Declarative and procedural memory systems and linguistic difficulty in instructed second language acquisition of English

Gervazio Tchesa

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Applied Language Studies and Linguistics, the University of Auckland, 2020

ABSTRACT

In contrast to first language (L1) acquisition and child second language (L2) learning which almost always lead to native-level proficiency in the language, L2 acquisition is highly variable in the rate of learning and the ultimate level of achievement. Very few learners achieve nativelike proficiency in L2 comprehension or production. To explain the modulating factors in L2 acquisition and the nature of L2 knowledge systems, current approaches that view language as inextricably intertwined with cognition have called for the investigation of the role of the two most important long-term and domain-general memory systems, the declarative and procedural memory systems, and of how these two memory systems are influenced by external factors such as exposure condition and the complexity of learning stimuli (e.g., DeKeyser, 2016; Hamrick, Lum, & Ullman, 2018; Housen & Simoens, 2016; M. Paradis, 2009; A. S. Reber, Walkenfeld, & Hernstadt, 1991). A great deal of L2 acquisition research has examined the role of learning conditions (e.g., Norris & Ortega, 2000; Spada and Tomita, 2010) and language aptitude (e.g., Carpenter, 2008; Dornyei, 2005; 2006; Grey, Williams, & Rebuschat, 2015; 2017; 2019; Li, 2018; Yalçın & Spada, 2016), largely ignoring the three-way interaction of the memory systems, learning conditions and linguistic complexity, which offers one promising way to characterise L2 acquisition.

Adopting an ex post facto design, this study investigated the role of the declarative and procedural memory systems in the acquisition of the knowledge of 14 grammatical structures of L2 English in a classroom environment in Malawi where grammar rules are explicitly taught. The goal was to present the overview of the educational effects on L2 English in Malawi. Quantitative data were collected from 103 L2 English learners at primary school, secondary school and university levels. In the absence of clear ways of distinguishing between simple and complex grammatical structures, a criterion based on Pienemann’s processability theory was proposed to evaluate the linguistic complexity of grammatical structures targeted in the present study. Declarative memory measures were the Continuous Visual Memory Test, the Llama-B test, the DAT Verbal Reasoning Test and the Three-Term Contingency Learning Task. Procedural memory was assessed using the Serial Reaction Times and Llama-D tests. The Timed Grammaticality Judgment Test and the Elicited Imitation Test assessed learners’ procedural language knowledge whereas the Untimed Grammaticality Judgment Test assessed learners’ declarative language knowledge. All the tests were administered using a computer except for the Untimed Grammaticality Judgment Test and the DAT Verbal Reasoning Test which were paper-and-pencil-delivered. The evidence of test validity was explored using correlational, exploratory and confirmatory factor analyses.

To simultaneously investigate various interrelated dependence relationships among variables, data were analysed using a series of structural equation modelling analyses and one-way between-group analyses of variance. The principal findings of the study were that (1) procedural memory played no role in the learning of both simple and complex grammatical structures; (2) linguistic complexity and exposure type appeared to modulate only the declarative learning processes; and (3) no group differences were found in the procedural memory system. These results suggest that the declarative/procedural learning and declarative/procedural memory distinction may indeed be a one-to-one relationship, without the conscious, declarative learning processes interfacing with the unconscious, procedural processes. They also suggest that procedural processes display tighter distributions in a population when compared with declarative systems. Methodological implications and practical suggestions for pedagogy and future research are identified as well.

ACKNOWLEDGEMENTS

I am deeply grateful to many people who provided support, inspiration and encouragement during the completion of this thesis. In particular, I would like to express my most sincere thanks to my supervisors, Dr Rosemary Erlam and Dr Yalu Wen. To Dr Erlam, I greatly appreciate for your dedication to supervision and for your careful attention to detail and argument, which pushed further my understanding of the L2 acquisition literature, methods and my own study findings. To Dr Wen, I am sincerely thankful for your guidance and advice on the various statistical analyses I conducted in this research. I also extend my big thank you to Dr Natsuko Shintani and Dr Shaofeng Li who not only supervised me at the start of this study but continued to act as advisors and offered valuable comments on this study's findings and discussion. My sincere thanks also go to Prof. Lawrence Zhang who also supervised me at the very beginning of this study. How privileged I was to have all of you!

My special thank you go to my generous funding source, the University of Auckland, for the award of the University of Auckland Doctoral Scholarship. Without this financial support, this study would not have been possible. I will forever remain grateful for this. I would also like to extend my thanks to Dr S. B. Kaufman for sharing with me his SRT test, 3-Term test and example DAT-V test items, and to Dr K. Morgan-Short for sharing her version of the CVMT and the CVMT sample analysis documents.

Thank you to all my fellow doctoral students and friends at the University of Auckland, and very specially, Joohyun Park, Lily Wang, Orlando Chaves, Maria Guapacha Chamorro, Imdad Ullah and many others in doctoral hub N452, at Epsom, for various kinds of support over the four years of my study. And to John Finlay, at the Faculty of Education for always making sure that I had a conducive reading space at the faculty despite that I had transferred to the Faculty of Arts. I am also grateful to my participants at Chikamveka Primary School, Mulunguzi Secondary School and Chancellor College in Zomba District for their time to participate in this study and to the headteachers of the schools for not only allowing me to conduct this research in their respective schools but also providing the necessary support throughout the data collection process. The data collection was also made possible by the tremendous support from my two research assistants. I am further grateful to Hilary van Uden (of Hone) for proofreading

some parts of this thesis, effectively improving tremendously the clarity of my writing. To my wife, Joyce, a giant thank you for the love, support, encouragement and understanding during the stressful days and the long absences.

DEDICATION

This thesis is dedicated to the University of Auckland

TABLE OF CONTENTS

Abstract ii

Acknowledgements iv

Dedication vi

Table of Contents vii

Chapter 1. Introduction 1

- 1.1 Overview 1
- 1.2 The Purpose of the Present Study 2
- 1.3 English in Malawi 4

Chapter 2. Literature Review 11

- 2.1 Nature of language knowledge 14
  - 2.1.1 Implicit and explicit language knowledge. 14
  - 2.1.2 The interface hypothesis. 16
  - 2.1.3 Controlled and automatic processing 18
  - 2.1.4 Measurement of language knowledge 19
- 2.2 Declarative and Procedural (DP) Memory Systems 24
  - 2.2.1 Declarative Memory. 25
  - 2.2.2 Procedural memory 26
  - 2.2.3 Declarative and Procedural (DP) Models and L2 Acquisition. 29
  - 2.2.4 Measures of memory systems. 35
- 2.3 Grammar Rule Complexity 43
  - 2.3.1 The Psycholinguistic Perspectives. 44
  - 2.3.2 Pedagogical Perspective. 44
  - 2.3.3 A Linguistic Perspective 45
- 2.4 Explicit and Implicit Learning 48
- 2.5 Empirical Studies: Learning Domain in L2 Acquisition 53
  - 2.5.1 Robinson (1996; 1997). 53
  - 2.5.2 Norris and Ortega (2000). 55
  - 2.5.3 Spada and Tomita (2010). 55
  - 2.5.4 DeKeyser (1995). 56
  - 2.5.5 Morgan-Short et al. (2010; 2012). 57
  - 2.5.6 Aptitude and working memory research. 58
- 2.6 Empirical Studies: Memory Systems and L2 Acquisition 60
  - 2.6.1 Memory, exposure and language knowledge. 61
  - 2.6.2 Memory, linguistic difficulty and exposure conditions. 69

2.7 Summary and rationale 80
2.8 Research Questions and Hypotheses 84
Chapter 3. Methods 88
3.1 Research Design 88
3.2 Participants 88
3.3 Operationalisation 89
3.4 Target Structures 90
3.4.1 Processability theory (PT) 91
3.4.2 Target Functors 95
3.4.3 The Syntactic Structures 103
3.4.4 Difficulty, complexity and learner groups 115
3.4.5 Exposure conditions 116
3.5 Materials and procedures 122
3.5.1 Memory ability measures 122
3.5.2 Procedural memory ability measures 123
3.5.3 Declarative memory ability measures 127
3.5.4 Grammar knowledge measures 135
3.5.5 Declarative knowledge test 135
3.5.6 Procedural knowledge tests 136
3.5.7 Background Questionnaire 140
3.5.8 General procedure 140
3.5.9 Piloting of test materials 141
3.6 Data analysis 147
3.6.1 Scoring of Memory Ability Measures 147
3.6.2 Scoring of language knowledge Tests 157
3.6.3 Statistical Analyses 160
Chapter 4. Results 166
4.1 Descriptive Statistics 166
4.1.1 Memory Ability Measures 166
4.1.2 Declarative Memory Ability Measures. 166
4.1.3 Procedural memory measures 170
4.1.4 Language Knowledge Measures 172
4.2 Validation of Test Instruments 177
4.2.1 Exploration of Memory Ability Test Instruments 178
4.2.2 Exploratory Factor Analysis for memory measures 179

4.2.3 Confirmatory factor analysis for memory ability measures 181
4.2.4 Validating language knowledge test instruments 187
4.3 Present study findings 191
4.3.1 Memory and language knowledge type 191
4.3.2 Memory, Linguistic Difficulty and Language Knowledge 205
4.3.3 Exposure Length, Memory and Language Knowledge 221
Chapter 5. Discussion 228
5.1 Memory, exposure and language knowledge type 228
5.1.1 The association between memory and language knowledge. 229
5.1.2 Effects of exposure type on memory and language knowledge. 232
5.2 Memory, difficulty, exposure and knowledge type 236
5.2.1 Memory, difficulty and language knowledge. 237
5.2.2 Memory, difficulty, exposure and language knowledge 239
5.3 Exposure length, memory and language knowledge 243
5.3.1 Between-group differences in language knowledge. 243
5.3.2 Between-group differences in memory abilities. 246
Chapter 6. Conclusion 249
6.1 Summary of key findings 251
6.2 Implications 253
6.2.1 Pedagogical. 253
6.2.2 Theoretical implication. 255
6.2.3 Methodological. 257
6.3 Limitations and future research 259
6.4 Conclusion 264
References 265
Appendix A: Background questionnaire/ Biodata Form 290
Appendix B: Test materials 293
Appendix C: Study results tables 304
Appendix D: Sample of Malawi's education system syllabus with required information in all columns 311

CHAPTER 1. INTRODUCTION

1.1 Overview

The goal of language learning for many learners is to use the language for authentic and effective communication. A high level of language knowledge and ability is required to be able to authentically and effectively communicate. Advanced-level fluency and comprehensibility are some significant attributes of high-level language knowledge ability (Robinson, 2005a). However, second language (L2) acquisition is highly variable in the rate of learning and in the ultimate level of achievement, with grammar learning posing a particular difficulty for learners (R. Ellis, 2005b; Weber-Fox & Neville, 1996 cited in Ettlinger, Bradlow, & Wong, 2014; Saville-Troike & Barto, 2017). The widely differential levels of L2 acquisition outcomes range from nativelike to a far more limited L2 proficiency, even when learners share commonalities such as native language, educational level, and experience with the L2 (Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014). In contrast to first language (L1) acquisition and child L2 learning before the age of 4 or 5 (M. Paradis, 2009) which almost always lead to native-level proficiency in the language, research suggests that very few learners achieve nativelike proficiency in L2 comprehension or production (Faretta-Stutenberg & Morgan-Short, 2018; Linck et al., 2013; Saville-Troike & Barto, 2017; Selinker & Gass, 2008; Tagarelli, Ruiz, Vega, & Rebuschat, 2016), with the estimated proportion of highly proficient adult learners ranging between zero and about 5% of learners (see Linck et al., 2013).

A great deal of L2 acquisition research has examined the factors that account for the limited success in, and the differential levels of, L2 attainment. In addition to the influential role of the context in which an L2 is learned (i.e., naturalistic or instructed) as well as the social and linguistic background of the language learner (Norris & Ortega, 2000; M. Paradis, 2009; Saville-Troike & Barto, 2017; Spada and Tomita, 2010), the research has found that a large amount of the outcome variance of L2 acquisition is attributable, all other things being held constant, to the impact of learner individual differences (IDs, i.e., the internal factors that vary by learner in the process of L2 acquisition) in personality, language aptitude, motivation, learning styles and learning strategies (Dornyei, 2005; 2006; see also Carpenter, 2008; Carroll, 1958; 1981; Dörnyei & Skehan, 2003; Erlam, 2005; Grey et al., 2015; Li, 2016; Mackey, Adams, Stafford, & Winke, 2010; Miyake & Friedman, 1998; J. Paradis, 2011; Robinson, 2002a; 2002b; 2003; Segalowitz & Frenkiel-Fishman, 2005; Skehan, 1991; Winke, 2005; Yalçın & Spada, 2016). The findings of this research have been invaluable and crucial for the theoretical and pedagogical understanding of L2 acquisition. However, "we are still a long way from understanding how [L2s] are learned, why many individuals have difficulty in reaching high levels of proficiency in [an L2], or even what the best pedagogical approach might be" (Gass & Mackey, 2012, p. 1). The reason for this is that L2 acquisition research has largely neglected to examine the interplay of the sources (i.e., constraining factors) and nature (i.e., the types or representation) of the underlying L2 knowledge system. Doughty and Long (2003) argue that the ultimate goal of L2 acquisition investigation is to understand the underlying L2 knowledge system.

Two language knowledge systems are distinguished: the explicit (or declarative) and implicit (or procedural) language knowledge types. Declarative language knowledge is conscious, analysed and may be verbally described. Procedural language knowledge is the learner's implicit linguistic competence, which is intuitive and unconscious, and which enables spontaneous and fluent use of the language characterised by fluency and the type of control characteristic of one's L1 (N. C. Ellis, 2008; R. Ellis, 2009a; M. Paradis, 2009). Therefore, the primary goal of L2 learning is the development of procedural knowledge (R. Ellis, 2005b). In instructed settings, declarative learning and knowledge in L2 acquisition are considered by some to have a facilitative role in the development of L2 linguistic competence (Akakura, 2009; N. C. Ellis, 2008; R. Ellis, 2002). As the debate regarding the role of grammar instruction in L2 acquisition continues, the examination of the interaction the sources and the nature of L2 knowledge systems offers one promising way to understand L2 acquisition.

1.2 The Purpose of the Present Study

With the overarching goal to explore, and present the overview of, the educational effects on L2 English in a small south-east African nation of Malawi, the present study is aimed at understanding the sources and nature of L2 knowledge systems. This goal is achieved by investigating not only the learner's internal and external factors that constrain L2 acquisition but also the types of the underlying L2 knowledge system as the function of the constraining factors.

To explain the modulating factors in L2 acquisition, current approaches in cognitive psychology which view language as inextricably intertwined with cognition have called for the investigation of the role of the two most important long-term and domain-general memory systems, namely the declarative and procedural memory systems (Hamrick et al., 2018; M. Paradis, 2004; 2009; A. S. Reber, 1989; A. S. Reber et al., 1991; Ullman, 2001a; 2001b; 2004; 2015; 2016). The research has implicated the memory systems in (non-)language functions both in humans and animals. The declarative and procedural memory systems differ with reference to their relationships with awareness, the computations they perform, and the neural substrates subserving them (Foerde, Knowlton, & Poldrack, 2006; Morgan-Short & Ullman, 2012; M. Paradis, 2004; 2009; Ullman, 2001a; 2001b; 2004; 2015; 2016). The declarative memory system has been found to play a role in the appropriation and representation of conscious, declarative language processes (e.g., Carpenter, 2008; Hamrick, 2015; Hamrick et al., 2018; Lum & Kidd, 2012). The procedural memory system has been found to underlie the appropriation and representation of unconscious, procedural processes (e.g., Carpenter, 2008; Faretta-Stutenberg & Morgan-Short, 2018; Granena & Long, 2013; Hamrick, 2015; Hamrick et al., 2018). In contrast to declarative memory, the information in the procedural memory system is said to be encapsulated and difficult to verbalise and to access via introspection (Ullman, 2004).

In his implicit learning theory, A.S. Reber (1989; 1992; A. S. Reber et al., 1991) proposes that unconscious, implicit processes are unaffected by conscious, explicit learning processes, suggesting that implicit learning and knowledge function independently from explicit learning and knowledge. This non-interface approach to the distinction between implicit learning and explicit learning, on the one hand, and implicit knowledge and explicit knowledge, on the other hand, implies that the two learning conditions have moderating effects on the two long-term memory systems. In fact, there is now empirical evidence from the research assessing L2 processing and the role of cognitive abilities (e.g., working memory and language aptitude) that, even though explicit instruction is more effective than implicit instruction, the two learning conditions appear to be fundamentally different with only implicitly trained learners relying on native-like language learning processes (Carpenter, 2008; DeKeyser, 1995; Erlam, 2005; Ettlinger et al., 2014; Faretta-Stutenberg & Morgan-Short, 2018; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Norris & Ortega, 2000; A. S. Reber et al., 1991; Robinson, 1996; 1997; 2002a; 2005a; 2005b; Spada and Tomita, 2010; Tagarelli, Mota, & Rebuschat, 2015; Tagarelli et al., 2016). Reber's implicit learning theory further claims that the unconscious, implicit learning or memory is not an ability and is more effective when learning complex, rule-governed knowledge (Kaufman et al., 2010; A. S. Reber et al., 1991). Therefore, Reber's implicit learning theory posits that the two long-term memory systems are constrained by learning conditions and the level of difficulty of the learning stimuli. Indeed, this is in tandem with current L2 acquisition research that calls for the examination of the interplay between L2 learner internal abilities and external factors: namely, IDs, learning conditions and linguistic complexity (DeKeyser, 2016; Housen & Simoens, 2016). It is argued that such a research approach offers a promising way "to address the dynamic relationships that may be relevant for [L2 acquisition]" (Sanz 2005, cited in Faretta-Stutenberg, 2014, p. 2).

Very little research has investigated the role of the declarative and procedural memory systems in L2 acquisition and how learning conditions and linguistic complexity moderate the effects of the two learning mechanisms. Thus, the predictive role of the declarative and procedural memory systems for L2 learning and the moderating effects of learning conditions and linguistic complexity remain unclear (e.g., Carpenter, 2008; Ettlinger et al., 2014; Tagarelli et al., 2016). Further, there is a paucity of research examining the interplay of memory systems, learning conditions and linguistic complexity in L2 acquisition in real-world settings such as immersion and classroom settings. L2 acquisition research in real-world settings is necessary in order to maximise the ecological validity of the findings in the field (Faretta-Stutenberg, 2014).

This thesis addresses these gaps by examining the interplay of the declarative and procedural memory systems and linguistic complexity in the instructed L2 English acquisition of grammatical structures in classroom settings. The goal is to understand the interplay of the sources and nature of the language knowledge systems that proceed from instruction in classroom settings.

1.3 English in Malawi

This thesis research was conducted in Malawi, the home country of the researcher. Like many other Malawians, English is the L2 of the researcher, while Chichewa, a Bantu language, is his mother tongue. The researcher completed his primary school to university education in Malawi, and he graduated with a bachelor’s degree in English linguistics. Thus, the researcher was educated through the same educational system as his research participants in the present study.

The researcher’s interest in investigating the acquisition of L2 English in Malawi began about a decade ago when he was a language teacher at one government secondary school in Malawi. First, despite studying English for many years from primary school all the way to university, the researcher, like many other teachers, felt lacking in communicative competency during the delivery of English lessons. This was not the case for lessons in Chichewa language, the researcher’s L1. Second, one could easily notice how difficult speaking English was not only for the researcher’s secondary school students but for most secondary school and university students country wide. These students had studied English for years from primary school, but still had very limited communicative proficiency in English. They were only able to answer some simple questions; they could barely sustain a conversation in their L2 English.

Owing to his tertiary educational background (i.e., as a “pure linguist”) and to grammatical errors that the secondary school English learners were observed to make (e.g., in the use of third person singular marker “-s” and possessive determiners/pronouns “her(s)/his” among others), the researcher had the conviction that some of the problems the learners encountered were deeply rooted in their L1 influence. The researcher therefore elected to explore this thesis during his master’s degree. However, the results of the study did not speak much to the issue of the deplorably low levels of the learners’ English communicative competence.

Having been introduced to the notions of explicit and implicit language learning and knowledge at the beginning of this doctoral study, the researcher was motivated and challenged to investigate the effects of instructed language learning by examining the nature of English language knowledge the learners have in Malawi. Reflecting on his English language teaching approach, the researcher found that he was largely influenced by how he himself had been taught English when he was a student. Generally, in English grammar lessons in a Malawian classroom, teaching is teacher centred. A teacher first presents a grammatical rule followed by detailed explanation which may include exceptions to the grammatical rule, if any. Then learners are given a task to determine the grammaticality of example sentences based on the rule they have just learned. Often, the questions that learners are asked during the lessons and the grammar tasks they do at the end of the lessons require a single word answer. Comprehension, composition, note making, and, of late, literature lessons are usually taught separately from grammar. In comprehension lessons, learners are asked to read a passage and then answer questions afterwards. The answers to the questions are usually a sentence long. Due to the overwhelming class sizes, free writing (that includes composition writing) is a daunting task to grade. Thus, it is not done often. Such a teaching approach should limit learners' communicative abilities and linguistic competence in L2 English because it gives learners minimal opportunity to use the language in the most authentic, creative and meaningful manner.

Malawi is a relatively small landlocked country that lies in the south eastern part of the African continent. It borders Zambia to the west, Tanzania to the north, and Mozambique to the south, southeast and southwest. The country has more than 10 different ethnic groups, with roughly sixteen African languages and a population of about 19 million (Chiyembekezo, Kondowe, & Ngwira, 2019; Kamwendo, 2016). From 1891 to the early 1960s, Malawi was under British rule. It became an independent country in 1964. Because the country is multilingual, the language situation has been dominated by two languages: English, the official language whose relatively long history in Malawi is traced back to the mid-19th century; and Chichewa, a Bantu and national language of the country. As an official language, English is used as the language for administration, commerce and trade, employment, the judiciary, the legislature, and for scientific and technological advancement. Further, English has an innovative function in Malawi, where most of the literary genres are in English. Thus, because of the high status of English in Malawi, learners are highly motivated to learn the language.

In education, English serves an instrumental function (as the medium of learning). The education system in Malawi consists of eight years of primary education (from Standard 1 to Standard 8), four years of secondary education (from Form 1 to Form 4), and 3 to 4 years of tertiary education. Learners enter school at the age of 6. All primary and secondary schools offer the same curriculum. The curriculum is determined by the Ministry of Education, Science and Technology. English is used as the medium of instruction from Standard 5 up to tertiary education in all subjects except in the

Chichewa subject. The use of Chichewa as the language of instruction in the first four grades of primary education is thought to act as a prerequisite for learning English. However, currently the Ministry of Education, Science and Technology has a new language-of-instruction policy to replace Chichewa with English in the first four years of primary school so that English is used as a language of instruction all the way from Standard 1 (Kamwendo, 2016).

English is also a compulsory subject from Standard 1 and a compulsory course (i.e., as English for Academic Purposes) for all first-year university students. In primary and secondary schools, English is allocated more time than any other subject; it is taught almost every day with about seven class-schedule time slots as compared to an average of three class-schedule time slots for each of the other subjects. In addition, English is a passing subject (i.e., a subject that qualifies one to get a certificate if passed) in all the three national examinations: namely, the Primary School Leaving Certificate Examinations (PSLCE) in grade 8, the final year of primary school; the Junior Certificate Examinations (JCE, now phased out from 2017) in the second year of secondary education (or grade 10 in some other education systems); and the Malawi School Certificate of Education examinations (MSCE) in the final and fourth year of secondary school (or grade 12). The Malawi National Examinations Board (MANEB), a body under the Ministry of Education, Science and Technology, is charged with assessing the performance of all primary and secondary school students. It administers all three standardised examinations.

Although tremendous emphasis is placed on the teaching and learning of English in Malawi, empirical research has shown that English learners' achievements both in written and spoken skills are very low (Kamwendo, 2003; Mmela, 2006). Mmela (2006) reports an average of only about 8% of primary school students pass the PSLCE and are selected to go to secondary school; the rest, over 90%, are left behind. Of the many reasons, "failure to pass English is the highest contributing factor" (Mmela, 2006, p. 4). Mmela (2006) also reports several studies carried out between 1990 and 2005 that investigated the L1 and L2 language reading proficiency of Standards 3, 4 and 6 children in Malawi. All the studies found that learners' achievements were critically low in English. Similar results are reported in Kunje, Selemani-Meke, and Ogawa (2009) and Maganga, Mwale, Mapondera, and Saka (2010). Investigating how school, classroom and pupil factors influence Standards 5 and 7 pupil achievement in mathematics, English and Chichewa, Kunje, Selemani-Meke, and Ogawa (2009) found low achievements in English and mathematics. Maganga et al. (2010) assessed Standards 3 and 7 learners’ achievement levels in mathematics, Chichewa, English and life skills in Malawi. The results showed that Standard 3 learners performed relatively better in numeracy and mathematics, but the performance was the poorest in English in both the baseline and end-line studies. At tertiary level, the Department of English at Chancellor College, one of the constituent colleges of University of the Malawi has recently introduced four English grammar courses, one course for each year, with an aim of improving students’ abilities in English language. This suggests that at every level of the education system in Malawi all the way to university, the students’ English abilities are generally deplorable.

While it is very clear that primary to university students’ speaking and writing achievements in L2 English in Malawi are deplorably low, it is not clear what the exact sources of this poor performance are. Several factors have been identified, including the lack of teaching and learning materials, poor education facilities, and very high teacher-pupil ratios in rural areas (Chimpololo, 2010; Kayambazinthu, 1998; Mmela, 2006). In addition, teaching approaches such as teacher-centred and examinations-orientated instruction have shown to be major factors contributing to the low-level English achievements (Chiuye, 2005). Although the curriculum emphasises learner-centred teaching, the traditional teacher-centred approach dominates the teaching of English in Malawi schools. Evidence suggests that teachers pay “lip service to learner-centred teaching which encourages participatory and active learning, but in reality, use a transmission style” (Mmela, 2006, p. 9). This observation is supported by Mizrachi, Padilla, and Susuwele-Banda (2010) who point out that in Malawi there is a divide between policy and practice: while policies have been put in place to support active-learning approaches or student-centred approaches, teachers still rely on the traditional approach of grammar teaching. The traditional teacher-centred approach is “characterized by grammatical analysis, reading without comprehension, and patterned drills resulting in students’ scoring well on grammar tests but failing to communicate in the target language” (Mmela, 2006, p. 7). School in Malawi represents the only context where children learn English (Mmela, 2006). Outside the classroom, learners are not exposed to English. This means that L2 English grammar learning in Malawi may be said to proceed mainly from instruction received at school.

The situation in Malawi, where English language teaching involves explicit instruction and learners' writing and speaking achievements are deplorably low, and the conflicting results in L2 acquisition research, which has consistently shown that explicit instruction is superior to implicit instruction (e.g., Norris & Ortega's 2000; Robinson 2002; 1996; Spada and Tomita 2010; see section 2.5 for the discussion of this research), raise questions about how explicit instructional type influences language learning and how the knowledge acquired in this way is represented. To the knowledge of the researcher for the current study, no study has investigated the nature of learners' knowledge of L2 English in Malawi. The present study seeks to address this gap; it aims to establish whether there will be a relationship between learners' two long-term memory systems, their language knowledge that proceeds from explicit instruction, and the linguistic complexity of the grammatical structures.

The findings of this thesis research have theoretical, pedagogical and methodological significance for the field of L2 acquisition. Insights into whether the instructed learners' L2 knowledge system is represented as either declarative or procedural, or both, and whether the level of difficulty of grammatical structures has moderating effects on the knowledge system, are of interest to L2 acquisition theorists as well as to L2 teachers, teacher trainers, learners, and curriculum and material designers. Chiuye (2005, p. 202) points out that in a country like Malawi "where proficiency in English determines livelihood, educational opportunities, and socio-political participation, it is critical that students develop English proficiencies from their schooling experiences". The findings of this study also have methodological significance. The adoption of the research design features such as the triangulation of measures, natural learning context, and factor and structural analyses is hoped to maximise not only the validation of target constructs but also the ecological validity of outcomes.

This thesis consists of five chapters. This chapter has provided the background to the study within which the study's primary objective is stated. Chapter 2 presents a theoretical framework for the study by reviewing extant literature on language knowledge types, long-term memory systems (i.e., declarative and procedural memory systems) and proposed learning models, learning conditions and linguistic complexity. The chapter further presents major findings from empirical research studies on the interplay of the two long-term memory systems, language learning contexts and linguistic complexity. Gaps in the previous literature are subsequently identified and research questions are raised for investigation. The methodological approach adopted in the study is presented in Chapter 3. The study site and participants, the operationalisation of declarative and procedural learning ability, declarative and procedural language knowledge and of linguistic complexity are presented in the chapter. Finally, instruments for data collection and procedures followed in collecting and analysing data are presented. Data collection methods in the present study were triangulated to offer the possibility of providing results that complement, elaborate and confirm each other. Chapter 4 reports the results of the study, including descriptive statistics for all the tests, and the key findings from correlational, factor and structural modelling equation analyses. The discussion and conclusion in Chapter 5 includes a detailed summary and interpretation of the findings of the study, with reference to each of the research questions raised and in relation to relevant previous research findings. The concluding section summarises the findings, draws conclusions from those findings, and indicates some of the implications of the findings before the limitations of the study and suggestions for further research in the field are considered.

CHAPTER 2. LITERATURE REVIEW

It is not until the 1960s that scholars began to formulate systematic L2 acquisition theories and models to address the basic questions about *what* the L2 learner knows, *how* the learner acquires this knowledge, and *why* some learners are more successful than others (Saville-Troike & Barto, 2017). Since then, a considerable number of theoretical frameworks of L2 acquisition have been proposed based on linguistic, psychological and social perspectives (Table 2.1).

Table 2.1. Important Theoretical Frameworks for Study of L2 Acquisition

Perspective	Focus	Framework
Linguistic	Internal	Transformational-Generative Grammar Principles and Parameters Models Minimalist Program
	External	Functionalism
Psychological	Languages and the Brain	Neurolinguistics
	Learning Processes	Information processing Processability Connectionism
	Individual Differences	Humanistic Models
Social	Microsocial	Variation Theory Accommodation Theory Sociocultural Theory Computer-Mediated Communication
	Macrosocial	Ethnography of Communication Acculturation Theory Social Psychology

Adapted from Saville-Troike & Barto (2017, p. 26)

These theoretical frameworks offer insights into L2 acquisition. However, they differ in the questions they ask and investigate. Social theoretical frameworks underline the importance of social context for language acquisition and use. In these theoretical frameworks, what is being learned is not only the language itself but also the social and cultural knowledge infused in the language. Ultimately, (all of) language learning is seen as a social process where group membership and identity dictate what is learned, how it is acquired, and why some learners are more successful than others (Saville-Troike & Barto, 2017). Vygotsky's (1962; 1978) Sociocultural Theory views inter- and intra-personal interaction not only as a facilitating factor to language learning but also as the essential genesis of language (Saville-Troike & Barto, 2017).

Linguistic theoretical approaches view what is being acquired as the underlying knowledge of highly abstract linguistic principles and constraints (Chomsky, 1995; Saville-Troike & Barto, 2017; White, 2003; 2015). To explain how L2 is learned, these theoretical models differ in their emphasis on continued innate universal grammar (UG) capacity for language learning. UG is a biologically given cognitive structure the role of which is to enable children to "so reliably and effortlessly" project grammars from the impoverished primary linguistic data (PLD) to which they are exposed (Boeckx, 2006, p. 19). L1 learners have full access to UG. In L2 acquisition, learners have four possibilities in that they may retain (1) *full access* to UG (Epstein et al., 1998; Epstein, Flynn, & Martohardjono, 1996; Schwartz & Sprouse, 1996); (2) *partial access* to UG, keeping some of its components but not others (Eubank, 1994; Vainikka & Young-Scholten, 1994); (3) *indirect access* to UG through knowledge that is already realised in their L1 but to which they have no remaining direct access (see White, 2003); (4) *no access* to UG and must learn L2 via entirely different means to those through which they learn their L1 (Bley-Vroman, 1989; Saville-Troike & Barto, 2017). Linguistic theoretical frameworks account for L2 individual differences by implicating factors which are largely internal to language and mind such as the varying degrees of access to UG or specifications of lexical features. Further, factors to do with L2 input such as qualitatively different L2 input or differential L1-L2 transfer or interference are also implicated (Saville-Troike & Barto, 2017). Saville-Troike and Barto (2017, p. 66) point out that "[purely] linguistic approaches, though, have largely excluded psychological and social factors" in their attempt to explain L2 acquisition.

The present study is couched within the psychological or cognitive theoretical approaches. Psychological approaches are informed by both linguistics and psychology. They essentially view language as inextricably intertwined in complex and dynamic ways with cognition. On *what* is acquired in L2 acquisition, psychological theoretical frameworks point to "additions or changes that occur in neurological makeup, and on how the multilingual brain is organised" (Saville-Troike & Barto, 2017, p. 100). One major finding in cognitive linguistics is that the L2 physical representation in the brain is not very different from the L1's, but that the differences in brain organisation are related not only to the L2 learners' proficiency but also to how the acquirers learned the L2. In these theoretical models, language knowledge is declarative (i.e., 'controlled', conscious and explicit) and procedural (i.e., "uncontrolled", unconscious and implicit). In contrast to Chomsky's proposal that there is a species-specific language acquisition device (LAD), the psychological approaches generally follow Anderson (1983; 1976) in viewing L2s as learned according to the same processes as the acquisition of other areas of complex knowledge and skills from which rules and principles are abstracted, and/or neurological associative networks and connections are developed. Declarative memory and procedural memory are the two most important long-term memory systems in the brain, not only in terms of the range of functions and domains that they subserve but also in terms of the key roles they play in language in ways that are analogous to how they function in other domains (Ullman, 2015; 2016). To explain *why* some learners are more successful than others, psychological approaches implicate learners' IDs influenced by age, sex, personality, aptitude, motivation, individuals' learning styles and strategies. Further, the complexity of stimuli, the condition of learning and the role of input are debated and viewed to significantly constrain L2 acquisition (DeKeyser, 1995; Ettlinger et al., 2014; Granena, 2013b; Housen & Simoens, 2016; Morgan-Short et al., 2012; M. Paradis, 2009; A. S. Reber et al., 1991; Robinson, 1996; 1997; 2002b). The present study's focus is on the role of the declarative and procedural memory systems and the learning environment in L2 acquisition.

This chapter synthesises the literature on L2 acquisition as conceived in psychological approaches. To address *what* the L2 learner knows, section 2.1 reviews literature on the nature of language knowledge. The section also provides the overview of the common ways of operationalising the two types of language knowledge. To address *how* the learner acquires the L2 and *why* some learners are more successful than others, section 2.2 reviews literature on declarative and procedural memory systems and on the three proposed models that explain different ways in which the memory systems may be involved in L2 acquisition. The section further provides the overview of how the memory systems (i) may account for learners' IDs and (ii) have been operationalised in previous research. The learning domain has been implicated as external constraints on how the memory systems are engaged in L2 acquisition. It consists of linguistic difficulty and learning conditions discussed in sections 2.3 and 2.4, respectively. Section 2.5 reviews empirical research on whether explicit and implicit learning conditions are fundamentally different. This will be followed by the review of empirical studies on the interplay of the two long-term memory systems, language exposure type and linguistic complexity in L2 acquisition in section 2.6. Summary and rationale are presented in section 2.7. The chapter concludes with the presentation of research questions and hypotheses in section 2.8.

2.1 Nature of language knowledge

Within the psychological frameworks, language acquisition, like the acquisition processes of any other complex knowledge and skills, is seen as drawing on both implicit and explicit learning mechanisms (e.g., Anderson, 1976; A. S. Reber, 1967; 1989; A. S. Reber et al., 1991; Saville-Troike & Barto, 2017; Ullman, 2001a; 2001b; 2004; 2005; 2015; 2016). In turn, the implicit and explicit learning mechanisms result in implicit and explicit language knowledge types. In L2 acquisition research, the idea that learners may possess two types of language knowledge underlies two early language learning theories: first, Krashen's (1981) monitor theory that proposes that the learners' "acquired system" arises from subconscious or implicit learning processes while a "learned system" results from paying conscious attention to language and memorising rules; and second, Bialystok's (1978) theory of L2 learning which proposes that exposure to communicative or naturalistic language use results in implicit knowledge while explicit knowledge arises from conscious exposure to a language (Erlam, 2006). Erlam (2006) points out that the existence of both types of language knowledge is commonly held and widely accepted in cognitive psychology.

2.1.1 Implicit and explicit language knowledge.

Implicit language knowledge is intuitive and procedural, that is, it is accessed instantaneously and easily, and is available for use in rapid, fluent communication (R. Ellis, 2009a). Further, implicit language knowledge is held unconsciously and can only be verbalised if it is made explicit. This means that implicit knowledge is only evident in learners' verbal behaviour and 'it exists in the form of statistically weighted connections between memory nodes' whose "regularities are only manifest in actual language use" (R. Ellis, 2009a, p. 13). It is the view of most researchers that competence in an L2 largely depends on the amount of implicit language knowledge a language user has (Bialystok, 1978; R. Ellis, 2004; 2005a; 2009a). That is, the ability to confidently and fluently communicate depends on one’s implicit knowledge. Thus, implicit language knowledge constitutes linguistic competence, irrespective of whether linguistic competence is described according to innatist theory as a biological capacity for acquiring language, or whether it is defined as abstraction of rules and principles in tandem with the development of neurological associative networks and connections. In neurolinguistic, psycholinguistic, and L2 literature, there have been several terms used to refer to implicit language knowledge: unanalysed knowledge (Bialystok, 1978); unconscious knowledge (A. S. Reber et al., 1991; Schmidt, 1990); acquired knowledge (Krashen, 1981); procedural knowledge (DeKeyser, 1998; M. Paradis, 1994; 2004; 2009); tacit knowledge (A. S. Reber, 1989; A. S. Reber et al., 1991); and implicit linguistic competence (Roehr, 2008). In the present study, these terms are used interchangeably.

Explicit language knowledge is knowledge about language, and it is the declarative or conscious knowledge of the phonological, lexical, grammatical, pragmatic and other features of an L2 together with the metalanguage for labelling this knowledge (R. Ellis, 2004). Explicit knowledge is held consciously, is verbalisable and is typically accessed through controlled processing in cases where learners experience linguistic difficulty in the use of the L2 (R. Ellis, 2009a). Analysed explicit language knowledge is distinguished from metalingual explanations. Analysed language knowledge refers to a conscious awareness of how a structural feature works while metalinguistic knowledge consists of knowledge of grammatical metalanguage and the ability to understand explanations of rules (R. Ellis, 2009a). Simply put, metalanguage is the technical terminology needed to describe language. A distinction between analysed and metalinguistic explicit knowledge entails that not all explicit language knowledge is necessarily manifested in technical metalinguistic terminology such as *present progressive tense* or *definite article*. Structural rule descriptions (e.g., the verb -ing form is used when the action is ongoing) are an acceptable display of explicit knowledge even though metalingual terms (e.g., present-progressive) are not used (Akakura, 2009). In cognitive psychology, the explicit knowledge of one’s L1 is consciously acquired later than the implicit knowledge thereof (Reber, 1989). In contrast, whichever L2 knowledge type comes first is dependent on the learning context. The following terms have been used in literature to refer to explicit language knowledge: analyzed knowledge (Bialystok, 1978); conscious knowledge (A. S. Reber et al., 1991; Schmidt, 1990); learned knowledge (Krashen, 1981); declarative knowledge (DeKeyser, 1998;

Paradis, 1994; 2004; 2009); explicit knowledge (A. S. Reber, 1989; A. S. Reber et al., 1991); and explicit metalinguistic knowledge (Roehr, 2008). These terms will be used interchangeably to refer to explicit language knowledge in the present study.

