmidt

Department of Second Language Studies
University of Hawaiʻi at Manoa
1890 East-West Road
Honolulu, HI 96822

You are invited to take part in a study on individual differences in second language learning.

Purpose of the study: The main purpose of this study is to investigate individual factors in how people learn new languages. This study is being conducted as part of the investigator's degree requirements.

Procedures: If you agree to participate, you will be asked to (a) perform a language learning task, (b) provide information about your background and experience, and (c) take tests of language aptitude and working memory. The study is expected to last approximately 1 hour.

Participation: Your participation is completely voluntary. There is no penalty if you (a) decide not to participate or (b) decide to withdraw from the study. You may withdraw at any time.

Confidentiality: Data and records gathered for this study will be kept confidential to the extent allowed by law. Any report based on this study will protect the anonymity of participants.

Risks and benefits: There are no foreseeable risks or discomforts associated with this study. A potential benefit is that you may learn about language learning research. In addition, the responses you provide may ultimately contribute to our understanding of second language learning and teaching.

APPENDIX A (CONTINUED)

Contact information: If you have any questions about this study, you may ask now or contact Daniel Jackson at danieloj@hawaii.edu or by calling (808) 956-2467. Should you have questions or concerns regarding your rights as a subject in this study, you may contact the Committee on Human Studies, University of Hawaiʻi at uhirb@hawaii.edu or by phone at (808) 956-5007.

If you agree to participate, please sign below and return this page to the lab assistant. Please keep this information for your reference.

Individual differences in second language learning

Consent statement: I have read and understood the above information and have received answers to any questions that I asked. I agree to take part in the study.

Signature: ____ Date: _________

Printed Name: ____

APPENDIX B

Images used during the experiment (training and testing phases)

		Singular	Dual	Plural
Animate	Bird			
	Cat			
	Dog			
	Fish			
	Horse			
	Monkey			

APPENDIX B (CONTINUED)

		Singular	Dual	Plural
Inanimate	Book			
	Box			
	Chair			
	Key			
	Telephone			
	Umbrella			

APPENDIX C

Attention-focusing task performance: Word-copying and meaning trials

Table C1. *Means for Word-copying Trials by Item and Condition*

Item	Word	Lexical	Morpheme	Scrambled	Item totals
1	blonmi	3.93	3.07	3.83	3.61
2	huilku	3.83	3.33	3.93	3.70
3	trem	3.93	3.87	3.73	3.84
4	plin	3.90	3.50	3.93	3.78
5	teunku	3.77	3.53	3.80	3.70
6	doir	3.73	3.67	3.97	3.79
7	fralba	3.33	3.73	3.83	3.63
8	haunzo	3.73	3.87	3.87	3.82
9	crefmi	3.87	4.00	3.90	3.92
10	flis	3.70	3.87	3.90	3.82
11	seorzo	3.80	3.87	3.73	3.80
12	drosba	3.87	3.93	3.90	3.90

APPENDIX C (CONTINUED)

Table C2. *Means for Picture-naming Trials by Item and Condition*

Item	Word	Meaning assigned to stem	Most frequently used label*	Alternate labels	Lexical	Morpheme	Scrambled	Item totals
1	teun	Dog	Dog	N/A	63.33	70.00	63.33	65.55
2	crefku	Monkey	Monkey	Chimpanzee, chimp, ape	83.33	80.00	83.33	82.22
3	haunba	Box	Box	N/A	83.33	83.33	86.67	84.44
4	fralzo	Book	Book	Journal, notebook	86.67	86.67	86.67	86.67
5	huilmi	Bird	Bird	Hummingbird, parakeet	90.00	86.67	90.00	88.89
6	doirmi	Fish	Fish	Goldfish	90.00	83.33	86.67	86.67
7	tremba	Key	Key	N/A	90.00	90.00	90.00	90.00
8	seor	Chair	Chair	N/A	90.00	93.33	86.67	90.00
9	blon	Cat	Cat	N/A	96.67	93.33	90.00	93.33
10	flisku	Horse	Horse	N/A	93.33	93.33	90.00	92.22
11	dros	Phone	Phone	Cell phone	100.00	90.00	93.33	94.44
12	plinzo	Umbrella	Umbrella	N/A	100.00	93.33	83.33	92.22

*Disregarding any determiners or plural inflection added to the word

APPENDIX D

Picture-word matching task test items and statistics: Trained and untrained subtests

Table D1. *Trained Test Items and Item Statistics*

Item	Item Type	Referent shown	Referent meaning	Referent animacy	Referent number	Word shown	Item facility	Item difficulty
1	Correct	fralzo	Book	Inanimate	Plural	fralzo	0.84	0.37
2	Wrong Number	teun	Dog	Animate	Singular	teunmi	0.83	0.37
3	Wrong Animacy	crefmi	Monkey	Animate	Dual	crefba	0.52	0.17
4	Correct	haunba	Box	Inanimate	Dual	haunba	0.91	0.07
5	Correct	huilku	Bird	Animate	Plural	huilku	0.92	0.17
6	Wrong Number	trem	Key	Inanimate	Singular	tremba	0.68	0.53
7	Wrong Number	drosba	Phone	Inanimate	Dual	droszo	0.64	0.57
8	Correct	plin	Umbrella	Inanimate	Singular	plin	0.98	0.07
9	Correct	doirmi	Fish	Animate	Dual	doirmi	0.94	0.03
10	Wrong Number	flisku	Horse	Animate	Plural	flismi	0.79	0.37
11	Correct	blon	Cat	Animate	Singular	blon	1.00	0.00
12	Wrong Number	seorzo	Chair	Inanimate	Plural	seorba	0.68	0.63

APPENDIX D (CONTINUED)

Table D2. *Untrained Test Items and Item Statistics*

Item	Item Type	Referent shown	Referent meaning	Referent animacy	Referent number	Word shown	Item facility	Item difficulty
1	Correct	flismi	Horse	Animate	Dual	flismi	0.72	0.33
2	Correct	doirku	Fish	Animate	Plural	doirku	0.56	0.53
3	Correct	huil	Bird	Animate	Singular	huil	0.87	0.30
4	Wrong Number	cref	Monkey	Animate	Singular	crefmi	0.61	0.67
5	Correct	fral	Book	Inanimate	Singular	fral	0.80	0.43
6	Wrong Animacy	teunmi	Dog	Animate	Dual	teunba	0.39	-0.03
7	Wrong Animacy	blonku	Cat	Animate	Plural	blonzo	0.40	0.27
8	Wrong Number	haun	Box	Inanimate	Singular	haunku	0.76	0.50
9	Correct	seorba	Chair	Inanimate	Dual	seorba	0.70	0.30
10	Wrong Animacy	tremzo	Key	Inanimate	Plural	tremku	0.42	-0.17
11	Wrong Number	plinba	Umbrella	Inanimate	Dual	plinzo	0.48	0.57
12	Correct	droszo	Phone	Inanimate	Plural	droszo	0.72	0.33

APPENDIX E

Written questionnaire: Awareness, personality, and background items

In this part of the study, you respond to a series of surveys. Your responses are important to the research questions in this study and will be kept completely anonymous.

*** 1. Enter your participant number (please ask if you do not know it).**

APPENDIX E (CONTINUED)

Part 1. Post-experiment Questionnaire

Please answer all of the following questions. There are several brief, open-ended questions in this part.

***3. Did you notice anything about the words and pictures you saw during the experiment? What did you notice?**

APPENDIX E (CONTINUED)

*4. Are you aware of any relationships between the word endings in this experiment and the pictures? If so, what are they?

***5. What do you think each of the word endings (-mi, -ba, -ku, -zo) means, specifically?**

For each of the four word endings, please provide as much information as you can. Feel free to guess if you are uncertain about the meaning of any of the endings. Your answers are important to this research and will remain anonymous.

***6. Did you try to learn? What strategies, if any, did you use during the experiment?**

APPENDIX E (CONTINUED)

*7. What motivated you to join this study? How motivated were you during the experiment?

APPENDIX E (CONTINUED)

Part 2. Personality

On the next page, you will see a number of characteristics that may or may not describe you. For example, do you agree that you seldom daydream, compared to most other people? Please check the button that best indicates the extent to which you agree or disagree with each statement listed below. Be as honest as possible, but rely on your initial feeling and do not think too much about each item.

***8. Use the following scale to answer all items.**

	1 - Strongly disagree	2	3 - Neither agree nor disagree	4	5 - Strongly agree
I am quick to understand things.	○	○	○	○	○
I enjoy the beauty of nature.	○	○	○	○	○
I have difficulty understanding abstract ideas.	○	○	○	○	○
I believe in the importance of art.	○	○	○	○	○
I can handle a lot of information.	○	○	○	○	○
I love to reflect on things.	○	○	○	○	○
I like to solve complex problems.	○	○	○	○	○
I get deeply immersed in music.	○	○	○	○	○
I avoid philosophical discussions.	○	○	○	○	○
I do not like poetry.	○	○	○	○	○
I avoid difficult reading material.	○	○	○	○	○
I seldom notice the emotional aspects of paintings and pictures.	○	○	○	○	○
I have a rich vocabulary.	○	○	○	○	○
I need a creative outlet.	○	○	○	○	○
I think quickly.	○	○	○	○	○
I seldom get lost in thought.	○	○	○	○	○
I learn things slowly.	○	○	○	○	○
I seldom daydream.	○	○	○	○	○
I formulate ideas clearly.	○	○	○	○	○
I see beauty in things that others might not notice.	○	○	○	○	○

Part 3. Background and Experience Questionnaire

As part of this study, you are asked to provide information about your background and experience. This information will help the researcher to understand your responses to the other tasks used in the study. Thank you for your assistance in providing this information.