2.1.2 The interface hypothesis.

However, the relationship between explicit and implicit knowledge is a matter of considerable debate. For instance, it is possible that through enough practice, explicit knowledge may become automatised and accessed for rapid online processing in much the same way as implicit knowledge (R. Ellis, 2009a). In fact, DeKeyser (2003, cited in R. Ellis, 2009a, p. 12) proposed that automatized explicit knowledge can be considered "'functionally equivalent' to implicit knowledge". However, it is possible that what appears to be the automatization of explicit knowledge is the automatization of implicit knowledge (R. Ellis, 2009a; Hulstijn, 2002). This debate on the relationship between explicit and implicit knowledge has led to what has become to be known as the interface hypothesis. The interface hypothesis addresses several key questions of both theoretical and practical importance for L2 acquisition and pedagogy: "To what extent and in what ways are implicit and explicit learning related? Does explicit knowledge convert into or facilitate the acquisition of implicit knowledge? Does explicit instruction result in the acquisition of implicit as well as explicit knowledge?" (R. Ellis, 2009a, pp. 20-21).

Three competing positions about the role of explicit instruction in the process of acquiring L2 knowledge have emerged from this debate: (1) the non-interface position, (2) the strong interface position and (3) the weak interface position. The non-interface position's basic claim is that learned knowledge cannot become acquired knowledge. The proponents of this position hold that there is an absolute distinction between implicit and explicit knowledge since they involve very different acquisition mechanisms, are stored in different parts of the brain, and are accessed differently for performance (R. Ellis, 1993; Kaufman et al., 2010; S. D. Krashen, 1981; M. Paradis, 1994; 2009; A. S. Reber et al., 1991). Explicit knowledge or conscious learning is available just as a monitor for performance (Krashen, 1981). The non-interface position is also referred to as the strong non-interventionist position (Long, Inagaki, & Ortega, 1998) as it limits the contribution of classroom intervention as providing only "comprehensible input that might not otherwise be available outside the classroom"

(Krashen, 1985, cited in Akakura, 2009, p. 20). Therefore, under the non-interface position, L2 and L1 acquisition processes are considered similar with no role for explicit language knowledge.

The strong interface position is in polar opposition to the non-interface position. The strong interface position, usually associated with DeKeyser (1998; 2003; 2007) and Sharwood Smith (1981), claims that not only can explicit knowledge be derived from implicit knowledge, but also that explicit knowledge can be converted into implicit knowledge through practice. According to the strong interface position, L2 knowledge is first gained in explicit form and then converted into implicit form through communicative practice while, conversely, the initial implicit knowledge is transformed into explicit knowledge through reflecting on and analysing the output generated by means of implicit knowledge (DeKeyser, 1998, Milasi & Pishghadam, 2007 cited in Zhang, 2015). The strong interface position, therefore, posits a significant role for explicit instruction in L2 acquisition, leading to a theoretical basis for many studies investigating the direct consequences of explicit instruction on L2 acquisition.

Lastly, the weak interface position is an in-between position. Its basic claim is that explicit knowledge does not have a causal relationship with implicit knowledge, but instead, it only triggers or speeds up the implicit learning process, which subsequently leads to the generation of implicit knowledge (Vafaee, Suzuki, & Kachisnke, 2017). Therefore, like the strong interface position, the weak interface position provides theoretical basis for experimental studies that manipulate instruction to investigate its effectiveness in L2 acquisition.

Three versions of the weak interface position have emerged. The first version espoused by N. C. Ellis (2005; 2008; 2015) posits that explicit knowledge indirectly contributes to the acquisition of implicit knowledge by promoting some implicit learning processes. According to this version, explicit knowledge can make linguistic features salient, thereby enabling learners to notice them and recognize the gap between the input and the linguistic knowledge they already possess. This position suggests that explicit and implicit learning processes work in tandem in L2 acquisition, resulting in a dynamic interaction between them for the consolidation of implicit knowledge (R. Ellis, 1993; 1994; Vafaee et al., 2017). The second version espoused by R. Ellis (1993) posits that explicit knowledge about a language structure may facilitate the development of implicit knowledge only if the learner is developmentally ready to acquire the linguistic form, and when the instruction primes several key acquisitional processes such as noticing the gap (e.g., Schmidt, 1990). This version of the weak interface draws on notions of learnability and the attested developmental sequences in L2 acquisition (e.g., Pienemann, 1989; 1998a; 1998b; 2005). The learnability/teachability hypothesis claims that L2 learners can benefit from classroom instruction only when they are psycholinguistically ready for target language features (Pienemann, 1989). The third version of the weak interface position, as espoused by Schmidt & Frota (1986) and Sharwood Smith (1981), claims that L2 learners use "their explicit linguistic knowledge to produce (presumably planned) output" (Bowles, 2011, p. 249), which, in turn, becomes auto-input for their implicit system.

2.1.3 Controlled and automatic processing

Finally, there is no a one-to-one relationship between explicit/implicit language knowledge distinction and controlled/automatic language processing. Explicit and implicit language knowledge need not be conflated with and equated to the notions of controlled and automatic language processing. Based on information processing perspective, the acquisition of an L2 initially "requires the use of controlled processes with focal attention to task demands but as performance improves, attention demands are eased, and automatic processes develop, allowing other controlled operations to be carried out in parallel with automatic processes" (McLaughlin et al., 1983, cited in Akakura, 2009, p. 15). Therefore, the nature of language knowledge and the process of acquiring that knowledge may best be represented as two separate (dichotomous) dimensions: implicit ↔ explicit and controlled ↔ automatic. Consequently, both implicit and explicit language knowledge may either be controlled or automatic. According to R. Ellis (2009b), the L2 information processing (Table 2.2) proceeds as follows:

> "Explicit knowledge is used initially with deliberate effort (A) but may later be used with less effort and relative speed (B), provided the L2 user is developmentally ready. Novel implicit knowledge is slow and inconsistent at first (C) but may later become effortless (D) after form-focused practice or meaningful communication. (Akakura, 2009, p. 16)

Table 2.2. The Difference between Explicit/Implicit Knowledge and Controlled/Automatic Processing

Type of Knowledge	**Controlled Processing**	**Automatic Processing**
Explicit	A A new explicit rule is used consciously and with deliberate effort	B An old explicit rule is used consciously but with relative speed.
Implicit	C A new implicit rule is used without awareness but is accessed slowly and inconsistently	D A fully learnt rule is used without awareness and without effort.

Adopted from Akakura (2009, p. 15)

In summary, the distinction between implicit and explicit language knowledge with the related three hypothetical interface positions has significant theoretical and pedagogical implications for the role of explicit knowledge, exposure condition and linguistic difficulty in L2 acquisition. To empirically investigate the effects of explicit instruction in terms of explicit and implicit language knowledge, test measures that are designed to distinguish between the knowledge types are required. These test measures are discussed next.

2.1.4 Measurement of language knowledge

Researchers acknowledge that though the development of implicit language knowledge is the primary goal of L2 acquisition, most L2 learners possess both implicit and explicit language knowledge (R. Ellis & Roever, 2018). Consequently, it is difficult to determine the type of knowledge they "deploy on particular occasions" (R. Ellis & Roever, 2018, p. 3). This problem is further aggravated by the lack of direct ways to see how language knowledge is represented in the learners' mind. However, by examining learners' linguistic behaviour, researchers make inferences about the type of knowledge that a learner draws on during any language performance. Based on the differences in the representation and processing of implicit and explicit knowledge, the extent to which the two knowledge types are involved in L2 performance is dependent on such factors as task modality, time pressure, task requirements, proficiency level, and length and type of prior L2 study (R. Ellis, 2005a; Rebuschat, 2013; Roehr, 2008; Spada, Shiu, & Tomita, 2015). For instance, "a written task administered without time pressure is likely to result in greater access to and use of explicit knowledge, whereas an oral task administered with time pressure is likely to encourage greater use of implicit knowledge" (Spada et al., 2015, p. 724). Further, oral production tasks (OPTs) that impose relatively few constraints and encourage spontaneous and communicative use of target language features are the most effective ways of eliciting the use of implicit knowledge (Doughty, 2003; Norris & Ortega, 2003; both cited in Spada et al., 2015). On the contrary, the following linguistic tasks focus learners' attention on the language itself: "(a) judging the overall grammaticality of a sentence, (b) identifying an erroneous form of a sentence, (c) correcting an erroneous form of a sentence, (d) explaining why a form is erroneous, (e) identifying named parts of speech in a sentence" (Spada et al., 2015, p. 725), with (a) requiring comparatively the least and (d) requiring comparatively the highest degree of the use of explicit knowledge. R. Ellis (2005a) has presented what has been referred to as concrete proposals on how to operationalize implicit and explicit language knowledge by means of various tests (Akakura, 2009), and these proposals have been hailed as "a crucial moment in rendering theories of implicit and explicit knowledge and learning testable" (Hulstijn, 2005, cited in Akakura, 2009, p. 25).

R. Ellis (2005a; R. Ellis & Loewen, 2007), building on Han and Ellis (Han & Ellis, 1998) and other previous studies (e.g., Birdsong, 1999; R. Ellis, 2004; Han, 1996; Long, 2007; Singleton & Ryan, 2004), proposed the following three behavioural criteria that are hypothesized to translate into how tests can be created so that they provide relatively separate measures of the two knowledge types: the amount of *time available,* with time pressure (implicit) vs. no pressure (explicit); the *focus of attention*, with primary focus on meaning (implicit) vs. primary focus on form (explicit); and the *utility of metalanguage*, not required (implicit) vs. encouraged (explicit). Further, R. Ellis (2005a) proposed the following additional conditions that were hypothesized to "provide supporting evidence that the test was in fact measuring what it purported to measure" (Akakura, 2009, p. 26): the *degree of awareness,* responses by feel (implicit) vs. responses by rule (explicit); *systematicity,* with consistent responses (implicit) vs. variable responses (explicit); the degree of *certainty in response*, high (implicit) vs. low (explicit); and *learnability*, with early learning favoured (implicit) vs. later form-focused instruction favoured (explicit). The summary of these criteria is in Table 2.3.

Table 2.3. Criteria for Measuring Implicit and Explicit Knowledge

Criteria suited for	Criterion	Implicit knowledge	Explicit knowledge	Current understanding
Test design				
1	Primary focus of attention	Meaning	Form	Empirical support (R. Ellis, Loewen, & Erlam, 2006)
2	Time available	Restricted	Unrestricted	Empirical support (Han & Ellis, 1998; R. Ellis, 2005a). Insufficient control as explicit knowledge may not be totally excluded (e.g. de Graaff, 1997). Difficult to impose consistently, particularly in writing tasks (Erlam, 2003a; 2003b).
3	Metalinguistic knowledge	Not required	Encouraged	Theoretical support (Elder & Manwaring, 2004)
Supporting evidence				
1	Degree of awareness	Response according to feel	Response according to rule	Unreliable as dependent on self-report.
2	Systematicity of response	Consistent	Variable	Empirical evidence for variable explicit knowledge (Han, 1996)
3	Degree of certainty in response	High	Low	Empirically unsupported (R. Ellis, 2005a; Roehr, 2006)

4	Learnability	Early learning favoured	Late explicit instruction favoured	Theoretical support (Birdsong, 1999; Long, 2007; Singleton & Ryan, 2004)

From Akakura (2009, p. 27).

In accordance with four of the criteria for distinguishing explicit and implicit language knowledge, R. Ellis (2005a) developed the following five tests such that each was predicted to “provide a relatively separate measure of either implicit or explicit knowledge according to how it mapped out on these criteria” (p. 157; Table 2.4): the Elicited Imitation Test (EIT), the Oral Narrative Test (ONT), the Timed Grammaticality Judgment Test (TGJT), the Untimed Grammaticality Judgment Test (UGJT), and the Metalinguistic Knowledge Test (MKT). Both the EIT and the ONT were predicted to measure implicit knowledge. In these tests, responses “rely predominantly on feel”. In addition to responding under time pressure, participants are also focused primarily on meaning. Such design features are aimed at reducing the opportunity to access explicit knowledge. Because the MKT is unpressured, focuses on form, and requires a very high degree of awareness and use of metalinguistic knowledge, it was predicted to be the best measure of explicit language knowledge. In the TGJT and UGJT, judgements on the correctness of sentences require participants to focus attention primarily on form. However, because the time-pressured TGJT encourages the use of feel with perhaps little opportunity to access explicit knowledge, it was predicted to measure primarily implicit knowledge whereas the unpressured UGJT, like the MKT, was predicted to measure primarily explicit knowledge.

Table 2.4. Design Features of the Tests of Explicit and Implicit Language Knowledge

Criterion	EIT	ONT	TGJT	UGJT	MKT
Degree of awareness	Feel	Feel	Feel	Rule	Rule
Time available	Pressured	Pressured	Pressured	Unpressured	Unpressured
Focus of attention	Meaning	Meaning	Form	Form	Form
Metalinguistic knowledge	No	No	No	Yes	Yes

From R. Ellis (2005a, p. 157)
EIT = Elicited Imitation Test; ONT = Oral Narrative Test; TGJT = Timed Grammaticality Judgment Test; UGJT = Untimed Grammaticality Judgment Test; MKT = Metalinguistic Knowledge Test.

Using principal component and confirmatory factor analyses, the test scores for 91 L2 participants and 20 L1 participants on 17 English constructions deemed difficult by L2 users were analysed to determine whether there are two underlying dimensions (implicit and explicit) to L2 knowledge. The results revealed that there were indeed two separate factors the five tests loaded onto: the EIT, ONT and TGJT which required unplanned language use under speeded conditions loaded on one factor while the UGJT and MKT which encouraged the use of analysed verbalizable explicit language knowledge loaded on another. These results provided strong evidence that it is indeed possible to measure explicit and implicit knowledge relatively separately by manipulating test conditions to elicit one type of language knowledge over the other.

Several studies have been conducted to validate R. Ellis' (2005a) and R. Ellis & Loewen's (2007) results. Erlam (2006), Bowles (2011), Erçetin and Alptekin (2013), Gutiérrez (2013), Spada et al. (2015), Zhang (2015), and Godfroid et al. (2015) provide further empirical support for the construct validity of the test battery in R. Ellis (2005a; R. Ellis & Loewen 2007) as consisting of measures of implicit and explicit knowledge. Further, R. Ellis (2005a; 2009a), R. Ellis and Loewen (2007), Zhang (2015), and Godfroid et al. (2015) have also demonstrated that the UGJT grammatical and ungrammatical sentences appear to measure different constructs, with grammatical sentences drawing on implicit knowledge and ungrammatical sentences tapping into explicit knowledge. Gutiérrez (2013) and Zhang (2015) have further shown that irrespective of the time condition chosen, GJTs' grammatical sentences tap into implicit language knowledge whereas their ungrammatical sentences draw on explicit language knowledge. Unlike Gutiérrez (2013), Godfroid et al. (2015) found the effect of grammaticality only in the untimed condition of the GJTs. However, some research has empirically shown that GJTs and EIT, because they allow for conscious attention to language, are too coarse to be measures of implicit language knowledge, but that they measure different levels of explicit language knowledge, namely automatized explicit knowledge and less automatized explicit knowledge (Suzuki, 2017; Suzuki & DeKeyser, 2015; 2017; Vafaee et al., 2017).

While the MKT is seen as the "best" measure of explicit language knowledge (R. Ellis, 2009c, p. 59), OPTs such as ONT in R. Ellis' (2005a) test battery have, as pointed out above, the greatest face and construct validity as measures of implicit knowledge because they involve freely constructed responses involving the spontaneous use of an

L2 in “continuous discourse and in a context that resembles real-life communication” (R. Ellis, 2015a, p. 427). However, OPTs have a major drawback. In addition to the fact that some language features are more difficult to naturally elicit than others (Loschky & Bley-Vroman, 1993, cited in Spada et al., 2015), it is challenging to design a structured OPT that successfully elicits the grammatical structures targeted by form-focused instruction (R. Ellis, 2015a). As the number of target grammatical structures included in an investigation increases, it becomes increasingly difficult to design OPTs that successfully elicit all the structures. In the present study, a relatively large number of grammatical structures, 14 in total, were targeted. Consequently, the EIT and TGJT were employed as measures of implicit knowledge and the UGJT as the measure of explicit knowledge. Reviewing a number of studies on the validation of implicit and explicit knowledge measures, R. Ellis (2015a) points out that the EIT does not only provide the most effective and convenient way of measuring learners’ implicit knowledge but it also allows a task design that can elicit learner’s oral production on any (number of) grammatical structures targeted. Despite the current unavailability of pure measures of the explicit and implicit language knowledge, research has linked the acquisition of the two language knowledge types to domain-general acquisition processes as described in the next section.

2.2 Declarative and Procedural (DP) Memory Systems

Two different types of approaches, the single processes and the dual processes, have dominated the neurocognitive research on memory systems. On the one hand, the proponents of the single processes approach to memory system argue that there is only one memory system that subserves all experiences (Lovibond & Shanks, 2002; Mitchell, De Houwer, & Lovibond, 2009; Whittlesea & Price, 2001). On the other hand, the dual-processes approach to memory holds that experiences are subserved by multiple memory systems with functionally and anatomically separable neural circuits (Evans, 2003; Foerde et al., 2006a; Lum, Kidd, Davis, & Conti-Ramsden, 2010; McLaren et al., 2014; P. J. Reber, Knowlton, & Squire, 1996; Squire, 1992; 2004; Squire & Zola, 1996; Ullman, 2001a; Ullman, 2001b). One model within the dual-processes approach distinguishes between the declarative memory and procedural memory (DP) systems. Squire (2004, p. 171) points out that a large amount of research from about 1980 supports the view of multiple memory systems, with “evidence from normal subjects, amnesic patients, and experimental animals demonstrating a fundamental distinction between the kind of memory that is accessible to conscious recollection and another kind that is not".

The DP memory systems are the most important two long-term memory systems in the brain that support both language and non-language functions in humans and animals. In contrast to working memory which holds information on the order of (a) second (s), the DP memory systems can store information that can last from minutes to years (Lum & Kidd, 2012). These two long-term memory systems are distinguished on two dimensions: first, based on their functional differences in learning and memory processes; and second, based on their dependency on distinct neurological structures or brain systems. Regarding the origins of the DP memory systems, it is suggested within cognitive neuroscience that the two long-term memory systems evolved because they serve distinct and functionally incompatible purposes (Squire, 2004). To provide a framework for the research design and interpretation of results, what follows is the discussion of various conceptualisations of the DP memory systems and their implications for L2 acquisition.

2.2.1 Declarative Memory.

Functionally, the declarative memory system is "the archetype of mainstream notions of *memory*" (Carpenter, 2008, p. 3). It subserves the encoding, storage and retrieval of semantic (facts) and episodic (events) knowledge (Lum et al., 2010; Squire, 1992; 2004). Subserving knowledge that is amenable to conscious reflection, the declarative memory system is also referred to as explicit memory. The memory system encodes information through creating associations between a representation and the related prior knowledge (Eichenbaum & Cohen, 2001). Frequent exposure to a representation leads to increased synaptic activation which in turn results in stronger representation and faster and more accurate recall (Lovett et al., 1999; Cowan, 1999; Eichenbaum, 2002; Engle, Kane, & Tuholski, 1999, all cited in Carpenter, 2008). Neurological research suggests that "the ability to make associations is mediated by the hippocampus" (Carpenter, 2008, p. 3). Because the declarative memory system has the ability "to detect and encode what is unique about a single event" (Squire, 2004, p. 174), the memory system is responsible for very rapid learning.

The declarative memory system is the storehouse for both short-term and long-term representations in which the initial storage relies on medial temporal lobe structures while long-term storage is subserved by neocortical areas (Foerde et al., 2006a; Lum & Kidd, 2012; Manns & Squire, 2001; P. J. Reber et al., 1996). The process to consolidate or establish long-term representations from short-term stores may take months or even years (Carpenter, 2008). The declarative memory system may have a limitless storage capacity for representations (Carpenter, 2008). The retrieval of the stored representations is dependent on the association of the representations: recalling a fact or an event activates the associated information, making it more easily accessible (Cowan, 1999 cited in Carpenter, 2008; Eichenbaum, 2002). "Retrieval from the declarative system, putatively a function of working memory, is thought to become more rapid and less effortful with practice" (Carpenter, 2008, p. 4).

The primary neural correlates of the declarative system are the hippocampus and parahippocampal area in the medial temporal lobe as well as in neocortical regions (Eichenbaum & Cohen, 2001; Squire, Stark, & Clark, 2004). These correlates are also responsible for the changes over time that characterise the declarative memory system (Eichenbaum & Cohen, 2001; Eichenbaum, 2002; Fernández et al., 1999; Sagarra & Herschensohn, 2010; Schacter & Wagner, 1999). Initially, memory is mediated by the parahippocampus. In the intermediate and final stages, memory is encoded and sustained through a parahippocampal buffer and the hippocampus, respectively. In normally developing individuals, IDs in declarative memory may arise from the activation levels of neural substrates, lateralization (bilateral versus dominant in one hemisphere), hormonal and gender differences (see Carpenter, 2008, for a brief discussion on this).

2.2.2 Procedural memory

The following provides a working definition of, and an overview on encoding, storage, retrieval processes, neural substrates, the process of change over time, and IDs in, the procedural memory system.

The procedural memory system is defined differently in the fields of cognitive neuroscience, psychology, and applied linguistics (Table 2.5). In this study, the procedural memory system refers to a habit or skill system specializing in sequencing and serial processing such as playing the piano, riding a bicycle, skilled game playing, swimming, and driving a car, and not to other implicit memory systems such as those subserving conditioning, priming, and reflexive learning (Squire & Knowlton, 1994; Ullman, 2001a; 2004). The procedural memory ability is often referred to as implicit memory because neither the learning (encoding) nor the remembering (retrieval) of procedures appear to be accessible to conscious memory (Lum et al., 2010; Manns & Squire, 2001; P. J. Reber et al., 1996; Ullman, 2001a; 2001b)

Table 2.5. Definitions of the Procedural Memory System across Research Paradigms

Anderson (1993), Cognitive Psychology	The ACT [i.e., Adaptive Control of Thought] model series posits any manipulation of declarative units is handled by a procedural system. The system functions through goal-driven behaviour. Production templates for procedures are strengthened with practice. The ACT procedural system handles a broad range of functions. It is not tied to specified neural correlates.
Flanagan et al. (2000), Psychometrics	“Procedural knowledge refers to the process of reasoning with previously learned procedures in order to transform knowledge” (p. 30). This definition encompasses a wide range of mental operations. It is used in psychometric paradigms involving intelligence testing.
M. Paradis (2004) Psychology	Any implicit learning is characterized as “procedural,” including lexical learning.
Skehan (1998), SLA	An L2 process has become “proceduralised” when the learner has achieved automaticity in using it (pp. 60-61). With practice, learners use this non-conscious processing in comprehending or producing the L2. Proceduralisation is not tied to neural systems.
Squire and Knowlton (2000), Neuroscience	Many non-explicit memory systems are often termed procedural, notably: skill and habit learning (striatal), priming (neocortex), basic associative learning (amygdala or cerebellar), and non-associative learning (reflex pathways). However, procedural learning is used to refer solely to the skill and habit learning system.

Adapted from Carpenter (2008, pp. 7-8)

Procedural representations (knowledge) consist of procedures or routines (Conway & Christiansen, 2001; Eichenbaum, 2002; Ullman, 2004) and, relative to declarative memory, these routines are gradually encoded, with practice and without the benefit of feedback (Carpenter, 2008; Lum et al., 2010). Once a skill is learned, it applies automatically, rapidly, and reliably because the response is triggered by the stimulus rather than being under any conscious control. Further, procedural routines tend to operate not only in parallel without interfering with themselves or with explicit processing but also without putting a heavy load on central processing capacity (K. E. Stanovich, 2011). Thus, procedural knowledge is unconsciously retrieved while performing the routine (Tulving, 2000). Because the knowledge is unconsciously accessible, learners cannot accurately report or consistently replicate patterns learned (Carpenter, 2008). The consolidation of procedural procedures is mediated by the cerebellum and/or interactions between striatal and frontal lobe regions (see Carpenter, 2008). The neural correlates of the procedural memory system are in the basal ganglia, cerebellum, and frontal, parietal, and temporal lobes (Squire & Zola, 1996; Ullman, 2001b; 2004; see Carpenter, 2008 for a brief discussion of these neural correlates). While the initial procedural learning is described as *data-driven* and characterised by errors as learners are simply reacting to events as they happen, the latter stage is referred to as *knowledge-driven* and characterised by the emergence of pattern learning with better accuracy and faster response times (Howard et al. 2004, cited in Carpenter, 2008).

Whether IDs exist in the procedural memory system is still contentious. On one hand, there is research positing that IDs in the procedural memory system are minimal relative to IDs in the declarative memory system. For instance, in his proposal of the tri-process theory of mind, Stanovich (K. Stanovich, 2009; 2011) distinguishes the two types of the (conscious) rational mind (i.e., the ‘‘algorithmic mind” and the ‘‘reflective mind”) from the unconscious, implicit “autonomous mind”. IDs in the autonomous mind are few relative to IDs in the rational mind. The few IDs in the autonomous mind “largely reflect damage to cognitive modules that result in very discontinuous cognitive dysfunction such as autism or the agnosias and alexias” (K. Stanovich, 2009, p. 59). Similarly, Reber and his colleagues (A. S. Reber, 1989; 1992; A. S. Reber et al., 1991 also see P. J Reber et al., 1996, Lum & Kidd 2012, and Lum et al., 2010) argue that “implicit cognition” (i.e., procedural memory) is an ability without or with minimal meaningful IDs relative to IDs in explicit learning. In their experimental study investigating IDs in the implicit learning of an artificial grammar and the explicit learning of a series-completion problem-solving task, A. S. Reber et al. (1991) found substantial IDs between subjects on the explicit task but relatively small IDs in the implicit task. They further found that their participants' performance on the explicit task correlated strongly with intelligence quotient while performance on the implicit task did not.

On the other hand, research both in linguistic and non-linguist domains suggest that procedural memory is an ability with meaningful IDs. Some experimental studies that have employed behavioural measures (e.g.,Carpenter, 2008; Ettlinger et al., 2014; Granena, 2013b; Kaufman et al., 2010; Kidd, 2012; Morgan-Short et al., 2014; Tagarelli et al., 2016) provide considerable support for the fact that IDs in procedural memory are meaningful. Neurological research also provides substantial evidence suggesting IDs in the degree and distribution of neural activation of procedural structures or substrates.[1] The research has shown that "[neural] activation in successful procedural learners, relative to unsuccessful procedural learners, may increase, decrease, or be redistributed" (Carpenter, 2008, p. 21). The following presents three proposed DP models: Ullman's DP model (e.g., Morgan-Short & Ullman, 2012; Ullman, 2001a; 2001b; 2004; 2015; 2016; Ullman et al., 1997), Paradis' DP model (e.g., 1994; 2004; 2009), and Reber's Implicit Learning Theory (e.g., A. S. Reber, 1989; 1992; A. S. Reber et al., 1991). The claims they make about the two memory systems have different implications for L2 acquisition.

2.2.3 Declarative and Procedural (DP) Models and L2 Acquisition.

2.2.3.1 Ullman's DP model.

Ullman's DP model (e.g., Morgan-Short & Ullman, 2012; Ullman, 2001a; 2001b; 2015; 2016) makes two important claims regarding L2 acquisition. First, the model posits that at the initial stage of L2 acquisition, the learning, representation and processing of arbitrary language (i.e., the mental lexicon) and rule-based language (i.e., grammar) are subserved by the declarative memory system and that with less experience (practice), increasing age of exposure and strong declarative memory abilities, the grammar rule

---

[1] Carpenter (2008) provides a brief discussion on this research.

learning dependency on declarative memory may persist for some time and possibly forever. In this model, the mental lexicon refers to the conceptual meanings of words, their phonological forms and grammatical specifications (e.g., irregular morphology and argument structure) and grammatically complex structures which are memorised as chunks. On the other hand, rule-based grammar refers to all rule-computed structures, across grammatical sub-domains, including phonology, morphology, and syntax. For L1, the declarative memory system is posited to underlie the learning, representation and processing of the mental lexicon while the procedural memory system underlies rule-based grammar. This first claim of Ullman's DP model suggests that only declarative memory involvement will be evident in early stages of L2 acquisition while evidence for procedural memory will be found only in data that isolate rule-governed grammar at later stages of L2 acquisition (Carpenter, 2008).

The model also proposes that the declarative/procedural memory distinction and the explicit/implicit knowledge distinction are not a one-to-one relationship. Although the information learnt through procedural memory is implicit, knowledge learnt and stored in the declarative memory need not be explicit. The declarative memory system may also underlie implicit learning and knowledge. This is based on the redundancy hypothesis which proposes that the two memory systems play at least partly redundant roles such that "humans can learn sequences, rules, and categories implicitly in procedural memory or explicitly (and perhaps also implicitly) in declarative memory" (Ullman, 2016, p. 957). This claim suggests that the declarative memory system involvement will be evident not only in the appropriation of declarative knowledge but also in the representation of procedural language knowledge.

There is considerable empirical support for Ullman's DP model. Carpenter (2008) points out that the claims of the DP model were first articulated in L1 research in patient populations. Ullman et al.'s (1997) study provides initial evidence for the DP model; it found a double dissociation where impairment to the procedural memory system (basal ganglia) as the result of Huntington's and Parkinson's diseases led to the impairment in rule-computed (regular) morphosyntax, while the impairment of declarative memory (hippocampus) as the result of Alzheimer's disease resulted in the impairment of memorised (irregular) morphosyntax. Since, neurological and behavioural research have provided further evidence for the DP model in L1 and L2 acquisition. For overviews of this research see Carpenter (2008) and Ullman (2004). In relatively recent times, empirical research (e.g., Hamrick, 2015; Kidd & Kirjavainen, 2011; Lum & Kidd, 2012; Morgan-Short et al., 2010; Morgan-Short et al., 2012; Morgan-Short et al., 2014, discussed in section 2.6) provides further support for Ullman's DP model claims for L2 acquisition.

2.2.3.2 Paradis's DP model

M. Paradis (1994; 2004; 2009) proposes a similar model to Ullman's. As in Ullman's DP model, the declarative memory system sustains L2 vocabulary and early grammar (i.e., grammar rules not yet internalised). In L1, vocabulary and grammar are subserved by declarative memory and procedural memory, respectively. However, Paradis' DP model makes two important claims. The core argument of Paradis's (2009) DP model is that "items that are sustained by procedural memory for language are subserved by neural substrates different from those that are sustained by declarative memory" (p. 13). In Paradis's DP perspective, unlike Ullman's DP model claim, all grammatical aspects, including the mental lexicon's phonological forms and grammatical specifications (e.g., argument structure) and grammatically complex structures are subserved by the procedural memory system. Paradis's DP model makes a distinction between vocabulary and lexicon. Vocabulary refers to sound-meaning correspondences of words whose appropriation and representation is subserved by the declarative memory system. The lexicon consists of grammatical features such as morphophonological rules and argument structure, which form part of the speaker's implicit competence (i.e., grammar) and, like the rest of the grammar, they are sustained by the procedural memory system. This claim implies that the explicit-implicit, declarative-procedural distinction is a one-to-one relationship and so is the explicit-implicit knowledge, declarative-procedural memory distinction (i.e., all language knowledge appropriated and represented in declarative memory is explicit while all language knowledge learned and represented in procedural memory is implicit). Consequently, regardless of how substantial the (in)direct influence may be, no interface exists between explicit metalinguistic knowledge and implicit linguistic knowledge (Paradis, 2009).

Another claim of Paradis' DP model is that learning condition is differential where "[incidental] acquisition through practice is the only way to internalize implicit linguistic competence" (Paradis, 2009, p. 7). For L1, all grammatical structures are learned incidentally and sustained by procedural memory. For explicitly instructed L2, Paradis (2009) argues, there is not sufficient evidence that learners actually do acquire (part of) the L2 grammar. Citing Roehr (2008) study that reports of a strong positive relationship between L2 proficiency and L2 metalinguistic knowledge, Paradis asserts that the speeded-up explicit metalinguistic knowledge accounts for the advanced learners' proficient use of L2 structures and vocabulary. This second claim therefore suggests that the knowledge of explicitly instructed grammatical features will be declarative and sustained by the declarative memory system.

2.2.3.3 Reber's Implicit Learning Theory.

A. S. Reber's (1989; 1992; 1993; A. S. Reber et al., 1991) implicit learning theory is rooted in research in cognitive psychology independently investigating implicit learning, implicit memory, and declarative and procedural knowledge. Like Ullman's and Paradis's DP models, A. S. Reber's implicit learning theory (forthwith Reber's implicit learning theory), proposes that declarative memory (or explicit cognition) and procedural memory (or implicit cognition) sustain language learning. The theory's core argument is that implicit cognition is automatically associative and evolutionarily older (by a considerable amount of time) than explicit cognition which is thought to have developed later only with the rise of Homo sapiens (Kaufman et al., 2010; A. S. Reber et al., 1991). Evolutionarily older forms and structures exhibit features that distinguish them from the forms that develop later. Further, the model assumes that unconscious, implicit processes are unaffected by the late arrival of explicit cognition and that these two sets of consciousness continue to function independently. Therefore, Reber's theory takes the non-interface approach to the distinction between implicit learning and explicit learning, on the one hand, and implicit (procedural) knowledge and explicit (or declarative) knowledge, on the other hand. Thus, as is the case with Paradis's DP model, the explicit-implicit memory and declarative-procedural knowledge distinction is one-to-one (with no interface) in Reber's theory of implicit memory.

Reber's implicit learning theory also claims that unconscious, implicit learning is not an ability. Since unconscious, implicit learning and memory are evolutionarily older cognitive processes, they are (in addition to being resistant to disruption of function by diseases and disorders and being insensitive to age and IQ [intelligence quotient] effects) stable with very few successful individual-to-individual differences relative to those in the declarative memory system. Thus, only declarative memory is an ability with meaningful IDs. This claim is consistent with Krashen’s (1981) point of view regarding L2 acquisition and learning contexts. Krashen proposed “that aptitude should only predict learner success when emphasis is placed on formal accuracy and metalinguistic explanations, which may promote more explicit learning processes and the development of explicit knowledge” (Tagarelli et al., 2015, p. 225). A contrary view was proposed by Skehan (2002) who suggested that aptitude is less important in controlled and structured classroom environments, but it is more important in naturalistic, informal and more demanding, environments as learners must rely on their individual cognitive capacities (see DeKeyser, 2000; Harley & Hart, 1997; 2002; Ross, Yoshinaga, & Sasaki, 2002)

Finally, Reber claims that implicit learning and memory are more effective when learning complex, rule-governed knowledge (Kaufman et al., 2010; A. S. Reber et al., 1991). This claim suggests that the procedural memory system will be more effective in acquiring complex grammatical structures. This stands in stark contrast with the semantically driven cognitive grammar proposition that “discovering and instructing the conceptualizations of the form ... facilitate L2 acquisition, particularly in instances when the L2 differs from the L1’ (Dirven, 1989, cited in Akakura, 2009, p. 11).

There is some empirical support for Reber’s implicit learning theory. A. S. Reber et al. (1991) provide the initial empirical support for Reber’s implicit learning theory. The study investigated the extent to which individual differences in performance were observed in a group of 20 undergraduate subjects. They used two different tasks: an implicit task, which was a standard artificial grammar learning task, and an explicit task, which was a series completion problem-solving task. A. S. Reber and his colleagues found minimal individual differences in subjects’ performance on the implicit learning task relative to the task involving explicit learning. Robinson’s (2005b) study confirmed these results. Reber et al. also found that the participants’ performance on the explicit task correlated strongly with intelligence quotient (IQ) while the performance on the implicit task did not. Similar results are reported in Gebauer and Mackintosh (2007) where learners’ performance on implicit learning measures did not correlate with measures of intelligence only when instruction was implicit.

However, in recent years, some research reports that procedural memory is an ability with meaningful IDs. For instance, Granena (2016), in a study in cognitive psychology, investigated whether aptitudes for implicit and explicit learning are differentially related to the two-main information-processing cognitive styles proposed by the dual-process theories in cognitive psychology, namely rational–analytical and experiential–intuitive. Rational-analytical information processing is assumed to be subserved by explicit aptitude whereas experiential-intuitive information processing is subserved by implicit cognitive processes (Granena, 2016). Aptitudes were measured by Llama subtests and SRT; information processing styles were measured by the Rational-Experiential Inventory (REI), a self-reported ability measure of the two independent processing modes. The results showed a relationship between a rational–analytical profile and explicit aptitude, as well as between an experiential–intuitive profile and implicit aptitude. Based on these results, Granena (2016) concluded that not only are implicit and explicit (learning or processing) systems dissociable but they are also abilities with meaningful IDs. Similar results are reported in studies such as Kaufman et al. (2010), Granena (2013b) and Robinson (2002a; 2005b). See section 2.6 for a discussion this research.

Since A. S. Reber et al.'s (1991) study, cognitive psychology and L2 acquisition research has seen a surge in the number of studies investigating the association of IDs in declarative and procedural memory abilities. It however remains an open question whether implicit cognition is an ability with or without meaningful IDs relative to IDs in the explicit memory system. Further, though the research has attempted to address Reber's claim that implicit learning is not an ability, it has largely ignored Reber's claim that a complex stimulus domain is amenable to implicit learning. Under the approach that assumes explicit and implicit learning as dissociable, processes subserved by implicit memory are assumed to be stored or represented only in implicit memory system. Hence, Reber's claim suggests not only that the representation of complex rules is subserved by implicit memory but also that complex rules are accessible only as implicit knowledge.

In summary, the present study aims to understand how these models and their associated predictions explain the role of the two long-term memory systems in the acquisition of L2 English grammar rules of varying complexity in instructed context. The preceding review of literature links explicit learning and knowledge to the conscious, declarative memory system and implicit learning and knowledge to the unconscious, implicit memory system. Reliable measures of declarative and procedural memory are required to meaningfully address the relationship between the two long-term memory systems and L2 acquisition. The following section presents the common tests for measuring the memory systems and some related methodological considerations.

2.2.4 Measures of memory systems.

The operationalisation of the two memory constructs is based on how information retrieval processes proceed in either long-term memory system. The tasks measuring the two memory systems are operationally distinguished by manipulation of the instructions given to participants in a way that affects the retrieval of information (Hulstijn, 2005). For the tasks designed to assess the declarative memory system, participants are explicitly asked to recall past events or to recognise previously studied events. Instructions for tasks designed to measure the procedural memory system simply ask participants to perform the task as accurately and quickly as possible with no reference to past events. Another important consideration regarding the testing of the two long-term memory systems concerns the extent to which the memory systems are domain specific (to language) or domain general (Carpenter, 2008). Described below are several widely used behavioural tests of the declarative and procedural memory systems, some are considered linguistic (i.e., verbal) measures and others considered non-linguistic (i.e., non-verbal) measures.

2.2.4.1 Declarative memory measures.

There are several behavioural tests that have been used to measure the declarative memory system. However, the MLAT5 Paired Associates (Memory) and Llama-B tests are probably the most widely used measures. These two tests are said to be similar because Llama-B is loosely based on the original vocabulary learning subtask of Carroll and Sapon’s (1959) MLAT5 Paired Associates (Meara, 2005).

2.2.4.1.1 MLAT5 Paired Associates (Memory).

Developed in the 1950s by Carroll and Sapon (Carrol & Sapon, 1959), the Modern Language Aptitude Test (MLAT) is perhaps the best-known test of language learning aptitude in L2 acquisition research. It is a paper-and-pencil test battery that consists of five subtests. One of the subtests is the MLAT5 Paired-Associates which involves word associations. It is the memory test of the ability to remember 24 Kurdish/English word pairs which are memorised within a 4-minute timeframe (Dornyei, 2005). Testing (which is untimed) involves presentation of a Kurdish word and five English alternatives from which test-takers must choose the proper equivalent. All the five alteratives are selected from the 24 words contained in the original list (Dornyei, 2005). Therefore, the test measures the learners' ability to form links in memory based on explicit knowledge. Investigating IDs in learners' language learning aptitudes in implicit, incidental, rule-search and instructed learning conditions, Robinson (1997) found that MLAT5 Paired-Associates significantly correlated with learners' performance in rule-search and instructed conditions, conditions that enhance the adoption of conscious learning strategies. In Carpenter (2008), the MLAT5 Paired-Associates obtained marginally significant to significant moderate correlations with other declarative and working memory measures and with learners' language knowledge scores at low proficiency when grammar learning in L2 is predicted to rely on declarative memory.

2.2.4.1.2 Llama-B.

The Llama-B language aptitude test is one of the four Llama sub-tests, namely Llama-B (a vocabulary learning test), Llama-D (a sound recognition test), Llama-E (a sound-symbol correspondence test) and Llama-F (a grammatical inferencing test, Meara, 2005). It is a recall and recognition verbal test. As described in the Llama manual (Meara, 2005), Llama-B is a simple vocabulary learning task which measures a learner's ability to learn relatively large amounts of vocabulary in a relatively short space of time. Using picture stimuli, learners are asked to learn real words taken from a Central American language which are arbitrarily assigned to the target images. After a training phase involving memorising object-word pairings, test-takers are asked to identify a correct image on a computer screen when presented with a word. Because

Llama tests do not require any L1 input, Llama-B is independent of the languages spoken by test-takers, and therefore suitable for use with learners of any L1.

In her exploratory validation study on the reliability of the underlying structure of Llama sub-tests, Granena (2013a) reports an acceptable Cronbach's alpha coefficient of .76 for Llama-B. Further, Granena notes that Llama-B, Llama-E and Llama-F loaded together with general intelligence (*g*), a construct that is biased towards attention-driven explicit processes. In an experimental study with an artificial language, Hamrick (2015) found that Llama-B correlated with learners' explicit knowledge of syntax. He interpreted the results as suggesting that declarative memory ability as measured by Llama mediated the learning of the syntax. Other studies that have used Llama include Granena (2016), Granena & Long (2013), and Abrahamsson & Hyltenstam (2008). In these studies, Llama-B, Llama-E and Llama-F, the Llama sub-tests that are considered measures of analysed, explicit learning ability, were found to correlate with the measures of language knowledge that require an ability to think analytically. These results provide strong evidence that the Llama-B task indexes the declarative learning ability.

2.2.4.1.3 Continuous Visual Memory Task.

The Continuous Visual Memory Task (CVMT, Trahan & Larrabee, 1988) is probably the most widely used non-linguistic measure of the declarative memory system. As a test of visual recognition declarative memory, the CVMT is designed "to minimise reliance on the verbal strategies or knowledge" (Morgan-Short et al., 2014, p. 60). During training, test-takers view a series of complex, abstract designs at the centre of a computer screen and then they are asked to indicate whether each complex abstract design was novel ("new") or had appeared previously ("old"). Some versions of the CVMT require participants to respond orally while others require responses by pressing a response key on a computer keyboard.

The CVMT has been used quite a lot in clinical studies. See Strong and Donders (2008) for the discussion of the validity of the CVMT with participants with traumatic brain injury; Vasterling et al. (2002) on attention, learning, and memory performances; and Trahan et al. (1990) on the CVMT visual recognition memory in normal adults and patients with unilateral vascular lesions. Investigating how individual differences in cognitive abilities account for variance in the attainment level of adult L2 syntactic development, Morgan-Short et al. (2014) found that the CVMT was related to the MLAT-V, another declarative learning ability measure while they were no such relationships with procedural learning ability measures namely the Tower of London task (TOL, cognitive skill learning) and the Weather Prediction Task (WPT, probabilistic). Carpenter (2008) cites (i) clinical studies linking the CVMT to a wide range of impairments involving declarative memory, such as Alzheimer’s disease, and (ii) neuroimaging studies that link the CVMT performance to medial temporal lobe structures and declarative memory. In Carpenter (2008), the CVMT scores (visual learning) correlated moderately with total scores on the California Verbal Learning Test-II (CVLT-II), the measure that indexes working memory, verbal learning, verbal memory, semantic clustering, and serial clustering and was used to assess verbal learning subserved by declarative memory. Further, Carpenter found that the CVMT scores were highly related to scores on the MLAT5 Paired Associates test, another declarative learning measure.

Other measures of declarative memory include linguistic or verbal measures such as the Cognitive Ability for Novelty in Acquisition of Language (CANAL-F) developed by Grigorenko, Sternberg, and Ehrman (2000) which includes testing the ability to cope with novelty in learning; the High-Level Language Aptitude Battery (Hi-LAB, Linck et al., 2013), a test for advanced proficiency; the Three-Term Contingency Learning test (B. A. Williams & Pearlberg, 2006); and the verbal reasoning section of the Differential Aptitudes Test (DAT-V, Corporation, 1995).

2.2.4.2 Procedural memory measures.

The skills associated with procedural memory have been commonly assessed using the following: (the dual-task version of) the Weather Prediction Task (WPT, Foerde, Knowlton, & Poldrack, 2006; Knowlton, Squire, & Gluck, 1994) and the Serial Reaction Time test (SRT, e.g., Nissen & Bullemer, 1987) for probabilistic learning, and the Artificial Grammar Learning (AGL, e.g., A. S. Reber, 1967; 1989) and the Tower of London task (Kaller, Unterrainer, & Stahl, 2012; Kaller, Rahm, Köstering, & Unterrainer, 2011; Ouellet, Beauchamp, Owen, & Doyon, 2004; Unterrainer, Rahm, Leonhart, Ruff, & Halsband, 2003) for cognitive skill acquisition. Several studies have also used Llama-D to index procedural memory.

2.2.4.2.1 The WPT and TOL

The WPT assesses the test-takers’ knowledge of probabilistic weather outcomes (i.e., the probability of sunshine or rain) based on combinations of cue cards. For instance, a card with circles combined with a card with squares may be associated with an 80% chance of sunshine (Buffington & Morgan-Short, 2018). To reduce participants’ reliance on the declarative memory in the dual-task version of the test, a secondary task such as asking participants to keep track of the number of high tones is introduced. Learning is assessed as accuracy of weather prediction. In the TOL task, participants are asked to match a goal configuration of coloured circles resting on pegs. At the beginning of the task, participants are instructed to plan the sequence of moves. They are also asked to complete the goal configuration in the stated number of moves in each trial. Generally, the measure of learning is the average reaction time to match goal configuration in trials.

2.2.4.2.2 Artificial Grammar Learning (AGL)

Probably, the most widely used tests of procedural memory are the AGL and the SRT. The AGL test assesses participants' ability to judge the grammaticality of letter sequences such as VXXVS and TPPPTS that are generated by an artificial finite-state grammar (e.g., Bowles, 2003; Gebauer & Mackintosh, 2007; Pothos, 2005; 2007; 2010; Poletiek, 2002; Reber,1967; Schiff & Katan, 2014; Van den Bos & Poletiek 2015; Ziori & Pothos, 2015). During training, participants are instructed to memorise the letter sequences determined by an artificial grammar. They are also told about a later memory test. During testing, participants are asked to perform a GJT on old and new grammatical and ungrammatical letter strings. Correctly judged test sequences of letters constitute the measure of learning.

2.2.4.2.3 Serial Reaction Time (SRT)

The SRT task (Nissen & Bullemer, 1987) is a visual and non-linguistic measure of the procedural memory system. In the SRT, a visual stimulus such as an asterisk (*) or a red dot repeatedly appears in one of four designated spatial locations on a computer screen. Test-takers’ only task is to press as quickly and accurately as possible the corresponding button on a response keyboard. Unbeknownst to the participants, on some blocks the stimulus follows a sequence and on other blocks the sequence is violated. Testing is based on reaction times (RTs) based on the assumption that when implicit learning develops, RTs will decrease as participants respond to the repeating sequences. In contrast, RTs will increase when participants are presented with a block where the visual stimulus appears randomly.

The SRT and AGL tests have become the “paradigmatic methods of studying implicit learning” (Shanks 2005, cited in Kaufman et al., 2010, p. 325). The following two considerations suggest that the SRT is a better measure of implicit learning (Kaufman et al., 2010; also see Pretz, Totz, & Kaufman, 2010; 2014; Urry, Burns, & Baetu, 2015). First, sequence learning in the SRT test is more incidental than in the AGL. While in AGL participants are explicitly told to memorise letter strings, learning in the SRT task is an incidental result of responding to visual stimuli with no any instructions to memorise the sequences or look for underlying rules. Second, unlike in SRT, there is a clear separation between the acquisition and test phases in AGL where “participants are typically informed about the existence of a structure and are told to try to exploit it” in the test phase (Kaufman et al., 2010, p. 325).

There are two SRT versions. In contrast to the original deterministic SRT version (Nissen & Bullemer, 1987) where every stimulus follows the predetermined pattern in all blocks except one, the probabilistic version of the SRT (Schvaneveldt & Gomez, 1998) has control trials interspersed with sequence trials in every block. The interspersing structured with control trials does not only allow implicit learning to be measured online (i.e., during the training phase) but it also makes it

> more difficult for participants to explicitly discover the existence of a sequence and of making the task more ecologically valid: implicit learning in the real world often happens under conditions of uncertainty, where information to be learned is noisy and probabilistic instead of deterministic (Kaufman et al., 2010, p. 325).

This makes the probabilistic version of SRT the best measure of implicit learning currently available. The probabilistic version of SRT is like the Alternating Serial Reaction Time (ASRT, Howard Jr & Howard, 1997) where patterned trials alternate with random trials. Research has demonstrated that sequence learning tasks such as the SRT engage procedural memory neural circuits (e.g., Fletcher et al., 2005; Rauch et al., 1997; both cited in Buffington & Morgan-Short, 2018). Further, investigating the construct validity of procedural memory tests, Buffington and Morgan-Short (2018) found that the ASRT as well as the WPT and TOL showed some discriminant validity from declarative and working memory. But like Gebauer and Mackintosh (2007), Buffington & Morgan-Short (2018) found no evidence for convergent validity, suggesting that procedural learning may not be a unitary ability.

2.2.4.2.4 Llama-D.

Lastly, a few studies have used Llama-D as a measure of the procedural memory system. As pointed out above, Llama-D is a subtest of Llama. It is a verbal and auditory task which measures the ability to recognise short pieces of previously heard sound sequences in new sequences (Granena, 2013a; 2013b; Meara, 2005; Yalçın & Spada, 2016). It is the only subtest of Llama that does not have a default study phase (Granena, 2013a). The lack of default study phase in this subtest creates an incidental learning condition that minimises the adoption of explicit strategies (Granena, 2013b). Participants listen carefully to the string of 10 sound sequences which is played once on a computer. Soon after, a recognition test is presented, and participants must discriminate between old and novel sound sequences.

In an exploratory validation study, Granena (2013a) found that Llama-D correlated weakly with Llama-B, Llama-E and Llama-F. Further, Llama-D and SRT loaded more strongly on one component, suggesting that these two tests measure the same underlying factor. Furthermore, in her study investigating the association of IDs in language aptitudes for implicit learning and for explicit learning with two cognitive styles for information processing, rational–analytical and experiential–intuitive, Granena (2016) found that scores on Llama-B, Llama-E, and Llama-F significantly correlated with the rational–analytical style, a skill that relies on explicit language learning ability and rote learning ability. There were no significant correlations between rational ability/engagement and performance on Llama-D. Granena (2016) interpreted these results as suggesting that Llama-D does not seem to draw on participants' analytical problem-solving abilities. Interestingly in the same study, while the experiential-intuitive processing modality was significantly related to implicit language aptitude as measured by SRT, it was not significantly related to participants' performance on Llama-D. Finally, Granena (2013b) used the Llama-D and SRT tests to investigate the relationship between sequence (implicit or procedural) learning ability and morphosyntactic L2 attainment by early and late learners. The results of this showed that Llama-D and SRT behaved differently: while Llama-D was related to L2 attainment in the early group, the SRT was related to L2 attainment in the late group. Granena points out that these results could suggest that different cognitive tasks are differentially sensitive to L2 outcomes.

Generally, all the tests described in the preceding discussion are widely considered to be valid and reliable measures of declarative and procedural memory. However, no single test can provide a perfect measure of the construct that it purportedly represents. As results from Granena's studies (e.g., 2013a; 2013b; 2016) suggest, the Llama-D and SRT tests do not seem to measure the exact same underlying learning skills despite the fact that these tasks have consistently shown a lack of significant correlations with measures that tap into learners' analytical or explicit processes. Therefore, administering multiple linguistic and non-linguistic tasks to measure the same construct and then form a latent variable should be a rigorous research method to obtain a comprehensive measure of the construct.

The present study has included various tasks, linguistic and non-linguistic, to measure each of the two long-term memory systems. The Llama-B, CVMT, Three Term Contingency Learning and DAT-V tests are used as measures of declarative memory. The Llama-B test and CVMT (the only non-linguistic measures in this set) are widely used measures and are also language independent. These two tests were used in Kaufman et al. (2010) as declarative memory measures. Further, in Kaufman et al.'s (2010) study, the DAT-V correlated with the SRT test. The probabilistic version of SRT (non-linguistic) and the Llama-D (linguistic) test are used as measures of the procedural memory systems. As pointed out above, the probabilistic version of SRT is probably the best measure of procedural memory presently. The Llama-D is included because it is a verbal measure that has been widely used in L2 acquisition research (e.g., Granena 2013a; 2013b; 2016). Further, it is language independent and easy to use like all the other Llama subtests. It is hoped that such triangulation of the memory systems measures should shed more light regarding the (differences in) underlying components that each test is designed to tap into.

However, the understanding of the existence of other influences on the memory systems may have far-reaching implications on the interpretation of the results from these memory systems measures. Two external factors, linguistic complexity and exposure condition have particularly been implicated in modulating the declarative and procedural memory systems. Defining the construct of linguistic complexity (also known as linguistic difficulty or grammar rule complexity) has been contentious in L2 acquisition research. The following section reviews this research.

2.3 Grammar Rule Complexity

The assumed differential nature of the learning conditions and the learning of complex and simple rules leads to another daunting issue in second language acquisition research: the question of what makes a grammar rule simple or complex. The term "complexity" does not have a commonly accepted definition despite its important position in contemporary science (Bulté & Housen, 2012). At the most basic level, complexity is defined as the "property or quality of a phenomenon or entity in terms of (1) the number and the nature of the discrete components that the entity consists of, and (2) the number and the nature of the relationships between the constituent components" (Bulté & Housen, 2012, p. 23).

The notion of complexity in L2 acquisition has not been explicitly addressed. Though the relevance of the distinction between levels of complexity of the rules is acknowledged by many researchers in L2 acquisition, no clear criteria for distinguishing the complexity of L2 rules have emerged (Robinson, 1996). In recent times, some studies (e.g., DeKeyser, 2005; 2016; R. Ellis, 2006; 2009c; Goldschneider & DeKeyser, 2001; Housen & Simoens, 2016; Spada and Tomita, 2010) have tempted to address the issue of complexity or the related construct "learning difficulty" of grammar rules of English. However, the determination of complex and simple rules or what makes some rules difficult to learn in L2 acquisition remains unclear. Spada and Tomita (2010, see also Housen & Simoens, 2016) observe that because of the lack of consensus in how grammar rule complexity is conceptualised, L2 acquisition research has been characterised by inconsistencies in the way the terms *simple* and *complex* are operationalised. This lack of consensus originates from the fact that complexity is viewed from different perspectives. Collins et al. (2009) identify and detail four broad approaches to the issue, namely (a) a focus on learner behaviour, the acquisition perspective; (b) a focus on language characteristics, the linguistic perspective; (c) a focus on teacher explanations on rules, the pedagogical perspective; and (d) a focus on the interaction between learners and language input, the psycholinguistic perspective. Spada and Tomita (2010) on the other hand, identify three different perspectives of complexity, namely psycholinguistic, linguistic, and pedagogical.

2.3.1 The Psycholinguistic Perspectives.

Complexity within the psycholinguistic perspective is conceptualised in terms of whether a feature is acquired early or late or is more or less difficult to process. The predictable developmental stages that L2 learners go through (see, Pienemann, 1989; 2005; 2015, in Chapter 3 below, or see Lightbown 1980, Ravem 1973, Schumann 1979, Wode 1976, all cited in Spada and Tomita 2010) result from learners' psycholinguistic processing abilities. It is argued that learners cannot progress to the next developmental stage unless they are able to cognitively process the structures at earlier stages (Spada and Tomita, 2010). As Spada and Tomita (2010) point out, grammatical difficulty "arises when learners are expected to learn grammatical structures that they are not developmentally ready to learn" (p. 267). The problem with the psycholinguistic perspective of complexity rests in the fact that it defines complexity based on developmental orders of acquisition which leads to circularity; that is, the language feature is acquired late, therefore it must be complex (Spada and Tomita, 2010)**.**

2.3.2 Pedagogical Perspective.

In this perspective, complexity is described in terms of whether learners find a grammatical feature easy or difficult to understand and learn (Spada and Tomita, 2010). The pedagogical perspective places focus on learners' performance errors. A difficult grammatical feature is one that L2 learners fail to systematically use accurately in their production. It is mainly teachers that identify easy and difficult grammatical rules by observing the performance of their learners. The drawback with the pedagogical perspective of rule complexity is that it does not consider aspects of learners' IDs which may include factors such as a learner's aptitude or L1 background. As a result of IDs, learners are expected to behave differently in response to the same stimulus. Therefore, if a learner finds a specific rule difficult, it does not mean that is the case with other learners.

2.3.3 A Linguistic Perspective.

A linguistic perspective of complexity arises from the linguistic aspects of a grammatical feature such as the number of transformational rules to arrive at a target feature or whether a feature is marked or unmarked (Spada and Tomita, 2010). DeKeyser (2005) claims that at least three factors determine linguistic difficulty: complexity of form, complexity of meaning, and complexity of form-meaning mapping. Complexity of form arises when a grammatical feature has more than one realisation. An example is the use of the correct plural “s” morpheme in the correct place in English. Complexity of meaning may arise from the grammatical form’s abstractness. For instance, articles in English are considered too abstract for learners to infer from the input, and explicit instruction on article use is often not effective. Complexity of form-meaning mapping comes about (i) when a form is semantically unessential (e.g., third-person singular in English), (ii) when a form is non-mandatory (e.g., null subjects in Spanish and Italian), and (iii) when the correlation between form and meaning is low as with the “s” morpheme in English, which serves several functions (e.g., plural of noun, third-person singular).

Other determinants of linguistic difficulty include perceptibility, the learner’s L1 and the communicative value of a language form. A form which is salient in the input is noticed and easily learned. A meta-analysis investigating the factors determining the “natural” order of morpheme acquisition in English (Goldschneider & DeKeyser, 2001) supports this view. Regarding the communicative value of a language form, it is claimed that errors that interfere with meaning (e.g., incorrect use of possessive pronouns “his/her”) may be easier to correct than errors that do not interfere with meaning (e.g., the absence of inversion in questions such as “What she is reading”) and thus are more difficult to learn (Spada and Tomita, 2010).

The problem with the linguistic perspective, that defines linguistic complexity based on the number of transformational rules, is that what is easy to describe is not necessarily easy to learn. Spada and Tomita (2010) give an example of third-person singular “s” in English, which is relatively easy to describe but has proven to be generally difficult to acquire by L2 learners. However, Spada and Tomita (2010) adopted the same transformational rules criterion that relies on derivational rules to determine simple and complex rules. The transformational rules criterion is relatively straightforward to apply and categorise complexity across a wide range of different language features. Secondly, the conceptualisation of the criterion has a cognitive, linguistic, and pedagogical value.

However, the use of the transformational rules criterion in Spada and Tomita's (2010) study resulted in the following two categories of the rules: all those grammatical rules pertaining to functors such as the plural -*s*, articles and possessive determiners were considered simple rules; and all syntactic structures were considered complex rules. Evidence suggests that some functors such as articles, classifiers, and grammatical gender are strongly resistant to instructional treatments (DeKeyser, 2005). These functors express highly abstract notions that are extremely hard to infer, implicitly or explicitly, from the input (DeKeyser, 2005). Consequently, these functors are intensely hard to learn especially for native speakers of L1s that do not have them or that use a very different system; in some cases, the learning problem is serious and long-lasting (DeKeyser, 2005). Further, research in morphosyntax singles out morphology as hard to acquire in comparison with syntax, such that learners acquire the syntactic features easily but continue to have problems with their morphological instantiation (DeKeyser, 2005). Therefore, the assumption that grammatical structures that involve a lesser number of, or no, linguistic transformations are simple and those involving more linguistic transformational rules are complex is misleading.

The present study adopted the linguistic perspective of complexity to determine simple and complex grammatical structures. Literature on L2 acquisition difficulty makes a distinction between objective and subjective difficulty. Subjective difficulty, also referred to as intra-individual difficulty, is learner related. It arises as the result of the encounter of language features with the learner's individual abilities such as cognitive abilities, previous knowledge, overall L2 proficiency, stage of L2 development, and some socio-affective and personality factors (Housen & Simoens, 2016). Since, for various reasons, learners' learning capabilities differ, then a linguistic feature that is difficult for one learner might be not difficult for another learner. On the other hand, some language features are more cognitively demanding for all language learners, irrespective of their individual learner characteristics (Housen & Simoens, 2016). This is referred to as objective, or inter-individual, difficulty. It arises from the target L2 feature's inherent or intrinsic properties. Housen and Simoens (2016) refer to this kind of conceptualisation of objective difficulty as the *intrinsic structural complexity* of L2 features.

Further, two dimensions of objective difficulty are distinguished in L2 acquisition research, namely formal and functional dimensions (Housen & Simoens, 2016). The formal dimension concerns such factors as the structural “substance” of a linguistic element which includes (i) the number and nature of its constituent components (e.g., English *-ing* vs. *-s)*, (ii) the number of positional variants of a feature (e.g., *-ing* has no allomorphs, whereas *-s* has three), and (iii) the number of operations needed to derive a target form from a base form (e.g., forming passive clauses from underlying active structures). The functional (or conceptual or semantic) dimension concerns the number and nature of meanings and functions that linguistic features express. Therefore, the characterisation of objective difficulty in terms of formal and functional dimensions affords the understanding that linguistic complexity arises from either the form of a linguistic feature or the meaning(s) the linguistic feature expresses or its form-meaning mapping or any combination of these. The linguistic perspective of complexity that focuses only on form leaves out the most important aspect of language acquisition: the communicative aspect of language.

The present study focuses on how L2 learners’ long-term memory systems, exposure conditions and the level of objective difficulty of English grammatical structures modulate learning outcomes. In the absence of the comprehensive accounts in L2 acquisition research of the exact nature and the relative weight of the various factors that contribute to feature-related difficulty of both the morphology and syntax of English language, the present study makes an attempt to categorise the functors targeted in the study on one hand, and the syntactic structures, on the other hand, into either simple or complex grammar structures on the basis of both their formal and functional features. Drawing on the aspects of Pienemann’s processability theory (1989; 2005; 2015), the present study proposes a criterion for determining simple and complex grammar structures in English.

Some early research suggests that the complexity of linguistic structures interacts with the type of instruction. For instance, explicit learning may be effective when learning simple rules (e.g., S. Krashen, 1994), whereas patterned domain stimuli that are sufficiently complex are more likely to be acquired by an implicit learning system (e.g.,

A. S. Reber, 1989; 1993). Re-examining the question of what makes some grammatical structures more difficult to learn than others, R. Ellis (2006) found that structures that were easy in terms of implicit knowledge were often difficult in terms of explicit knowledge and vice versa for some other structures. Since empirical support for the role of linguistic complexity in L2 acquisition is linked to exposure condition, more of this research is therefore discussed in section 2.5 after a look at what exposure condition types are in the next section 2.4.

2.4 Explicit and Implicit Learning

The terms *implicit learning* and *explicit learning* were first employed by A. S. Reber (1967) in cognitive psychology (Rebuschat, 2009). In L2 acquisition research, however, the interest in the distinction between explicit and implicit learning is traced back to Krashen's (1982; 1985) proposal that L2 development relies on incidental processes that result in implicit linguistic knowledge and that the explicit processes have very little role to play (Andringa & Rebuschat, 2015). Hulstijn (2005) observes that learning is often defined in relation to the resultant knowledge. Hulstijn (2005, p. 131), therefore, defines explicit learning as "input processing with the conscious intention to find out whether the input information contains regularities and, if so, to work out the concepts and rules with which these regularities can be captured". Conversely, "implicit learning is input processing without such an intention, taking place unconsciously".

The distinctions explicit/implicit instruction and intentional/incidental learning are related to the explicit/implicit learning distinction. Explicit and implicit instruction in the pedagogical literature refer to the presence or absence of information about the rules underlying the input, respectively (Hulstijn, 2005). Learners may receive language input with rules explicitly explained either prior to, or after working on, various examples. This is a further dimension to explicit learning that involves a distinction between deductive and inductive learning. In deductive learning, the presentation of examples is preceded by the presentation of rules. Conversely, inductive learning results from working with examples to find the rules by consciously determining the regularities in the examples.

The incidental/intentional learning distinction loosely parallels the distinction between implicit/explicit learning. Intentional learning refers to the type of learning where participants are informed, prior to their engagement in a learning task, that they will be tested afterward on their retention of some information (Hulstijn, 2005). Informing participants that the task before them is meant for them to learn and that they will be tested at the end encourages conscious learning and retaining of the information in preparation for the test. On the other hand, incidental learning has been defined in various ways. With reference to experimental designs, incidental learning refers to the type of learning where subjects are not forewarned of an upcoming retention test for the learning targets (Bell, 2015; Hulstijn, 2005). Hulstijn (2003) further points out that in the psychological literature and in the incidental learning condition, the experiment may not even be explicitly presented as a "learning experiment", because the word "learning" itself may already lead to testing expectancies among participants and hence to subject generated information-processing strategies unwanted by the experimenter (p. 356).

In more general terms, incidental learning is defined variously as the unintentional picking up of information; the learning of one stimulus aspect while paying attention to another stimulus aspect; the learning of formal features through a focus of attention on semantic features; or the learning of grammatical structures without exposure to instances of these (Hulstijn, 2003; 2005). In all these definitions, a common aspect is that learners' attention is drawn to something else and not to what the investigator intends the learners to learn. One important question that arises is whether incidental and intentional learning are synonymous with implicit and explicit learning, respectively. Incidental learning is always implicated in implicit learning, but that incidental learning just picks out part of implicit learning, likewise intentional learning in relationship explicit learning (Hulstijn, 2003). Implicit learning happens when there is no conscious awareness, while incidental learning occurs by not focusing attention on what is being internalised as in learning a grammatical form while focusing on the meaning. Explicit learning involves awareness at the point of learning (e.g., by trying to understand what the function of a certain language form is), while intentional learning involves a deliberate attempt to commit new information to memory (e.g., by applying rehearsal and/or mnemonic techniques; Hulstijn, 2003). In the present study, explicit learning is conceived of broadly as learning as the result of conscious awareness, whether intentional, deductive or not. Implicit learning, on the other hand, is defined generally as the unconscious learning of grammatical features, whether incidental or not.

The distinction of explicit and implicit is central to almost every major theme central to L2 acquisition theory construction. It is relevant to the understanding of (i) language learning trajectories, (ii) child and adult L2 acquisition differences, (iii) the amenability of language features to various L2 instructional treatments and how it interacts with IDs, and (iv) how development is affected by language learning conditions (Andringa & Rebuschat, 2015). However, Andringa and Rebuschat (2015) point out that using the term *regularities* in defining explicit and implicit learning, as in the definition by Hulstijn's (2005) definitions presented above, limits the constructs to only regular language aspects. They therefore suggest that it is more accurate to say that both regular and irregular aspects can be learned explicitly and implicitly. "Learners can explicitly learn the exceptions to a rule, and they can learn to behave according to these exceptions without being consciously aware of them" (Andringa & Rebuschat, 2015, p. 187). This position deals, however, with the question of whether explicit knowledge interfaces implicit knowledge, a notion that is controversial in language acquisition research.

There is a further confusion regarding the definitions of the two learning conditions. As pointed out by Hulstijn (2005), researchers on the theory of explicit and implicit learning have tended to conflate explicit and implicit learning as a process i.e. as the *how*, or as a product i.e., as the *what.* As the process *how*, the two constructs refer to two different ways of input processing, that is, explicitly, consciously with awareness or implicitly, unconsciously, without awareness. Andringa and Rebuschat (2015) state that it is problematic to define the constructs in terms of the *how* especially because it is difficult to determine the exact input processes that are involved. A suggested way to address this problem, however, is to invoke theoretical underpinnings of the learning process in L2 acquisition which allows for the situating of the role of implicit and explicit learning along the stages of the language learning process (Leow, 2015). The language learning process postulated for L2 acquisition consists of several major stages (Figure 2.1). The take-off point is the L2 (Input) stage which involves exposure to language input. Then the next stage is the (Intake) stage at which point some of the input is picked up by the learner. What follows the (Intake) stage is the (Internal system) stage at which point some of the intake is internalised and becomes L2 knowledge in the internal system. The (Output) is the final stage where the internalised (or representative of the) L2 knowledge is produced.

INPUT	{ >	**INTAKE**	>	**INTERNAL SYSTEM**	>}	**OUTPUT**
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	
(Product)	(**process**)	(product)	(**process**)	(product)	(**process**)	(product)
(*input*)	(*input*)	(*intake*)	(*intake*)	(*L2 knowledge*)	(*L2-knowledge/ output*)	(*representative L2 knowledge*)

Figure 2.1. Stages of the learning process in SLA: Of processes and products. Adapted from Leow (2015, p. 49)

As illustrated above, the language learning process consists of two external products, Input and Output and five internal stages comprising three processes: input processing, intake processing, and output processing, and two resultant products: intake product and L2 knowledge. Thus, there is the process of converting input into intake, then converting intake into the internal system where some type of knowledge, implicit or explicit, is assumed and there is also the process of producing output (Leow, 2015). Ideally, explicit or implicit learning may occur at any of processing Stages 1, 3 or 5. As opposed to offline measures, concurrent data elicitation measures such as verbal reports or think aloud protocols are appropriate to determine the learning processes at any of these internal stages (Leow, 2015). When either of the learning conditions is described as the product *what*, it means it is looked at as an output, the representative L2 knowledge which could either be declarative or procedural. The output is measured by non-concurrent or offline measures. What that means is that the output as the product obscures the various learning processes that have occurred internally at the three processing stages. If the explicit/implicit learning and declarative/procedural knowledge distinction is non-orthogonal as proposed by Ullman (e.g., 2001a), then the output does not speak much about whether learning proceeded explicitly or implicitly. The present study, however, adopts a definition of explicit and implicit language learning as the product *what*. The study assumes a one-to-one relationship between explicit/implicit learning and explicit/implicit knowledge distinction, as postulated in Paradis's DP model and Reber's implicit learning theory.

A decision to assume a one-to-one relationship between explicit/implicit language learning (or memory) and explicit/implicit knowledge distinction was based on two reasons. First, the present study assumes that this position signifies an idealisation of the questions about the interface between explicit knowledge and implicit knowledge, on one hand, and the one-to-one relationship between explicit/implicit learning distinction and explicit/implicit knowledge distinction on the other hand. Following Hooker (1994), Ellis (2015b) characterises idealisation as a statement or a law - about a phenomenon – which is a simplification that does not account for the full complexity of a phenomenon. However, the statement is theoretically tenable as regards the rational basis for proposing it, practically manageable, intelligible, and "empirically accurate to a degree in as much as it can account for the characteristics of the phenomenon but not necessarily for all characteristics of the phenomenon" (Ellis, 2015b, pp. 189-190). In this regard, the non-interface position is a simplification of the various processes in both declarative and procedural memory; it is theoretically tenable in that it has a rational basis as espoused in Paradis' DP model and Reber's implicit learning theory; and manageable and intelligible in that it allows empirical verification on the basis of well-defined constructs as proposed in the DP model. Ellis (2015b) points out that because idealisations help to focus on important issues, they are fundamental to scientific inquiry.

The second reason for assuming the non-interface position is that there are no reliable online or concurrent measures of awareness that could be used at any of the three internal processing stages in the learning process framework without either influencing subjects' type of learning or knowledge or negatively impacting on the subjects' responses. For instance, Rebuschat et al. (2015) carried out a validation study on the usefulness of measures of awareness - concurrent verbal reports (i.e., think-aloud protocols), retrospective verbal reports, and subjective measures - in the investigation of implicit and explicit learning. They found clear evidence of the interference of concurrent verbal reports and subjective measures with subjects' performance: they found that only the participants who did not think aloud during exposure were able to generalise the acquired knowledge to novel instances. Rebuschat et al. (2015) point out that their results demonstrate a potential drawback in the use of think-aloud protocols. They further observe that even asking participants to choose rule knowledge as the source of their answers influenced the participants to become aware that there was a rule. Rebuschat et al. (2015) note that many subjects in the think-aloud group in their study had shown no evidence of awareness during the exposure phase but began "to search for a rule once they realized that they did not know how to answer the new type of item presented in the test phase" (p, 329). These results show that though the concurrent measures would help to establish learning processes, an investigator risks collecting unreliable data.

In summary, explicit and implicit learning, best defined in relation to the resultant knowledge, refer respectively to conscious and unconscious ways of language input processing at various stages of the learning process in L2 acquisition. As Hulstijn (2005, p. 129) puts it, there are “good theoretical and educational reasons to place matters of implicit and explicit learning high on the agenda” for L2 acquisition research if indeed L2 development relies only on implicit linguistic knowledge. Indeed, as shown in the next sections, there has been a great deal of research investigating the role of the two learning conditions.

2.5 Empirical Studies: Learning Domain in L2 Acquisition

A large body of research has investigated the differential nature of the two learning conditions and the complexity of grammatical structures. Generally, behavioural research has shown that explicit instruction is more effective than implicit instruction, with explicitly trained groups always outperforming implicitly trained groups regardless of the complexity of the targeted grammatical structures. Neurocognitive and aptitude research, however, provide some strong support for the differential nature of explicit and implicit learning conditions and of the complexity of grammatical structures. The review of the behavioural research is presented first. DeKeyser (1995) is the only study reported here whose behavioural results appear to suggest the differential nature of the two learning conditions

2.5.1 Robinson (1996; 1997).

Included in Norris and Ortega (2000) and Spada and Tomita’s (2010) meta-analyses discussed below are Robinson’s (1996; 1997) studies. In these studies, Robinson investigated A. S. Reber’s (1989, A. S. Reber et al., 1991) and Krashen’s (1985) claims that implicit learning (under implicit and incidental learning conditions) is more effective than explicit learning (under rule-search and instructed conditions) when the stimulus domain is complex. A total of 104 L1 Japanese, Mandarin, and Korean (ages 19-34) learners of English as a second language were randomly assigned to one of the four different learning conditions: implicit, where the learners memorised word order without any explanation; incidental, in which the learners read sentences in order to answer comprehension (yes/no) questions; rule search, where the learners looked for rules from sentence stimuli; and instructed learning condition, in which learners read explanations of rules, accompanied by a spoken explanation from the researcher. Robinson found no support for A. S. Reber's and Krashen’s claims that complex rules are most effectively learned under unconscious conditions: both the implicit condition and incidental condition were not effective for learning either the complex rule (defined as pseudo-clefts formation rules) or simple rule (defined as rules involving locative inversion). The implicit condition and the incidental condition groups performed comparably and there were no significant differences in their English grammar knowledge as measured by the UGJT (but response times were recorded). It was, however, the instructed condition group that outperformed all the others in learning simple rules. Further, instructed learners were significantly more accurate than rule-search learners in performance on both types of rules. Overall, these results are similar to those obtained in Norris & Ortega’s (2000, Section 2.5.2) meta-analysis where explicit learning is seen as superior over implicit or incidental learning.

However, several caveats need to be noted regarding Robinson’s (1996) study. Firstly, there were only two training sessions, which might have favoured groups in instructed and rule-search groups. As Robinson (1996) points out, longer periods of exposure to greater quantities of input could have also improved implicit and incidental learning. Secondly, UGJT was used as the only measure of language knowledge, and as experienced L2 learners, the participants may have relied on explicit knowledge especially considering that they were asked "Can you say what the rules were?" under each question from the beginning of the test. In this condition, the learners from the start of the test became aware that they had to (consciously) search for rules when judging the sentences. Lastly, Robinson distinguished between simple and complex rules by asking teachers to rate the complexity of the rules. The rule for describing how to form pseudo-clefts of location, such as “Where Mary and John live is in Chicago not in New York” was considered the hard rule and therefore a complex rule. The rule describing the fact that subject-verb (SV) inversion is allowed in sentences where adverbials of movement or location are fronted, such as, “Into the house John ran/ran John” was considered easy, therefore simple, rule. Lacking a theoretical basis, it is not clear whether the two sets of rules were indeed different in terms of complexity.

2.5.2 Norris and Ortega (2000).

Norris and Ortega (2000) carried out a synthesis of 49 primary studies published between 1980 and 1998 on the effectiveness of L2 instruction. One general finding was that explicit instruction, whether deductive and inductive, “leads to more substantial effects than implicit instruction (with average effect sizes differing by 0.59 standard deviation units), and this is a probabilistically trustworthy difference” (p. 500). Further, average L2 instruction effect sizes for post-tests remained relatively large, indicating that the effects of L2 instruction are durable. These results suggest the superiority of explicit learning over implicit learning regardless of whether the stimulus domain is complex or not. However, there are several caveats regarding these results. First, in this meta-analysis, the role of the complexity of stimulus domain was not considered. Second, Norris and Ortega point out several problems, for example regarding the primary study design features where only 18% of 78 sample studies operationalised true control conditions, and regarding data analysis where 91% of the reviewed study reports reported interpretations of quantitative findings according to results of statistical significance tests. These might have had significant impact on the (interpretation of the) primary results and hence on the meta-analysis results.

2.5.3 Spada and Tomita (2010).

Similar results to Norris and Ortega (2000) have been reported in Spada and Tomita’s (2010) meta-analysis, which was conducted to investigate the effects of explicit and implicit instruction on the acquisition of simple and complex grammatical features of English. Unlike Norris and Ortega (2000), Spada and Tomita (2010), following Hulstijn and de Graaff (1994), categorised the target features in their 41 study-reports into simple or complex based on the number of criteria applied to arrive at the correct target form (i.e., based on the number of transformational rules involved to derive a structure). A grammatical feature involving only one transformational rule was designated as simple while a feature with more than one transformation was described as complex. As in Norris and Ortega (2000), the results indicated consistently larger effect sizes for explicit over implicit instruction, for both simple and complex features, suggesting that “explicit instruction positively contributes to learners’controlled knowledge and spontaneous use of complex and simple forms” (Spada and Tomita, 2010, p. 263).

Other studies such as Tagarelli et al. (2016) discussed in detail in section 2.6, report similar results where explicit instruction is shown to be more effective than implicit instruction. Tagarelli et al. (2016) found that the instructed group significantly outperformed the incidental group in terms of overall scores on simple and complex grammatical structures. They further found that on simple and complex rules the instructed group outperformed the incidental group, and some of the comparisons reached or approached significance. Further, investigating only incidentally trained learners of a semi-artificial language, Ettlinger et al. (2014, see section 2.6) found that the learners performed better on a simple morphophonological rule than they did on complex rules. These results are contrary to A. S. Reber's (1989; 1993; A. S. Reber et al., 1991) implicit learning model claim that complex rules are amenable to implicit or incidental learning.