APPENDIX E (CONTINUED)

***9. Gender**

- ☐ Male
- ☐ Female

***10. Are you...**

- ☐ Left-handed
- ☐ Right-handed
- ☐ Use both

***11. What is your date of birth (MM/DD/YYYY)?**

____

***12. What is your major/department at UH Manoa?**

____

***13. What is your current standing?**

- ☐ Undergraduate 1st year
- ☐ Undergraduate 2nd year
- ☐ Undergraduate 3rd year
- ☐ Undergraduate 4th year
- ☐ Other (see below)

If you chose "other", what is your program and year?

____

14. Have you ever had a (check all that apply):

- ☐ Vision problem
- ☐ Hearing impairment
- ☐ Learning disability
- ☐ Language disability

***15. Please enter your first language:**

____

APPENDIX E (CONTINUED)

16. Please enter the names of all languages other than English that you have learned or studied in the boxes below, in order of how well you know them (enter the strongest language other than English as Language 1).

Language 1

Language 2

Language 3

Language 4

Language 5

Language 6

Language 7

***17. May we email you in the future to ask whether you would be interested in participating in other research studies on language learning?**

- Yes
- No

APPENDIX F

Histograms for the picture-word matching task: Trained and untrained subtests

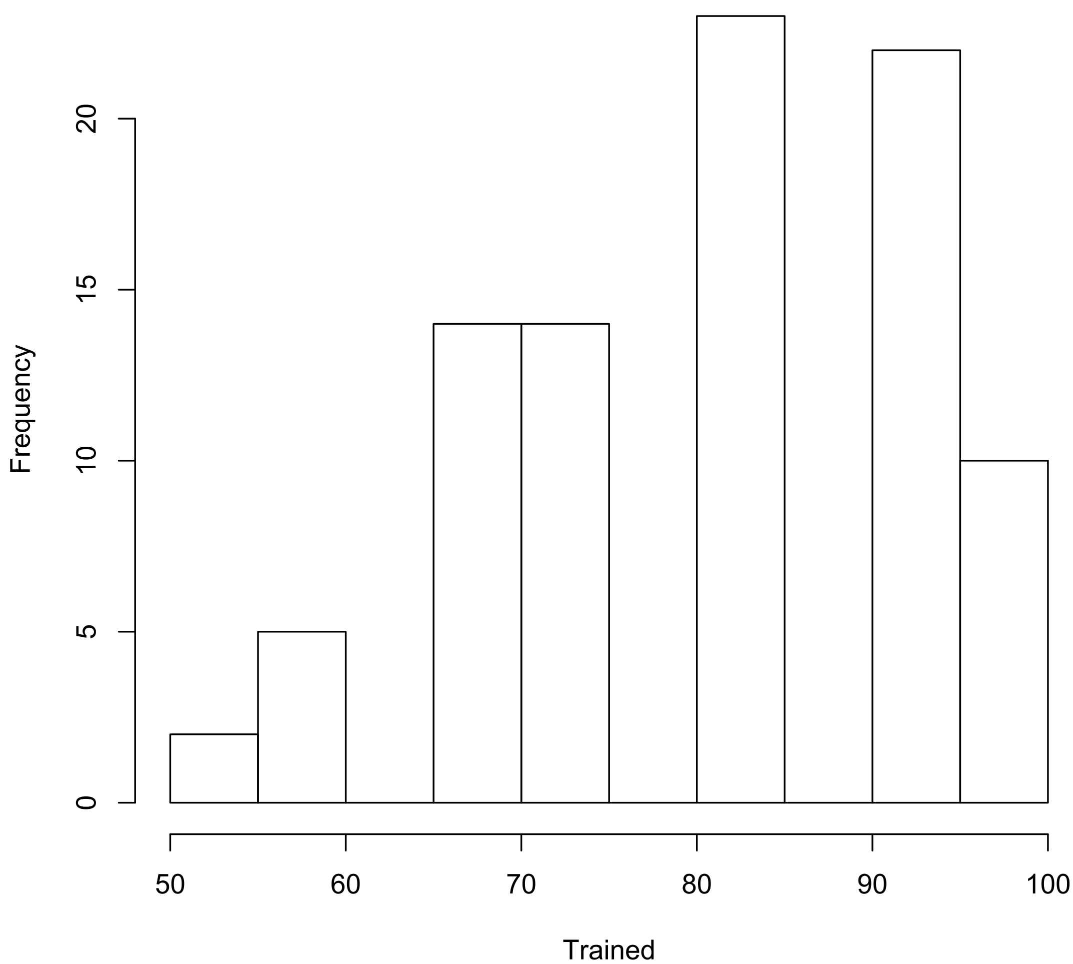

Figure F1. Histogram of trained subtest scores.

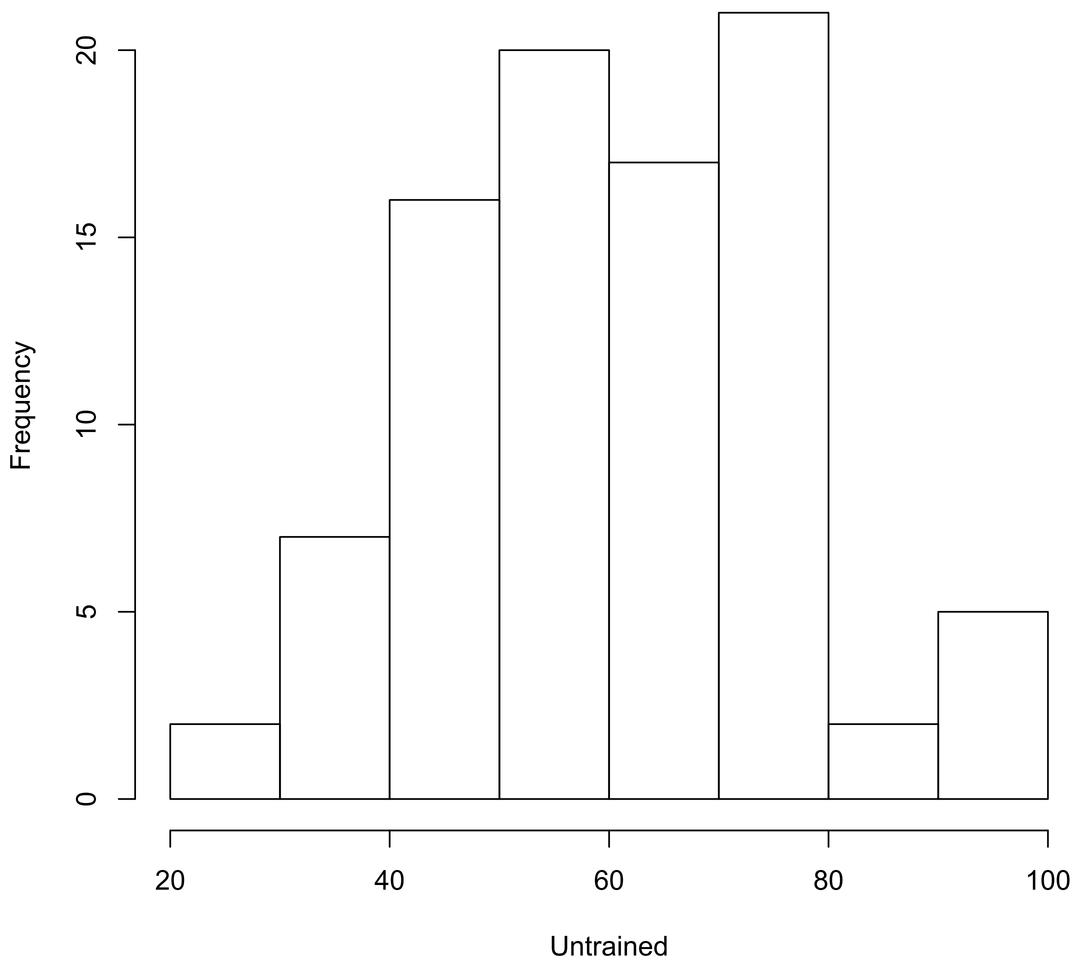

Figure F2. Histogram of untrained subtest scores.