2.5.4 DeKeyser (1995).

From the behavioural research stream, DeKeyser's (1995) study results suggest that the two learning conditions may indeed be fundamentally different. Using a computerised experiment with the two varieties of Implexan, an artificial language, DeKeyser investigated the role of explicit and implicit learning of straightforward (categorical) and fuzzy (proto-typicality patterns i.e., irregular forms that are not merely exceptions but are determined by the linguistic environment such as the type of final consonant and initial consonant) morphological rules. A total of 61 participants, consisting of 51 undergraduates 10 graduate students, completed the study. The Modern Language Aptitude Test (MLAT, Carroll & Sapon, 1959) was administered to all subjects to ensure the outcomes were not "unduly influenced by differences in language learning aptitude among the various groups" (DeKeyser, 1995, p. 389).

The results of the production test revealed that subjects in the explicit learning condition learned categorical rules better (than those in the implicit learning condition), and this relationship was confirmed through statistically significant results. Conversely, the results of the production test showed that the subjects in the implicit learning condition learned prototypicality patterns better (than those in the explicit learning condition) although this relationship was not confirmed through statistically significant results.

Morgan-Short et al.’s (2010; 2012) studies are the only two studies in neurolinguistics that have investigated the differential nature of explicit and implicit learning conditions. The results in these studies provide strong evidence suggesting the two learning conditions and simple and complex rules are fundamentally different. These studies are discussed next.

2.5.5 Morgan-Short et al. (2010; 2012).

Using a combined behavioural-and-event-related potential (ERP) assessment approach, Morgan-Short et al. (2010) investigated the neurocognition of the processing of noun-article and noun-adjective gender agreement in the artificial language, Brocanto2, in 41 adult subjects that were split into two groups, one trained explicitly, and the other, implicitly. The study found no evidence to suggest that the two learning conditions are fundamentally different. However, the results suggested the interactions among linguistic structure, proficiency level, and type of training. Though the results at low proficiency level showed that neither group processed the two types of agreement in the same way as they are processed in L1, implicitly trained learners appeared to rely on lexical/semantic processes – and likely on declarative memory – for both adjective and article gender agreement processing. Explicitly trained learners also appeared to rely on lexical/semantic processes at low proficiency only for adjective (not article) agreement. At high proficiency level, the results showed that noun-article agreement processing depended, at least to some extent, on L1 processing mechanisms for both explicitly and implicitly trained subjects while noun-adjective agreement processing depended on lexical/semantic processes. The targeted grammatical structures were considered difficult for the learners. This was based on the research that has shown inflectional morphology to be particularly difficult for late L2 learners to acquire (e.g., Montrul, 2004, and Montrul, Foote, Perpinan, Thornhill & Vidal, 2008 cited in Morgan-Short et al., 2010). However, the results of this study must be interpreted with caution: behavioural results showed that implicitly trained learners performed as well at both low proficiency and end-of-practice as explicitly trained learners which is not consistent with the majority of L2 research.

Morgan-Short et al. (2012) found some evidence suggesting that explicit and implicit learning conditions are fundamentally different. Using the same artificial language, Brocanto2, Morgan-Short et al. investigated longitudinally (at both low and high proficiency) whether explicit or implicit training conditions differentially affect neural (event-related potential, ERP) and behavioural (performance) measures of L2 syntactic processing. Their study was motivated by previous claims in cognitive psychology that even aspects of grammar that are thought to be difficult, “L1-like brain processing may eventually be attained” (p.933). After three practice sessions in their study, the results from 30 adult participants (average age = 24.25 years) fluent in English showed that both explicitly and implicitly trained participants attained statistically indistinguishable performance at both low and high proficiency. However, the electrophysiological measures revealed that the implicit training condition (but not the explicit training condition) showed the full spectrum of ERP components typically found for L1 syntactic processing. These results suggest that L2 adult learners can come to rely on native-like language brain mechanisms only when learning proceeds implicitly.

2.5.6 Aptitude and working memory research.

Further strong evidence in support of the differential nature of explicit and implicit learning comes from L2 acquisition research on the role of aptitude and working memory (WM) in L2 acquisition. L2 learning aptitude, defined as individual learners’ strengths in cognitive abilities required for information processing during language learning, has been found to be consistently and strongly predictive of L2 learning outcomes (Dörnyei & Skehan, 2003; Li, 2017; 2018; 2019; Robinson, 2005a). Importantly, the aptitude-treatment interaction (ATI) research, which is a strand of research investigating how aptitude interfaces with the type of instruction has demonstrated that the predictive nature of aptitude for L2 acquisition is mediated by learning contexts. For instance, in Carpenter’s (2008) study, a significant correlation was found between learners’ scores on aptitude and scores in the explicit condition where learners were asked to learn artificial grammar rules. No significant correlations were found between aptitude and the effects of an implicit treatment with no rule explanations provided. Similar results are reported in Robinson (1995; 2002b; 2005b) showing stronger effect for aptitude under explicit conditions than under conditions engaging implicit learning. Further, like Reber et al. (1991), Gebauer & Mackintosh (2007) and Kaufman et al. (2010) have shown that psychometric intelligence is strongly related to explicit learning and not to implicit learning, providing further evidence that the two learning modes are distinct.

However, some studies provide contradictory findings. First, de Graaff (1997) found significant correlations between an aptitude test and learners’ performance for both explicit and implicit conditions. However, it has been pointed out that the implicit treatment in de Graaff’s (1997) study may have encouraged learners to consciously process the L2 (Carpenter, 2008; Li, 2018; 2019). Second, studies investigating the role of cognitive abilities in L2 acquisition in immersion settings have consistently shown that aptitude is strongly predictive of learning (DeKeyser, 2000; Harley & Hart, 1997; 2002). See Carpenter (2008) for a detailed discussion of this research. These findings are consistent with Skehan’s (2002) proposal that the more demanding immersion environments which may promote more implicit processes make aptitude much more important than the more controlled, structured explicit learning contexts. However, a critical problem with this stream of research is that the reported findings may arise from the lack of control over the quality of instruction.

WM is simply a learner’s ability to ‘simultaneously process and store information’ (Faretta-Stutenberg & Morgan-Short, 2018, p. 69). In addition to findings suggesting a strong and significant relationship between WM and L2 abilities, learner proficiency and learning targets (e.g., Linck, Osthus, Koeth, & Bunting, 2014; McDonald, 2006; Sagarra, 2007; Sagarra & Herschensohn, 2010), a great deal of this research has shown that IDs in working memory predict L2 development only in more explicit, intentional learning conditions (Brooks, Kempe, & Sionov, 2006; Brooks & Kempe, 2013; Brooks, Kwoka, & Kempe, 2017; Conway, Bauernschmidt, Huang, & Pisoni, 2010; Grey et al., 2015; Kaufman et al., 2010; Kempe & Brooks, 2008; Martin & Ellis, 2012; Tagarelli et al., 2015; Tagarelli, Mota, & Rebuschat, 2011; J. N. Williams, 2012; Yang & Li, 2012). Contradictory results come mainly from studies that have explored the facilitative role of WM in L2 acquisition in immersion settings (LaBrozzi, 2009; 2012; Sunderman & Kroll, 2009; Tokowicz, Michael, & Kroll, 2004). For a relatively detailed discussion on this research see Faretta-Stutenberg (2014). By generally comparing learners with and without immersion experiences, these studies have found that WM is predictive in learners with immersion experiences, but not for learners without immersion experiences. Further, some experimental studies (e.g., Denhovska, Serratrice, & Payne, 2016; Robinson, 2002a; 2005b) found that the interaction between WM and explicit and implicit/incidental learning conditions were not differential.

To summarise, although behavioural research has generally shown explicit instruction to be more effective than implicit learning, neurocognitive and aptitude research provides considerably strong evidence suggesting that the two learning conditions and the levels of linguistic difficulty are differential. However, as Morgan-Short et al. (2012) points out, the behavioural research has several limitations that bias the results to favour explicit instruction. For instance, quite small and short amounts of training, the provision of explicit information in addition to the stimuli in the explicit conditions, and the use of language knowledge measures available to conscious awareness have generally biased the results toward an advantage for explicit training in the studies. The next section reviews research that has examined how the interplay of the two long-term memory systems, learning conditions and level of linguistic difficulty modulate L2 acquisition.

2.6 Empirical Studies: Memory Systems and L2 Acquisition

Since Carpenter (2008, discussed in detail below) directly examined the role of the declarative and procedural memory systems in L2 acquisition, there has been a surge in studies exploring the predictive nature of the two long-term memory systems and how exposure type and linguistic complexity mediate L2 acquisition. Two streams of this research can be distinguished: first, that which has examined the role of the two long-term memory systems, with only exposure type as a mediating factor; and second, that investigating the role of the memory systems as modulated by the level of linguistic difficulty (or linguistic complexity) and exposure type. In each stream of research, the investigations have adopted either an ex post facto or experimental design. Two further strands of experimental studies are distinguished. While some have used natural language grammars in their investigations, others have used (semi-)artificial grammars. An artificial language is “a model linguistic system composed of a small, novel lexicon and a few grammatical rules that are consistent with natural language rules” (Tagarelli et al., 2016, p. 297). The small, novel lexicon and grammar of artificial languages allow them to be learned relatively quickly in highly controlled conditions (Tagarelli et al. 2016). On the other hand, a semi-artificial language consists of “lexical information from the participants’ L1 and grammatical information from another language” (Tagarelli et al., 2016, p. 297). Learning is facilitated even more in semi-artificial languages because participants do not need to learn new vocabulary. The following sections review all this research, beginning with the research on the role of the two long-term memory systems and exposure on L2 acquisition in section 2.6.1. Section 2.6.2 reviews the research on the three-way interplay of declarative and procedural memory, exposure, and linguistic complexity in L2 acquisition. The review includes the examination of instruments and procedures used in the investigations.

2.6.1 Memory, exposure and language knowledge.

Adopting the ex post facto design to examine the role of the declarative and procedural memory systems and exposure, previous research reports conflicting results. Some studies have found no predictive role for the procedural memory system in language acquisition (e.g., Kidd & Kirjavainen, 2011; Lum & Kidd, 2012; Pretz et al., 2010; 2014). However, Kaufman et al. (2010) found that both the declarative and the procedural memory systems predicted L2 learning. These studies are detailed below.

2.6.1.1 Lum and Kidd (2012).

Lum and Kidd (2012) investigated whether declarative and procedural memory systems were related to typically developing children’s past tense and lexical knowledge. The study tested Ullman’s DP model which proposes that, in L1, procedural memory supports rule-based language acquisition and use while declarative memory supports the learning and use of the non-rule language (i.e., irregular verbs and vocabulary). Participants were 58-typically developing monolingual English children, approximately 5 years of age, in elementary schools in the United Kingdom. Lum and Kidd tested the children’s procedural memory, declarative memory, nonverbal intelligence and short-term working memory. The children’s vocabulary and past tense knowledge were measured by the British Picture Vocabulary Scale (a word-picture matching) and a past tense task (involving sentence completion), respectively. The SRT, adapted for children, was used as a procedural memory measure. Declarative memory for verbal information was measured by the Word Pairs subtest from the Children’s Memory Scales (CMS), a task like Llama-B. In CMS, participants have a study period where they are asked to learn a list of orally presented word-pairs.

Using bivariate and partial correlational analyses, the study found no correlations between the scores of the measures of declarative and procedural memory and scores on either regular or irregular past tense use. However, a significant relation was observed between declarative memory and vocabulary. Lum and Kidd interpreted the results as partly supporting Ullman’s DP model where declarative memory supports vocabulary in this age group. Overall, Lum and Kidd (2012) question the role of declarative and procedural memory in L1 English acquisition of regular and irregular past tense in 5-year-olds. However, Lum and Kidd did not use separate tests to measure the learners’ declarative and procedural language knowledge of the past tense (regular and irregular). The use of multiple tests for each of the memory systems and language knowledge types would shed more light on these results.

2.6.1.2 Kidd and Kirjavainen (2011).

Similar results to Lum and Kidd (2012) are reported in Kidd and Kirjavainen (2011). The study examined the role of procedural and declarative memory in the acquisition of L1 Finnish past tense morphology. Two competing models were tested: Ullman’s (2004) declarative/procedural model, and the single-route model which predicts that declarative memory should support lexical learning, which in turn should predict morphological acquisition (Lum & Kidd 2012). A total of 124 learners aged between 4.0 and 6.7 from kindergartens in Southern Finland completed a battery of tests. As in Lum and Kidd (2012), a children’s version of the SRT and the Finnish-translated CMS test were used to measure declarative and procedural memory, respectively. There were also tests of vocabulary knowledge, nonverbal ability, and a ‘‘wug’’-style past tense elicited production (Kidd & Kirjavainen, 2011, p. 13). In the wug-test, the children saw on a computer a series of characters performing actions. At the same time, the children heard a pre-recorded test sentence describing the picture. They were then asked to complete related sentences with a past tense verb.

Kidd and Kirjavainen (2011) found that procedural memory was not significantly correlated with the children’s performance on the measure of past tense. Instead, a direct positive and significant relationship was observed between declarative memory and vocabulary size, which in turn predicted the children’s performance on every verb type. Further, there was an indirect significant relationship between declarative memory and learners’ past tense knowledge. Kidd and Kirjavainen interpreted these results as consistent with the single-route approach. They suggested that procedural processes might be more likely to play a role in the sequencing of linguistic information at the level of the sentence rather than the word as is the case with inflectional morphology (e.g., past tense). However, though the wug-test used is an oral production test (OPT), the extent to which it indexes procedural learning processes needs to be validated. Further, the learners might have had an opportunity for reflection when, in cases of non-response or incorrect response, the sentences were repeated, and the learners were asked to use the past form of the verb they had just heard.

2.6.1.3 Pretz et al. (2010; 2014).

Pretz et al. (2010; 2014) report similar results to those in Lum and Kidd (2012) and Kidd and Kirjavainen (2011). Pretz et al. (2010; 2014) examined the effects of mood, cognitive style, and cognitive ability on procedural learning. The SRT and Artificial Grammar Learning (AGL) were used as procedural memory measures. In addition to measures of cognitive style and mood, the study used standardised test scores on four subtests of the American College Testing (ACT; English, Mathematics, Reading, and Science) as a proxy for cognitive ability. ACT scores are multiple-choice test scores that measure cognitive abilities such as critical thinking, reasoning, and problem solving in each of the subject areas. Because all these cognitive abilities call for reflection, ACT scores should index declarative learning processes. With 109 undergraduates (mean age=19.29, SD=1.27)[2], Pretz et al. (2010; 2014) found that procedural memory was not correlated with ACT English scores (see Pretz et al. 2014, corrigendum). This lack of correlation is unsurprising because ACT English language tests must have allowed test-takers to heavily rely on their declarative processes and not procedural processes. Considering that in this study there were no direct measures of declarative memory and language knowledge as well as of procedural language knowledge, it is not possible to draw a clear conclusion regarding the role of procedural memory in language acquisition.

2.6.1.4 Kaufman et al. (2010).

Among studies adopting the ex post facto design to examine the role of the declarative and procedural memory systems and exposure, Kaufman et al. (2010) is the only study that found both declarative and procedural memory to be significantly associated with

---

[2] No further detail is given regarding the location of the study as well as whether the participants were L1 or L2 English learners.

L2 scores. Kaufman et al. (2010) investigated the association of IDs in implicit learning with a variety of cognitive and personality abilities and academic achievements in 153 high achieving English 16-18-year-old students attending a selective sixth form college in Cambridge, England. Declarative memory (i.e., explicit associative learning) was assessed by the Three-Term Contingency Learning (3-Term test) and Paired-associates (PA) learning (B. A. Williams & Pearlberg, 2006). Procedural memory (i.e., implicit learning) was assessed by the SRT test. There were also measures of psychometric intelligence, elementary cognitive abilities (i.e., working memory), processing speed and personality. For the academic performance on L1 English, Math, and Science and L2 French and L2 German, the participants reported General Certificate of Secondary Education (GCSE) exam scores. GCSE exams are national, subject-based exams taken by students in England between the ages of 15–16 (i.e., 11th year of schooling) before entry to sixth form. Generally, GCSE involves some combination of coursework and written, listening, speaking, and reading examinations.

The study found that declarative memory (likewise psychometric intelligence and working memory) was significantly correlated with GCSE L1 English and GCSE L2 French scores but not with GCSE L2 German scores. Controlling for declarative memory, psychometric intelligence, working memory, and processing speed, the correlation between procedural memory and L2 French and German scores was statistically significant. Although Kaufman et al. (2010) do not provide information about the exposure type to the two foreign languages investigated in the study and about the nature of the GCSE language tests (i.e., if they allow learners access to explicit and implicit language knowledge), the significant correlations between procedural learning and the two foreign language scores indicate that IDs in the procedural memory system are meaningful and play a role in L2 acquisition. Further, consistent with the results reported in Kidd and Kirjavainen (2011), Lum and Kidd (2012) and Pretz et al. (2010; 2014) studies reviewed above, L1 English scores in this study were related only to declarative memory. These results appear to suggest that the procedural memory system may probably be relevant only in L2 acquisition.

In highly controlled settings, several experimental studies have explored the role of the two long-term memory systems and exposure type in L2 acquisition. Overall, the findings of this research show that declarative memory plays a role if L2 learners are explicitly trained and/or if they remain at lower levels of proficiency. The procedural memory system, on the other hand, appears to support implicit learning processes but only at advanced proficiency. What follows is the discussion of this research.

2.6.1.5 Morgan-Short et al. (2014).

To affirm the results reported in Carpenter (2008, discussed below) where declarative and procedural memory differentially predicted L2 proficiency depending on the training condition, Morgan-Short et al. (2014) investigated the role of declarative and procedural memory on the learning of an artificial language called Brocanto2 under implicit training and with additional practice time. Brocanto2 is an artificial language with no writing system (see Carpenter, 2008; Morgan-Short et al., 2012). Unlike other artificial grammars, it is suggested that Brocanto2 has a productive structure allowing generation of contextually meaningful novel sentences as is the case in natural languages. Brocanto2 has also been shown to produce brain activity characteristic of natural language processing. Further, the use of Brocanto2 allows for strict control over the type and amount of exposure to the L2 and the (dis)similarity between the L2 and the learners' L1. Fourteen participants in the study were all L1 English speaking, right-handed, healthy adults between the ages of 18 and 30 (*M* = 22.21, *SD* = 2.72). They were either enrolled in college or held a bachelor's degree in a non-language-related field in the USA.

Seven sessions were conducted, with four language training and practice sessions and three test sessions. The participants trained and practised on Brocanto2 vocabulary and word-order rules. Prior to testing participants were told they would be learning an artificial language and that they would be tested. Tests consisted of the CVMT and MLAT-V as measures of declarative memory, and the Weather Prediction Task (WPT, probabilistic) and the Tower of London task (TOL, cognitive skill learning) as measures of procedural memory. The TGJT accuracy *d*' scores (i.e., d-prime scores, see section 3.6.1.3 for a discussion of these scores) were used as a measure of participants' grammar knowledge of the artificial language. Morgan-Short et al. (2014) found that both declarative and procedural memory systems were abilities with meaningful IDs even when instruction was implicit. There were positive relationships between declarative learning ability and syntactic development at early stages of acquisition and between procedural learning ability and development at later stages of acquisition. Morgan-Short et al. noted that the IDs in the two long-term memory abilities accounted for a large amount of variance at both stages of development. These results were interpreted as providing support to Ullman's (2001a; 2004; 2005) DP model where domain-general cognitive abilities, namely declarative and procedural memory, are hypothesised to be significant predictors of development at early and later stages of L2 grammar acquisition, respectively. This study was, however, limited by its small sample size. Further, the learners might have reflected on their learning since they were made aware at the beginning of the study of learning and of a test afterwards. There was also only one training condition (i.e., implicit condition) and one measure of language knowledge.

2.6.1.6 Brill-Schuetz and Morgan-Short (2014).

Brill-Schuetz and Morgan-Short (2014) examined the role of procedural memory in adult L2 development under either explicit or implicit conditions. Twenty-six university students (18–24 years old) at a large midwestern university were trained on an artificial language, Brocanto2 (see Morgan-Short et al., 2014 above). First, participants trained on Brocanto2 vocabulary and word-order rules. In the explicit training condition, participants were auditorily presented with the specific rules and examples of phrases and sentences in Brocanto2. In the implicit training condition, however, no rules or explanations were given. Instead, participants received repeated, aural examples of the language that ranged from simple noun phrases to complete sentences. Then, during practice, participants used Brocanto2 to play a computer-based board game in which they either heard sentences and made the corresponding move on the game board (comprehension) or saw a move and orally described it (production). Participants completed 20 practice modules (10 production and 10 comprehension), with 20 novel sentences presented in each module for a total of 400 practice items.

Using an aurally presented GJT to measure language knowledge at two time points and the Alternating Serial Reaction Task (ASRT, Howard Jr & Howard, 1997) and the WPT (Knowlton et al., 1994) to measure procedural memory, Brill-Schuetz and Morgan-Short (2014) found no differences based on training condition. However, using a computed procedural memory composite score, Brill-Schuetz and Morgan-Short found that participants with high procedural memory ability performed better on final grammar assessments than participants with low procedural memory ability when they had been trained under implicit conditions. These results suggest that higher procedural memory ability leads to higher levels of L2 development, under implicit conditions. This finding was consistent with the findings in Carpenter (2008, see below) and Morgan-Short et al. (2014) which have reported the predictive effects of procedural memory in implicit language training conditions at later stages of L2 development.

However, Brill-Schuetz and Morgan-Short's (2014) study was limited in three main ways. First, as pointed out by Brill-Schuetz and Morgan-Short (2014), the learners had not reached a high level of proficiency in consecutive modules. This should explain the lack of a significant relationship between language knowledge scores and the measures of procedural memory. Second, there were no separate measures for language knowledge types. The GJT used in this study is not clearly described. No information is given as to whether it was speeded or un-speeded. Thirdly, declarative memory was not investigated.

2.6.1.7 Hamrick (2015).

Motivated by the lack of rigorous experimental design in Morgan-Short et al.'s (2014) study, Hamrick (2015) investigated whether IDs in declarative and procedural memory abilities mediate the learning and retention of L2 syntactic structures under incidental conditions. He adopted a "true" experimental design where the incidental learning condition was controlled for awareness of learning of the rules. To control for learners' explicit reflection on learning, the 31 monolingual speakers of English (25 undergraduates and 6 graduate students, with mean age of 21.4) in the USA were asked to participate in a study about meaning comprehension when reading scrambled sentences. They were never told that they were to be tested immediately after, and again, after about two weeks.

The study employed a semi-artificial language consisting of words from the participants' native language (English) and three Persian word order-rules.[3] In the

---

[3] As Hamrick (2015) points out, semi-artificial language paradigm "circumvents the need for vocabulary pre-training and allows researchers to more easily misdirect participants about the nature of the exposure phase, thereby reducing the likelihood that participants will engage in intentional, strategic learning of the target structures" (p. 11).

exposure phase, participants were incidentally exposed to each target structure three times. In a recognition task, the measure of the learners' language knowledge, participants were asked to discriminate previously seen (i.e., "old") syntactic structures from previously unseen (i.e., "new") syntactic structures. Both categories of syntactic structures consisted of "core" sentences (i.e., words and meanings) from the exposure phase, suggesting that old sentence structures were "exactly the same in every way as in the exposure phase, while the other half were only the same in terms of their lexical and semantic content, but had different syntactic structures" (Hamrick 2015, p. 11). Previous research (e.g., Hamrick, 2013; 2014a; 2014b) has shown that, unlike conventional GJTs, a recognition task is a more sensitive measure in incidental conditions than GJTs, probably because it does not require testtakers to consciously attend to language forms. Participants completed an immediate and a delayed 2-week recognition task. Using data from 18 participants with Llama-B and SRT as measures of declarative and procedural memory, respectively, Hamrick found that declarative memory was positively correlated with performance on the immediate, but not delayed, recognition task, whereas procedural memory abilities were significantly related to performance on the delayed, but not immediate, recognition task. These results were consistent with those found in Morgan-Short et al. (2014) and Ullman's DP model, even in cases where instruction was incidental. As in Morgan-Short et al. (2014) and Brill-Schuetz and Morgan-Short (2014), no separate language knowledge measures were used in Hamrick's (2015) study.

2.6.1.8 Faretta-Stutanberg and Morgan-Short (2018).

Faretta-Stutanberg & Morgan-Short (2018) adopted a quasi-experimental short-term longitudinal design, to examine the role of the two long-term memory systems and different natural learning contexts in the acquisition of natural L2 grammar. The study investigated L1 English learners of L2 Spanish in "at-home" (i.e., a traditional university classroom) and "study-abroad" (i.e. an immersion or naturalist) settings. In the "at-home" group, a total of 29 participants studying L2 Spanish at university level enrolled in one to three Spanish content courses at a fifth-semester level or above (e.g., Spanish grammar review, introductory linguistics, literary analysis) at a large public university. Data for only 17 participants were used for analysis. In the "study-abroad" group, 20 learners studying L2 Spanish in a Spanish-speaking country were enrolled in 12- to-15-week study-abroad programmes for a semester. They completed four or five university-level courses taught in Spanish (Mean weekly classroom hour = 15.5, SD = 3.5). Data for only 13 female participants was used for analysis. To participate in the study, a participant's experience with Spanish had to be classroom-based, with no previous substantial immersion experience and no substantial exposure to the language before the age of 12.

All the participants completed several cognitive assessments: declarative memory, MLAT5 and CVMT; procedural memory, the ASRT and WPT; and WM measures. The L2 behavioural language measure was a UGJT. During a participant's performance of the UGJT, electroencephalogram (EEG) data were collected as the measures of L2 neurocognitive processing. The tests for language development were administered twice, as a baseline at the beginning of a semester and as follow-up at the end of the semester. The 'at-home' learners showed behavioural gains, with no detected predictive role for IDs in cognitive abilities. The 'study-abroad' learners showed some behavioural gains and processing changes that were partially accounted for by procedural learning ability and WM. These results are partially consistent with results found in previous research with artificial languages, where learners with high procedural memory experienced advantages when trained under implicit, but not explicit, conditions (Brill-Schuetz & Morgan-Short, 2014). These results provide further evidence to support the claim that the role of procedural memory may be enhanced under less explicit, more exposure-based contexts. However, this contradicts research in L1 acquisition with children of around 5 years of age (e.g., Kidd & Kirjavainen, 2011; Lum & Kidd, 2012, reported above). L1 learning is naturalistic and rule-based language is implicitly learned. However, this L1 acquisition research has found no predictive effects of the procedural memory system. More acquisition research of natural language grammars in naturalistic contexts is required to further validate these results.

2.6.2 Memory, linguistic difficulty and exposure conditions.

Several studies have examined the three-way interplay of the two long-term memory systems, linguistic complexity and exposure condition in L2 acquisition. Experimental studies have dominated this research, with some adopting the artificial language paradigm (Carpenter, 2008; Ettlinger et al., 2014; Tagarelli et al., 2016) and others using natural language grammars (Robinson, 1997; Yalçın & Spada, 2016). However, these studies report conflicting results regarding the moderating effects of the complexity of the learning domain and of exposure type on the association of the declarative and procedural memory systems with language knowledge. Only Granena (2013b; 2014) adopted the ex post facto design to examine the role of the procedural memory systems and linguistic difficulty in L2 acquisition in immersion settings. Granena’s (2013b; 2014) results are reported last in Section 2.6.2.6.

2.6.2.1 Robinson (1997).

Robinson (1997) is the earliest study to explore the three-way interaction of cognitive abilities, learning contexts and linguistic complexity. He examined Krashen’s (1981; 1985) and A. S. Reber’s (1989; 1993; A. S. Reber et al., 1991) claims that unconscious learning under implicit and incidental conditions, especially when a stimulus domain is complex, is insensitive to measures of IDs in cognitive abilities, in contrast to conscious learning under rule-search and instructed conditions. Of the two long-term memory systems, only declarative memory was investigated along with general language learning aptitude. These two cognitive abilities were investigated under the term *language learning aptitude*.

A total of 104-L1 Japanese, Korean and Mandarin Chinese learners of English between 19 and 34 years old participated in the research. All the participants were enrolled in language programmes at intermediate level in Hawai’i. Through a placement test, potential participants had to demonstrate that they were unfamiliar with the English target structures in the study. The participants were then randomly assigned to one of the four learning conditions: implicit, where the learners memorised word order without any explanation; incidental, in which participants read sentences in order to answer comprehension (yes/no) questions; rule search, where the learners looked for rules from sentence stimuli; and instructed learning condition, in which learners read explanations of rules, accompanied by a spoken explanation from the researcher. The goal was to learn two different rules, one “easy” (involving subject-verb inversion where adverbials of movement/location are fronted, as in *Into the house ran John* or *Into the house John ran*) and one “hard” (involving pseudo-clefts of location as in, *Where Mary and John live is in Chicago and not in New York*). To assess aptitude, all learners completed the MLAT5 Paired Associates (Memory) and MLAT4 Words in Sentences (Grammatical Sensitivity) prior to commencing learning of the target rules. The UGJT was used to assess L2 learning.

Strong positive correlations were found between aptitude and learning of both easy and hard rules in implicit, rule-search, and instructed conditions, but not between aptitude and incidental learning. The learners' scores on the MLAT4 Words in Sentences (Grammatical Sensitivity) were strongly and significantly correlated with performance on both easy and hard rules in the implicit and instructed conditions and only with the easy rule in the rule-search condition. The MLAT5 Paired Associates was significantly related to both rule types in the instructed condition and only with the hard rule in the rule-search learning condition. Contrary to Krashen (1981; 1985) and A. S. Reber (1989; 1993; A. S. Reber et al., 1991), these results were interpreted as indicating that aptitude can influence learning in at least some implicit learning conditions. Consistent with Krashen (1981) and A. S. Reber (1989; 1993; A. S. Reber et al., 1991), aptitude did not appear to influence learning in meaning-based incidental learning (where comprehension is paramount). Further, the results suggest that declarative memory, as indexed by MLAT5 Paired Associates task, can predicate explicit learning of both easy and hard grammatical rules under instructed and rule-search learning conditions. However, several limitations are associated with this study. First, there were only two training sessions, which might have favoured instructed and rule-search groups. Second, the UGJT was the only measure of language knowledge, and as experienced L2 learners they may have relied on explicit knowledge especially considering that they were asked "Can you say what the rules were?" under each question of the test. Third, the distinction between easy and hard rules was based on expert knowledge thereby lacking the theoretical basis as to whether the two sets of rules were indeed different in terms of complexity. Lastly, the inclusion of measures of procedural memory must have shed more light on the role of memory, learning condition and linguistic complexity as moderating variables in L2 acquisition.

2.6.2.2 Carpenter (2008).

Using behavioural and neurological measures, Carpenter (2008) is the first study to directly examine the contributions of individual differences in declarative, procedural and WM memory to L2 acquisition and how exposure type and various grammatical structures mediated L2 learning at different stages of proficiency. A total of 29 adult (ages 18-37; average age 23) learners of an artificial L2, Brocanto2, were trained under either explicit condition, where grammatical rules were explained, or implicit condition, where exposure to the language lacked any grammatical rule explanation. The participants were all native English speakers (in the USA) with very limited or no prior exposure to Romance languages and without memory or language disorders.

Using a computer program, participants in both explicit and implicit conditions were aurally introduced to the names of game-tokens (i.e., vocabulary) of a chess-like game by clicking on pictures of them and eliciting a recording of the word. Three grammatical aspects were targeted: morphosyntax was indexed by agreement (noun-determiner and noun determiner-adjective); syntax was indexed by phrase structure (S-V-O), and lexical aspects of grammar were indexed by verb argument (transitivity). The grammar training consisted of exposure to 127 meaningful examples in the implicit condition and the exposure to metalinguistic information on the functions and rules of nouns, verbs, determiners, adjectives, and adverbs, and only 33 meaningful examples in explicit condition. Following grammar training were practice sessions involving self-instructional, computer-based, alternating comprehension and production activities, with two comprehension blocks then two production blocks, and so on. To control for input, no corrections were provided during practice. Low and advanced proficiency benchmarks were respectively at 45 and 95% accuracy on two consecutive comprehension blocks. L2 performance and neurocognitive processes were assessed with TGJT tasks (at low and advanced proficiency) and event-related potentials (ERPs), respectively. Declarative memory was assessed using the CVMT (visual), the California Verbal Learning Test-II (CVLT-II), MLAT5 Paired Associates and the Balanced Chunk Strength Artificial Grammar (AG; probabilistic) in which high-frequency chunks are potentially memorisable allowing participants to “sidestep implicit rule-learning processes” and instead rely on the knowledge of exemplars (i.e., explicit knowledge) during a transfer task (Carpenter, 2008, p. 113). Procedural memory was assessed using a dual task version of the WPT (probabilistic) and the Low Frequency AG task (probabilistic).

Regression analyses indicated that declarative and procedural memory differentially predicted L2 proficiency depending on the training condition. The explicitly trained learners relied on declarative memory during the learning of all the three grammatical structures at both low and advanced proficiency. For the implicit condition, declarative memory predicted agreement and verb argument scores only at low proficiency. Procedural memory predicted accuracy on agreement and phrase structure (both rule based in Brocanto2) for the implicitly trained learners only when proficiency approached the most advanced stages. Procedural memory did not predict accuracy on the memorisable argument structure, “which is not rule-concatenated [in] Brocanto2” (Carpenter, 2008, p. 297). These results were interpreted as supporting Ullman’s DP model where initial learning is predicted to be supported by the declarative memory and reliance on the memory may continue into the later stages of acquisition. Reliance on the procedural memory is predicted to manifest only at advanced proficiency. The results also suggest that linguistic complexity moderated the learning of Brocanto2 only in the implicitly trained group where only the rule-based language was predicted by the declarative memory at low proficiency and by the procedural memory at advanced proficiency. However, the study did not have separate measures of language knowledge types. Carpenter (2008) also points out that the addition of the SRT test which indexes complex sequencing abilities and is the best measure of implicit learning (see section 2.2.4) was desirable. Further, in addition to the limited sample size in the study (N = 29), the level of linguistic difficulty of the targeted rule-based grammatical structures was not determined. It is therefore not possible to make inferences regarding the role of simple and complex rule-based structures in L2 acquisition based on the study’s results.

2.6.2.3 Ettlinger, Bradlow and Wong (2014).

Ettlinger et al. (2014) investigated the association of memory and the knowledge of incidentally acquired simple and complex morphophonology of the semi-artificial language based on the grammar of Shimakonde, a Bantu language spoken in Mozambique. A simple morphophonological rule was defined as one in which the phonological realisation of a morpheme was consistently determined by context (i.e., *i*-stems and *a*-stems). A complex morphophonological rule was a diminutive plural form for *e*-stems where a vowel harmony rule is applied to change the vowel in the plural suffix followed by a reduction rule that is triggered by the diminutive prefix which changes the stem vowels.

Ettlinger et al. (2014) hypothesised that the simple and complex morphophonological patterns of word formation would correlate with standardised measures of procedural and declarative memory, respectively. Using data from 31 native English-speaking students (mean age 20.9) at Northwestern University in the USA, Ettlinger et al. (2014) found that declarative memory was associated with acquiring the complex pattern, whereas procedural memory was associated with the acquisition of both simple and complex patterns. In the control experiment with 18 native English-speaking students (mean age = 21.1 years; *SD* = 2.5), there were no significant correlations between learners' performance on language and any of the memory measures. These results suggest that simple morphological rules were subserved by the procedural memory system, contrary to Reber's proposal. In this study, a visual–auditory learning subtest of Woodcock–Johnson III Tests of Cognitive Ability and the Tower of London (TOL) task were used to measure declarative memory and procedural memory, respectively. The learners' morphophonological knowledge was assessed by a version of a "wug" test.

However, the results of this study need to be interpreted with caution. As in Morgan-Short et al.'s (2014, see above) study, participants were told at the beginning of the study that they were to be exposed to a new language and then tested on what they learned afterwards. Consequently, the subjects, though trained incidentally, must have consciously learned the rules of the semi-artificial language. Separate and multiple measures of declarative and procedural language knowledge and memory systems should have validated the results in this study.

2.6.2.4 Tagarelli et al. (2016).

Tagarelli et al. (2016) examined the relationship between IDs in cognitive abilities, exposure conditions, and linguistic complexity in L2 learners. Tagarelli et al. were motivated by the fact that research had not yet established how exactly these factors interact to influence language learning. The semi-artificial language they used consisted of an English lexicon and a German syntax for three verb-placement rules of varying levels of difficulty. The difficulty of the target structures was defined as linguistic and cognitive complexity. Linguistic complexity was operationalised as the number of clauses per T-units (T-units are defined as "one main clause with all subordinate clauses attached to it" (Hunt, 1965, cited in Tagarelli et al., 2016, p. 298). Based on this criterion, Tagarelli et al. (2016, p. 298) categorised the three verb-placement rules as follows, with previous research demonstrating that L1 and L2 German acquisition is characterized by an increased production of sentences with subordinate clauses for structure types 2 and 3:

1. Simple: The finite verb is placed in the second phrasal position of main clauses that are not preceded by a subordinate clause (V2 pattern).
2. Complex 1: The finite verb is placed in the final position of all subordinate clauses (V2-VF pattern).
3. Complex 2: When a subordinate clause precedes a main clause, the finite verb is placed in the first position of the main clause and the final position of the subordinate clause (VF-V1 pattern).

Data were collected from 51 one native speakers of English (mean age = 19) in the UK and USA and they had no background in German or any other V2 language. Participants randomly assigned to an incidental condition were not informed of the linguistic target nor that there would be a testing phase. In contrast, those randomly assigned to an instructed exposure condition were explicitly taught the target rule system before being exposed to the language. In the incidental condition, there were four practice trials and 120 randomly presented training sentences, with 40 sentences for each syntactic pattern.

An UGJT was used to assess language knowledge. Procedural memory and WM were assessed by a SRT test (or ASRT) and a reading span (Rspan) test, respectively. There were no measures for declarative memory. The study found no relationship between exposure condition and syntactic complexity. The simple rule was learned better in both conditions, although the instructed group performed better overall. This finding was unexpected; it had been predicted that linguistically simple rules would be easier to learn in the instructed condition and that complex rules would be easier to learn in the incidental condition. Regarding the relationship between cognitive abilities and learning outcomes in the two exposure conditions, the study found that only the procedural memory system was negatively but strongly related to the UGJT's *d'* scores for the incidental group (r = -.586, p = .003). Finally, regression and correlational analyses showed that procedural learning abilities were significantly and negatively related to the learners' outcome scores on Complex 2 sentences in the incidental condition, suggesting that the learners with better procedural memory abilities performed worse. From these results, no clear conclusions can be made regarding the three-way interaction in L2 acquisition. Because only the UGJT was used as a language knowledge measure, the study was weighted in favour of explicit learning processes. Further, the UGJT was scored in terms of *d'*. Though *d'* takes response bias into account, its scoring does not allow grammatical and ungrammatical sentences to be scored separately. Research on the measures of explicit and implicit language knowledge has shown that participants draw on different language knowledge types when responding to grammatical and ungrammatical sentences in a GJT (see section 2.1). The study examined only procedural memory and WM. Declarative memory was not investigated. Further, there was a significant correlation between RSpan (measure for WM) and SRT (procedural memory measure) scores in the incidental group, but not in the instructed group. The use of a small number of tests made it difficult for the measures to be validated to establish if they were indeed measuring the constructs they were designed for.