APPENDIX G

Histograms for individual difference measures

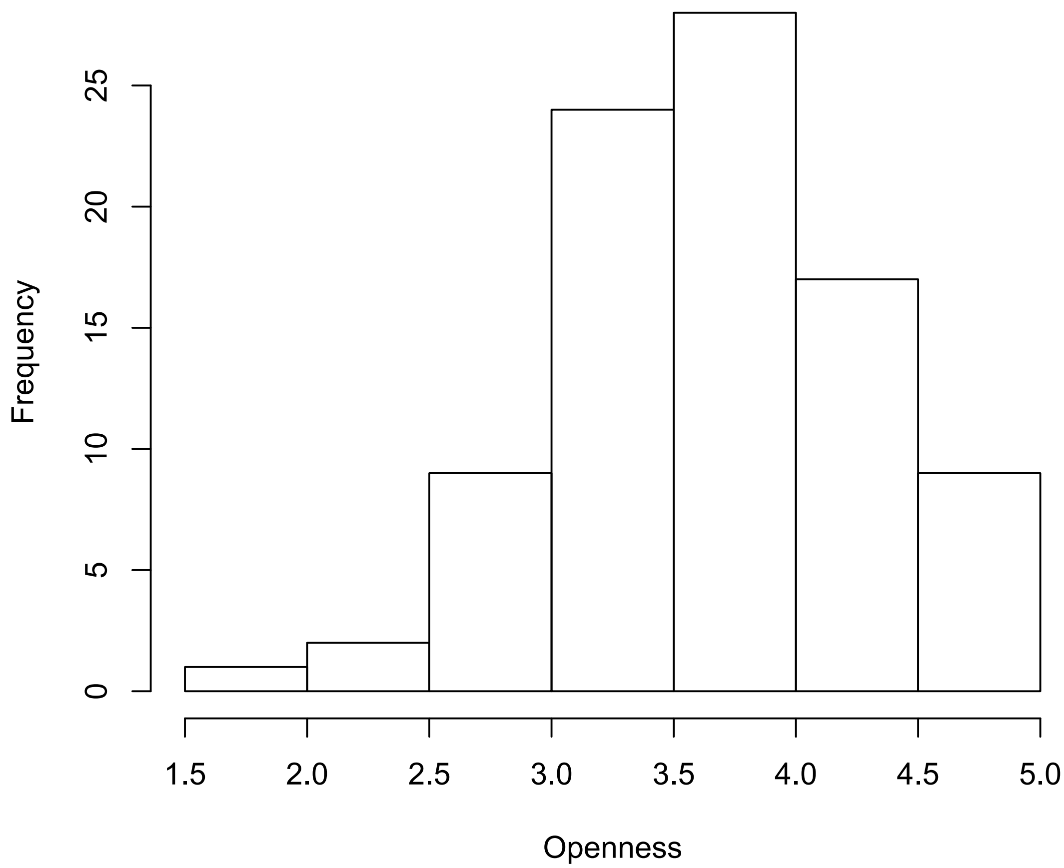

Figure G1. Histogram of openness scores from the BFAS.

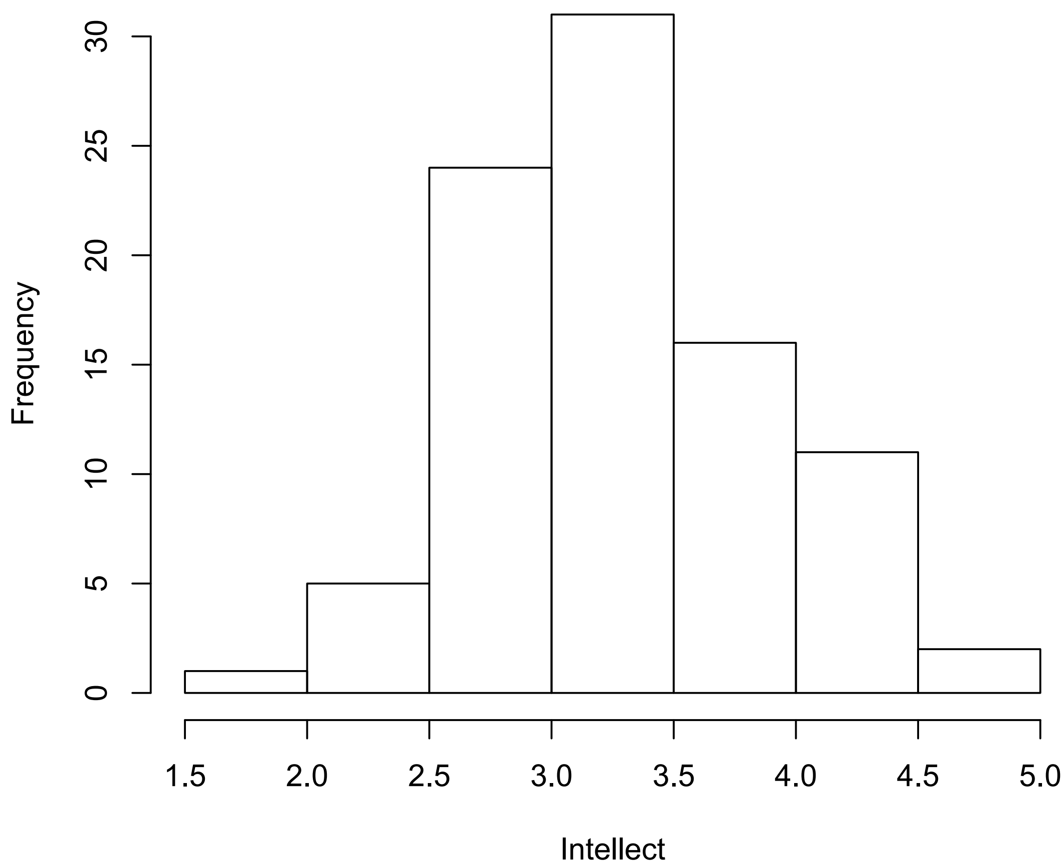

Figure G2. Histogram of intellect scores from the BFAS.

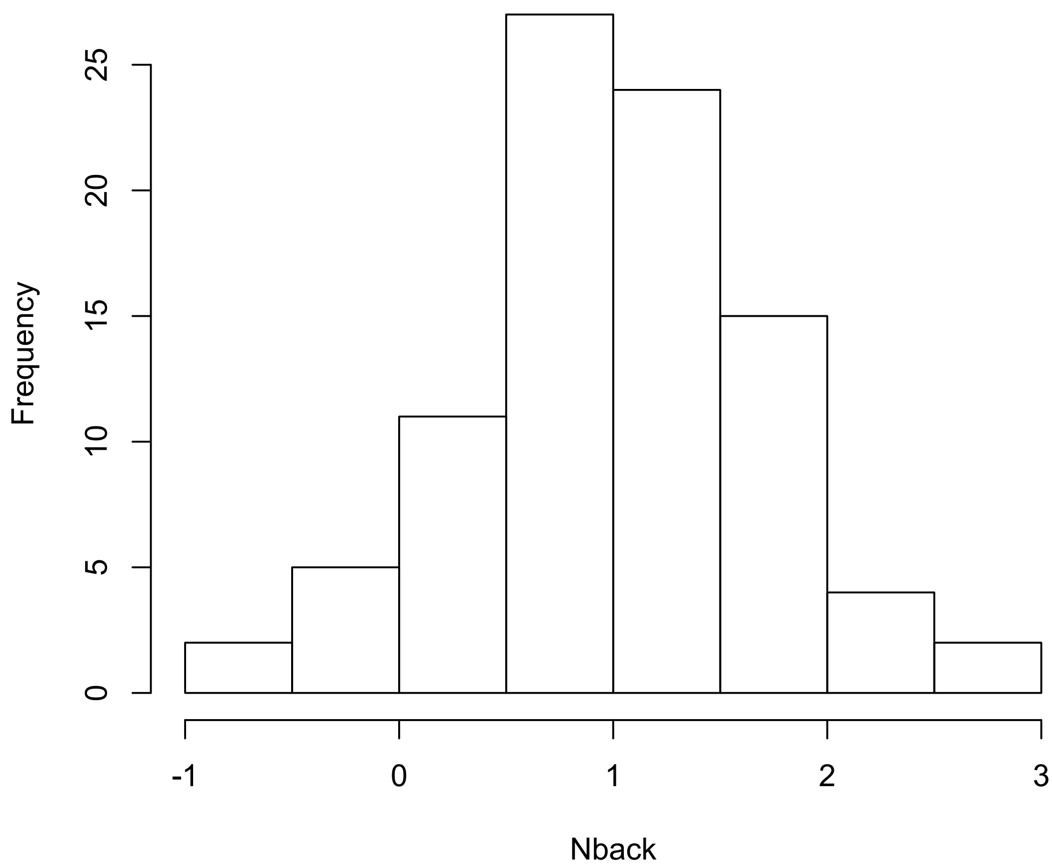

Figure G3. Histogram of dual 3-back scores (*d*-prime scores).