2.6.2.5 Yalçın and Spada (2016).

Lastly, Yalçın and Spada (2016) is another more recent study to experimentally investigate the role of memory and natural language grammar features. Yalçın and Spada (2016) elected to examine the role of aptitude in the learning of two L2 English features that differ in terms of their difficulty. Thus, like Robinson (1997), Yalçın and Spada (2016) did not directly investigate the role of declarative and procedural memory on L2 acquisition. Instead, they investigated the role of aptitude as measured by the four subtests of the Llama Aptitude Test (Meara, 2005). As pointed out in section 2.2.4, there is evidence suggesting that Llama-B, Llama-E and Llama-F index aptitude for explicit language learning with Llama-B as a measure of declarative (or rote or associative) memory, while Llama-D is considered to index implicit language learning abilities.

A total of 66 L2 English learners, at pre-intermediate level of proficiency and enrolled in the eighth grade (i.e., 13 to 14 years old) at a private secondary school in Turkey, received 4 hours of instruction on the passive (a difficult structure) and the progressive (an easy structure). The past progressive was characterised as an ‘easy’ structure because of its transparent form-meaning relationship, high frequency in the input, and its high transparency because it is realised by a free morpheme ‘was/were’ and a syllabic bound morpheme with no allomorphs. The passive was characterised as difficult because it involves many grammatical operations (i.e., many transformational or derivational rules are required to arrive at the target form) and is infrequent in the input. Further, the passive is acquired relatively late for L1 learners of English.

Participants’ opinions also indicated that the passive voice was more difficult than the progressive. Exposure and practice involved the provision of explicit information about the formal properties of the target grammatical structures as well as both implicit and explicit corrective feedback during all communicative activities. A written UGJT and an oral production task (OPT) were used to assess the learners’ explicit and implicit L2 knowledge of the target features, respectively.

Grammatical inferencing ability (i.e., Llama-F) was strongly related to the performance on the passive voice UGJT but the same aptitude subcomponent was not a significant factor affecting learners’ performance on the past progressive UGJT. On the other hand, Llama-B (i.e., associative memory ability) was related to learners’ use of the past progressive on the OPT but there was no role for this aptitude on the UGJT. Multiple regression analyses revealed that Llama-F predicted the passive UGJT scores while the progressive OPT scores were predicted by Llama-B. The Llama-D test did not correlate with any grammatical structure scores most likely because these structures were learned explicitly. While the other three Llama subtests (Llama-B, Llama-E and Llama-F) positively and moderately correlated with each other, Llama-D did not correlate with any of them. Yalçın and Spada interpreted these results as suggesting that different aptitude components contribute to the learning of difficult and easy L2 structures in different ways under the same learning conditions.

However, one limiting factor of this study was that all the participants had prior knowledge of the progressive but not of the passive structure, suggesting that the results for the progressive may not be explained by the intervention alone. Futhermore, Yalçın and Spada (2016) had designed for only one language learning condition (i.e., the explicit condition) and used only the Llama subtests to measure aptitude. As argued by Granena (2013a) Llama subtests generally measure the same underlying factor (i.e., conscious, explicit processing of information.) To sum up, the fact that Llama-F and Llama-B, but not Llama-D, predicted the explicit learning of easy and difficult grammatical structures in this study is unsurprising because both measures are generally considered to index explicit learning abilities. The use of separate and multiple measures of declarative and procedural learning mechanisms could have validated these results.

2.6.2.6 Granena (2013b; 2014).

Granena (2013b), adopting an ex post facto design, examined the three-way interplay of memory, exposure type and morphosyntactic L2 attainment by 18-year-old Chinese learners of L2 Spanish in a naturalistic learning context in Spain. A total of 100 learners participated. These learners had all lived in Spain for at least five years with educational level of no less than high school. Half of them were early learners with ages of onset between 3 and 6, and half were late learners with ages of onset 16 and older. “Age of onset was operationalized as the beginning of a sustained process of language acquisition as the result of migration or the commencement of a formal Spanish language program” (Granena, 2013b, p. 675). A group of 20 Spanish native speakers was included as a control group.

Six Spanish grammatical structures were investigated: three grammatical agreement relations (i.e., agreement structures: noun–adjective gender agreement, subject–verb agreement, and noun–adjective number agreement), and three structures that make essential contributions to meaning (i.e., non-agreement structures: subjunctive mood, perfective/imperfective aspect, and Spanish passive). In L1 Spanish, grammatical agreement structures are said to be acquired with almost 100% accuracy by age 3 whereas those involving meaning are not mastered until age 7 due to the semantic complexity. Though the two sets of grammatical structures are acquired at different time points in L1, Granena, following Meisel (2009), hypothesised that all the six structures would be difficult for the L2 learners because the aspects of inflectional morphology (e.g., gender agreement) are affected by maturational changes as early as age 3 in early childhood L2 acquisition. Moreover, Chinese is an isolating language without agreement markers and mood distinction while the passive is marked differently.

Procedural memory was assessed by the SRT and Llama-D tests. The learners’ declarative and procedural L2 Spanish knowledge were assessed respectively by the Metalinguistic Knowledge Test (MKT) and Word Monitoring Task (WMT, i.e., an online word monitoring and comprehension test). The role of declarative memory was not tested. Using a repeated-measures analysis of covariance (ANCOVA), the study found that the knowledge of the two sets of grammatical structures was mediated in different ways despite both sets of structures being hypothesised as difficult for the L2 learners. There was a significant relationship between procedural learning ability and early and late L2 learners' grammatical sensitivity to the agreement structures only. While Llama-D was found to be significantly related to early L2 learners' scores on agreement structures in the MKT, the SRT was significantly related to the late L2 learners' grammatical sensitivity to agreement structures in the WMT. Procedural memory did not modulate learning in the NS group. Granena interpreted the results as suggesting that the L2 Spanish learners might have relied on additional explicit knowledge which accounted for the different type of variability that sequence learning ability was unable to account for in the agreement structures. However, because the Llama-D test, measure of procedural memory, was significantly correlated with the MKT, measure of declarative knowledge, Granena's results need to be validated with separate and multiple measures of both declarative and procedural memory abilities and language knowledge types. Further, the two types of the target grammatical structures were linguistically complex for the learners, suggesting that no inferences can be made from Granena's results regarding the modulating effects of simple grammatical structures on L2 acquisition in immersion contexts.

Similarly, Granena (2014), adopting Granena's (2013b) study's design and L2 learners, found that language aptitude, as measured by the four Llama subtests (Llama-B, Llama-D, Llama-E, and Llama-F), was related to early learners' attainment in agreement structures on the auditory UGJT (not in the auditory TGJT) scores. No procedural memory measure was included. Granena (2014) interpreted the results as either suggesting that aptitude is not related to linguistic competence, understood as implicit language knowledge (i.e., as measured by the TGJT) that can be used automatically, or suggesting that a different type of aptitude correlates with more spontaneous use of L2 knowledge than with controlled language use. In this study, a Llama overall score was computed from all its four subtests, including Llama-D which is a procedural memory measure (see, Granena, 2012; 2016). Because Llama is generally considered a declarative memory measure, Granena's (2014) results suggests that declarative memory played a role in the acquisition of agreement structures but not in the non-agreement structures. These are the same grammatical structures which were found to be significantly related to the procedural memory system in Granena's (2013b) study. Thus, though the results of the two Granena (2013b; 2014) studies show that linguistic difficulty modulated L2 Spanish acquisition in immersion settings, both long-term memory systems predicted the learning of the same grammatical structures, the grammatical agreement relations. As a function of the complexity of the learning domain as proposed by A. S. Reber (1993; A. S. Reber et al., 1991), one would expect dissociation in the predictive effects of the memory systems.

However, Granena's (2013b; 2014) results are not surprising. The adoption of the *ex post facto* design meant there was no control over the mode of learning. Further, as pointed out above, Granena's (2013b; 2014) determination of complexity was based on the psycholinguistic perspective in which complexity is seen in terms of developmental stages, that is, whether a feature is acquired early on or late. Finally, the two targeted structural types were both considered difficult for L2 learners.

2.7 Summary and rationale

To summarise, this chapter has reviewed literature that addresses what, how and why L2 learners know what they know as conceptualised in psychological or cognitive theoretical approaches where language is viewed as inseparably linked to cognition. The product of language learning is either declarative or procedural knowledge. The question as to whether the two language knowledge types interface with each other is still contentious. Cognitive theoretical accounts attribute language learning to memory systems, influenced by two external factors, namely learning conditions and the level of linguistic difficulty. Within the dual route approach, three models have been proposed to explain the involvement of the declarative and procedural systems in language learning. Unlike Ullman's DP model (2001a; 2001b; 2004; 2015), Paradis DP model (2009) and Reber's implicit learning theory (e.g., A. S. Reber, 1989; A. S. Reber et al., 1991) propose that the procedural learning processes are independent of and not affected by declarative learning processes. Only in Ullman's DP model is the declarative memory system expected to predict implicit knowledge.

The literature review has shown that a great deal of research has directly investigated the effectiveness and the differential nature of exposure conditions. Behavioural results from this research indicate that explicit learning is more effective than implicit learning (e.g., Norris & Ortega, 2000; Spada & Tomita, 2010). Neurocognitive results, however, strongly suggest that the two exposure types are differential, with the implicit learning condition leading to L2 brain activation found in L1 syntactic processing. Empirical research on the role of aptitude and WM in L2 acquisition provides further evidence for the differential nature of the learning condition, with aptitude and WM being found predictive of language learning for explicitly trained learner-groups. The review has also shown that the research examining the interaction of linguistic complexity and learning conditions in L2 acquisition is lacking. The few studies that have explored this interaction appear to suggest that grammatical structures are differentially amenable to declarative and procedural learning. Early research in this area suggests that declarative learning may be effective when learning simple rules while sufficiently complex grammatical structures are more likely to be acquired by an implicit learning system (e.g., Krashen, 1994; A. S. Reber, 1989; 1993). However, relatively recent research appears to suggest that both linguistically simple and complex structures are best learned as declarative knowledge (Ettlinger et al., 2014; Robinson, 1996; 1997; Tagarelli et al., 2016).

The literature review has also shown that systemic L2 acquisition research on the role of the declarative and procedural memory systems has largely been ignored. The few studies that have experimentally investigated the two long-term memory systems and learning conditions as moderating variables in L2 acquisitions have reported findings suggesting that declarative memory plays a role if L2 learners are explicitly trained and/or if they remain at lower levels of proficiency. The procedural memory system, on the other hand, supports implicit learning processes but only at advanced proficiency.

Finally, the review has also shown that very little empirical research has examined the interplay in L2 acquisition of linguistic complexity, learning conditions and the two long-term memory systems, despite calls for such investigations (DeKeyser, 2016; Housen & Simoens, 2016). The findings of this research show that the declarative and procedural memory system are predictive of L2 learning, but the role of learning conditions and linguistic complexity as moderating variables in L2 acquisition is not clear. On the one hand, research by Robinson (1997) and Yalçın and Spada (2016) suggests that declarative memory is predictive of explicit learning of both linguistically easy and hard grammatical rules. On the other hand, research by Ettlinger et al. (2014) suggests that declarative memory subserves the incidental learning of complex grammatical structures while the procedural memory system supports the incidental learning of both simple and complex structures. Yet Tagarelli et al.’s (2016) results suggest that the procedural memory system is only associated with the incidental learning of complex grammatical structures, which is consistent with Krashen’s (1994) and Reber’s (1989; 1993) claim regarding the learning of complex stimuli. Several reasons account for these conflicting results.

First, one of the difficulties in this line of research has been the operationalisation of linguistic complexity or difficulty. While Carpenter (2008) and Granena (2013b; 2014) did not distinguish between the levels of difficulty of the grammatical structures for the L2 learners, the research has tended to determine simple and complex grammatical structures based on either subjective criteria or the number of transformational rules to arrive at the target structure (see section 2.3). Such operationalisation of difficulty does not appear to accurately determine the level of difficulty of grammatical structures. Second, these studies have tended to separately examine the memory systems (see Granena, 2013b; 2014; Robinson, 1997; Tagarelli et al., 2016) or the language knowledge types and usually using only one test measure of memory or of language knowledge (see., Carpenter, 2008; Ettlinger et al., 2014; Granena, 2014; Robinson, 1997; Tagarelli et al., 2016). Such single-task designs do not provide indication of variation between tasks and, consequently, may have limited generalisability (Granena, 2012). Further, though grammatical and ungrammatical test items in GJTs have been shown to be differential in allowing test-takers access to explicit and implicit knowledge (see section 2.1), the empirical research on L2 acquisition and the role of the three-interaction of memory, linguistic difficulty and learning conditions have tended to use GJTs’ total scores computed from grammatical and ungrammatical test items (see, Carpenter, 2008; Granena, 2014; Robinson, 1997; Tagarelli et al., 2016; Yalçın & Spada, 2016). The current study addresses these issues. As adopted in the present study, the multi-tasks research design, where both memory systems and language knowledge types are examined using multiple test measures for each construct in a single study, offers the opportunity of obtaining more insightful results and corroborating evidence for construct validity of the measures.

In addition to demonstrating that no comprehensive study has explored the three-way interplay in L2 acquisition, the review of the empirical studies has shown the lack of L2 acquisition research in real-world settings. Granena (2013b; 2014) are the only studies that have examined the role of memory and natural language grammatical structures in naturalistic settings. However, as pointed out above, the role of linguistic complexity in language acquisition was not the focus of these studies and was therefore not clearly addressed. Further, the studies involved independent analyses of the role of declarative and procedural memory in L2 acquisition of natural language grammatical structures in naturalistic settings (Granena, 2013b; 2014).

A number of these studies (i.e., Carpenter, 2008; Ettlinger et al., 2014; Tagarelli et al., 2016) adopted the artificial language paradigm in their exploration of the three-way interaction. Though artificial language learning (ALL) experiments have become an important tool in exploring language acquisition processes, it is not well documented whether “ALL engages the linguistic system and whether ALL studies are ecologically valid assessments of natural language ability” (Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, 2015, p. 822). In their study, Ettlinger et al. (2015) examined the relationship between performance in an ALL task (see Ettlinger et al., 2014) and L2 Spanish learning ability. A strong relationship between performance on the ALL tasks and L2 learning was found. Further, only participants’ performance on the complex ALL morphophonology showed the strongest relationship with more objective L2 Spanish measures. Ettlinger et al. interpreted these results as suggesting that “success in ALL experiments, *particularly more complex artificial languages* [emphasis added], correlates positively with indices of L2 learning even after controlling for IQ” (p. 822). Therefore, only artificial language grammars with more complex grammatical systems may closely resemble natural language grammars. As DeKeyser (1995) points out, agreement rules in artificial grammars are always categorical; whenever a noun is object, or a verb is feminine, for instance, it always carries the corresponding morpheme. Such grammars therefore do not typify natural language grammars which are complex. Natural language grammars “are multifaceted phenomena that defy simple definitions” with no “a one-to-one relationship between form and meaning, from the morpheme level all the way up to the text level” (Hulstijn, 2005, p. 134). Indeed, Robinson (2005b, p. 235) found little evidence for the “content generalisability” of the ALL findings to the incidental learning of Samoan.

Another potential area of reactivity of (semi-) artificial research design rests on the goal of language learning. Language is a tool for communication. However, this is not the case with artificial language learning. The artificial language is not viewed as a tool for communication but as an object used for the purposes of research. R. Ellis (2015a) points out that when older, cognitively mature learners view a target language as an object rather than as a tool for communicating they tend to reflect on their use of specific linguistic features even if those features were acquired implicitly. These observations about the characteristics of (semi-) artificial languages demonstrate that (semi-)artificial languages do not accurately capture natural language phenomena. However, while research designs involving controlled settings and artificial language paradigm are still essential in examining the interaction of the three factors in L2 acquisition, “it is also critical that research in more [natural settings] be conducted in order to maximize the ecological validity of the findings in the field” (Faretta-Stutenberg, 2014, p. 58).

2.8 Research Questions and Hypotheses

The present study adopted an ex post facto design to investigate the role of the declarative and procedural memory systems in the acquisition of 14 simple and complex L2 English grammatical structures in an English-medium instructional context. To the knowledge of the researcher of the present study, no study has examined the three-way interplay of memory, linguistic complexity and the classroom contexts where grammatical structures are generally explicitly taught.

To address previous research difficulty in the operationalisation of linguistic complexity, the present study draws on the core aspects of Pienemann’s processability theory (PT, Pienemann, 1998a; 1998b; 2015; 2005; Pienemann & Keßler, 2012; Pienemann & Lenzing, 2015) to propose a criterion that focuses on the intrinsic, structural complexity of grammatical features to objectively categorise the targeted L2 English grammatical structures into simple and complex (see section 3.4). Thus, grammatical structures categorised as simple structures in the present study are those whose processing procedures appear at the lower level of the PT’s proposed universal ‘Processability Hierarchy’ (see section 3.4 for a detailed discussion). These structures either require no grammatical information exchange between and within constituents or involve default mapping processes between the levels of linguistic representation. On the contrary, structures categorised as complex are those whose processing computational routines appear at the higher level of the ‘Processability Hierarchy’ and require either grammatical information exchange between and within constituents or non-linear mapping processes between the levels of linguistic representation (see section 3.4 for a detailed discussion).

The targeted grammatical structures are further differentiated based on whether they are explicitly taught or not explicitly taught in schools (see section 3.4). The document analysis of the teaching syllabuses of L2 English was carried out to determine what structures were explicitly taught. The present study's researcher's experience of the educational system complemented this process. Explicitly taught grammatical structures are those that appeared to have been explicitly exposed to the learners.

The present study hopes to gain answers to the following three research questions (RQs) in order to examine the role of the three-way interaction of the two long-term memory systems, linguistic complexity and grammar-rule learning in L2 acquisition in an instructed context:

RQ1: Do instructed L2 English learners' declarative and procedural memory systems predict the knowledge of grammatical structures, as measured by tests of implicit and explicit language knowledge?

RQ2: Does the level of difficulty of grammatical structures have a differential effect on the nature of the relationship between the learners' memory systems and their language knowledge?

RQ3: Do L2 instructed learners from different age groups and educational levels show different patterns in (a) their declarative and procedural memory systems, and (b) their linguistic knowledge?

Consistent with the decision taken in section 2.4 invoking the concept of idealisation with regard to the language learning processes and the resulting knowledge, three important postulations in Reber's implicit learning model are relevant to the formulation of the present study's hypotheses: first, that substantial IDs characterise only declarative memory such that procedural memory is not an ability; second, that there is a one-to-one relationship between procedural/declarative memory and knowledge distinctions such that the unconscious, implicit processes are unaffected by, and independent of, explicit cognition; and third, that procedural learning is the process whereby a complex, rule-governed knowledge base is acquired (A. S. Reber, 1989; 1993; A. S. Reber et al., 1991; Kaufman et al., 2010). However, contrary to the first postulation above, and following evidence from recent research (see sections 2.2.3.3 and 2.6), the current study conceptualises both declarative and procedural memory as abilities and investigates the relation of IDs in both memory abilities to the two knowledge types of L2 English grammatical structures varying in complexity. Further, consistent with the second and third postulations in Reber's implicit learning model as pointed out above, procedural and declarative processes were conceptualised as non-interfacing processes, where the acquisition of complex stimulus proceeds procedurally.

Therefore, regarding Research Question 1, the assumption that declarative and procedural learning processes are independent of each other and the fact that there are no pure or direct measures of language knowledge (see sections 2.1.4 and 4.2.4), the following hypotheses are made:

Hypothesis 1a. Declarative memory will strongly predict learners' explicit language knowledge while procedural memory will strongly predict implicit knowledge.

Hypothesis 1b. Declarative memory will be strongly related to learners' performance on the grammatical structures that are explicitly taught while the learners' L2 English knowledge of grammatical structures that are learned in a more implicit manner will be strongly predicted by procedural memory.

Regarding Research Question 2, the fact that there are no pure or direct measures of language knowledge and the assumption that declarative and procedural learning processes are independent of each other, with procedural learning and memory as more effective when learning complex, rule-governed knowledge, the following hypotheses are made:

Hypothesis 2a. Grammatical structures will have a differential effect on the nature of the relationship between the learners' memory systems and their grammar knowledge.

Hypothesis 2b. Declarative memory will be strongly predictive of linguistically simple grammatical structures while procedural memory will be strongly predictive of linguistically complex structures.

Hypothesis 2c. Procedural memory will strongly predict the learners' performance on linguistically complex structures regardless of whether they are explicitly taught or not. However, this relationship will be stronger for the grammatical structures where learning proceeds implicitly than for structures acquired explicitly.

Research Question 3 is largely exploratory. However, one specific prediction about language knowledge is made. It is expected that the younger primary school learners need not to have as much language knowledge (be it declarative or procedural) as the older secondary and university students. Thus, the following hypothesis is made:

Hypothesis 3a. It is hypothesised that learners will differ in their language knowledge. It is expected that as a function of the length of exposure, university and secondary school learners will perform better on both explicit and implicit language knowledge measures than primary school learners.

Finally, it is hoped that the triangulation of measures (see section 3.5) will allow for more sophisticated statistical analyses (namely exploratory factor analysis, confirmatory factor analysis and structural equation modelling) that examine the underlying construct design features of test measures.

CHAPTER 3. METHODS

This chapter provides a detailed description of the design of the study. First, section 3.1 provides the description of the overall research design. The participants are described in Section 3.2. Section 3.3 provides the description of how the four constructs, namely declarative memory, procedural memory, second language knowledge type, and grammar rule complexity, have been operationalised in the study. The subsequent sections in the chapter describe each of the remaining elements of the design in detail as follows: target grammatical structures, the determination of the level of difficulty and exposure type for each structure are described in Section 3.4; Section 3.5 presents test materials, the procedures in data collection and a description of how test materials were piloted; and, finally, Section 3.6 provides the scoring of tests and the description of the statistical analyses that were performed to address each of the research questions.

3.1 Research Design

To investigate the association between the two long-term memory systems and the knowledge of grammatical structures in L2 English learners in an instructed context, the study adopted an *ex post facto* design with the following four constructs: (i) the memory system variable (i.e., the learning ability variable), constituting the declarative and procedural memory systems; (ii) the language knowledge variable, comprising declarative and procedural knowledge of fourteen L2 English grammatical structures; (iii) the linguistic (or grammar rule) complexity (difficulty) variable, consisting of simple and complex grammatical features; (iv) the age-educational level variable, consisting of three groups, namely primary school, secondary school and university L2 English learners; and finally (v) learning condition variable, consisting of explicitly taught and not explicitly taught grammatical structures.

3.2 Participants

A total of 103 learners participated in this study: forty-three learners in their last two years of primary education were recruited from one public primary school (males 20 and females 23, mean age 14.1, SD 2. 05); twenty-nine from one public secondary school in the second and third years of the four-year secondary school education (males 15 and females 14, mean age 16.4, SD 1.78); and lastly, thirty-two from a constituent college of the University of Malawi, in their first, second and third years of their tertiary education (males 22 and females 10, mean age 23.9, SD 3.14). In Malawi, English language instruction begin as soon as the learners begin their year one of primary school at the age of 6 (see section 1.3). The learners were all residing in a small city called Zomba in the Southern Region of Malawi at the time of data collection. The study location was chosen on basis of convenience. In addition to other requirements such as being in primary school, secondary school or university, one other important requirement to participate in the study was that the learner had acquired or was acquiring English in an instructed context with English as a medium of instruction. University students studying courses offered by either the English Department or the African Languages and Linguistics Department were excluded from the study because it was considered that they could be too knowledgeable about English grammar.

Based on background questionnaire information (Appendix A), the data of one participant in the primary school group was excluded from all analyses because she had learned English at a very young age in a naturalistic environment in South Africa. Therefore, data for 102 participants were analysed. The participants were recruited by flyers which were pasted on noticeboards at the respective participating institutions. To participate in the research, each participant, parent of primary school participants, and head-teachers for the participating schools were asked to give their consent by signing a Participant Information Sheet and a Consent Form which were in English.[4] For the primary school participants and their parents, the forms were translated into Chichewa, the Malawi's national language. For their time in participating in the research project, the participants got a compensation of 8 New Zealand Dollars (about 4, 195.24 Malawi Kwacha).

3.3 Operationalisation

Declarative memory ability was operationalised as the participants' performance on four tests: Llama-B, Three-Term Contingency Learning task (3-Term), Differential Aptitudes Test-Verbal Reasoning (DAT-V) and Continuous Visual Memory Test

---

[4] Because these forms were quite many, they have not been included in the appendixes due to limited space. However, they are available upon request.

(CVMT). Llama-B, DAT-V and 3-Term are all verbal measures. CVMT, on the other hand, is a non-linguistic measure. Procedural memory ability was operationalised as the performance of learners on Serial Reaction Time (SRT) and Llama-D. While SRT is non-linguistic, Llama-D is a verbal measure.

Another construct that was targeted in this investigation was language knowledge. Language knowledge consisted of two aspects: language knowledge type and language difficulty. The language knowledge type consisted of declarative and procedural language knowledge types, with the declarative knowledge defined as explicit, conscious, and verbalizable knowledge, and procedural knowledge type as implicit, incidental, unconscious and unverbalizable knowledge. Declarative language knowledge type was operationalised as the participants' performance on an Untimed Grammaticality Judgment Test (UGJT) whereas procedural knowledge type was operationalised as the learners' performance on a Timed Grammaticality Judgment Test (TGJT) and an Elicited Imitation Test (EIT).

Linguistic difficulty or complexity was defined as the objective difficulty that arises from the intrinsic properties of a target L2 feature. The study targeted various grammatical structures of English. Some structures were considered as simple and likely to be more amenable to explicit, conscious learning but less susceptible to implicit learning, and others as complex and therefore likely to be more amenable to implicit, unconscious learning but less susceptible to explicit learning.

3.4 Target Structures

Table 3.1. shows fourteen English grammatical structures targeted in the present study. They comprised seven functors and seven syntactic structures. The functors targeted were (i) plural -*s*, (ii) possessive -*s*, (iii) present progressive –*ing*, (iv) regular past -*ed*, (v) definite article *the*, (vi) 3rd person singular –*s*, and (vii) possessive determiner *his/her*. The targeted syntactic structures comprised (i) passive, (ii) *wh*-question, (iii) pseudo-cleft, (iv) locative inversion, (v) adverb placement, (vi) dative alternation, and (vii) structural parallelism.

Table 3.1. List of Targeted English Functors and Syntactic Structures

No.	Functors	Syntactic Structures
1	plural -*s*	passive
2	possessive -*s*	*wh*-question
3	progressive -*ing*	pseudo-cleft
4	regular past -*ed*	locative inversion
5	definite article *the*	adverb placement
6	3rd person singular -*s*	dative alternation
7	possessive determiner *his/her*	structural parallelism

The decision to include these structures was driven by several reasons. First, the goal of the present study motivated the inclusion of a wide range of grammatical structures. The present study's objective is to explore and propose/present the overview of the educational effects in Malawi. It was thought that the greater range of grammatical structures included in the study the more generalizable the results would be. Second, the structures were chosen to include both morphological (i.e., functors) and syntactic features. Lastly, it was also decided to include a wide range of grammatical structures that are known to cause problems to learners universally and have been included in previous studies such as Goldschneider & DeKeyser (2001), R. Ellis (2005a; 2006; 2009b), Robinson (1997), and Spada and Tomita (2010).

3.4.1 Processability theory (PT)

To determine simple and complex grammatical structures, the present study drew on Processability Theory (PT; Pienemann, 1998a; 1998b; 2005; 2015; Pienemann, Di Biase, & Kawaguchi, 2005; Pienemann & Keßler, 2012; Pienemann & Lenzing, 2015). PT is defined as "a psycholinguistic theory of L2 grammar acquisition that offers an account of the stages learners go through in learning to process L2 morphosyntactic structures" (Buyl & Housen, 2015, p. 525). Not only does PT provide a psycholinguistically plausible framework for the sequences and stages learners go through in L2 grammar acquisition, but it is also a cross-linguistically applicable approach which can account for the entire systems of morphosyntax or grammar (Buyl & Housen, 2015; Pienemman, 2015).

The human language processing architecture forms the centre of PT. Making minimal assumptions about the innate linguistic knowledge where only the basic notion of constituency and the one-to-one mapping of semantic roles (i.e., agent, patient, etc.) are assumed, PT views language learning as a "logico mathematical hypothesis space" which is constrained by the architecture of human language processing computations or routines that are separate from linguistic knowledge (e.g., Pienemman, 2005; 2015; Pienemann & Keßler, 2012). The possible structural options of a grammar will therefore be learned only when the necessary processing resources or computations are available to the learner. In other words, a language learner will produce only those grammatical structures for which the necessary processing resources are available. From this viewpoint, PT holds as follows (e.g., Pienemman, 2005; 2015). First, language acquisition incorporates the gradual acquisition of those very computational routines or procedural skills needed for the processing of the language. Second, the sequence in which the learning unfolds in the learner is determined by the sequence in which the necessary processing routines develop. Third, because several key psychological factors in human language processing are at play, language learning follows describable developmental routes.

PT's approach to language processing mechanisms (or procedures) is based on Levelt's (1989) framework to language production (for details see Pienemann, 2005). In PT, Levelt's (1989) approach to language production is modelled using a typologically and psychologically plausible theory of grammar, namely Lexical-Functional Grammar (LFG; e.g., Bresnan, 2001). Pienemann (2005, p. 15) cites three LFG key features that relate to the procedural accounts of language generation: "(i) the assumption that grammars are lexically driven and (ii) the assumption that functional annotations of phrases (e.g. 'subject of ') assume the status of primitives and (iii) the mechanism of feature matching".

Dalrymple (2001, p. 1) defines LFG as a "linguistic theory which studies the various aspects of linguistic structures and the relations between them". In its current conceptualisation, LFG assumes that the three levels of linguistic representation, namely c(onstituent)-structure, f(unctional)-structure, and a(rgument)-structure and their relations, are relevant to the description of the grammar of language (Dalrymple, 2001). A c-structure consists of word order and phrasal groupings. An f-structure represents grammatical functions like subject, topic, focus and object as well as features such as tense, case, person, and number. An a-structure represents a predicator with its argument roles. Furthermore, LFG is lexical: the linguistic generalisations are explained not in terms of transformations but in terms of a richly structured lexicon (Dalrymple, 2001). Therefore, the c-structure, the f-structure, and the a-structure and their relationships are all determined by this well-structured lexicon.

Within the LFG-based perspective, PT accounts for developmental stages through (1) transfer of information within and between constituents and (2) several mapping processes between levels of linguistic representation (Pienemann, 2015). Terms such as information exchange, feature matching or information transfer are used to refer to feature unification within and between constituents. The process of feature unification concerns the c-structure and ensures the fitting together of different parts that constitute a sentence. Mapping processes connect the semantic roles, constituent structure, and grammatical functions of constituents in a sentence. The following example illustrates how feature and point of unification operate to reflect the time course of real time processing (Pienemann & Keßler, 2012, p. 233):

> In the sentence “He talks” the insertion of the verbal affix “–s” relies on information contained in the subject-noun phrase, namely the features PERS(ON) and NUM(BER) and their values PERS=3 and NUM=SG. These features are unified in [S(entence)] .... In other words, the need to store grammatical information on PERS and NUM during sentence generation illustrates the non-linearity of this morphological process.

PT makes a distinction between individual words that belong to categories such as “noun” and “verb,” and category procedures which are the memory stores that hold grammatical information such as “singular” or “past.” PT posits a time sequence in the matching of grammatical information such that lemma (or no) procedures appear before category procedures, which are assembled before phrase procedures, which are assembled before sentences. Because the processing procedures are implicationally ordered (i.e., every procedure is a prerequisite for the next procedure) and they mirror the time-course in language generation, then language learning unfolds along this hierarchy leaving learners with no choice. Further, because of the implicational nature of the hierarchy, no processing procedures are expected to develop before all other requisite procedures have developed (Pienemann, 2005). PT therefore proposes a

‘Processability Hierarchy’ consisting of hierarchically ranked developmental stages as presented in Table 3.2. While the processing procedures in the Processability Hierarchy are considered universal, “the resulting developmental schedules (i.e. which grammatical structures arise at each stage) are language-specific” (Buyl & Housen, 2015, p. 526).

Table 3.2. Universal processability hierarchy in L2 acquisition as proposed in PT

1. The lemma procedure activates the lexical items.
2. The category procedure accesses the categorical information associated with the activated lemmas.
3. The phrasal procedure builds phrases by unifying information between constituents of the same phrase.
4. The S-procedure exchanges information between phrases in a sentence and accesses the target word order rules.
5. The subordinate clause procedure or S'-clause procedure operates on subordinate clauses, allowing learners to produce target-like word orders which are specific to such clauses.

From Buyl & Housen (2015, p. 526)

Buyl & Housen (2015, p. 527) state as follows regarding PT as a theory of L2 grammar acquisition: “PT is noteworthy for making explicit and falsifiable predictions …, which (a) concern both morphology and syntax, (b) involve both orders and sequences of acquisition, and (c) which are cross-linguistically valid”. It is argued that PT is more comprehensive in scope than other accounts of developmental stages, with empirical support from child and adult learners of different L2s and different L1 backgrounds (for the discussion of such research see e.g., Buyl & Housen, 2015; Pienemann, 2005; Pienemann & Lenzing, 2015).

Simply put, and a very important note for the present study, the foregoing proposals made in PT strongly suggest that the hurdles that L2 learners encounter during acquisition are essentially determined by the processing requirements that intrinsic features of grammatical structures place on the developing human language processing mechanisms. Based on the implicational nature of the Processability Hierarchy as construed in PT, one expects that grammatical features that involve processing procedures at the lower level of the hierarchy are handled by relatively more simplified human language processing computations or routines that develop early on in language production and acquisition process. In contrast, higher level computational routines must consist of relatively complex processing procedures. Therefore, defining an L2 grammatical structure as simple or complex in terms of PT’s processing procedures affords the determination of difficulty or complexity based on the intrinsic properties of the target L2 grammatical structure.

3.4.2 Target Functors

Except for the possessive determiner *his/her*, the other six functors targeted in this study are those that Goldschneider & DeKeyser’s (2001) meta-analysis investigated. In LFG, the structural type which is of relevance to the issue of functors is the c-structure. The c-structure “represents the concrete phrasal expression of sentential constituents, governed by language-particular constraints on word order and phrase structure” (Dalrymple, 2001, p. 1). In many languages such as English, two sorts of categories constitute the c-structure. First are a group of lexical categories, comprising of items such as the Noun, Verb, Adjective, Adverb and Preposition. The second sort of categories consists of functional categories. They include Inflection, the Complementizer, and the Determiner. Dalrymple asserts that these functional categories “play an organizing role in syntax and are either associated with closed-class categories such as complementizers or are filled with subtypes of particular lexical categories” (2001, p. 3). Consequently, feature unification occurs at various points in the c-structure among various items in lexical and functional categories.

The point of feature unification in the c-structure is used to distinguish three groups of structures (Pienemann & Keßler, 2012). The first set involves structures that do not require exchange of grammatical information. Such structures do not involve the unification of features in c-structure. For instance, in the sentence ‘He talked’, the grammatical information ‘past’ need not to be exchanged between the two sentence constituents, the Noun Phrase ‘He’ and the Verb Phrase ‘talked’. The second set of structures involves the exchange of grammatical information within phrases. In the sentence ‘He has two kids’, the information ‘plural’ is exchanged between the determiner ‘two’ and the noun ‘kids’ (Pienemann, 2015, p. 128) within the Noun (or Determiner) Phrase. The last set of structures involves exchange of grammatical information within the sentence. For instance, the sentence “He talks” consists of two elements, namely the Noun Phrase ‘He’ which is the subject of the sentence and the Verb Phrase ‘talks’ the sentence’s predicate. The verbal affix ‘–s’ has four different types of grammatical information: first, the information PERS with the value third participant (PERS=3); second, the information NUM with the value singular (NUM=SG); third, the information TENSE with the value PRESENT; and fourth, the information ASPECT with the value non-progressive. While the TENSE and ASPECT features need not to be exchanged between the two sentence elements, the features PERS and NUM and their values PERS=3 and NUM=SG are initially passed on to the Noun Phrase procedure and the Verb Phrase procedure, and thereafter, the information is passed on to the sentence procedure for feature unification or matching.

These three sets of structures are summarised as below (Pienemann & Keßler, 2012, p. 233):

- No exchange of grammatical information (= no unification of features),
- Exchange of grammatical information within phrase,
- Exchange of grammatical information within sentence.

3.4.2.1 Simple functors

The notion of the ‘point of feature unification’ as described in the preceding section was therefore applied to distinguish between simple and complex functors targeted in the present study. The assumption is that the objective difficult or linguistic complexity of the functors depends on the points of feature unification. Consequently, the following functors were considered simple structures because they do not involve feature unification: the present progressive *–ing*, the plural *-s* as in the noun ‘books’, and the regular past *–ed*. In a lexically driven LFG, these functors contain information grammaticalised in the lexicon or word categories, but almost invariably, this grammatical information need not to spill over to other items in the constituent. Consequently, no feature matching or unification is required in the structures involving these functors.

3.4.2.2 Complex functors

The rest of the functors targeted, namely the plural *–s* as in ‘two kids’, the possessive *-s*, the 3rd person singular *–s*, the possessive determiner *his/her*, and the definite article *the* were considered complex. They involve feature unification either at phrase level or sentence level. How feature unification proceeds for the 3rd person singular –*s* has been detailed above. Feature unification for the plural –*s* in Noun Phrases with determiners, the possessive -*s*, the possessive determiner *his/her*, and the definite article *the* proceed at Noun Phrase level. For Noun Phrases involving the plural -*s* and a determiner for example, the number feature is matched between that of the noun and that of the determiner.

In the PT literature, there are no (clear) descriptions regarding how feature unification proceeds in structures or Noun Phrases containing grammatical features such as the possessive -*s*, the possessive determiner *his/her*, and the definite article *the*. In fact, the proponents of PT treat the possessive –*s* as involving a category procedure (Buyl & Housen, 2015), suggesting that no feature unification takes place between the possessor noun and the noun heading the Noun Phrase. An attempt is therefore made in the present study to propose what features and how the features are matched in the structures involving these functors.

3.4.2.2.1 The definite article *the*

The conceptualisation that the features pertaining to the plural –*s* (e.g., ‘two kids’), the possessive -*s*, the possessive determiner *his/her*, and the definite article *the* are valuated and matched at phrasal level motivates the analysis of the internal structure of a Noun Phrase in English. Literature on the syntax of Noun Phrases demonstrates that the Noun Phrase is in fact a Determiner Phrase (DP), according to which a determiner, and not a noun, heads the phrase. The DP is thought to be the most complex of all the clause elements (Bohnacker, 1997). Figure 3.1 is the simplified structure of DP in English as proposed in the generative tradition.