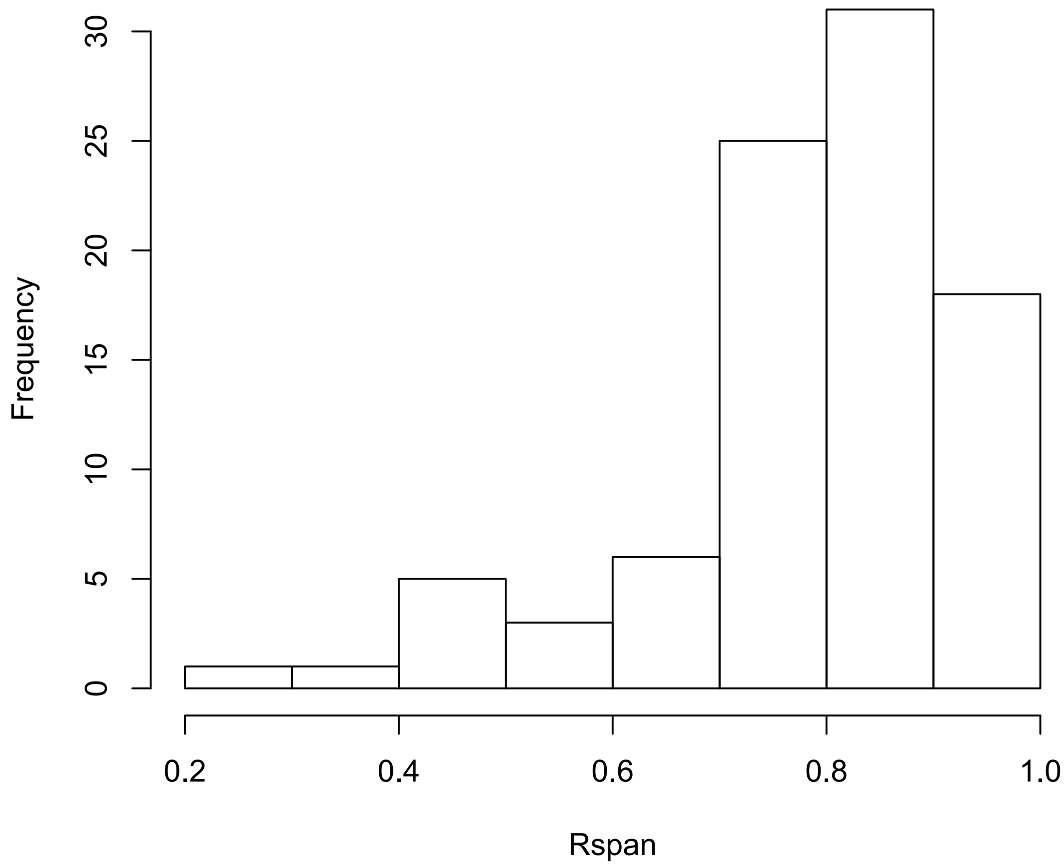

Figure G4. Histogram of reading span scores (partial credit unit scoring).

APPENDIX H

Inter-correlations between dependent and independent measures (PWM = Picture-word matching task)

Table H. *Inter-correlations between Dependent and Independent Measures*

Measure	1	2	3	4	5	6
1. Trained PWM	—					
2. Untrained PWM	.41***	—				
3. Openness	.09	.27**	—			
4. Intellect	.09	.14	.36***	—		
5. Dual 3-back	.29**	.27*	.10	-.04	—	
6. Reading span	.17	.27*	.23	.15	.20*	—

Note. * = $p < .05$; ** = $p < .01$; *** = $p < .001$ (one-tailed)

APPENDIX I

Source attribution data

Source attributions grouped by item type and awareness level, indicating proportion of responses in each attribution category (Prop.), percentage of accurate responses (Acc.), and row totals for all attributions (Tot.).

Table I1. *Source Attribution Data based on the Trained Subtest*

	Source Attributions								
	Guess		Intuition		Memory		Rule		
Item	Prop.	Acc.	Prop.	Acc.	Prop.	Acc.	Prop.	Acc.	Tot.
Correct									
None	11	50	17	100	72	100	0	n/a	18
NL	7	78	17	81	70	97	6	79	312
NH	2	100	7	100	59	97	31	100	54
BL	0	n/a	17	100	67	100	17	100	18
BM	6	100	0	n/a	67	100	28	100	18
BH	3	50	7	100	60	96	30	89	120
Number									
None	20	33	13	100	67	50	0	n/a	15
NL	15	51	17	70	52	56	17	77	260
NH	2	0	4	50	31	100	62	100	45
BL	0	n/a	40	83	40	67	20	67	15
BM	20	100	0	n/a	47	86	33	100	15
BH	3	67	6	67	37	89	54	98	100
Animacy									
None	33	0	33	0	33	0	0	n/a	3
NL	29	53	13	71	50	46	8	25	52
NH	0	n/a	11	0	44	75	44	50	9
BL	0	n/a	0	n/a	100	100	0	n/a	3
BM	0	n/a	0	n/a	67	50	33	0	3
BH	5	0	15	33	35	86	45	56	20
Total	9	59	13	77	57	84	21	87	1080

APPENDIX I (CONTINUED)

Table I2. *Source Attribution Data based on the Untrained Subtest*

	Source Attributions								
	Guess		Intuition		Memory		Rule		
Item	Prop.	Acc.	Prop.	Acc.	Prop.	Acc.	Prop.	Acc.	Tot.
Correct									
None	33	83	17	33	50	44	0	n/a	18
NL	16	58	22	74	38	62	24	75	312
NH	4	100	9	40	13	100	74	95	54
BL	11	50	39	86	33	67	17	100	18
BM	17	33	22	100	28	40	33	83	18
BH	10	50	9	64	16	84	65	90	120
Number									
None	44	50	22	0	33	33	0	n/a	9
NL	13	29	18	43	43	33	26	88	156
NH	4	100	0	n/a	11	100	85	100	27
BL	11	100	11	0	56	20	22	100	9
BM	11	100	11	100	33	33	44	75	9
BH	7	25	7	50	13	63	73	98	60
Animacy									
None	22	0	22	50	56	100	0	n/a	9
NL	16	56	25	36	33	48	26	30	156
NH	4	0	7	50	7	0	81	5	27
BL	0	n/a	33	67	56	0	11	0	9
BM	22	50	22	0	33	33	22	0	9
BH	10	67	13	50	7	50	70	52	60
Total	13	52	18	57	30	53	39	74	1080

REFERENCES

Achard, M., & Niemeier, S. (Eds.). (2004). *Cognitive linguistics, second language acquisition, and foreign language teaching*. Berlin, Germany: Mouton de Gruyter.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227–257.

Ackerman, P. L. (1999). Traits and knowledge as determinants of learning and individual differences: Putting it all together. In P. L. Ackerman, P. C. Kyllonen & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 437–462). Washington, DC: APA.

Ackerman, P. L. (2013). Personality and cognition. In S. Kreitler (Ed.), *Cognition and motivation: Forging an interdisciplinary perspective* (pp. 62–75). New York, NY: Cambridge University Press.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30–60.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219–245.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu, HI: University of Hawai‘i, Second Language Teaching & Curriculum Center.

Albert, Á. (2011). When individual differences come into play: The effect of learner creativity on simple and complex task performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 239–265). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Albert, Á., & Kormos, J. (2004). Creativity and narrative task performance. *Language Learning, 54*, 277–310.

Allin, T. R. (1976). *A grammar of Resígaro.* Buckinghamshire, UK.

Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly, 45*, 235–266.

Andersen, R. W. (1989). The theoretical status of variation in interlanguage development. In S. Gass, C. Madden, D. Preston & L. Selinker (Eds.), *Variation in second language acquisition. Volume II: Psycholinguistic issues* (pp. 46–64). Clevedon, UK: Multilingual Matters.

Andrade, J., & Baddeley, A. (2011). The contribution of phonological short-term memory to artificial grammar learning. *The Quarterly Journal of Experimental Psychology, 64*, 960–974.

Aslin, R. N., & Newport, E. L. (2009). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research disorders* (pp. 15–29). New York, NY: Lawrence Erlbaum.

Atkins, P. W. B., & Baddeley, A. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19*, 537–552.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and patten perception. *Psychological Bulletin, 53*, 452–471.

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. Oxford, UK: Oxford University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford, UK: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation, Vol. 8* (pp. 47–89). New York, NY: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York, NY: Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language Is a complex adaptive system: Position paper. *Language Learning, 59*(Supplement 1), 1–26.

Bell, P. (2009). Le cadeau or la cadeau? The role of aptitude in learner awareness of gender distinctions in French. *The Canadian Modern Language Review / La revue canadienne des langues vivantes, 65*, 615–643.

Bergsleithner, J. M., Frota, S. N., & Yoshioka, J. K. (Eds.). (2013). *Noticing and second language acquistion: Studies in honor of Richard Schmidt*. Honolulu, HI: National Foreign Language Resource Center.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16*, 157–168.

Bialystok, E. (2001). Metalinguistic aspects of bilinguals processing. *Annual Review of Applied Lingustics, 21*, 169–181.

Bialystok, E. (2011). How analysis and control lead to advantages and disadvantages in bilingual processing. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 49–58). Washington, DC: Georgetown University Press.

Bialystok, E., & Hakuta, K. (1994). *In other words: The science and psychology of second-language acquisition*. New York, NY: BasicBooks.

Bickerton, D. (2001). Okay for content words, but what about functional items? *Behavioral and Brain Sciences, 24*, 1104–1105.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41–68). New York, NY: Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3–49.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*, 175–198.

Block, N. (1995). On a confusion about a function of consciousness. *Behaviorial and Brain Sciences, 18*, 227–287.

Block, N. (1996). How can we find the neural correlate of consciousness? *Trends in Neurosciences, 19*, 456–459.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*, 481–548.

Blom, E., Paradis, J., & Sorenson Duncan, T. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular –s in child L2 English. *Language Learning, 62*, 965–994.

Bo, J., Jennett, S., & Seidler, R. D. (2011). Working memory capacity correlates with implict serial reaction time task performance. *Experimental Brain Research, 214*, 73–81.

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York, NY: Routledge.

Brewer, W. F. (1974). There is no convincing evidence for operant or classical conditioning in adult humans. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the Symbolic Processes* (pp. 1–42). Hillsdale, NJ: Erlbaum.

Briggs Myers, I. (1998). *Introduction to type: A guide to understanding your results on the Myers-Briggs Type Indicator* (6th ed.). Palo Alto, CA: Consulting Psychologists Press.

Brodsky, P., Waterfall, H., & Edelman, S. (2007). Characterizing motherese: On the computational structure of child-directed language. In D. McNamara & J. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society Conference* (pp. 833–838). Austin, TX: Cognitive Science Society.

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: the mediating effect of metalinguistic awareness. *Memory & Cognition, 41*, 281–296.

Brooks, P. J., Kempe, V., & Sionev, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27*, 185–209.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge, UK: Cambridge.

Brown, J. D., Robson, G., & Rosenkjar, P. R. (2001). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 361–398). Honolulu, HI: University of Hawaiʻi, Second Language Teaching and Curriculum Center.

Busch, D. (1982). Introversion-Extraversion and the EFL proficiency of Japanese students. *Language Learning, 32*, 109–132.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*, 381–410.

Cadierno, T. (2004). Expressing motion events in a second language: A cognitive typological perspective. In M. Achard & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 13–49). Berlin, Germany: Mouton de Gruyter.

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences: Applications to syntax. *Studies in Second Language Acquisition, 16*, 205–230.

Carrell, P. L., Prince, M. S., & Astika, G. G. (1996). Personality types and language learning in an EFL context. *Language Learning, 46*, 75–99.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11–29). Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.

Carroll, S. E. (1999). Putting 'input' in its proper place. *Second Language Research, 15*, 337–388.

Carroll, S. E. (2006). Salience, awareness, and SLA. In M. G. O'Brien, C. Shea & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 17–24). Somerville, MA: Cascadilla Proceedings Project.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8*, 500–508.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153–193.

Chamorro-Premuzic, T., & Furnham, A. (2005). *Personality and intellectual competence*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chan, R., & Leung, J. (2012). Implicit learning of L2 word stress rules. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 192–197). Austin, TX: Cognitive Science Society.

Chapelle, C. A. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal, 93*, 741–753.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.

Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., & Dienes, Z. (2011). Unconscious structural knowledge of form-meaning connections. *Consciousness and Cognition, 20*, 1751–1760.

Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37*, 739–748.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, UK: Cambridge University Press.

Christiansen, M. H., Onnis, L., & Hockema, S. A. (2009). The secret is in the sound: from unsegmented speech to lexical categories. *Developmental Science, 12*, 388–395.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*, 170–177.

Clark-Carter, D. (2003). Effect size: The missing piece in the jigsaw. *The Psychologist, 16*, 636–638.

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences, 16*, 411–417.

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). Response to Tsuchiya et al.: Considering endogenous and exogenous attention. *Trends in Cognitive Sciences, 16*, 528.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*, 354–373.

Conway, A. R. A., Getz, S. J., Macnamara, B., & Engel de Abreu, P. M. J. (2011). Working memory and intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (Vol. 394–418). Cambridge, UK: Cambridge University Press.

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. (Eds.). (2008). *Variation in working memory*. New York, NY: Oxford University Press.

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

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356–371.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17*, 905–912.

Corbett, G. G. (2000). *Number*. Cambridge, UK: Cambridge.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Lingustics in Language Teaching, 4*, 161–170.

Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., et al. (2002). The cognitive-affective-conative triad. In L. J. Cronbach (Ed.), *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow* (pp. 165–188). Mahwah, NJ: Lawrence Erlbaum.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*, 163–191.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Cowan, N. (2005). *Working memory capacity*. Hove, East Sussex, UK: Psychology Press.

Cowan, N. (2008). Working memory. In N. J. Salkind (Ed.), *Encyclopedia of Educational Psychology, Vol. 2* (pp. 1015–1016). London, UK: Sage.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Craik, F. I. M., & Lockhart, R. S. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology, 44*, 87–112.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford, UK: Oxford University Press.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interaction*. New York, NY: Irvington.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*, 369–382.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*, 279–281.

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

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433.

de Graaff, R. (1997). The *Experanto* experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249–276.

de Zeeuw, M., Verhoeven, L., & Schreuder, R. (2012). Morphological family size effects in young first and second language learners: Evidence of cross-language semantic activation in visual word recognition. *Language Learning, 61*, 68–92.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*, 103–132.

Dehaene, S., & Changeux, J.-P. (2004). Neural mechanisms for access to consciousness. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1145–1157). Cambridge, MA: MIT Press.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences, 10*, 204–211.

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Malden, MA: Blackwell.

DeKeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "Processing instruction: An update". *Language Learning, 52*, 805–823.

DeKeyser, R. M. (1994). How implicit can adult second language learning be? *AILA Review, 11*, 83–96.

DeKeyser, R. M. (1995). Learning second language grammar rules: an experiement with a miniature linguistic system. *Studies in Second Language Acquisition, 17*, 379–410.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195–221.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge, UK: Cambridge University Press.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge, UK: Cambridge University Press.

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 119–138). Malden, MA: Wiley-Blackwell.

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

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8*, 343–350.

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning, 49*, 509–544.

Dewaele, M. (2005). Investigating the psychological and emotional dimensions in instructed language learning: Obstacles and possibilies. *The Modern Language Journal, 89*, 367–380.

DeYoung, C. G. (2011). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 711–737). Cambridge, UK: Cambridge University Press.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73*, 825–858.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880–896.

DeYoung, C. G., Shamosh, N. A., Green, A. E., Braver, T. S., & Gray, J. R. (2009). Intellect as distinct from openness: Differences revealed by fMRI of working memory. *Journal of Personality and Social Psychology, 97*, 883–892.

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquistion* (pp. 337–364). Berlin, Germany: Mouton de Gruyter.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1322–1338.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*, 3–23.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning Memory, and Cognition, 17*, 875–887.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*, 338–351.

Dirven, R. (2001). English phrasal verbs: Theory and didactic application. In M. Pütz, S. Niemeier & R. Dirven (Eds.), *Applied cognitive linguistics II: Language Pedagogy* (pp. 3–27). Berlin, Germany: Mouton de Gruyter.

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

Dörnyei, Z. (2009b). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning, 59*, 230–248.

Dörnyei, Z. (2009a). *The psychology of second language acquisition*. Oxford, UK: Oxford University Press.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Malden, MA: Blackwell.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). Cambridge, UK: Cambridge University Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). Cambridge, UK: Cambridge University Press.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge, UK: Cambridge Univerisity Press.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Malden, MA: Blackwell.

Doughty, C. J. (2004). Effects of instruction on learning a second language: A critique of instructed SLA research. In B. VanPatten, J. Williams, S. Rott & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 181–202). Mahwah, NJ: Lawrence Erlbaum Associates.

Doughty, C. J., Campbell, S. G., Mislevy, M. A., Bunting, M. F., Bowles, A. R., & Koeth, J. (2010). Predicting near-native ability: The factor structure and reliability of Hi-LAB. In M. T. Prior, Y. Watanabe & S.-K. Lee (Eds.), *Selected proceedings of the 2008 Second Language Research Forum* (pp. 10–31). Somerville, MA: Cascadilla Proceedings Project.

Edelman, S. (2008). *Computing the mind: How the mind really works*. New York, NY: Oxford.

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *The Modern Language Journal, 79*, 67–89.

Eitam, B., Hassin, R. R., & Schul, Y. (2008). Nonconscious goal pursuit in novel environments: The case of implicit learning. *Psychological Science, 19*, 261–267.

Eitam, B., Schul, Y., & Hassin, R. R. (2009). Goal relevance and artificial grammar learning. *The Quarterly Journal of Experimental Psychology, 62*, 228–238.

Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5*, 289–318.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91–126.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning, 48*, 631–664.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge, UK: Cambridge University Press.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden, MA: Blackwell.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305–352.

Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics, 27*, 1–24.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*, 164–194.

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). New York, NY and London, UK: Routledge.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35–47). Washington, D.C. : Georgetown University Press.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London, UK: Academic Press.

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

Ellis, N. C., & Robinson, P. (2008). An introduction to cognitive linguistics, second language acquisition, and language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 3–24). New York, NY: Routledge.

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA *Studies in Second Language Acquisition, 19*, 145–171.

Ellis, N. C., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes, 13*, 307–336.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49*, 234–250.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters.