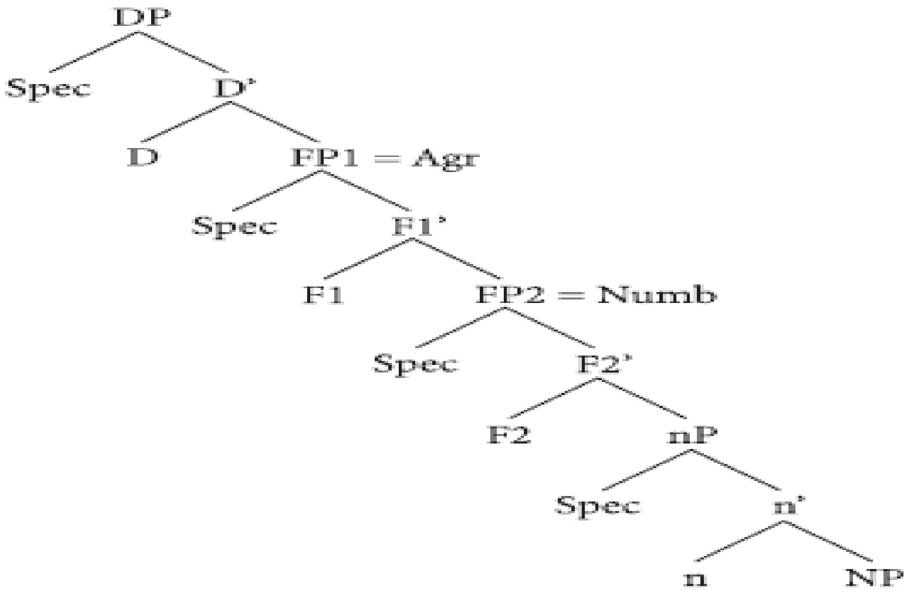


Figure 3.1. The structure of a Determiner Phrase
Adapted from Alexiadou (2004, p. 34)

Three assumptions in the generative tradition (Alexiadou, 2004) are relevant to the proposal made in this study. First is the assumption that only DPs can appear in argument positions in a clause, and the function of D, the head position of DP, is to render the noun phrase into an argument. Second is the assumption that DPs contain semantic features: the projections DP and NumbP encode definiteness and number specification (i.e. plurality), respectively. In the DP projection, D is the locus of definiteness determination. Further, the nominal Infl/Agr is assumed to be "no longer visible' in Modern English DP and that its properties are 'resumed' by D and/or Number" (Alexiadou, 2004, p. 48). The third assumption is that the presence of an overt determiner in D or a demonstrative in Spec,DP renders the DP definite. Based on these assumptions, it can be concluded that Noun Phrases as arguments in clauses are DPs, having their definiteness and number valuated at D, and that DPs are either definite or indefinite depending on whether or not the overt determiner occurs in D, the head position of DP, or in Spec, DP. What this means is that DPs with an overt determiner in D involve the transfer of grammatical information, the definite feature, which must be matched between the determiner and the noun in the DP.

It can be assumed therefore that there is a difference in grammatical features involved in feature matching between Noun Phrases with a numeral determiner such as 'two kids' and those containing the definite article *the*, such as 'the kids/the two kids'. Both are DPs, having both the definiteness and numeral features valuated and matched among the elements of the structure. But those DPs with a numeral determiner are indefinite since the determiner occurs in NumbP, a position below D. Those with an article *the*, however, are definite since the article, which is inherently definite, appears in D. Therefore, grammatical information that is exchanged in DPs with the definite article *the* involves a definite feature. Considering this view from the LFG perspective, it can be assumed that determiners encode such grammatical information as definiteness and number, and that this information is unified in the DP. Thus, the definite article *the* was considered complex in the present study.

3.4.2.2.2 The pronominal possessive determiner *his/her*

Possessors are base generated in Spec,nP of DP (Alexiadou, 2004). They include possessive determiners such as ‘his’ and ‘my’ in the DPs ‘his book’ and ‘my book’, respectively. Such possessive determiners are also known as pronominal possessive pronouns. Functionally, they are like the definite article *the* in English. There is robust evidence demonstrating that in some languages, such as English, German and French, the presence of a pronominal possessive pronoun excludes the presence of a determiner while in languages such as Italian they freely co-occur with determiners (Alexiadou, 2004). This is illustrated in example sentence (1a-d), with examples from English, German, French, and Italian, respectively:

(1) a. (**the*) *my book* *English*
b. (**das*) *mein Buch* *German*
c. (**le*) *mon livre* *French*
d. *il mio libro* *Italian*
the my book
(From Alexiadou, 2004, pp. 31 and 32)

The pronominal possessive pronouns that do not co-occur with determiners such as those in English are termed possessive determiners whereas those that do freely co-occur with determiners are referred to as possessive adjectives (Alexiadou, 2004). The difference in terms of whether a pronominal possessive pronoun co-occurs with determiners has significant semantic repercussion for DPs: “the type of the [pronominal] possessive pronoun used in a language correlates with the definiteness of the whole noun phrase in this language” (Alexiadou, 2004, p. 32.). The whole DP is definite in languages which have possessive determiners and it may be indefinite in languages which have possessive adjectives (Alexiadou, 2004). As pointed out by Schoorlemmer (1998), possessive determiners are definite since they appear either in D or Spec,DP. Therefore, possessive determiners in English are definite just as is the case with the definite article *the*. However, while the definiteness of the article *the* arises from both its inherent and structural characteristics, definiteness in possessive constructions is related only to the structural position the possessor occupies, rather than its inherent feature specification (Alexiadou, 2004). However, relevant to the proposal advanced in the present study is the fact that pronominal possessive determiners are definite. In addition, these possessive determiners encode other grammatical information such as number, gender and person. But, of all these features, only the grammatical information related to definiteness is involved in feature unification with the noun in a DP. This is evidenced by the fact that a singular and feminine possessor can take either a plural or singular noun and vice versa, as illustrated in (2).

(2) a. her book
b. her books
c. their book
d. their books

Though DPs with pronominal possessive determiners are definite as is the case with DPs containing the definite article *the*, the two DP types differ in one important way. Ordinary definite DPs (i.e., those with a definite article) combine the two properties of specificity and uniqueness while article-less possessive DPs are only specific in that they display the definiteness effect but lack the property of uniqueness. For instance, as described by Schoorlemmer (1998, p. 60), “[if something is *my book* it certainly does not mean that this is my only book; however, if something is *the book* it is the only one (in the set that is referred to)”. From this distinction, it follows that only definite possessive DPs (i.e., DPs with a definite article in addition to the possessor) have both specificity and uniqueness properties. Thus, “possessors are not equivalent to definite determiners even when they seem to occupy the position of a determiner” (Schoorlemmer, 1998, p. 60).

This difference in feature properties between the definite article *the* and pronominal possessive determiners may lead to the following regarding the exchange of grammatical information in DPs: in DPs with the definite article, information involving specificity and uniqueness is shared among the elements of the structure; in DPs with a pronominal possessive determiner only specificity is exchanged between the elements, ‘determiner’ and ‘noun’. But the pronominal possessive determiner has number, gender and person features that are not matched with the noun in that DP. The question that arise therefore is how those features are valuated. In cognitive grammar, definiteness is analysed as “residing in the presupposition that the speaker and addressee have each established mental contact with the same target entity or can do so given the content of the noun phrase itself” (Langacker, 1995, p. 63). Therefore, the anaphoric relation between the pronominal possessive determiner *his*/*her* and its antecedent in which number, gender and person features are checked is responsible for the definiteness in the DP the pronominal possessive determiner occurs[5].

3.4.2.2.3 The possessive -*s*

The possessive -*s* is also referred to as the prenominal ‘*s*-genitive (Alexiadou & Wilder, 1998). Some scholars that have treated the English -*s* genitive as a real instantiation of genitive (e.g. Longobardi (1996) cited in Alexiadou, 2005) have proposed approaches that suggest that (in)definiteness does not spread in DPs involving the genitive. Instead, such scholars suggest that, as in Semitic languages with case marking, such as Hebrew, the grammatical information that is exchanged between structural elements e.g. in ‘John’s book’ is Case which is licensed by agreement (Alexiadou, 2005, p. 789).

However, strong evidence points to the fact that the ‘s-genitive DPs in English are distinct from genitive DPs in Semitic languages. Alexiadou notes that the English Noun Phrase underwent numerous changes in the Middle English period, which included (i) the emergence of the non-inflected determiner or article *the*, (ii) possessive -*s* generalisation to all noun classes and to both singular nouns and plural nouns, and (iii) the genitive’s loss of number and gender distinctions. The result of these changes was

---

[5] There are controversies regarding the analysis of pronominal possessive determiners as anaphors. The source of the contention is that pronominal possessive determiners either seem to have some form of constituent-commanding antecedent that binds them in the same local domain or have pragmatically controlled antecedent outside the local domain (Helke, 1970; Safir, 2004).

that the genitive -*s* “increasingly began to behave like a clitic, and not as a genitive suffix” (Alexiadou, 2005, p. 794). Further, the emergence of the non-inflicted determiner *the* resulted in the creation of the only structural position D that encodes (in)definiteness in English. Consequently, to mark definiteness, the structural position D must be filled either by a phrase occupying the specifier or by an article in D.

Literature on generative grammar demonstrates that noun phrases with prenominal ‘s-genitive are DPs: they may be analysed as an instantiation either of D, with the possessor DP in Spec,DP’ or of a projection below D (Alexiadou & Wilder, 1998). Such an analysis of the English ‘s-genitive augurs well with the reported two main classes of ‘s-genitive, namely specifying ‘s-genitives and classifying ‘s-genitives (Quirk, Greenbaum, Leech, Svartvik, & Crystal, 1985; Rosenbach, 2004). Specifying ‘s-genitives are also known as determinative. They function as determiners. In addition, they are in complementary distribution with determiners. Hence, they do not co-occur with determiners. Examples of specifying ‘s-genitive include: ‘Jenny’s (new) desk’, ‘my daughter’s (new) desk’, ‘some people's opinions’, ‘the Italian government's recent decision’ (Quirk et al., 1985, p. 326) and ‘a king’s daughter’ (Rosenbach, 2004, p. 83). In all these examples except the first, the possessor noun is preceded by a determiner. However, both the determiner and the possessor noun are thought to function as the determiner of the superordinate noun in the DP. Rosenbach (2004, p. 81) points out that the determinative function of specifying ‘s-genitives does not only make the possessor to be referential (i.e. to refer to a specific referent) but it also forces a definite interpretation of the whole DP.

Quirk et al. (1985) refer to classifying ‘s-genitives as modifiers since they have a classifying role like that of noun modifiers and some adjective modifiers. For instance, the ‘s-genitive ‘several women’s universities’ may mean ‘several universities for women’ (Quirk et al., 1985, p. 327), and ‘a fisherman’s cottage’ does not refer to a specific referent but rather describes what type of cottage is being referred to (Rosenbach, 2004, p. 81). It is evident that ‘several’ in the DP ‘several women’s universities’ describes universities and not women, suggesting that the ‘s-genitive is not determinative. On the assumption that D is the locus of definiteness, it can be concluded that classifying ‘s-genitives are indefinite and so is the whole DP in which they occur. What this discussion shows is that DPs with ‘s-genitives involve the exchange of grammatical information, namely (in)definiteness feature.

Table 3.6. presents the summary of the present study’s targeted grammatical structures, in terms of their relative linguistic complexity as well as in terms of their exposure type. With regard to the functors, the table provides information on points at which feature unification occurs for each functor and the kind of grammatical information that is exchanged between the elements in the structure in which the functor occurs. The two types of plural -*s* are included: one that is attached to bare DPs (i.e., phrases without determiners) for which the number feature is not exchanged; and one where the number feature is exchanged between the noun and the determiner in the DP.

3.4.3 The Syntactic Structures

The seven target syntactic structures are as follows: adverb placement, *wh*-question, dative alternation, passive, locative inversion, pseudo-cleft, and structural parallelism. As is the case with structures involving functors, there is no agreed upon criterion to determine simple and complex syntactic structures. For instance, Robinson (1996; 1997) used ESL experts to determine simple and complex rules. On the other hand, Spada and Tomita (2010), in their meta-analysis of 41 studies, categorised English grammar rules into simple and complex based on the number of linguistic transformational rules such that the rules involving no transformation were considered simple while those involving one or more transformational rules were considered complex. However, while some studies have shown that there is no direct correspondence between the number of syntactic transformations and the degree of difficulty or ease of acquisition (Goldschneider & DeKeyser 2001), others have reported that experts’ learning-difficulty-judgements’ lack significant predictions about learners’ knowledge (e.g., Silva & Roehr-Brackin, 2016). Therefore, what determines difficulty of syntactic structures remains an open question.

In the absence of a rigorous criteria for determining simple and complex syntactic structures, the present study defined the targeted syntactic structures as either simple or complex based on PT. Specifically, the study drew on two related notions, ‘mapping’ and ‘linguistic linearity’, as construed in PT. Both notions derive from LFG. As pointed out above, LFG, as currently construed, is based on the three independently motivated parallel levels of structural representation: c-structure, f-structure, and a-structure. The active sentence (3, adapted from Pienemann et al., 2005, p. 218) illustrates these three parallel structures.

(3) a. Peter sees the dog.
b. Active Verb: SEE <experiencer theme> → a-structure
SUBJ OBJ → f-structure
Peter the dog → c-structure

In the sentence, the c-structure consists of the verb ‘sees’ and the noun phrases ‘Peter’ and ‘dog’. In its simplified form, the f-structure consists of PRED(ICATE) ‘sees’, SUBJ(ECT) ‘Peter’ and OBJ(ECT) ‘dog’. Finally, the a-structure consists of EXPERIENCER ‘Peter’ and THEME ‘dog’. As illustrated in (3b), the c-structure constituents ‘Peter’ and ‘dog’ map onto f-structure and a-structure as SUBJ and OBJ and as EXPERIENCER and THEME, respectively. These parallel levels of representation must be mapped onto each other. Mapping can be conceived of as the correspondence or the relationship between the elements of the three parallel levels of representation (Pienemann et al., 2005). In relation to syntactic structures, the core of LFG is the understanding that “the mapping of a-structure onto f-structure and the mapping of c-structure onto f-structure is the driving force behind the grammatical formalism” (Pienemann et al., 2005, p. 207). Further, this mapping is conceptualised as either default or non-default. The notion of linearity relates to the ‘defaultness’ of mapping. Default mapping generates linearity whereas non-default mapping generates non-linearity. Syntactic structures involving default mapping were considered simple in the present study, and those that involve non-linearity were considered complex syntactic structures.

Default mapping arises from a one-to-one mapping onto each other of the three parallel levels of representation. Though default mapping is a one-to-one mapping, it is not entirely random. It is determined by several principles. In LFG, the list of grammatical functions contains the following: TOP, FOC, SUBJ, OBJ_1, OBJ_2, OBL, XCOMP, COMP, ADJ[6]. All of these grammatical functions, except TOP, FOC, and ADJ, have

---

[6] TOP=Topic, FOC=Focus, SUBJ=Subject, OBJ_1=Direct/Primary Object, OBJ_2=Secondary/Indirect Object, OBL=Oblique, XCOMP= Predicate complement, COMP=Complement, ADJ=Adjunct
Examples of XCOMP (Bresnan, 2001, p. 265f):

- Mary didn’t sound *ashamed of herself*.

argument functions (Bresnan, 2001), that is, they are governed by a predicate and they map directly to a-structure roles. Among the grammatical functions with argument functions, SUBJ, OBJ$_1$, and OBJ$_2$ are core functions. They are “associated with the central participants of the eventuality expressed by the verb”, and in English, core arguments have canonical c-structure positions (Bresnan, 2001, p. 96). Why core arguments are associated with central participants or have canonical c-structures is explained in terms of the notion of the universal hierarchy of thematic roles. Thematic hierarchy refers to the ordering - from left to right in the thematic hierarchy - of argument roles in a-structure relative to their prominence as given in (4, from Pienemann et al., 2005, p. 215).

(4) Thematic Hierarchy

agent > *beneficiary* > *experiencer/goal* > *instrument* > *patient/theme* > *locative*

Both a set of intrinsic role features and a set of default assignments of features constrain the assignment of theta roles: a default value depends on the thematic hierarchy, and the thematic role that is highest in the hierarchy receives the default value. The highest thematic role in the hierarchy, agent, is the most prominent role to receive the default value. What this means is that the initial argument of a predicator, in other words a SUBJ, is assigned the thematic role of agent. Therefore, the remaining argument(s) will be assigned a theta role that appears lower on the hierarchy and the most probable candidate - with a grammatical function of object - is theme/patient. Consequently, mapping onto each other of the three parallel levels of linguistic representation is one-to-one or linear and canonical in English as illustrated by the predicate ‘killed’ in (5).

(5) a. Peter killed the rat.
b. Active Verb: KILLED <agent theme> → a-structure
SUBJ OBJ → f-structure
Peter the rat → c-structure

- Louise struck me *as a fool*.
- Jogging keeps Susan *in a bad mood*.
- Linda will have your brother *working again*.

Examples of Adjuncts (Bresnan, 2001, p. 265f):

- Mary looked down, *ashamed of herself*.
- Louise enjoyed sports, naturally, *as a Southern California*.
- Susan arrived for lunch, *in a bad mood as usual***.**
- Linda found the money *walking our dog*.

3.4.3.1 Simple syntactic structures

In the present study, the syntactic structures concerning adverb placement and structural parallelism involve linear mapping of the three parallel levels of representation as described above. The core arguments in these structures have canonical c-structure positions as illustrated in (6) and (7) for adverb placement and structural parallelism, respectively:

(6) a. Peter killed the rat slowly.
b. Active Verb: <agent theme> → a-structure
SUBJ OBJ → f-structure
Peter the rat → c-structure

(7) a_1. Either Peter or Prisca killed the rat slowly.
b_1. Active Verb: KILLED <agent theme> → a-structure
SUBJ OBJ → f-structure
Either Peter or Prisca the rat → c-structure

a_2. Peter either killed or chased the rat.
b_2. Active Verb: KILLED/CHASED <agent theme> → a-structure
SUBJ OBJ → f-structure
Peter the rat → c-structure

3.4.3.2 Complex syntactic structures

Evident in the variation of syntactic structures in many languages, and in English in particular, is the fact that the relationship between a-structure, f-structure and c-structure need not be linear. Literature distinguishes two ways in which such linguistic non-linearity is created: (i) from the mapping of c-structure onto f-structure; and (ii) from the mapping of a-structure onto f-structure (Pienemann et al., 2005). Both types of mapping are constrained by grammar. In LGF, the correspondent relationship between c-structure and f-structure is constrained by "general principles for annotating c-structure with functional schemata" Pienemann et al., 2005, p. 208). As noted above, f-structure comprises TOP, FOC, SUBJ, OBJ_1, OBJ_2, OBL, XCOMP, COMP and ADJ. Generally, distinctions amongst these grammatical functions are made on two dimensions. First, on the dimension of whether a grammatical function has an argument function or not, as illustrated above and summarised in (8); and second, on the dimension of whether a grammatical function has discourse function as shown in (9).

(8) Argument functions
TOP, FOC, SUBJ, OBJ$_1$, OBJ$_2$, OBL, XCOMP, COMP, ADJUNCTS

(9) Non-discourse functions
TOP, FOC, SUBJ OBJ$_1$, OBJ$_2$, OBL, XCOMP, COMP, ADJ

(Adapted from Pienemann et al., 2005, pp. 209-210)

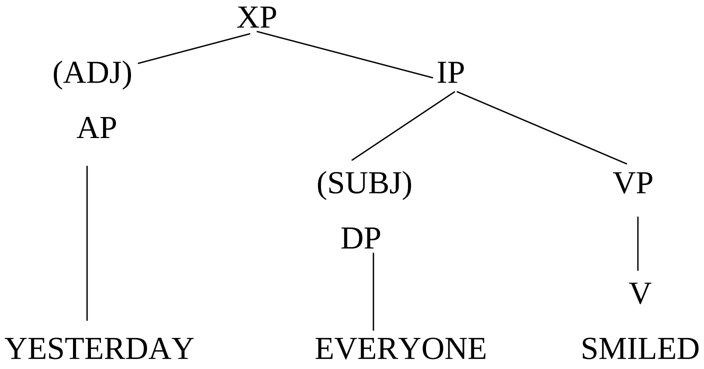
TOP, FOC, and ADJ have non-argument functions as shown in (8). Of all the grammatical functions, only TOP, FOC, and SUBJ have discourse functions as shown in (9).[7] These distinctions are relevant to c-structure onto f-structure mapping and the issue of linearity.

Two correspondence principles are linked to argument-non-argument distinction and discourse-non-discourse distinction. First is the principle that states that specifiers of functional projections are grammaticalised discourse markers, namely TOP, FOC or SUBJ (Pienemann et al., 2005). As illustrated in (10, adapted from Pienemann et al., 2005, p. 211), the specifier of the Inflectional Phrase (IP) is SUBJ and it maps onto the c-structure constituent, the DP, EVERYONE.

---

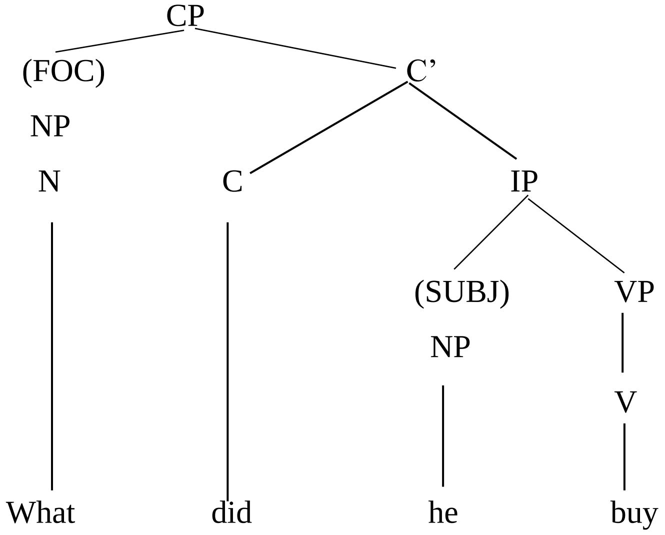
[7] Note that in Lexical Functional Grammar discourse roles such as TOPIC and FOCUS are syntacticised and therefore are represented in f-structure. Why they are syntacticised is demonstrated by the fact that these discourse roles are “subject to syntactic constraints in such cases as English interrogative clauses, cleft constructions and relative clauses” (Pienemann et al., 2005, p. 210).

(10) Yesterday everyone smiled.



The second principle states that constituents that are adjoined to XP are one of the non-argument functions, namely TOP, FOC or ADJUNCT (Pienemann et al., 2005). This principle explains, as illustrated in (10), why the constituent adjoined to XP is ADJ function which is expressed in terms of an Adverbial Phrase (AP). Hence, the c-structure elements YESTERDAY in the initial position and EVERYONE in second position map on to ADJ and SUBJ, respectively. The initial position therefore is filled by a non-subject, an ADJUNCT. This creates non-linearity since the first constituent of c-structure is now mapped onto an adjunct, leading to a non-canonical mapping. The initial position can be filled also by a discourse marker FOC, another non-argument function which can fill XP adjunction. As illustrated in (11, adapted from Pienemann et al., 2005, p. 211), the *Wh*-question constituent 'What' maps onto FOC function and it is in the initial position. The presence of FOC in the initial position, the default position of SUBJ, creates non-linearity.

(11) What did he buy?



Further, another source of non-linearity in English *Wh*-questions arises from the co-reference of the *Wh*-word. In (11), the predicate 'buy' categorises for both SUBJ and OBJ as arguments. This means the *Wh*-word 'What' is not only mapped onto FOC function, but it is also linked to OBJ function. Pienemann et al., (2005, p. 225) point out that "information about the link between [FOC] and OBJ needs to be exchanged between the two grammatical functions, and this information exchange constitutes the non-linearity that is present in English [*Wh*-questions]". Therefore, the non-linearity in the mapping of c-structure onto f-structure arises when grammatical functions with non-argument functions occur at the beginning of a sentence, a default position for SUBJ.

3.4.3.2.1 The *Wh*-question, locative inversion and pseudo-cleft

In the present study, three syntactic structures, *Wh*-question, locative inversion and pseudo-cleft violate the English canonical c-structure due to non-linearity in c-structure onto f-structure mapping. For *Wh*-question, an illustration is provided in example sentence (11). Literature on locative inversion and pseudo-clefts demonstrates that their c-structures are not canonical. In locative inversion, a 'locative phrase is preposed and the subject is postposed' (Bresnan & Kanerva, 1989, p. 2), as illustrated in (12) and (13). In both (12 a-b) and (13 a-b), the adverbial, a locative prepositional phrase, is fronted. The thematic subject of the sentence immediately follows the fronted adverbial but precedes the verb in (a).

(12) John ran *into the house* (Adverbial of direction)
a. *Into the house* John ran
b. *Into the house* ran John
(13) An elm tree stands *in the garden* (Adverbial of position)
a. *In the garden* an elm tree **stands/*is**
b. *In the garden* stands/is an elm tree
(From Celce-Murcia & Larsen-Freeman, 1999, p. 405)

In (b), the subject comes after the verb. Celce-Murcia & Larsen-Freeman observe that moving the adverbial to sentence-initial position gives it greater focus such that it expresses either emphasis or contrast in the discourse. This means that the locative phrase has a grammatical function FOC while appearing in the SUBJ default position. The thematic subject of the sentence appears either in IP's canonical SUBJ position as in (a) or remains in situ in presumably predicate-internal position as a 'light' thematic subject as in (b) (Rizzi & Shlonsky, 2006).

The present study draws on Boeckx's (2007) analysis of pseudo-clefts to motivate non-linearity in the mapping of c-structure onto f-structure in syntactic structures involving pseudo-clefts. In literature, the following three subclasses of pseudo-clefts in English are distinguished: *Wh*-clefts, *Th*-clefts, and *All*-clefts (P. C. Collins, 1991). The present study's focus is on *Wh*-clefts such as the one given in (14b) which, in Boeckx's syntactic analysis, derives from a sentence such as that in (14a). The *Wh*-cleft consists of three main parts: a copula; the post-copula element called 'counterweight'; and the pre-copula material in the form of a '*Wh*-clause' (Boeckx, 2007). In the *Wh*-cleft (14b), the elements 'is', 'in Chicago not in New York', and 'Where Mary and John live' are copula, counterweight and *Wh*-clause, respectively.

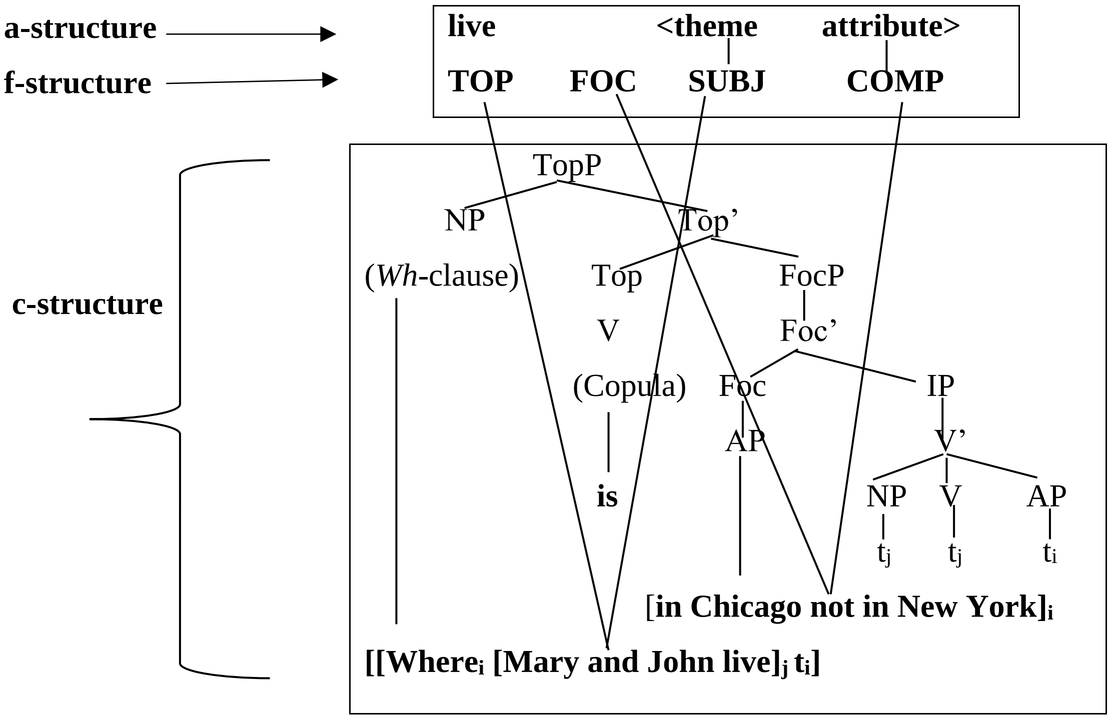
(14) a. Mary and John live in Chicago not in New York
b. Where Mary and John live is in Chicago not in New York

Boeckx (2007) observes that in pseudo-clefts it is the counterweight that receives a 'contrastive' focus. In other words, the counterweight such as 'in Chicago not in New York' in (14b), receives the grammatical function FOC. This process can be illustrated as the movement from IP to FocP. The assumption is that focused material must move to check a [+Foc] feature in Spec,FocP (Boeckx, 2007). Following Rizzi (1997), Boeckx (2007, p. 33) assumes that FocP is dominated by TopP which hosts "the known, given, presupposed part of the sentence". Hence, Spec,TopP should be the most probable landing site of the 'old' or known information, namely the element that remains after the focused element has moved. This element can be illustrated by the element 'Mary and John live' in (14a). Boeckx further assumes that TopP has a 'focus-reinforcing' role which motivates the movement of the remaining element. Consequently, the movement to Spec,TopP of the remaining element reinforces the focal interpretation of the focused element by realizing structurally the informational contrast between the two elements (Boeckx, 2007). Finally, the relativisation process does not only influence the reactivation of a copy of the focused material (usually a *wh*-word) but it also changes the whole TopP into a predicate. Because the copula verb 'be' is functional and predicative, it is assumed to be directly inserted in a functional projection. Since topicalization amounts to the formation of a 'higher' (CP-level, as opposed to the canonical IP-level) predication structure, then the predicative ‘be’ must be inserted in TopP in order to realize the head of TopP (Boeckx, 2007, p. 34). The final structure of the pseudo-cleft is presented in (15).

(15) [TopP [*Where*$_i$ Mary and John live t_i]$_j$ is [FocP *in Chicago not in New York* [IP [VP t_j]]]]
(Adapted from Boeckx, 2007, p. 34)

From the proposal that Boeckx (2007) advances on derivation of *Wh*-clefts, there is evidence to support the non-linearity in the mapping of c-structure onto f-function in pseudo-clefts. The structure of a pseudo-cleft, as illustrated in (15), shows that the cleft’s c-structure is not an IP; the initial element of a cleft is the complex topicalized *Wh*- relative clause whereas the last element is the focused counterweight. All the elements appear above IP. The structure-to-function correspondence principle states that only non-argument functions can fill an adjoining XP (Pienemann et al., 2005). In other words, grammatical functions above IP are non-argument functions. Consequently, both TOP and FOC functions are non-argument, unlike SUBJ function which has an argument function. The topicalized *Wh*-clause in the initial position of pseudo-cleft therefore has a TOPIC function. However, this is the default position of SUBJ. Further, well-formedness conditions require that each a-structure role must be associated with a unique function and that every predicator must have a subject (Pienemann et al., 2005). As illustrated in (16), the linking verb ‘live’ in (15) subcategorises for THEME and ATTRIBUTE, which must be mapped onto f-structure functions SUBJ and COMP, respectively. When focusing and topicalizing in pseudo-clefts take place, the counterweight has a FOC function while the *Wh*-clause has a TOPIC function. Hence focusing and topicalising render SUBJ and COMP positions in c-structure unoccupied. Satisfying unsatisfied arguments in the c-structure involves grammatical information exchange (Pienemann et al., 2005). Therefore, to satisfy the unsatisfied argument (THEME and ATTRIBUTE) functions in (15), grammatical information about the link between TOP and SUBJ functions, on one hand, and between FOC and COMP functions, on the other hand, must be exchanged. This kind of mapping of grammatical functions onto c-structure and the exchange of grammatical information, as shown in (16), creates non-linearity in pseudo-clefts.

(16) Where Mary and John live is in Chicago not in New York



3.4.3.2.2 The passive and dative alternation

The mapping of a-structure onto f-structure also results in linguistic non-linearity. The mapping of a-structure onto f-structure refers to the correspondence between argument structure and functional structure (Pienemann et al., 2005). In LFG, Lexical Mapping Theory explains this mapping of arguments onto grammatical functions on the basis of the assumption that "a-structure contains the lexical information about type and number of arguments that allows it to be mapped onto syntactic structure" (Pienemann et al., 2005, p. 212). Lexical Mapping Theory consists of four components: (a) hierarchically ordered semantic role structures, (b) a classification of syntactic functions, (c) principles of lexical mapping from semantic roles to functions, and (d) well-formedness conditions on lexical forms (Bresnan & Kanerva, 1989). To systematically regulate the type of association possible between argument roles and grammatical functions, Lexical Mapping Theory utilises three lexical mapping principles: (i) intrinsic role classifications, (ii) morpho-lexical operations and (iii) default classifications (Pienemann et al., 2005).

Morpho-lexical operations concern non-default verb forms (e.g. passives and causative constructions) and exceptional lexical entries (e.g. 'receive', 'please') and affect lexical argument structures by adding and suppressing thematic roles (Bresnan & Kanerva, 1989; Pienemann et al., 2005). Pienemann et al. (2005) point out that in both these cases, semantic roles are mapped onto non-default grammatical functions. For instance, the passive suppresses the agent role, the highest thematic role in lexical argument structure. While a theme or patient is promoted to SUBJ function, the agent phrase is realised optionally as an adjunct. Example sentence (17) illustrates this operation. The predicator for the passive form 'seen' in (b) has two arguments, EXPERIENCER and THEME, as is the case with the active form in (a). The passive form however promotes the THEME to SUBJ function, leading to non-default argument-function mapping as in (b).

(17) (a) **Active:** Peter sees a dog
See <experiencer, theme>
SUBJ OBJ

(b) **Passive:** A dog is seen by Peter.
Seen <experiencer, theme>
Ø SUBJ (ADJ)
(From Pienemann et al., 2005, p. 213)

A similar operation where an argument is suppressed, as is the case in the passive, arises from the inherent nature of some lexical entries such as 'please'. The use of the verb 'please' in (18) results in non-default argument-function mapping: the EXPERIENCER 'him' is mapped onto the grammatical object whereas the THEME 'the results' is mapped onto the subject (Pienemann et al., 2005).

(18) The result pleased him.

Some operations add an extra argument to the lexical argument structure. Pienemann et al. (2005) give an example of the ditransitive predicate 'cook for' which has a thematic role of BENEFICIARY (e.g., 'Jane') in addition to AGENT and PATIENT (e.g., 'Peter' and 'dinner', respectively) in (19).

(19) Peter will cook dinner for Jane

It must be noted that cases of morpho-lexical operations, as exemplified above, yield surface grammatical structures similar to canonical c-structure in English. It is however the mapping of a-structure onto f-structure which is non-canonical. The English canonical a-to-f-structure mapping involves an AGENT and a THEME being mapped onto SUBJ function and OBJ function, respectively. In the present study, passives and dative alternation are two sets of structures where non-linearity is a result of the non-canonical mapping of a-structure onto f-structure. In a passive structure as given in (17b) above, the verb form constrains a-to-f structure mapping, yielding non-default mapping in the two levels of linguistic representation. Lexical entries that involve dative alternation are exemplified by the verb 'cook for'. Such verbs subcategorise for three arguments, namely AGENT, PATIENT and BENEFICIARY. As illustrated in (20), some of these verbs such as 'cook' in (a) allow for the alternation of the thematic roles PATIENT and BENEFICIARY while other verbs such as 'explain' in (b) do not allow the alternation.

(20) a. **Cook**: Peter will cook dinner for Jane.
Peter will cook Jane dinner.
b. **Donate**: Peter donated a painting to the museum.
* Peter donated the museum a painting.

The foregoing discussion has shown that adverb placement and structural parallelism involve linear mapping of the three parallel levels of linguistic representation and that they do not alter the canonical c-structure, especially with regard to the position of SUBJ. These two structures were therefore operationalised as simple structures. On the other hand, *wh*-question, locative inversion and pseudo-cleft alter the English canonical c-structure. In these syntactic structures, a focused material or a topicalised element is sentence-initial, a default position of SUBJ. This non-linearity therefore arises from the uncanonical c-to-f structure mapping. Lastly, for the passive and dative structures, despite their canonical c-structure where the initial element of a sentence has SUB function, the a-to-f structure mapping is non-canonical. Common to the syntactic structures involving *wh*-question, locative inversion, pseudo-cleft, passive and dative alternation is the fact that at some level of linguistic representation non-linearity in mapping onto each other of the three-parallel structural representation and exchange of grammatical information come into play. Therefore, these structures were defined as complex. The alterations of the relationship between argument roles and syntactic functions constitute a deviation from default canonical mapping where the non-canonical mapping must be established only by assembling information about the constituents at the S-node (Pienemann et al., 2005). In Table 3.6, the summary of the present study's targeted grammatical structures includes information regarding simple and complex syntactic structures and the linguistic linearity type involved in each structure as discussed above.

3.4.4 Difficulty, complexity and learner groups

An important aspect in the operationalisation of complexity in the present study was to disentangle complexity from difficulty. From the discussion in section 2.3, the following is obvious as regards the use of the term ‘complexity’ in the present study: that complexity is defined, first, in objective, quantitative terms, and second, in terms of a learning problem, not as a desirable aspect of proficiency (see eg., Bulté & Housen, 2012; Housen & Simoens, 2016; Housen, Kuiken, and Vedder, 2012). Thus, complexity (also referred to as *linguistic* or *structural* or *absolute* complexity) has been defined as objective, or inter-individual, difficulty that arises from the target L2 feature’s inherent or intrinsic properties (see section 2.3). As such, a relatively complex grammatical structure should be more cognitively demanding for all language learners, irrespective of their individual learner characteristics. In contrast, difficulty (also known as *cognitive* or *relative* complexity; Housen & Simoens, 2016) has been defined as subjective difficulty which is learner-related as the result of the encounter of language features with the learner’s individual abilities (see section 2.3).

On the basis of the design adopted in the present study, the questions that may arise concern (i) the extent to which one would claim that a simple structure for the most proficient group (i.e., college students) was also simple for the least proficient group (i.e., primary school learners), and (ii) whether a structure may be difficult for (and may not have been acquired by) learners who are not developmentally ready. The criterion used in the present study to determine complexity addresses both questions. First, the determination of complexity involves objective considerations where the difficulty of learning a grammatical structure should arise from intrinsic properties of the structure. This assumes inter-individual difficulty regardless of individual learner characteristics.

Second, the use of PT’s processing computational routines to determine objective complexity of the targeted grammatical structures allows for clear predictions to be made regarding learners’ developmental stages based on the proposed five levels in the universal processability hierarchy (see Table 3.2). Pienemann (2015, p. 129) observes that though there is no standard time that “a learner requires to traverse” the five levels of the processability hierarchy, primary school students “have been demonstrated to reach level 2/3 of this framework (on average) by the end of 2 years of English L2 teaching”. In Malawi, L2 English is learned as a subject from year one of primary school and is used as the medium of instruction from year five of primary school (see section 1.3). Further, the document analysis of the teaching syllabuses in Malawi primary and secondary schools (see section 3.4.5) appear to suggest that primary school Years 7 and 8 L2 English learners in the present study must have already reached level five of the processability hierarchy as proposed in PT.

3.4.5 Exposure conditions

The fourteen grammatical structures targeted in the present study were also differentiated based on whether they are explicitly taught or not. This categorisation was largely based on the document analysis of syllabuses. The present study’s researcher’s experience of the education system in Malawi was complementary to the document analysis. As pointed out in Section 1.3, the researcher of the present study has had experience of how L2 English in Malawi is taught, having done all his primary school through secondary school to tertiary education in the country followed by his English language teaching experience that spans about eight years. This experience was applied to the process of determining the exposure type of the target grammatical structures, especially for the structures which were not explicitly listed in the syllabuses reviewed. The following provides the details of the document analysis that was conducted.