Ellis, R. (Ed.). (1999). *Learning a second language through interaction*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol, UK: Multilingual Matters.

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

Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A resource account of inhibition. *Psychological Science, 6*, 122–125.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York, NY: Cambridge University Press.

E-Prime (Version 2.0 Professional) [Computer software]. Sharpsburg, PA: Psychology Software Tools, Inc.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. . In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257–297). New York, NY: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research, 9*, 147–171.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429–492.

Eysenck, H. J. (1947). *Dimensions of personality: A record of research carried out in collaboration with H.T. Himmelweit [and others]*. London, UK: Routledge.

Eysenck, H. J. (1952). *The scientific study of personality*. London, UK: Routledge.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York, NY: Plenum Press.

Faretta-Stutenberg, M., & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams (2005). In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. Prieto Botana & E. Rhoades (Eds.), *Selected Proceedings for the 2010 Second Language Research Forum* (pp. 18–28). Somerville, MA: Cascadilla Proceedings Project.

Fillmore, C. J. (2003). *Form and meaning in language: Papers on semantic roles*. Stanford, CA: CSLI Publications.

Frankish, K., & Evans, J. S. B. T. (2009). The duality of mind: An historical perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1–28). Oxford, UK: Oxford University Press.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*, 1243–1252.

Furnham, A. (1996). The big five versus the big four: The relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences, 21*, 303–307.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum

Gass, S. M., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition: Introduction to the special issue. *The Modern Language Journal, 82*, 299–307.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439–454.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Lawrence Erlbaum Associates.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 34–54.

Genesee, F., & Hamayan, E. (1980). Individual differences in second language learning. *Applied Psycholinguistics, 1*, 95–110.

Geva, E., & Ryan, E. B. (1993). Linguistic and cognitive correlates of academic skills in first and second languages. *Language Learning, 43*, 5–42.

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 169–197). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaning the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51*, 1–50.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., et al. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences, 14*, 249–258.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178–186.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*, 445–474.

Gor, K. (2010). Introduction. Beyond the obvious: Do second language learners process inflectional morphology? *Language Learning, 60*, 1–20.

Granena, G. (2012). *Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3517774)

Granena, G., & Long, M. (Eds.). (2013). *Senitive periods, language aptitude, and ultimate L2 attainment*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*, 316–322.

Gries, S. T. (2008). Corpus-based methods in analyses of SLA data. In N. C. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 406–431). New York, NY: Routledge.

Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction*. Berlin, Germany: Mouton de Gruyter.

Griffiths, C. (Ed.). (2008). *Lessons from good language learners*. Cambridge, UK: Cambridge University Press.

Guo, X., Zheng, L., Zhu, L., Yang, Z., Chen, C., Zhang, L., et al. (2011). Acquisition of conscious and unconscious knowledge of semantic prosody. *Consciousness and Cognition, 20*, 417–425.

Haatveit, B. C., Sundet, K., Hugdahl, K., Ueland, T., Melle, I., & Andreassen, O. A. (2010). The validity of d prime as a working memory index: Results from the "Bergen n-back task". *Journal of Clinical and Experimental Neuropsychology, 32*, 871–880.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited. *Studies in Second Language Acquisition, 32*, 465–491.

Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2005). The role of working memory in higher-level cognition: Domain-specific versus domain-general perspectives. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 104–121). Cambridge, UK: Cambridge University Press.

Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning? In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365–382). Berlin, Germany: Walter de Gruyter.

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

Harris, Z. S. (1946). From morpheme to utterance. *Language, 22*, 161–183.

Hitch, G. J. (2006). Working memory in children: A cognitive approach. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 112–127). New York, NY: Oxford University Press.

Hsiao, A. T., & Reber, A. S. (1998). The role of attention in implicit sequence learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 471–494). Thousand Oaks: Sage.

Hulstijn, J. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H.-W. Dechert & M. Raupach (Eds.), *Interlingual processes* (pp. 49–73). Tübingen, Germany: Gunter Narr Verlag.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge, UK: Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27*, 129–140.

Izumi, S. (2002). Output, input, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition, 24*, 541–577.

Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly, 34*, 239–278.

Jackson, D. O. (2013). Implicit second language learning and individual differences. In J. M. Bergsleithner, S. N. Frota & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 271–287). Honolulu, HI: University of Hawai‘i, National Foreign Language Resource Center.

Jackson, D. O. (in press). Learner differences in metalinguistic awareness: Exploring the influence of cognitive abilities and language experience. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. Tseng, A. Tuninetti & D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges Between Disciplines*. Somerville, MA: Cascadilla Proceedings Project.

Jacobs, J. (1887). Experiments on "prehension". *Mind, 12*, 75–79.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*, 6829–6833.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*, 394–412.

James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York, NY: Routledge.

Jessner, U. (1999). Metalinguistic awareness in multilinguals: Cognitive aspects of third language learning. *Language Awareness, 8*, 201–209.

Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh, UK: Edinburgh University Press.

Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York, NY: Routledge.

Jiang, Y., & Chun, M. M. (2003). Contexual cueing: Reciprocal influences between attention and implicit learning. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 277–296). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Johnson, K. (2008). *Quantitative methods in linguistics*. Malden, MA: Blackwell.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193–224.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., et al. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*, 462–475.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363*, 623–625.

Jordan, G. (2004). *Theory construction in second language acquisition*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

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

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*, 137–166.

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

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

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 615–622.

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

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*, 321–340.

Keck, C. M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. vii–xxviii). Stanford, CA: CSLI Publications.

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning, 58*, 703–746.

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning, 60*, 127–153.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*, 171–184.

Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F., & Salthouse, T. A. (1999). Models of working memory: Eight questions and some general issues. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 412–441). New York, NY: Cambridge.

Kiyokawa, S., Dienes, Z., Tanaka, D., Yamada, A., & Crowe, L. (2012). Cross cultural differences in unconscious knowledge. *Cognition, 124*, 16–24.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, CO: Roberts and Company.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11*, 16–22.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*(Supplement 1), 1–44.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*, 261–271.

Kövecses, Z. (2001). A cognitive linguistic view of learning idioms in an FLT context. In M. Pütz, S. Niemeier & R. Dirven (Eds.), *Applied cognitive linguistics II: Language Pedagogy* (pp. 87–115). Berlin, Germany: Mouton de Gruyter.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.

Küntay, A., & Slobin, D. I. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. I. Slobin, J. Gerhardt, A. Kyratzis & J. Guo (Eds.), *Social interaction, social context, and language* (pp. 265–286). Mahwah, NJ: Lawrence Erlbaum.

Küntay, A. C., & Slobin, D. I. (2002). Putting interaction back into child language: Examples from Turkish. *Psychology of Language and Communication, 6*, 5–14.

Lado, R. (1964). *Language teaching: A scientific approach*. New York, NY: McGraw-Hill, Inc.

Laine, M. (1999). Meaning analysis of inflected words. *The Quarterly Journal of Experimental Psychology, 52*, 253–259.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago.

Lamme, V. A. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks, 17*, 861–872.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*, 12–17.

Langacker, R. W. (2008b). *Cognitive grammar: A basic introduction*. Oxford, UK: Oxford.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London, UK: Longman.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47*, 467–505.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*, 557–584.

Leow, R. P. (2007). Input in the L2 classroom: An attentional perspective on receptive practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 21–50). New York, NY: Cambridge University Press.

Leow, R. P., & Bowles, M. A. (2005a). Attention and awareness in SLA. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (pp. 179–203). Washington, DC: Georgetown University Press.

Leow, R. P., & Bowles, M. A. (2005b). Reactivity and type of verbal report in SLA research methodology. *Studies in Second Language Acquisition, 27*, 415–440.

Leow, R. P., & Hama, M. (2013). Implicit learning in SLA and the issue of internal validity: A response to Leung and Williams's (2011) "The implicit learning of mappings between forms and contextually derived meanings" *Studies in Second Language Acquisition, 35*, 545–557.

Leow, R. P., Hsieh, H.-C., & Moreno, N. (2008). Attention to form and meaning revisited. *Language Learning, 58*, 665–695.

Leow, R. P., Johnson, E., & Záráte-Sandez, G. (2011). Getting a grip on the slippery construct of awareness: Toward a finer-grained methodological perspective. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 61–72). Washington, D.C. : Georgetown University Press.

Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition, 33*, 33–55.

Leung, J. H. C., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, 634–662.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levy, M., & Stockwell, G. (2006). *CALL Dimensions: Options and issues in computer-assisted language learning*. Mahwah, NJ: Lawrence Erlbaum.

Littlemore, J., & Juchem-Grundmann, C. (2010). Introduction to the interplay between cognitive linguistics and second language learning and teaching. *AILA Review, 23*, 1–6.

Long, M. H. (1983a). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*, 177–193.

Long, M. H. (1983b). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126–141.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York, NY: Academic Press.

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). Cambridge, UK: Cambridge University Press.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition, 16*, 303–323.

Luk, Z. P., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles, and possessive ’s. *Language Learning, 59*, 721–754.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition, 20*, 51–81.

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

MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 45–68). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing WTC in an L2. *The Modern Language Journal, 82*, 545–562.

MacIntyre, P. D., & Gardner, R. C. (1994). The effects of induced anxiety on cognitive processing in computerised vocabulary learning. *Studies in Second Language Acquisition, 16*, 1–17.

MacIntyre, P. D., & Legatto, J. J. (2011). A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics, 32*, 149–171.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*, 557–587.

Mackey, A. (Ed.). (2007). *Conversational interaction in second language acquisition*. Oxford, UK: Oxford University Press.

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

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed second language learning* (pp. 181–209). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Mackey, A., & Polio, C. (Eds.). (2009). *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass*. New York, NY: Routledge.

Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, 704–740.

MacWhinney, B. (1983). Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by adults and children. *Journal of Psycholinguistic Research, 12*, 467–478.

MacWhinney, B. (1997). Implicit and explicit processes: Commentary. *Studies in Second Language Acquisition, 19*, 277–281.

Mānoa Institutional Research. (2010). *Total Headcount Enrollment by Geographic Origin (Permanent Address)*. Retreived from http://manoa.hawaii.edu/ovcaa/mir/?page=fac_enr.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*, 339–366.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Marsden, E., Williams, J., & Liu, X. (2013). Learning Novel Morphology: The role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, 1–36.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*, 379–413.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1083–1100.

McCaulley, M. H. (1990). The Myers-Briggs type indicator: A measure for individuals and groups. *Measurement & Evaluation in Counseling & Development, 22*, 181–196.

McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81–90.

McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the Five-Factor Model of personality. *Journal of Personality, 57*, 17–40.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509–516.

McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective*. New York, NY: Guilford Press.

McLaughlin, B., & Nayak, N. (1989). Processing a new language: Does knowing other languages make a difference? In H.-W. Dechert & M. Raupach (Eds.), *Interlingual processes* (pp. 5–16). Tübingen, Germany: Gunter Narr Verlag.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin, 89*, 575–588.

Michelon, P., & Zacks, J. M. (2003). What is primed in priming from imagery? *Psychological Research, 67*, 71–79.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101*, 343–352.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*, 678–686.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 302–331.

Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (2012). The potential contribution of statistical learning to second language acquisition. In P. Rebuschat & J. N. Williams (Eds.), *Statistical-sequential learning in development* (pp. 13–54). Berlin, Germany: Mouton de Gruyter.

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

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

Miyata, M. (2011). *The effects of frequency, distribution, mode of presentation, and first language on learning an artificial language* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3475720)

Modern Language Association. (2010). MLA Historical Enrollment Data, 2002-2009. Retrieved from http://www.mla.org/about_search_flsurvey.

Moody, R. (1988). Personality preferences and foreign language learning. *The Modern Language Journal, 72*, 389–401.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language, 28*, 360–374.

Morgan-Short, K., Heil, J., Botero-Moriarty, A., & Ebert, S. (2012). Allocation of attention to second language form and meaning. *Studies in Second Language Acquisition, 34*, 659-685.

Naiman, N., Fröhlich, M., Stern, H. H., & Todesco, A. (1978). *The good language learner*. Toronto, Canada: Ontario Institute for Studies in Education.

Nakamura, D. (2012). Input skewedness, consistency, and order of frequent verbs in frequency-driven second language construction learning: A replication and extension of Casenhiser and Goldberg (2005) to adult second language acquisition. *IRAL, 50*, 1–37.

Nation, R., & McLaughlin, B. (1986). Novices and experts: An information processing approach to the "good language learner" problem. *Applied Psycholinguistics, 7*, 41–56.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Norman, E., & Price, M. C. (2010). Measuring "intuition" in the SRT generation task. *Consciousness and Cognition, 19*, 475–477.

Norman, E., Price, M. C., & Duff, S. C. (2006). Fringe consciousness in sequence learning: The influence of individual differences. *Consciousness and Cognition, 15*, 723–760.

Norman, E., Price, M. C., Duff, S. C., & Mentzoni, R. A. (2007). Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition, 16*, 809–837.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 717–761). Malden, MA: Blackwell.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *NeuroImage, 11*, 424–446.

O'Grady, W. (2008). The emergentist program. *Lingua, 118*, 447–464.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51*, 45–100.

Ockey, G. (2011). Self-consciousness and assertiveness as explanatory variables of L2 oral ability: A latent variable approach. *Language Learning, 61*, 968–989.

Onnis, L. (2012). The potential contribution of statistical learning to second language acquisition. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquistion* (pp. 203–235). Berlin, Germany: Mouton de Gruyter.

Onnis, L. (2013). Artificial language learning (ALL). In P. Robinson (Ed.), *Routledge encyclopedia of second language acquistion* (pp. 35–39). New York, NY: Routledge.

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 886–891). Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., Edelman, S., & Waterfall, H. (2011). Local statistical learning under cross-situational uncertainty. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society Conference*. Austin, TX: Cognitive Science Society.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53*, 225–237.

Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition, 109*, 423–430.

Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (2000). Cognitive, affective, personality, and demographic predictors of foreign-language achievement. *The Journal of Educational Research, 94*, 3–15.

Ortega, L. (2009). *Understanding second language acquisition*. London, UK: Hodder Education.

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

Osaka, M., & Osaka, N. (2007). Neural bases of focusing attention in working memory: An fMRI study based on individual differences. In N. Osaka, R. H. Logie & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 99–117). New York, NY: Oxford University Press.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46–59.

Ożańska-Ponikwia, K., & Dewaele, J.-M. (2012). Personality and L2 use: The advantage of being openminded and self-confident in an immigration context. *EUROSLA Yearbook, 12*, 112–134.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Personality Processes and Individual Differences, 76*, 972–987.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Park, D. C., & Payer, D. (2006). Working memory in children: A cognitive approach. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 128–142). New York, NY: Oxford University Press.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–225). London, UK: Longman.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233–238.

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences, 25*, 297–388.

Peters, A. M. (1983). *The units of language acquisition*. Cambridge, UK: Cambridge University Press.

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proceedings of the National Academy of Sciences, 90*, 878–882.

Philp, J. (2003). Constraints on “noticing the gap”: Nonnative speakers’ noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition, 25*, 99–126.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning, 44*, 493–527.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly, 21*, 737–758.

Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning, 61*, 325–366.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York, NY and Evanston, IL: Harper & Row.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*, 227–244.

Pretz, J. E., & Sternberg, R. J. (2005). Unifying the field: Cognition and intelligence. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 306–318). Cambridge, UK: Cambridge University Press.

Pretz, J. E., Totz, K. S., & Kaufman, S. B. (2010). The effects of mood, cognitive style, and cognitive ability on implicit learning. *Learning and Individual Differences, 20*(3), 215–219.

PsyScope X (Version B 57) [Computer software]. Trieste, Italy: SISSA Language, Cognition and Development Lab.

Pütz, M., Niemeier, S., & Dirven, R. (Eds.). (2001). *Applied cognitive linguistics II: Language pedagogy*. Berlin, Germany: Mouton de Gruyter.

R: A language and environment for statistical computing (Version 3.0.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Ramírez-Esparza, N., Harris, K., Hellermann, J., Richard, C., Kuhl, P. K., & Reder, S. (2012). Socio-interactive practices and personality in adult learners of English with little formal education. *Language Learning, 62*, 541–570.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Learning and Verbal Behavior, 6*, 855–863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219–235.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford, UK: Oxford University Press.

Reber, A. S. (2011). An epitaph for grammar: An abridged history. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 23–34). Washington, DC: Georgetown University Press.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in syntactic grammar learning: A functionalist interpretation. *Cognition, 6*, 189–221.

Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227–247). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888–896.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*, 595–626.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2013). Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness. In J. M. Bergsleithner, S. N. Frota & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 249–269). Honolulu, HI: University of Hawai‘i, National Foreign Language Resource Center.

Rebuschat, P., & Williams, J. N. (2012b). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 829–856.

Rebuschat, P., & Williams, J. N. (Eds.). (2012a). *Statistical learning and language acquisition*. Berlin, Germany: Walter de Gruyter.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning Memory, and Cognition, 20*, 585–594.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning, 62*, 93–132.

Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working memory. *WIREs Cognitive Science, 1*, 573–585.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*, 573–598.

Robinson, P. (1994). *Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9506230)

Robinson, P. (1995a). Aptitude, awareness, and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 303–357). Honolulu, HI: University of Hawai'i Press.

Robinson, P. (1995b). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*, 283–331.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition, 18*, 27–67.

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

Robinson, P. (1997b). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–247.

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

Robinson, P. (2002b). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–266). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Robinson, P. (2002c). Learning conditions, aptitude complexes and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113–133). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

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

Robinson, P. (2005a). Aptitude and second language acquisition. *Annual Review of Applied Lingustics, 25*, 46–73.

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

Robinson, P. (2007a). Attention and awareness. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 6, pp. 133–142): Springer.

Robinson, P. (2007b). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and psychology* (pp. 256–286). Cambridge, UK: Cambridge University Press.

Robinson, P. (2007c). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL, 45*, 193–213.