The document analysis involved the review of L2 English teaching syllabuses for primary and secondary schools in Malawi. The syllabuses are public documents which are prepared by the Malawi Government through the Ministry of Education, Science and Technology (MoEST) through the Malawi Institute of Education (MIE). In the introduction to its forward to the publication of the National Education Standards for Malawi, MoEST spells out its mandate as to set and maintain national education standards that not only “specify expected outcomes for students which should be delivered by all education providers in public and private institutions” but also “identify the leadership, management and teaching processes which are essential to the achievement of the outcomes” (MoEST, 2015, p. ii). One way in which MoEST fulfils this mandate is through the developing the curriculum for primary and secondary schools. In addition to selecting and recommending teaching materials such as books, MoEST is also mandated to develop syllabuses for each grade (or class) from primary to secondary school. Syllabuses direct teachers on what and how to teach. For the purposes of the present study, therefore, these syllabuses were the most relevant documents to review to determine grammatical structures that are targeted for teaching in primary and secondary school. Except those enrolled in linguistics and literature courses, all university students sampled should have learned their L2 English grammar before entry to university. Because only students other than those studying linguistics and literature participated in the present research, the analysis of documents was only restricted to primary and secondary school syllabuses.

A total of six teaching syllabuses were reviewed: two syllabuses for the last two years (i.e., seventh and eighth years) of primary school; and all four syllabuses for the four years of secondary school. The primary school syllabuses were collected from the participating primary school with the assistance of a teacher. These syllabuses were reviewed and published by MoEST in 2004 (MIE, 2004a; 2004b). Similarly, the secondary school syllabuses, all reviewed and published by MoEST in 2013 (MIE, 2013a; 2013b), were collected from the participating secondary school with the help of a teacher also. The reason for analysing syllabuses for only the final two years of primary education was on the consideration that, in these last years, students must be exposed to all sorts of English grammar as they prepare for their national examination (i.e., the Primary School Leaving Certificate of Education, PSLCE) in the final eighth year of primary school. To avoid introducing a confounding variable where a structure would be explicitly exposed to secondary school learners but not to primary school learners, the researcher moved back-and-forth between the primary and secondary school syllabuses checking for any evidence of such.

The analysis was a two-stage process. In the first step, the syllabuses were examined for grammatical structures that were explicitly listed for teaching. Entries in the following three columns of the syllabuses were evaluated (see Appendix D): the ‘Success criteria’ column, describing learners expected general learning task or activity; the ‘Theme/topic’, containing the broader subject area; and the ‘Suggested teaching and learning activities’, containing suggested specific teaching and learning activities. The suggested teaching and learning activities provided detailed information about the theme or topic and was the major source of the information used to determine explicitly taught grammatical structures in the present study. As presented in Table 3.3, the definite is explicitly listed in the ‘Suggested teaching and learning activities’ column. As a result, it was categorised as explicitly taught grammatical feature.

Table 3.3. Extract for Suggested Teaching and Learning Activities for Articles in English

Success criteria	**Theme/Topic**	**Suggested teaching and learning activities**
1. Definite articles	Articles	Defining the [term] "articles" Brain storming articles Identifying articles in a text Discussing uses of determiners
2. Identify types of determiners, articles		Identifying types of articles Filling gaps with correct articles Differentiate definite from indefinite articles

Extracted from Secondary School Year One syllabus (MIE, 2013a, p. 77)

In the second step, grammatical structures that were not explicitly listed were scrutinised based on both the general information provided in the syllabuses and the present study’s researcher’s experience. In Tables 3.4 and 3.5 (also see Appendix D), for instance, there is a broader reference to the English language parts of speech and syntactic structures, suggesting that the learners had exposure to a wider range of parts of speech and syntactic structures. The present study’s researcher’s experience was applied in such circumstances to determine the present study’s targeted grammatical structures that were not explicitly listed in a particular syllabus, but which should have, in fact, been explicitly exposed to the learners.

The results of the document analysis of the syllabuses, complemented by the present study’s researcher’s experience, showed that many grammatical structures targeted in the present study were explicitly exposed to the learners (see Table 3.6). Only four out of fourteen grammatical structures appeared to have not been explicitly exposed to the learners (Table 3.6). The ‘explicitly taught’ grammatical structures were defined as those where evidence for instruction was found during the document analysis of the syllabuses while those classified as ‘not explicitly taught’ were defined as those where no evidence for instruction was found.

Table 3.4. Extract Showing a Primary School Syllabus with General Reference to Grammatical Structures

Success criteria	**Theme/Topic**	**Suggested activities**
Learners must be able to:		
1. Describe uses/functions of phrases and clauses	Language structure and grammar	Discussing uses/functions of various parts of speech as used in sentences and passages
2. Identify tenses in sentences	Tenses	Using correct verb tenses on oral and written narratives [e.g.,] Past Perfect Continuous Tense
3. Identify various structures	Sentence structures	Identifying sentences with various structures Discussing sentences with various structures Constructing own sentences using various structures Completing various sentence structures correctly e.g., No sooner had he … [than] …

Extracted from Primary School Year Eight syllabus (MIE, 2004b, pp. 91-92)

Table 3.5. Extract showing a primary school syllabus with general reference to grammatical structures

Success criteria	**Theme/Topic**	**Suggested activities**
Recognise complex sentences	**Simple and complex sentences**	Differentiating simple and complex sentences in different texts. Analysing simple sentences, e.g., into Subject Verb Object, Subject verb Complement Analysing complex sentences into nouns, adjective and adverb clauses
Use simple and complex sentences in oral and written texts		Using simple and complex sentences in oral and written texts Completing dialogues
Use phrasal verbs in oral and written texts	**Phrasal verbs**	Using phrasal verbs, e.g.: pull out, put forward, turn up, Using correctly words and phrases that have similar meanings, e.g.: reach/arrive at; steal/rob

Extracted from Primary School Year Seven syllabus (MIE, 2004b, pp. 71-72)

Table 3.6. The summary of targeted grammatical structures in terms of the level of complexity and the assumed exposure type

Grammatical structures				Complexity		Explicitly Taught
S/N	Structure	Example sentence	Functor	Complex	Type of grammatical information exchange/Mapping and linguistic linearity	
1	Progressive *-ing*	e.g., It is raining.	+	–	No exchange	+
2	Past *-ed*	e.g., Peter killed the rat.	+	–	No exchange	+
3a	Plural *-s*	e.g., Boys …	+	–	No exchange	+
3b	Plural *-s*	e.g., Two boys ….	+	+	Number (plural)	+
4	Adverb placement	e.g., Peter killed the rat slowly.	–	–	Canonical mapping	+
5	Structural parallelism	e.g., Either Peter or Pisca killed the rat.	–	–	Canonical mapping	–
6	Third person singular *-s*	e.g., He talks ….	+	+	NUM (singular); PER (3rd PER)	+
7	Possessive *-s*	e.g., John's books ….	+	+	Definiteness	+
8	Possessive determiner	e.g., His/her books ….	+	+	Definiteness: specificity	+
9	Definite article	e.g., The book ….	+	+	Definiteness: specificity & uniqueness	+
10	Locative inversion	e.g., Into the house John ran/ran John.	–	+	Non-canonical mapping	–
11	*Wh*-question	e.g., What did he buy?	–	+	Non-canonical mapping	+
12	Pseudo-cleft	e.g., Where she lives is here not there.	–	+	Non-canonical mapping	–
13	Dative alternation	e.g., Peter will cook dinner for Jane.	–	+	Non-canonical mapping	–
14	Passive	e.g., The rat was killed by Peter.	–	+	Non-canonical mapping	+

Note. NUM = number; PER = person; + = feature present; – = feature not present

3.5 Materials and procedures

This section presents a description of the data collection materials and procedures for each test. A battery of tests was administered to assess memory or learning abilities and declarative and procedural language knowledge types. The tests used in this study indexed the following abilities: declarative and procedural non-verbal learning, declarative and procedural verbal learning, and declarative and procedural language/verbal knowledge. Table 3.7 is the summary of the assessments used in the current study.

3.5.1 Memory ability measures

Memory ability measures used in the study were six in total. Four of these tested participants' declarative memory ability and two were measures of procedural memory ability. In either category, one measure indexed non-verbal learning and the others indexed verbal learning. The inclusion of both verbal and nonverbal learning domains in this study allowed, as Carpenter (2008, p. 108) points out, the examination of the extent to which L2 abilities are domain specific to language or domain general. First, declarative memory ability measures are described.

Table 3.7. Memory Ability and Language Knowledge Measures Used

Memory Ability Measures	***Procedural Memory***
	Serial Reaction Time
	Llama-D
	Declarative Memory
	Continuous Visual Memory Task
	The Differential Aptitudes, Verbal Reasoning Subtest
	Llama-B
	Three-Term Contingency Learning Task
Language Knowledge Measures	***Declarative Language Knowledge***
	Untimed Grammaticality Judgment Test
	Procedural Language knowledge
	Timed Grammaticality Judgment Test
	Elicited Imitation Task

3.5.2 Procedural memory ability measures

3.5.2.1 The Probabilistic Serial Reaction Times

The Probabilistic Serial Reaction Times test (SRT, Kaufman et al., 2010) was used as the measure of procedural language learning aptitude. The dot played out a sequence of positions that followed a probabilistic order. Two sequences with 12 elements each were used to generate either training or control trials. The two sequences used to generate training (A) and control (B) trials were balanced for simple location and transition frequency (see sample in Figure 3.2, from Kaufman et al., 2010, p. 326). The target sequence chosen was 1-2-1-4-3-2-4-1-3-4-2-3 (Sequence A), while the alternate (or control) sequence was 3-2-3-4-1-2-4-3-1-4-2-1 (Sequence B). Each location (i.e., 1, 2, 3, 4) occurs three times in each sequence, and each possible transition (e.g., 1-2, 1-3, 1-4) occurs once. For contexts made up of a single trial, that is, a single dot location (e.g., 1), the possible successors are the same in the two sequences (e.g., 1-2, 1-3, 1-4). Thus, each location and first-order transition appear with the same likelihood. The two sequences, however, differ in the second-order conditional information that they convey. This means that two consecutive locations lead to a different prediction (i.e., successor) for every sequence. For example, in Sequence A, 1-2 is always followed by 1, whereas in Sequence B, it is always followed by 4. Stimuli were congruent with the target sequence (Sequence A) 85% of the time and intermixed with the alternate sequence (Sequence B) 15% of the time. At any given trial, there was a 85% chance of the next stimulus conforming to sequence A and a 15% chance of it conforming to sequence B. This probability was computed at each given trial (stimulus). For example, because participants were trained in Sequence A, the most likely successor after 4-3 was 2, but on some trials the successor was 1 instead, which is the successor of 4-3 in Sequence B. Learning about this second-order conditional information leads to a difference in responding to training and control trials. Participants typically demonstrate faster response times on trials in which location follows the probable pattern, and slower response times on trials in which location follows the less probable pattern.

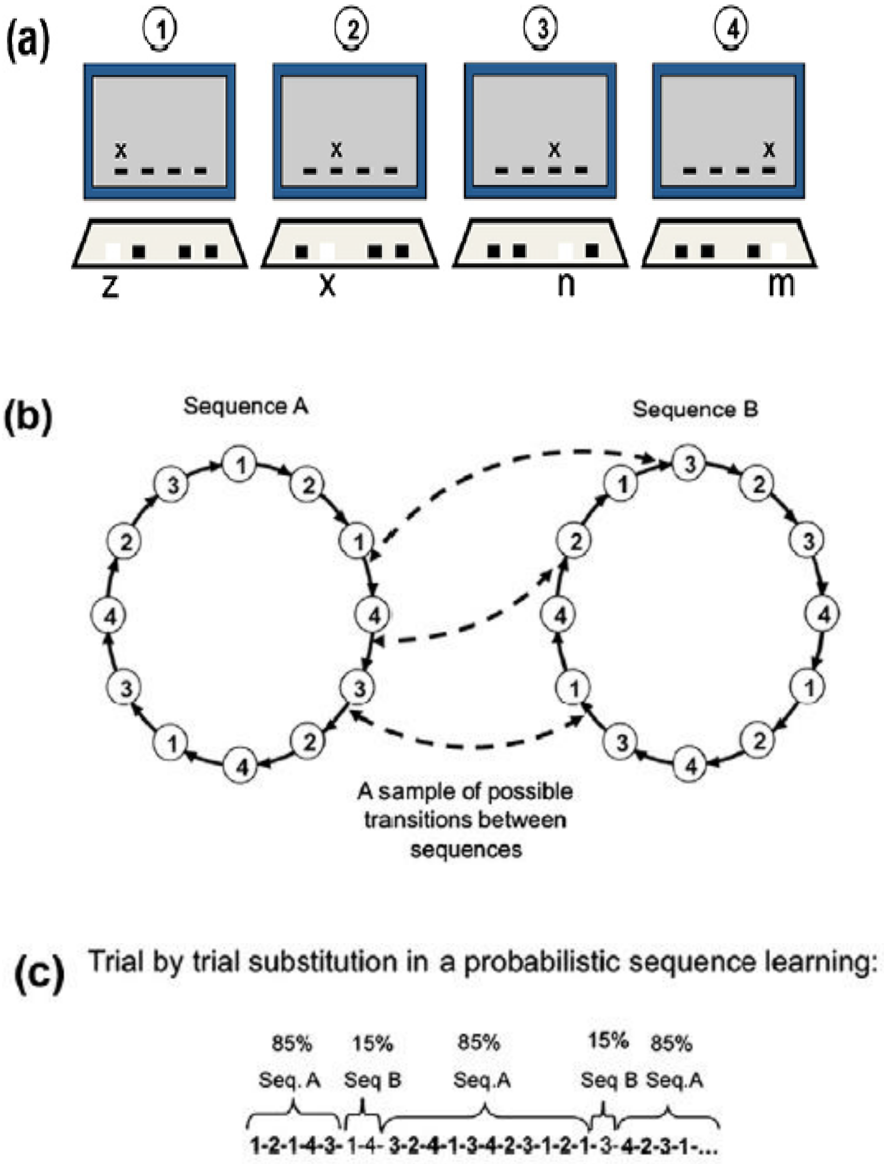


Figure 3.2: Structure of a probability sequence learning. **Note. (a)** Representation of the stimulus and the required keypresses for trial types 1, 2, 3, and 4, respectively. **(b)** Representation of the two 12-trial sequences used to generate either training or control trials for different participants (Sequences A and B). Both sequences are represented as recurrent structures to highlight that the sequences are continuously recycled, and that the starting point is randomly chosen on each block. A sample of possible transitions between sequences is also represented to illustrate that these transitions respect the second-order conditionals of the upcoming sequence. For instance, if a participant is trained with Sequence A and a control trial is scheduled to appear after the series 2–1–4, then the next trial would appear at location 2, which is the successor of the series 1–4 according to Sequence B. **(c)** Representation of the ‘‘trial by trial” substitution procedures employed in these probabilistic sequence learning tasks. Individual trials obeying either the training or the control conditionals are interspersed with each other in a way that respects the transitions between trials as well as the overall likelihood of

each type of trial (e.g., 85% of training vs. 15% of control trials). Notice that most trials obeying the unlikely control sequence would appear isolated, but small groups of control trials are also possible.

Procedure: The SRT task was individually administered on a computer using SRT computer software. The participants saw a visual cue (a dot) appear at one of four prescribed locations on a computer screen. The four locations were four horizontally arranged boxes in the middle of the screen. Participants were instructed to press a key corresponding to the location of the dot on the screen as fast and accurately as possible by placing their middle and index fingers of each hand on the keys marked “V,” “B,” “N” and “M” respectively. Unknown to the participants, the sequence of successive stimuli followed a repeating sequence intermixed 15% of the time with an alternate sequence. No instructions to memorize the series or look for underlying patterns were provided. The task started with a practice block (0) that included 48 trials where probable and improbable transitions had the same probability. That is, the next trial in sequence was equally likely to be determined by Sequence A as by Sequence B. After the practice block, participants completed eight training blocks of 120 trials each, 960 in total, with 15% control and 85% training trials interspersed in every block. There was a break between blocks and the participants had to press the spacebar to proceed to the next block. To maximize “the extent to which individual differences reflect trait differences rather than differences in item order”, all trials were initially randomized within each block and then presented in the same fixed order (Kaufman et al. 2010 p. 326). The task took about 8 minutes to be completed.

The most important part of a set of instructions was as follows:

The task requires you to respond as fast as you can. If you don’t hit the right key, you must try again, until you produce the correct response. After that, the dot will move onto another location, and you should continue the task by pressing on the key corresponding to the next location.

3.5.2.2 Llama-D

Llama-D (Meara, 2005), is a subtest of Llama tests and is a new task that does not appear in the work of Carroll and Sapon (1959). It is a sound recognition task: that is, it is a verbal and auditory task designed to measure the participants’ ability to recognise short stretches of previously heard sound sequences in new sequences. A total of 10 words are played followed by new sound sequences including the 10 sound sequences played a short while previously. The participants’ task is to discriminate those sound sequences previously heard from those not heard before.

Llama-D uses computer generated sound sequences. These sound sequences are based on words which are the names of flowers and natural objects in a British Columbian Indian language. But then these sound sequences have been synthesised using the AT&T Natural Voices (French), a text to speech technology. “This makes for a difficult set of stimuli which are unlikely to be recognised as belonging to any major language family” (Meara, 2005, p. 8). Of all the Llama subtests, Llama-D does not have a default study phase. The lack of the default study phase in this subtest creates an incidental learning condition that minimizes the adoption of explicit strategies (Granena, 2013a). Further, there are proposals that a key skill in language ability is one’s ability to recognise patterns in spoken language which is a prerequisite to word recognition, vocabulary acquisition and the recognition of “the small variations in endings that many languages use to signal grammatical features” (Meara, 2005, p. 8).

***Procedure**:* The task was presented to the participants individually on a computer using Llama software program (see Figure 3.3) and it took each participant about 5 minutes to be completed. The participants listened carefully to the string of 10 sound sequences which were played on a computer once, one after the other. Soon after, a 30-item recognition test was presented, and the participants discriminated between old and new sound sequences. There was no study time. The participants had to click on a happy face, if they recognized the word, or on a sad face if they thought the sound sequence was novel. The most important part of the instruction was as follows (not verbatim):

Your task is to listen carefully to a set of 10 words/sound sequences of a language unfamiliar to you. Immediately after that, you will hear those words alongside other words that you have not heard before. If you think it is a word that you have already heard, click the button with a happy face. Otherwise click on the sad face button. You will score points every time you are right, but you will lose points if you make a wrong judgement. You will hear a ding for a correct answer, and a bleep for a wrong answer. The screen will display your score as you do the test.

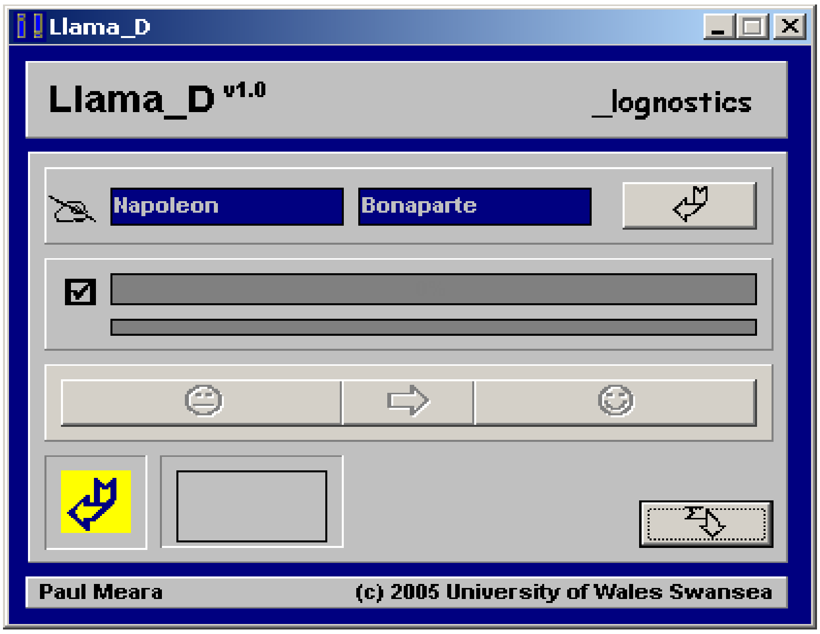


Figure 3.3. Graphic user interface for Llama-D program

3.5.3 Declarative memory ability measures

3.5.3.1 Continuous Visual Memory Test

The Continuous Visual Memory Test (CVMT, Larrabee, Trahan, & Curtiss, 1992; Trahan & Larrabee, 1988) and the version used by Carpenter (2008) and Morgan-Short et al. (2014) was the non-linguistic measure of declarative memory ability in the present investigation. The task indexes the ability to learn and remember visually presented information. It involves learning112 new and old, complex, abstract ambiguous drawings (e.g., 12-point polygons) and irregular nonsense figures not easily susceptible to verbal labelling/encoding which consequently minimises reliance on verbal strategies or knowledge (Morgan-Short et al., 2014). The designs, in the CVMT version used in this investigation, were organised in 2 blocks, with the first block consisting of 16 stimuli and the second block 96 stimuli. The first block stimuli were all new and different. They represented the input stimuli of which 7 appeared, at random intervals, 6 times each in the second block. Whenever these 7 reappeared they were considered "old" drawings since the other remaining drawings were entirely new stimuli as they had never appeared in the task before. The task of testtakers was to identify the recurring stimuli as "old" and the nonrecurring stimuli as "new". In this version of CVMT, the 96 items are scoreable while the first 16 are not scored as they represent the initial presentation of the 7 items that are later repeated. Of the 96 scorable items, 42 are repeating and 54 occur only once. Below are the some of the sample designs that were presented to participants for practice (Figure 3.4).

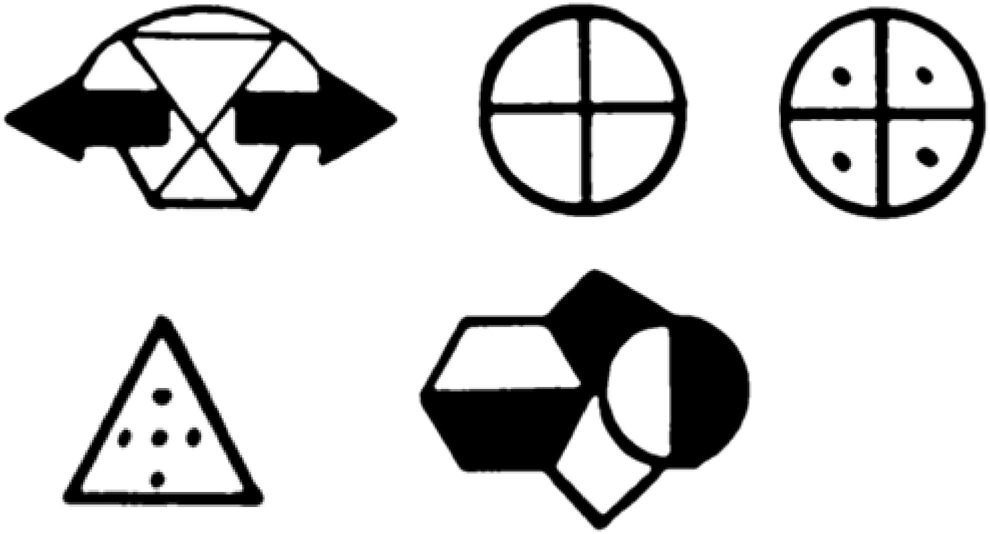


Figure 3.4. Sample CVMT designs

Procedure: The CVMT was individually administered on a computer using DMDX computer software. The participants were told that the task they were about to do was a visual memory task. Then the participants viewed a series of complex, abstract designs on a computer screen presented one after another. Their task was to indicate whether each design was novel ("new") or had appeared previously ("old") by pressing the response keys on the computer keyboard. The "old" items consisted of the seven target designs, presented seven times. These old items were interspersed among 54 "new" distractor items. The "new" and "old" items were presented in a randomized order, which was randomised for each participant. To further minimize verbal encoding, each design was displayed for only two seconds. The stimuli display was immediately followed by the display of the prompt "OLD or NEW?" for two seconds, during which the participants made their response. To indicate that the item was "old" the participants pressed the SHIFT button, and to indicate that the item was "new" they pressed the RIGHT SHIFT button. The prompt "OLD or NEW?" was displayed such that "OLD" appeared to the left and "NEW" to the right to correspond to the respective response keys. The reason was to minimise the confusion regarding the appropriate response key to which a response was mapped. After the participant responded, the presentation advanced to the next trial. The presentation of the items was continuous with no breaks between the blocks. The task began with 11 sample designs for the participants to practise making the responses, that is, indicating whether a design was "old" or "new". The computer program, DMDX, recorded the participants' responses and response times. The responses were used to calculate a CVMT d prime (d') score for each participant. The most important part of the instruction for the CVMT was as follows:

Your task is to try to remember each design you see. As each design is presented please look at it carefully. After you see a design you will need to indicate if the design is NEW or OLD. A design is NEW if you are seeing for the very first time. You will indicate that it is NEW by pressing the RIGHT SHIFT button. An OLD design is one that you have seen before. You will indicate that it is OLD by pressing the SHIFT button.

You should make your response as soon as the design disappears from the screen, and you see the NEW or OLD prompt. You will have 2 seconds to provide your response. Responses made during image presentation will not be recorded. Please wait for the prompt before making your response. The next image will appear after the 2-second response screen.

3.5.3.2 The DAT Verbal Reasoning Test

The verbal reasoning section of the Differential Aptitudes Test (DAT-V, The Psychological Corporation, 1995; Kaufman et al., 2010) is a linguistic (i.e., verbal) measure for psychometric intelligence. The DAT-V task consists of sentences where words are missing at the start and end of the sentences. The participants' task is to choose from answer options a pair of words that are analogically related to the words in the sentence to complete the sentence. Scoring is based on accuracy. The test used in this investigation was adapted to Kaufman et al. (2010).

Procedure: The DAT-V task was administered on paper to either individual participants or a small group ranging from 2 to 6 participants. The participants were told that the task they were about to do was aimed to test how well they can reason with words in English. They were then handed a small booklet containing 40 English sentences with four pairs of answer options for each. They were told that each sentence had two words missing, the first word and the last word. Their task was to choose from the four pairs of answer options the pair that would best complete the sentence. Each answer choice had a pair of words that were analogically related to the words in the sentence in some way. The first word of the correct answer option was to fit at the beginning of the sentence so that the first two words in the sentence were related to each other in a certain way. The second word of the correct answer option was to fit at the end of the sentence so that the second two words in the sentence were related to each other in the same way as the first two words. The participants were asked to encircle the letter of the pair they thought best completed a sentence. After three practice items, university students and secondary school participants had 15 minutes whereas primary school participants had 18 minutes to complete 40 problems. The differences in the time taken to complete the test were based on the results of the piloting of the data collection tools that was done at the beginning of data collection (see section 3.5.9). Two of the three examples included in the test, as well as the explanation regarding the reasoning process to arrive at the correct options to each example, are given below. At the beginning of the tests, the researcher discussed these examples with the participants. Both English and the learners' widely spoken language, Chichewa, were used for explanations.

Example 1: ... is to bark as cat is to ...

Answer options

a) miaow----kitten
b) dog-------miaow
c) dog-------scratch
d) seal--------kitten
e) tree--------scratch

Explanations to arrive at a correct answer

- In order to choose the correct answer for example 1, look carefully at each pair of words.
- Only one pair of words will complete the sentence so that the first two words are related to each other in the same way as the last two words are.
- Answer (a) is wrong because miaow and bark are not related in the same way as cat is related to kitten.
- Answer (c) is wrong because although a dog barks and a cat scratches, the actions are not the same type.
- Answer (d) is wrong because although a seal barks, cat and kitten are not related in the same way.
- Answer (e) is wrong because tree and bark are not related in the same way as cat is related to scratch.

- The answer is (b) because dog is related to bark. The answer is (b) because dog and bark are related in the same way as cat and miaow are.

Example 2: ... is to right as west is to ...

Answer options

a) left-------north
b) direction------east
c) left-------south
d) wrong----direction
e) left--------east

Explanations to arrive at a correct answer

- In order to choose the correct answer for Example 2, look carefully at each pair of words.
- Which pair of words will complete the sentence so that the first two words are related in the same way as the last two words?
- The answer is (e) because left and right are related in the same way as west and east. Left is the opposite of right, and west is the opposite of east.

The most important part of the instruction of the test was as follows

Now it's time for the real test! You will have 15/18 minutes to complete 40 analogy problems. If you have difficulty with a particular problem, try your best answer and move on. Try not to spend too much time on any one particular problem.

3.5.3.3 Llama-B

Llama-B is a recall and recognition test based on picture stimuli. Llama-B is a vocabulary (thus, a verbal) learning task that measures participants' ability to learn relatively large amounts of vocabulary in a relatively short space of time (Meara, 2005). As described in the Llama Manual, the program is loosely based on Carroll and Sapon's (1959) original vocabulary learning subtask, but it uses a completely new interface. In its current version, Llama-B no longer requires any L1 input, so the test is suitable for use with testtakers of any L1. The words that are learned are real words taken from a

Central American language, and they are arbitrarily assigned to the target images (Meara, 2005, p.5). The test has two phases, a training phase and a test phase. The training phase involves memorising object-word pairings. Testtakers are asked to learn words associated with images of imaginary creatures. It lasts for 2 minutes. In the test phase, testtakers are asked to identify a correct image on a computer screen when presented with a word.

Procedure: The task was presented to the participants individually on a computer using the Llama software program (see the program's graphic user interface in Figure 3.5). The following is the most important part of the instruction in the acquisition phase, taken verbatim:

Your task is to learn the names of as many of the twenty objects as you can in 2 minutes. The program will display in the centre of the panel the name of the object that you clicked. You can click the objects as many times as you like, but you should not take notes as you work. The clock in the centre of the main panel shows how much time you have left to complete the task. When your time is up, the program will warn you by playing a bleep sound, and all the main buttons will be deactivated. Then there will be a test soon after.

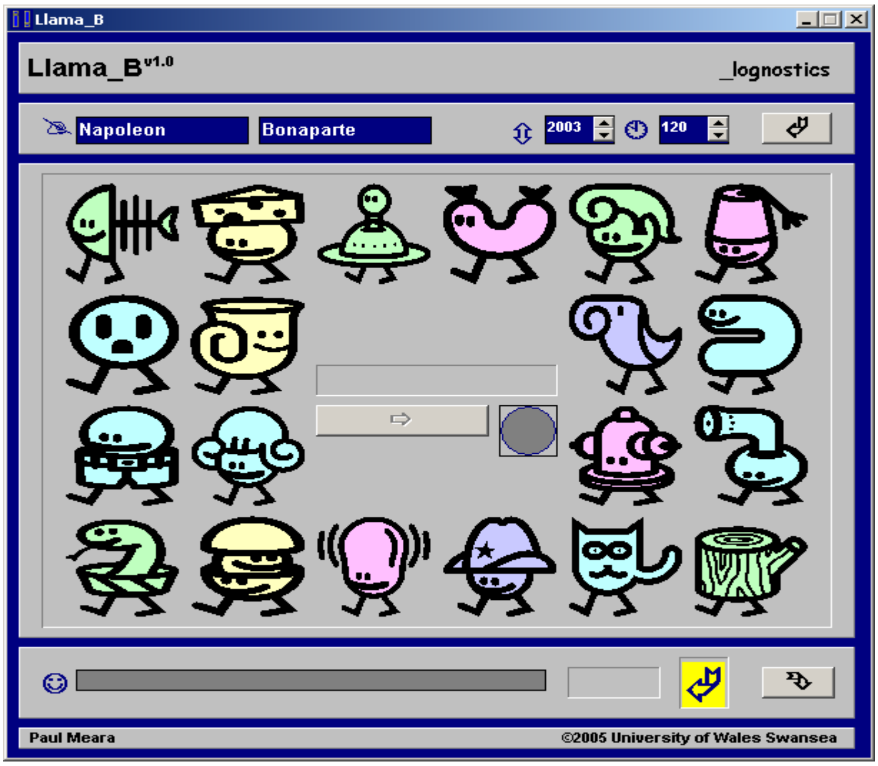


Figure 3.5. Llama-B computer program's graphic user interface

In the test phase, the computer program displayed the name of an image in its central panel and waited for the participants to identify the correct object by clicking on it. Feedback was in the form of a ding for a correct answer, and a bleep for a wrong answer. The participants' scores were displayed on the screen as they worked through the test. Scoring was based on accuracy. Both the training phase and test phase took about 7 minutes to be completed.

3.5.3.4 Three-Term Contingency Learning Task

The Three-Term Contingency Learning Task (the 3-Term, Williams & Pearlberg, 2006; Kaufman et al., 2010) was another linguistic measure of declarative memory ability used in this investigation. The task's goal is to learn the outcome of a specific response in the presence of a specific stimulus (i.e., S: R → O) (Williams and Pearlberg, 2006). The test is illustrated in Table 3.8. Each of the stimulus words (e.g., LAB) is individually presented on a computer screen along with instructions to press a response key (e.g., Key A) upon which a different word (the outcome, e.g., PUN) is produced. Testees are then instructed to press a second key (B), which is followed by a second outcome word (e.g., TRY), and then a third key (C), also followed by an outcome word (e.g., EGG) (Williams and Pearlberg, 2006). This process constitutes learning for each stimulus of three response–outcome contingencies. In the present investigation, the 3-term contingency task consisted of 10 stimulus words each with 3 response options that led to specific outcome words.

Procedure: The 3-Term test was presented to the participants individually on a computer using the 3-Term test software program. In the learning phase, 10 unique stimulus words were successively presented, with each requiring the participants to make 3 associated responses. The participants' task was to learn the outcome word that resulted from each stimulus–response pair. For instance, on one trial the word ‘‘LAB” showed on the screen with the letters ‘‘A”, ‘‘B”, and ‘‘C” listed underneath. When participants selected ‘‘A”, they saw one association (e.g., PUN), when they selected ‘‘B”, they saw a second association (e.g., TRY), and when they selected ‘‘C” they saw a third association (e.g., EGG). The exposure to each association was self-paced (max 2.5 s) with changeover intervals set at 0.2 s. The presentation of 10 stimulus words with the 30 outcome words was immediately followed by a test block. The test blocks were presented in a similar way to the learning blocks with one exception: instead of typing the letters ‘‘A”, ‘‘B”, or ‘‘C” to produce the outcome words on the screen, a stimulus word appeared on the screen along with one of ‘‘A”, ‘‘B”, or ‘‘C”, and participants were required to type in the outcome word corresponding to that stimulus–response pair. For each answer, the participants received feedback. When the response was correct, the phrase ‘You entered: X. You are correct’ appeared on the screen. If a participant entered an incorrect response, the phrase ‘You entered: Y. The correct word was X’ appeared on the screen. The participants’ answer and feedback remained on the screen until they pressed “ENTER” on the computer keyboard. All three response alternatives to a stimulus were exposed individually before the next stimulus word was presented. Pressing “ENTER” after the final response initiated a new test trial with the next stimulus word. Once the test block was completed, participants immediately moved to a second learning block in which the same stimulus words were presented in a different order. The task had four learning blocks with each followed immediately by a test block. A point was awarded for a correct response. The possible overall scores ranged from 0 to 120. The task took about 40 minutes to be completed.

Table 3.8. List of 3-term contingencies with outcomes of each stimulus-response combination

Stimulus	Response Alternative		
	A	B	C
LAB	PUN	TRY	EGG
FAR	WEB	ROB	BUG
RUM	SIT	CAN	HER
CUB	PEN	FED	CON
KIN	HAY	DIG	COP
GUM	SAD	PEG	AGE
TON	TAG	ALE	CAP
PEA	SUB	TIE	BED
HOT	DIE	SUN	ROT
HAD	WET	GUN	RUG

From Williams and Pearlberg (2006, p.179). NOTE: Words in columns A, B, and C are the outcomes of the different stimulus response combination

The following is the most important part of the instruction in the test phase:

You will now be tested for comprehension of the associations just presented. One of the stimulus words will appear on the screen along with a letter (a, b or c). Your task is to enter the associated word.

There is a dearth of studies in L2 research using the 3-Term test as a measure of language learning ability. Mackintosh (1998, cited in Williams and Pearlberg, 2006) argued for the existence of a general associative learning system that is largely independent of what is normally regarded as intelligence. Further, Williams and Pearlberg (2006) point out that research has shown that three-term contingencies are basic units of learning. In Kaufman et al. (2010) study, the 3-Term test significantly predicted explicit learning.

3.5.4 Grammar knowledge measures

The three tests namely the EIT, TGJT and UGJT detailed below, were constructed following Ellis (2005a; 2006; 2009c) and Erlam (2006; 2009). Further, the tests were constructed based on 84 test sentences, targeting all the fourteen grammatical structures in this study (see Appendix B). For the TGJT and UGJT, each targeted grammatical feature consisted of six test sentences, evenly divided between grammatical and ungrammatical. The ungrammatical sentences were systematic violations of the grammatical types.

3.5.5 Declarative knowledge test

3.5.5.1 The Untimed Grammaticality Judgment Test

The Untimed Grammaticality Judgment Test (UGJT) was used to test the ability of participants on how to discriminate grammatically correct structures from ungrammatical ones in a condition where there is no time pressure. The participants' task was to judge the grammaticality and to indicate not only the confidence level but also the knowledge attribution of their judgments.

Procedure: The UGJT was untimed and paper-and-pencil delivered. The participants were handed a pencil and the small test booklet containing the 84 sentences printed on both sides in Times New Roman, font size 12. The participants were asked to do the following. Firstly, they were asked to indicate the grammaticality of each of the sentences by selecting either ‘correct’ if the sentence were grammatical or ‘incorrect’ if the sentence were ungrammatical. Secondly, for each sentence, they indicated the degree of the certainty of their judgements on a scale from 0% to 100%, with 0% suggesting no confidence at all and 100% suggesting the highest level of confidence. Lastly, the participants self-reported the knowledge attribution of their judgments by indicating whether they used ‘rule’ or ‘feel’ to make their judgments. The sentences were randomised but presented in a fixed order to all the participants. At the beginning of the test, the participants were presented with six practice sentences and the researcher explained to the participants what their task was. The participants were told to choose ‘feel’ if they felt their response was correct, but they had no idea why, and to choose ‘rule’ if their response was based on some rule they had learned and that they could say it if asked. The participants were told not to change their initial response though they could go back to check if they had missed a test item unanswered. Sample test sentences are in Appendix B.