Robinson, P. (2010). Implicit artificial grammar and incidental natural second language learning: How comparable are they? *Language Learning, 60*(Supplement 2), 245–263.

Robinson, P. (2013). Aptitude in second language acquisition. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–5). Oxford, UK: Wiley-Blackwell.

Robinson, P. (Ed.). (2001a). *Cognition and second language instruction*. Cambridge, UK: Cambridge University Press.

Robinson, P. (Ed.). (2002a). *Individual differences and instructed language acquisition*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Robinson, P., & Ellis, N. C. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. New York, NY and London, UK: Routledge.

Robinson, P., & Ha, M. A. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition, 15*, 413–438.

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247–267). New York, NY: Routledge.

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

Roehr, K., & Gánem-Gutiérrez, G. A. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness, 18*, 165–181.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Rubin, J. (1975). What the “good language learner” can teach us. *TESOL Quarterly, 9*, 41–51.

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of experimental psychology: General, 136*, 485–501.

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL, 46*, 113–136.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 229–248). Oxford, UK: Oxford University Press.

Sagarra, N., & Abbuhl, R. (2013). Optimizing the noticing of recasts via computer-delivered feedback: Evidence that oral input enhancement and working memory help second language learning. *The Modern Language Journal, 97*, 196–216.

Sanz, C., & Leow, R. P. (Eds.). (2011). *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism*. Washington, D.C.: Georgetown University Press.

Sato, C. J. (1986). Conversation and interlanguage development: Rethinking the connection. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 5–22). Rowley, MA: Newbury House.

Sauro, S., & Smith, B. (2010). Investigating L2 performance in text chat. *Applied Linguistics, 31*, 554–577.

Saussure, F. (1972). *Course in general linguistics*. Chicago, IL: Open Court.

Schachter, J. (1998). Recent research in language learning studies: Promises and problems. *Language Learning, 48*, 557–583.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning Memory, and Cognition, 13*, 501–518.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137–174). Rowley, MA: Newbury House.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Lingustics, 13*, 206–226.

Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.

Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London, UK: Academic Press.

Schmidt, R. (1995b). Conscious in foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, HI: University of Hawaiʻi Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York, NY: Cambridge University Press.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (pp. 27–50). Boston, MA: De Gruyter.

Schmidt, R. (Ed.). (1995a). *Attention and awareness in foreign language learning*. Honolulu, HI: University of Hawai'i Press.

Schmidt, R. W., & Frota, S. N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Rowley, MA: Newbury House.

Seliger, H. W. (1977). Does practice make perfect? A study of interaction patterns and L2 competence. *Language Learning, 27*, 263–278.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology, 45*, 21–50.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shanks, D. R. (2003). Attention and awareness in "implicit" sequence learning. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 11–42). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London, UK: Sage.

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology, 61*, 273–301.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367–447.

Shin, H. B., & Kominski, R. A. (2010). *Language Use in the United States: 2007, American Community Survey Reports, ACS-12*. Washington, DC.: U.S. Census Bureau.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Slabakova, R. (2009). How is inflectional morphology learned? *EUROSLA Yearbook, 9*, 56–75.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York, NY: Holt, Rinehart & Winston.

Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology, 16*, 53–81.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist, 27*, 5–32.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3–37). New York, NY: Cambridge.

Sonbul, S., & Schmitt, N. (2012). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*, 121–159.

Soubelet, A., & Salthouse, T. A. (2011). Personality–cognition relations across adulthood. *Developmental Psychology, 47*, 303–310.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics, 25*, 293–321

Stansfield, C., & Winke, P. (2008). Testing aptitude for second language learning. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education, Second Edition, Volume 7* (pp. 81–94). New York, NY: Springer.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences, 53*, 44–49.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidelhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson* (pp. 125–144). Oxford, UK: Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.

Tagarelli, K. M., Borges-Mota, M., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society Conference.* (pp. 2061–2066). Austin, TX: Cognitive Science Society.

Taguchi, N. (2008). The effect of working memory, semantic access, and listening abilities on the comprehension of conversational implicatures in L2 English. *Pragmatics & Cognition, 16*, 517–539.

Talmy, L. (2000). *Toward a cognitive semantics: Typology and processes in concept structuring* (Vol. II). Cambridge, MA: MIT Press.

Thorndike, E. L., & Rock, R. T. (1934). Learning without awareness of what is being learned or intent to learn it. *Journal of Experimental Psychology, 17*, 1–19.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition, 16*, 183–203.

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171–195). Oxford, UK: Oxford University Press.

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness: The quest for a theoretical foundation. *Studies in Second Language Acquisition, 33*, 497–528.

Tsuchiya, N., Block, N., & Koch, C. (2012). Top-down attention and consciousness: comment on Cohen et al. *Trends in Cognitive Sciences, 16*, 527−528.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1–25.

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition, 31*, 1060–1071.

Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*. New York, NY: Routledge.

Tyler, A., & Evans, V. (2004). Applying cognitive linguistics to pedagogical grammar: The case of *over*. In M. Achard & S. Niemeier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 257–280). Berlin, Germany: Mouton de Gruyter.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second langauge acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (pp. 141–178). Washington, DC: Georgetown University Press.

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

Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., et al. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences, 30*, 194–200.

van Heuven, W. J. B., & Dijkstra, T. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*, 458–483.

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287–301.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

VanPatten, B. (2002a). Processing instruction: An update. *Language Learning, 52*, 755–803.

VanPatten, B. (2002b). Processing the content of input-processing and processing instruction research: A response to DeKeyser, Salaberry, Robinson, and Harrington. *Language Learning, 52*, 825–831.

VanPatten, B. (Ed.). (2004). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum Associates.

Varonis, E. M., & Gass, S. (1985). Non-native/Non-native conversations: A model for negotiation of meaning. *Applied Linguistics, 6*, 71–90.

Vatz, K., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Aptitude-treatement interaction studies in second language acquisition: Findings and methodology. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins.

Verhoeven, L., & Vermeer, A. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics, 23*, 361–374.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 255–274.

Wagner-Gough, J., & Hatch, E. (1975). The importance of input data in second language acquisition studies. *Language Learning, 25*, 297–308.

Watanabe, Y., & Bergsleithner, J. M. (2006). A systematic research synthesis of L2 WM measurements. In Z. Madden-Wood & K. Ueki (Eds.), *Selected Papers from the Tenth College-wide Conference for Students in Languages, Linguistics, and Literature* (pp. 47–60). Honolulu, HI: College of Languages, Linguistics, and Literature, University of Hawai‘i.

Waterfall, H. R. (2006). *A little change is a good thing: Feature theory, language acquisition, and variation sets* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3219602)

Waterfall, H. R., Sandbank, B., Onnis, L., & Edelman, S. (2010). An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language, 37*, 671–703.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*, 791–772.

Weissheimer, J., & Mota, M. B. (2009). Individual differences in working memory capacity and the development of L2 speech production. *Issues in Applied Linguistics, 17*, 93–112.

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics, 22*, 1–22.

Williams, J. N. (2004). Implicit learning of form-meaning connections. In B. VanPatten, J. Williams, S. Rott & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203–218). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*, 269–304.

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–353). Bingley, UK: Emerald Group Publishing Limited.

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). New York, NY: Routledge.

Williams, J. N. (2013). Attention, awareness, and noticing in language processing and learning. In J. M. Bergsleithner, S. N. Frota & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 39–57). Honolulu, HI: University of Hawai‘i, National Foreign Language Resource Center.

Williams, J. N., & Lovatt, P. J. (2003). Phonological Memory and Rule Learning. *Language Learning, 53*, 167–121.

Winter, B., & Reber, A. (1994). Implicit learning and the acquisition of natural languages. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 115–145). San Diego, CA: Academic Press.

Wolf, M. B., & Ackerman, P. L. (2005). Extraversion and intelligence: A meta-analytic investigation. *Personality and Individual Differences, 39*, 531–542.

Wong Fillmore, L. (1979). Individual differences in second language acquisition. In C. J. Fillmore, D. Kempler & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 203–228). New York, NY: Academic Press.

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition, 23*, 345–368.

Wood Bowden, H., Sanz, C., & Stafford, C. A. (2007). Individual differences: Age, sex, working memory, and prior knowledge. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (pp. 105–140). Washington, DC: Georgetown University Press.

Woolhouse, L. S., & Bayne, R. (2000). Personality and the use of intuition: Individual differences in strategy and performance on an implicit learning task. *European Journal of Personality, 14*, 157–169.

Xie, Q., Gao, X., & King, R. B. (2013). Thinking styles in implicit and explicit learning. *Learning and Individual Differences, 23*, 267–271.

Yang, L. R., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keck second language learning project. *Studies in Second Language Acquisition, 19*, 173–193.

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning, 44*, 189–219.

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal, 86*, 54–66.

Yim, D., & Rudoy, J. (2013). Implicit statistical learning and language skills in bilingual children. *Journal of Speech, Language, and Hearing Research, 56*, 310–322.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420.

Ziori, E., & Dienes, Z. (2006). Subjective measures of unconscious knowledge of concepts. *Mind & Society, 5*, 105–122.