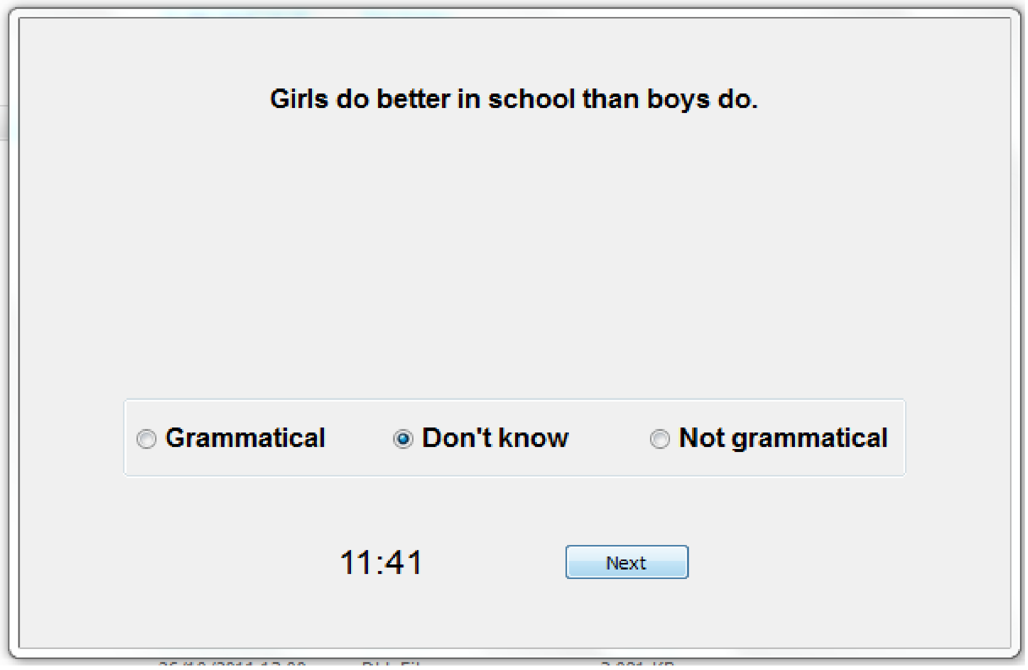
3.5.6 Procedural knowledge tests

3.5.6.1 The Timed Grammaticality Judgment Test

The Timed Grammaticality Judgment Test (TGJT) was computer-administered and timed. The participants were asked to indicate the grammaticality of each sentence by selecting, using a computer mouse, one of the three answer options namely ‘Grammatical’, ‘Don’t Know’ and ‘Not Grammatical’. These were displayed horizontally in that order at the bottom of the display panel. The aim of the task was to test participants’ ability to discriminate grammatically correct structures from ungrammatical ones in a time pressured condition. Figure 3.6 is the display panel of the TGJT computer program.

Procedure: The task was presented to the participants individually on a computer. The participants saw sentences on a computer screen, each displayed once one after the other. Each sentence appeared together with three answer options. To indicate their answer choices, the participants had to click on the options (i) ‘Grammatical’ if the sentence displaying was grammatically correct or (ii) ‘Not grammatical’ if the sentence was not grammatically correct or (iii) ‘Don’t Know’ if they did not know whether the sentence was grammatically correct. After making their response, participants were asked to click on the ‘Next’ button below the answer options to tell the program to move on and display the next sentence. Thus, there was no display time between sentences. The sentences were presented in a randomized order, which was constant for all the participants. At the beginning of the test, the participants were encouraged to judge the sentences based on their feeling as to whether a sentence was grammatical or not. Then the participants had six practice sentences. The participants were told that the time limit to complete the test was 13 minutes and that there were no breaks. They were then asked to respond as quickly as possible in order to complete the test. The time was determined in the piloting of the data collection tools (see section 3.5.9). An important part of the instruction was as follows (verbatim):

In this task you will have to read English sentences. You must decide if the sentence is GRAMMATICAL or if it is NOT GRAMMATICAL. If you don't know, then you can use the 'Don't Know' option. Indicate your choice by clicking the button next to your choice and then click ‘next’.



Girls do better in school than boys do.

○ Grammatical ● Don't know ○ Not grammatical

Figure 3.6. Display panel of the TGJT computer program

3.5.6.2 Elicited Imitation Test

The Elicited Imitation Test (EIT) was designed based on Erlam (2006). The test consisted of 57 items targeting 14 English grammar structures (see Appendix B). Each target structure was tested using a minimum of three items. The plural ‘-s’ and the third person ‘-s’ had the largest number of test items, with a total of ten items for the plural ‘-s’ and a total of 8 items for the third person ‘-s’. For each structure, there were two categories of items: grammatical items and ungrammatical items (Table 3.9). The grammatical items were aimed to test participants’ ability to repeat grammatically correct structures. The ungrammatical items aimed to test participants’ ability to correct ungrammatical structures. The three-seconds-time interval between the stimulus presentation and the elicited response was an approach that was used to minimise ‘parroting’ both with short and long sentences (see section 3.5.9.2). By design, the task aimed to test the participants’ English language ability of both simple and complex structures where length was an intrinsic feature of some, for instance, pseudo cleft. Thus, sentence length was not controlled. Further, the design of the EIT in the present study differed from the Erlam’s (2006) version in one important way: though it was time-pressured, it did not include the semantic aspect although it used the same test sentences as those in the original version. During the piloting of the test, done mainly to determine the quality of the test’s recorded audio and the time that it would take participants to complete the test (see section 3.5.9), it was discovered that primary school students found it difficult to repeat the sentence they had heard just before responding ‘true’ or ‘false’. Lastly, in the present study, the participants were not tested on their hearing problems for the purposes of this task.

Procedure: The task was orally presented to the participants individually using a computer software program. The participants listened carefully to 42 sentences containing grammatical and ungrammatical structures of the fourteen English grammar rules targeted in this study. Each sentence was played once on a computer using a woman’s voice with an American English accent. After each sentence, there was a three-seconds gap followed by a beep. The participants repeated the sentence soon after hearing the beep to minimise ‘parroting’ that could result from repeating the sentences soon after they were heard. The participants had nine seconds to repeat each sentence. The time pressure was intended to minimise participants’ reliance on their conscious knowledge as they repeated the sentences. The first six sentences were practice sentences to enable the participants to get used to the task. There was no break between the practice sentences and the test sentences. The EIT audio was presented via a laptop and the participants' responses were recorded using an Android mobile phone voice recorder application. The researcher made sure that neither the research assistant nor the participants paused the audio at any time. The most important part of the instruction was as follows (verbatim):

Repetition task. For this task you will be asked to first repeat sentences in English. Please pay careful attention to the instructions on the recording. Please do not take any notes during this exercise.

Now, let's begin. You are going to hear several sentences in English. After each sentence, there will be a short pause, followed by a tone sound [BEEP]. Your task is to try to repeat exactly what you hear in English. You will be given sufficient time to repeat the sentence. REPEAT AS MUCH AS YOU CAN. Remember, DON'T START REPEATING THE SENTENCE UNTIL YOU HEAR THE TONE SOUND. Now let's begin.

Table 3.9. Number of Grammatical and Ungrammatical Test Items in EIT

Rule	Grammatical	Ungrammatical	Total
Progressive '-ing'	1	2	3
Past '-ed'	2	2	4
Plural '–s'	8	2	10
Adverb placement	3	2	5
Structural parallelism	1	2	3
3rd person singular '–s'	5	3	8
Possessive '–s'	1	2	3
Possessive determiner 'her/his'	1	2	3
Definite article 'the'	1	2	3
Locative inversion	1	2	3
Wh-question	1	2	3
Pseudo-cleft	1	2	3
Dative alternation	1	2	3
Passive	1	2	3
Total	**28**	**29**	**57**

3.5.7 Background Questionnaire

The background questionnaire included questions about the participants' age, gender, and their language background and use (see Appendix A).

3.5.8 General procedure

The total time to complete all nine assessments took approximately four hours. The background questionnaire was completed on a different day. Two of the tests were paper-and-pencil. Except for the language knowledge and the DAT-V tests, all the other tests were randomised for each participant. The TGJT, EIT and UGJT were, however, administered in that fixed order. The reason was to minimise the effect of participants' reliance on explicit knowledge during the TGJT and EIT, which might have been the case had they started with the UGJT.

Testing proceeded in small groups. The group size for computer testing was limited by the number of computers available, with groups ranging between six and eight participants. Each participant was provided with a chair, desk, and a Windows PC. Testing was conducted in a computer room or at a quiet open space at the respective institutions of the participating students. Two research assistants helped the researcher to administer the tests. The EIT task was the only task that each participant completed one on one with the researcher or the research assistant in a secluded and quiet place as it required participants to repeat the sentence aloud so that they could be recorded. For all computer-administered tests, all participants were trained before the start of any test on key presses and how to use the mouse. Similarly, the paper-and-pencil tests were administered in small groups of participants, ranging between two and five in many cases. A pen or pencil was provided to participants for answering. The instructions in all tests were given both in English and in Chichewa (i.e., the local language of the participants) in both aural and written form. The participants could take a break.

The primary school group was the first to be tested in the first week of data collection. In the following two weeks, data were collected from the secondary school and university students. There was no alternative to beginning data collection with the primary school students. By the time data collection commenced, the primary school grade eight participants had only two weeks left before they were to start writing their final primary school national examinations.

3.5.9 Piloting of test materials

Before data collection, all the test materials were piloted in the months of March and April 2017. The purpose of the pilot study was mainly two-fold. First, it was to trial the test materials. As described above, a total of six tests (i.e.., SRT, Llama-D, CVMT, Llama-B, 3-Term and TGJT) were not only computer-administered but had the participants’ responses recorded by computer. The EIT was computer-administered while participants’ responses were audio recorded using a mobile phone voice recorder. Further, the CVMT was adapted to a new software (see section 3.5.5) whereas the EIT was administered using a computer program with a woman’s voice of American accent (see section 3.5.6). It was therefore imperative to test whether all these data collection tools would work and run on a computer as well as they were designed. Second, the pilot study aimed to determine the time it would take the participants to complete the TGJT, EIT and DAT-V tests. Like the DAT-V test, the TGJT had no display time between sentences (see sections 3.5.3 and 3.5.6). Thus, for both tests, there was a need to determine the total time that the participants would require to complete them. For the EIT, however, the pilot study was aimed at determining the response time to individual test sentences due to the continuous running of the computer program that was used. While all the nine tests in the study were trialled, the pilot study aimed to answer the following two specific questions:

1. How long will participants require to complete the DAT-V, TGJT and EIT tests?
2. Will all the computer test programs run and operate properly as designed?

3.5.9.1 Participants

A sample of 18 L2 English learners participated in the pilot study. These participants were comparable to the participants in the actual study in terms of language background, the length of exposure to L2 English and the study location. They consisted of 4 primary school students (male 1 and females 3, mean age 12.5, SD 1.80); 9 secondary school students (males 8 and female 1, mean age 15.4, SD 1.55), and 5 university students (all males, mean age 24.99, SD 1.85). The participants were drawn from three different primary schools and three secondary schools which did not participate in the actual study. However, because the participants in both the actual and pilot studies were drawn from the same university, those university students who took part in the pilot study did not participate in the actual study. The participants were recruited on basis of convenience. For their time in participating in the pilot study, the participants got a compensation of around 4 New Zealand Dollars (2000.00 Malawi Kwacha).

3.5.9.2 Materials and the resulting changes

All the nine instruments as described in Sections 3.5.1 – 3.5.6 were trialled: the SRT and Llama-D as tests for procedural memory; the CVMT, DAT-V, Llama-B and 3-Terms tests as measures of declarative memory; the TGJT and EIT tests as measures of procedural language knowledge; and the UGJT test as the measure of declarative language knowledge. Except for the DAT-V, TGJT and EIT tests, the design features of each test were as described above in Sections 3.5.1 – 3.5.6. During piloting, the general procedure in the administration of the tests was as described in Section 3.5.8, with the TGJT, EIT and UGJT administered in that fixed order. Testing took about three weeks. The two specific questions of the pilot study were assessed in various ways: first, by observing how the computer programs were running as well as whether the participants were following the instructions given to them at the beginning of each test; second, by evaluating the background operations of the computer programs after tests were administered; and finally, by imputing the average time the participants took to complete a test. The following reports and discusses the findings and changes to the tests in relation to the two specific questions.

3.5.9.2.1 Specific question 1

The first specific question asked how long it would take the participants to complete the DAT-V, TGJT and EIT tests. To establish the time limit of each of these tests, the researcher first observed the participants as they performed each test. After all the data were collected, the researcher calculated - where necessary - the average time the participants took. What follows is the detailed account of these analyses.

The DAT-V test

The DAT-V test in the pilot study had all other design features as described in Section 3.5.3.2 except that the participants were required to complete the test in 15 minutes. The time limit was established following Kaufman et al. (2010). In Kaufman et al. (2010), participants aged 16-18 and in their last two years of secondary education in Cambridge, England, were given 15 minutes to complete the 40 problems. During the administration of the test, the participants were reminded twice about the time they had already taken completing the test: first, after seven minutes, then when only five minutes was left. This was intended to encourage the participants to work as quickly as possible until they completed the test.

During the administration of the DAT-V test, the researcher checked and recorded the time when each participant completed the test. On several occasions, the researcher observed that participants were not able to complete the test within the provisional 15-minute time limit they were allowed. After data collection, it was observed that none of the primary school participants had completed the test within the provisional time limit. This was an indication that the younger participants must have taken a relatively longer time to read the test sentences and the answer options before they actually gave their responses by encircling the correct option on the test sheet. The primary school students were found to have taken an average of 18 minutes to complete the test. Secondary school and university students. Therefore, as a result of the pilot, a decision was then made to have two separate time limits, one for the primary school participants and the other for the secondary school and university students. The secondary school and university students were allowed the time limit of 15 minutes, as provisionally determined in Section 3.5.9.2 above.

The TGJT

Like the DAT-V test, the only design feature that distinguished the original TGJT version in the pilot study from the version in the actual study was the time required for the participants to complete the test. A total of 20 minutes was provided to complete the test at the trialling stage.

There was no easy way of establishing the actual time required to complete the TGJT especially because no control group of English native speakers was included in the study. In Ellis' (2005; see also Ellis 2006; 2009) study, university L2 English learners were allowed between 1.8 to 6.24 seconds to judge individual sentences in the TGJT. Ellis established the time limit for each sentence by timing English native speakers' performance on the sentences in a pilot study and then calculating an average response time for each sentence. Due to the slower processing speed of L2 learners, Ellis added an additional 20% of the time taken for each sentence. In the present study, the TGJT included some grammatical structures and test sentences that were not studied in Ellis' studies (2005; 2006; 2009; see section 3.4). Thus, the need to trial the test and establish the time limit for the present study. The provisional 20 minutes time limit was only an approximation. This time limit was about two times the product of the total number of the test sentences and the highest response time of 6.24 seconds as reported in Ellis (2005; 2006; 2009). To encourage the participants to respond as quickly as possible, the researcher told them that the test was timed but that there were many test sentences to respond to, without disclosing the actual number of the test sentences.

In the TGJT, the computer program recorded the reaction times (RTs) each participant took to respond to each of the 84 test sentences. To establish the time that the participants in the actual study would take to complete the test, the researcher simply averaged the RTs. This calculation resulted in the 13-minute time limit. As a result of the pilot, the participations in the actual study were allowed 13 minutes to complete the TGJT.

The EIT

There was only one difference between the original version of the EIT test used in the pilot study and the EIT version in the actual study (see section 3.5.6.2). In contrast to the EIT version used in the actual study, the original version of the EIT was designed to include the semantic aspect intended to focus the participants' attention on meaning. The participants were required to say first whether they agreed with, disagreed with, or were not sure about the content of each statement they heard. Thus, following Erlam (2006), the test sentences were constructed as statements where the participants could agree or disagree with. The participants were allowed three seconds to make their response before they heard a beep. McDade et al. (1982, cited in Erlam 2006 p. 469) ‘found that participants could repeat sentences they did not understand as long as imitation was immediate, but that after a three-seconds delay they were unable to’. Immediately after the beep, the participants were allowed nine seconds to repeat the statement they heard. When the nine seconds elapsed, the recording moved to the next statement. The pilot study was aimed at establishing the response time for the actual study. The following is the most important part of the instruction of the original EIT version used in the pilot study:

You are going to hear several sentences in English. Immediately after each sentence, you will have to state whether the statement you just heard is TRUE, NOT TRUE or NOT SURE. You will have a short time to give your answer. Soon after, you will hear a tone sound. Your task is to try to repeat exactly what you hear in English. You will be given sufficient time to repeat the sentence. REPEAT AS MUCH AS YOU CAN. Remember, DON'T START REPEATING THE SENTENCE UNTIL YOU HEAR THE TONE SOUND. Now let's begin.

The design of this original version of the EIT allowed for two response times: one, the three-seconds response time where the participants were required to give their belief response; and one, the nine-seconds response time in which they had to repeat the statement they heard. The assessment during and after recording suggested the nine-seconds time limit for repeating the statements was too short. The participants, more especially the primary school students were heard attempting to repeat the previously heard sentence while the test program had already moved to the next sentence. However, many who found the nine-seconds time limit not enough and complained that the woman in the test had not given them enough time to remember and repeat the sentences were observed to have hesitated and delayed beginning to repeat (i.e., they began to repeat some seconds after the beep was heard). All the primary school participants were observed to have this problem. When the researcher tried the EIT version without the semantic aspect (i.e., without asking the participants to provide their belief response) with the four primary school participants, tremendous improvements in their responses were observed. This finding suggested that the nine-seconds time limit for repeating the sentences was enough time for them to repeat the sentences. The finding further suggested that the EIT version without the semantic aspect would be best for the younger learners. Against this background, a decision was made to use the EIT version with no semantic aspect for all participants in the actual study. However, in retrospect, excluding the semantic aspect from the EIT was not the best decision because it removed a very important design feature of the test as the measure of procedural language knowledge. Although research suggests that after a three-seconds delay, participants are unable to imitate sentences they did not understand (McDade et al. 1982, cited in Erlam, 2006), the lack of the semantic aspect in the EIT in the present study should have negative implications on the validity of the test.

3.5.9.2.2 Specific question 2

The second specific question asked whether all the computer test programs would run and operate properly as designed. To assess this this specific question, the researcher observed how the computer program was running as well as whether the participants were following the instructions given at the beginning of each test. Soon after each test, participants were asked to describe how well they felt they had performed in the test and whether they faced any problems. (The debriefing was unfortunately not recorded.) The researcher also checked whether each computer test program had recorded each participant’s responses. The results of these analyses are as follows. It was found that all test programs were running smoothly. The assessment of the computer tests after the participants had completed the tests showed that each individual computer test program had automatically computed, recorded and saved each participant’s responses. However, the researcher’s observation during the performance of the EIT and the assessment of the recorded EIT responses revealed one general trend. All the younger participants, namely the primary school students, were observed to have difficulties in repeating the sentences they had heard before indicating their belief about the statement.

During debriefing, the participants said, on the one hand, that they enjoyed completing the UGJT most and that they wanted more of the test. On the other hand, the participants reported to have found the 3-Term test the hardest followed by the Llama-D test. They further complained that the ‘woman’ in the EIT (the EIT was recorded using a woman’s voice; see section 3.5.6.2) did not give them enough time to repeat the sentences they heard.

In conclusion, the results of the pilot study resulted in several important decisions, including determination of the time limits for the DAT-V, TGJT and the EIT. These time limits were used in the actual study. The pilot study also led to the decision to exclude the semantic aspect from the EIT in the actual study.

3.6 Data analysis

This section provides the details of the methods of analysis used in the research. Section 3.6.1 details the scoring of test data, beginning with memory ability tests. Section 3.6.2 presents the description of the statistical procedures used.

3.6.1 Scoring of Memory Ability Measures

The scoring of the data from the two procedural memory ability measures is described first.

3.6.1.1 Serial Reaction Times (SRT)

The SRT program automatically recorded accuracy and RTs (in milliseconds, ms) the participants took to press the correct key corresponding to the location of the visual stimulus. Scoring involved first assessing learning on the SRT. All data were combined as a single group and then scored. Data for 5 participants were missing. The computer program did not save the participants’ responses (accuracy scores) and RTs. All the participants with missing data were primary school students. The problem with the power supply (both from the small genset used and the national grid) experienced at one point on the day of data collection for this group might have caused the computer program to malfunction and fail to save the data. Thus, SRT data for 97 participants were scored. The scoring procedure began with discarding 10.4 % of trials on which participants gave incorrect responses. The high rate of error responses might have resulted from the fact that many primary school and secondary school students were not very computer literate. The SRT task required rapid pressing of the computer keys, which could lead to more errors for participants with little experience in using a computer.

The next step was to determine the cut-off to remove trials with ‘very’ short and long RTs before discarding outliers more than three standard deviations from the mean as computed individually for each block and participant. While there is no consensus regarding what the best cut-off for short and long RTs should be, there is no disputing the fact that some RTs are not indicative of reliable responses. Wall (2012) points out that most people cannot push a button in less than about 300ms in response to, say, a visual stimulus. This suggests that short RTs of, for example, less than 250ms would probably be initiated *before* the stimulus. He further notes that, very broadly, the mean RT for most ‘simple’ tasks tends to be around 400-700ms so that RTs longer than about 1000ms might reflect some other kind of process. For example, they might reflect the fact that the participant was bored, distracted, or temporarily forgot which button to push. He therefore recommends removing RTs that fall below 250ms, or above 1000ms. Whelan (2008) recommends eliminating very fast outliers using the cut-off of between 100ms and 200ms. This is based on the findings in Luce (1986, cited in Whelan 2008, p. 476) suggesting that genuine RTs have a minimum value of at least 100ms, that is, the “time needed for the physiological processes such as stimulus perception and for motor responses”. Lastly, Baayen & Milin (2010) argue that button presses within 5ms of stimulus onset reflect physically impossible short latencies while RTs exceeding 5 seconds in, for instance, a visual lexical decision task with unimpaired undergraduate subjects reflect absurdly long RTs that must be excluded. In the present study, the SRT task was simple, not requiring, for instance, a lexical decision. However, some of the participants were not only young (in primary and secondary school) but also not computer literate. These conditions could have conspired and led to relatively slow latencies, especially with primary and secondary school participants. Against this background, the cut-off allowing elimination of RTs less than 300ms and those longer than 3000ms was adopted. This procedure accounted for 3.4% of all the correct responses.

The distribution of the remaining RTs was examined. As shown in Figure 3.7, the RTs were not normally distributed, but substantially skewed to the right. A decision was made to transform the RTs using the log 10 transformation method. Larson-Hall (2010) has gathered Tabachnick & Fidell’s (2001) recommendations regarding the kinds of transformations to employ when data are not normally distributed. Log 10 transformation is recommended when data are substantially positively skewed. For the present study, the results of the log 10 transformation, as presented in Figure 3.8, showed that the RTs were now approximately normally distributed. With the transformed RTs, outliers more than three standard deviations from the mean which were computed individually for each block and participant were discarded, accounting for only 1.0% of trials.

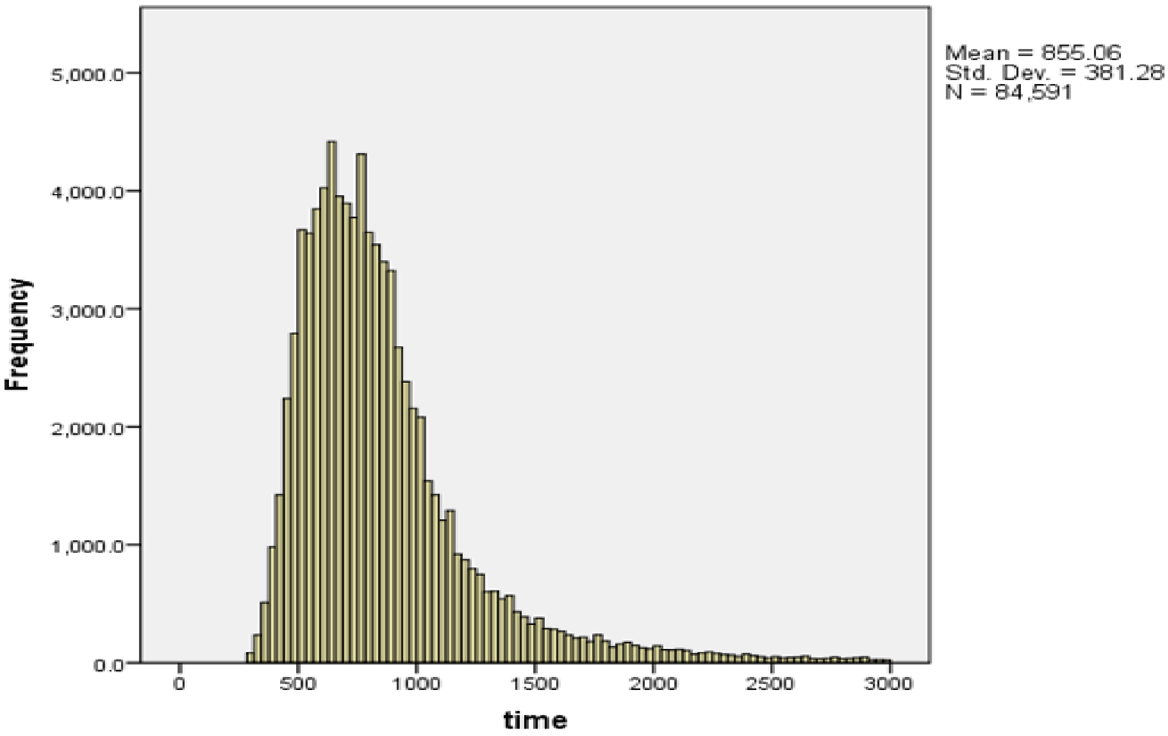


Figure 3.7. SRT's RTs in ms before transformation

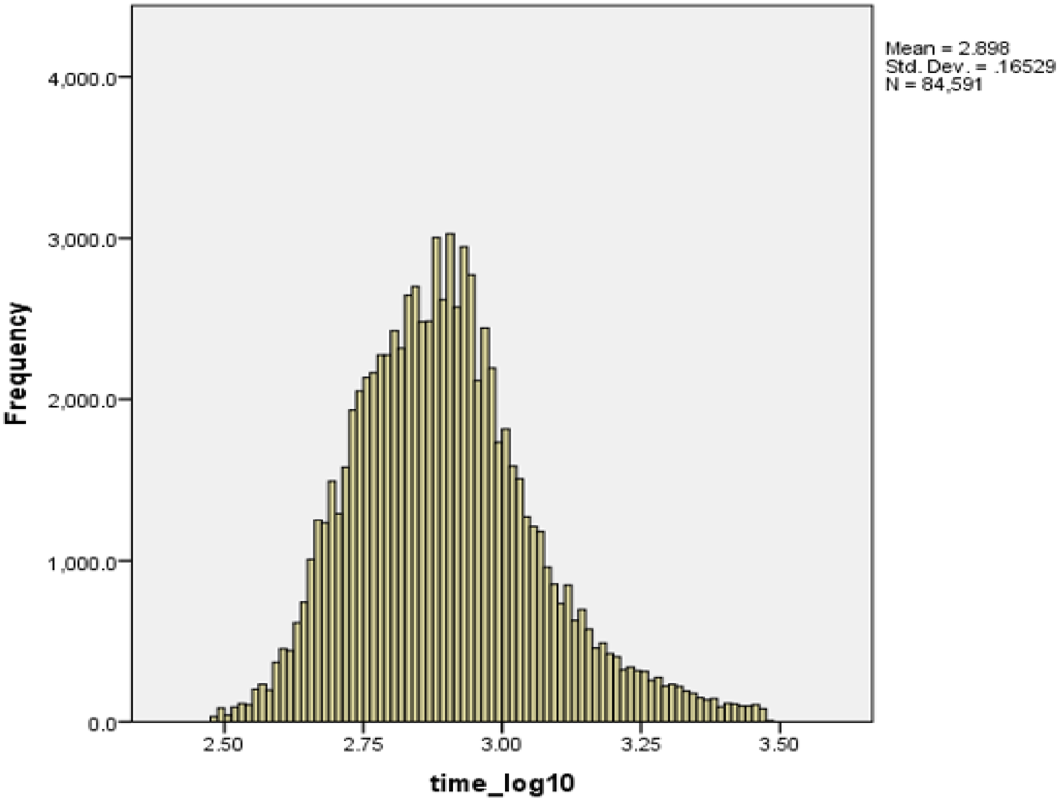


Figure 3.8. SRTs RTs after log 10 transformation

After all the instances of outliers were discarded, the mean was used as the measure of central tendency to score the data. The mean for each participant, block, and type of trial (i.e., probable/congruent/training vs improbable/incongruent/control) was calculated. To validate that implicit learning on probabilistic SRT task took place, the mean RTs for each participant, block, and type of trial were calculated and analysed to see if there were differences in RTs for probable and for improbable trials. Figure 3.9 shows learning on each block for the entire sample, comparing mean RT for trials that followed the most probable (85%) sequence with the mean RT for trials that did not follow the most probable sequence (15%).

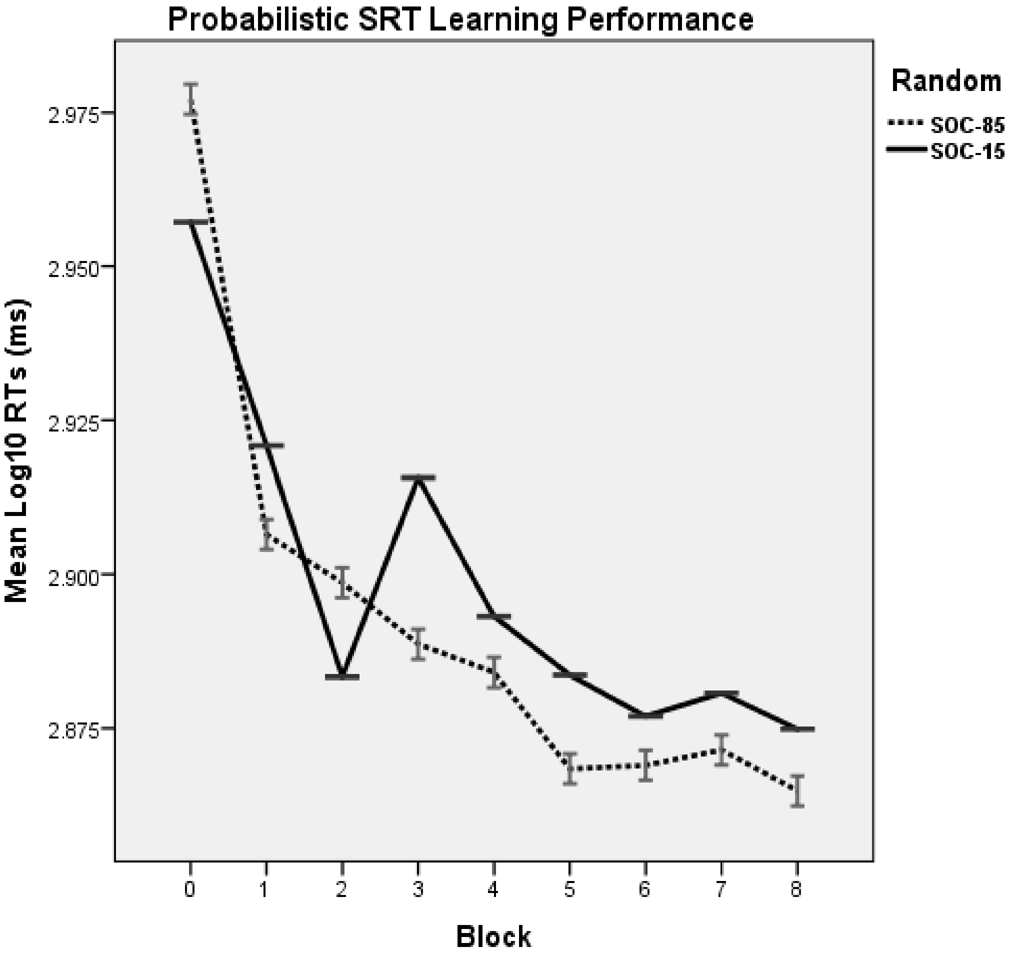


Figure 3.9. SRT learning performance for probable (SOC-85) and non-probable (SOC-15) trials across one practice and eight learning blocks. Note. N = 97; SOC = second-order condition

A mixed between-within group repeated-measures analysis of variance (RM ANOVA) with Group (3, primary school vs. secondary school vs. university students), Block (8) and type of Trial (2, training vs. control) was conducted on the measures of RT (note that block 0 was a practice block). Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the main effects of block, $\chi 2(27) = 308.099$, $p = .0001$, and for the interaction between block and random, $\chi 2(27) = 112.735$, $p = .0001$. Therefore, the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity: $\varepsilon = .441$ for the main effect of block and .736 for the main effect of block-random interaction. The results of the RM ANOVA's tests of within-subjects effects showed a significant effect of block, F(3.085, 289.958 = 8.852; *p* = .0001, partial eta-squared = .086, power = .996)[8], and type of trial, F(1, 94) = 19.992; *p* = .0001, partial eta-squared = .175, power = .993), as well as a significant interaction between block and type of trial, F(5.149, 484.020) = 6.250; *p* = .0001, partial eta-squared = .062, power = .997), indicating that the learning of the training sequence had occurred.

The mixed between-within group RM ANOVA also revealed that the amount of learning in the three groups of learners was comparable, as indicated by the non-significant results of the main effect of group (F(2, 94) = 2.002; *p* = .141, partial eta-squared = .041, power = .404); the interaction of group with block (F(6.169, 289.958) = 1.375; *p* = .223, partial eta-squared = .028, power = .543) or with type of trial (F(2, 94) = 2.357; *p* = .100, partial eta-squared = .048, power = .466); and the interaction of group with block and type of trial (F(10.298, 484.020) = 0 .445; *p* = .928, partial eta-squared = .009, power = .237).

The degree of learning was quantified as the mean difference in RT between correct responses to training and control trials for each block where the participants showed clear evidence of learning (i.e., where the RT for probable sequence was clearly less than the RT for improbable sequence. Despite the overall effect on learning of block, type of trial (random), and block-trial type interaction, the differences between responding to training and control trials became larger from block 3 (see Figure 3.9), indicating that learning was clearly established in the sample as whole across blocks 3-8. The SRT score was therefore the average of the mean differences of the last six

---

[8] Field (2009, p. 390) gives an idea as to how to interpret a type of partial eta squared as follows: "People normally report $\omega 2$ and it has been suggested that values of .01, .06 and .14 represent small, medium and large effects respectively."

learning blocks. Five SRT scores all in primary school were missing. These were for the participants whose SRT data were missing.

3.6.1.2 Llama-D

An accuracy score was automatically computed by the Llama-D software program for every participant. The scores ranged between 0 and 100. The dependent measure was the number correct minus the number incorrect.

3.6.1.3 Continuous Visual Memory Test (CVMT)

The CVMT was administered using the DMDX program. A positive reaction time (RT, in millisecond (ms) meant that a participant's response was correct. Hence if the correct response was 'old' and the respondent answered correctly as 'old' this was recorded as a positive RT. Conversely, a negative RT meant that the respondent's response was incorrect. That is, if the correct response to a trial was 'old' and a respondent answered 'new', DMDX recorded a negative RT. If a participant did not respond or did not respond within the response time of two seconds, the programme recorded the RT of -1990 ms before it moved on to display the next stimulus. Thus, all RTs of -1990 ms were coded as non-responses. Because the non-responses and responses made after response-time had elapsed were indistinguishable in the data, they were all filtered out during analysis. The fact that participants were not computer literate enough necessitated the decision to exclude non-responses from analysis in this task. The CVMT was designed as a continuous running computer task with a two-second response time. Though participants were trained at the beginning of the task on the response-key presses on computer, it was observed during the task that on some occasions some participants were seen briefly searching for response keys. Thus, the exclusion of the non-responses from further analysis was to avoid penalising such participants.

Signal Detection Theory

D-prime (signal detection) analysis, based on Signal Detection Theory, was used to score CVMT data in the present study. Signal detection theory (SDT, Green & Swets, 1988; Macmillan & Creelman, 2005), as defined by Bendixen & Andersen (2013, p.

928), “models situations in which an observer continuously monitors a technical (e.g., radar) system for incidents requiring active intervention”. The performance of the observer in distinguishing ‘true incidents’ (i.e., signals) from background noise is “evaluated in terms of his capacity for discrimination between the two events ([sensitivity]) and his tendency to report an incident rather than to refrain from doing so ([response bias or criterion])” (Bendixen & Andersen, 2013, p. 928). Thus, SDT attributes subjects’ responses to a combination of perceptual sensitivity to the stimuli presented and their decision strategy or bias toward saying something is there or not when they are in doubt.

But discrimination tasks are aimed at recovering sensitivity from bias (Keating, 2005). Thus, a subject in a discrimination task who shows 100% discrimination, might answer "Yes" to every item and get 100% correct on the signal trials, but she would of course also get 0% correct on the noise trials. In many studies, the noise trials are not analysed at all, and this response strategy would work well (Keating, 2005, p. 1). However, scoring 100% in this way does not mean the subject discriminated very well. One does not even have to attend to the stimuli to get 100%. Conversely, a better subject would be one who conservatively answers ‘Yes’ on signal trials and consistently responds ‘No’ on noise trials or when in doubt. Therefore, in SDT, the understanding is that percentage correct on signal trials alone is not a very meaningful measure of discrimination unless it is interpreted in terms of the listener's response bias, or tendency to respond "No" or "Yes", and the responses to the noise trials can be used as an indication of response bias (Keating, 2005, p. 1). Macmillan & Creelman (2005, p. 6) state that an ‘important characteristic of sensitivity is that it can only be measured between two alternative stimuli’, namely, signal trials and noise trials.

SDT can be applied to any task that involves two possible stimulus types. Consequently, the interpretation of *signal trials,* which present one or more signals, and *noise trials,* which present one or more noise stimuli, depends on the discrimination task. SDT has since been applied in many other areas. Stanislaw & Todorov (1999, pp. 137-138) illustrate the interpretation of *signals* (stimuli) and *noise* (no stimuli) labels in various tasks as follows: in “recognition memory (old and new items), lie detection (lies and truths), personnel selection (desirable and undesirable applicants), jury decision making (guilty and innocent defendants), medical diagnosis (diseased and well patients), industrial inspection (unacceptable and acceptable items)”.

Four possible outcomes (or answer types) to each of the many trials arise from a discrimination task which presents signals and non-signals to subjects. On signal trials (or *different) yes* responses are correct and are termed *hits* while *no* responses are incorrect and are termed *misses*. On noise trials (or *same*), *yes* responses are incorrect and are termed *false alarms* while *no* responses are correct and termed *correct rejections*. Below is the schematic diagram to illustration the different answer types where "Yes" represents the presence of a signal or difference to be detected (Table 3.10).

Table 3.10. Schematic diagram of answer types in SDT

	Response: Different (*yes*)	Response: Same (*no*)
Stimuli: **YES** (*different*)	**HIT**	**MISS**
Stimuli: **NO** (*same*)	**FALSE ALARM**	**CORRECT REJECTION**

From Keating (2005, p. 2).

A subject's performance is conventionally characterised in terms of Hits and False Alarms, and these are then given as proportions of the row totals, which are in turn viewed as estimates of probabilities of responses (Keating, 2005). On a yes/no task, the *hit rate (H)* is the probability of responding *yes* on signal trials and the *false-alarm rate (F)* is the probability of responding *yes* on noise trials. The best performing subject maximizes both H and the Correct Rejection rate. The larger the difference between H and F, the better the subject's sensitivity (Keating, 2005). The statistic ***d'*** (*d*-prime) is a measure of this difference. However, *d'* is not simply H-F; rather, "it is the difference between the z-transforms of these 2 rates" (Keating, 2005, p. 3):

$$d' = z(H) - z(F)$$

Neither H nor F can be 0 or 1. If so, a *d'* value is undefined and cannot be calculated because the inverse normal transform takes on an infinite value (Hautus, 1995). Therefore, to produce finite *d'* values, H and F of 0 or 1 must be adjusted slightly up or down. There are controversies regarding the best way to correct H and F of 0 or 1. One correction method involves discarding cases with H or F of 0 or 1 (Rotello, Masson, & Verde, 2008), which results in reducing the sample size. Another correction method involves combining the data from several subjects before calculating H and F (Stanislaw & Todorov, 1999). This procedure complicates statistical testing: it is impossible to have the individual subject's *d'* scores necessary for investigating single subject behaviour, and it also assumes that all subjects have comparable response biases and levels of sensitivity. The third method, the replacement method, involves adjusting only the extreme rates themselves, in which rates of 0 and 1 are replaced by 1/2N and 1-1/2N, respectively, where N is the number of trials on which the proportion is based (Macmillan & Creelman, 2005). Lastly, the loglinear correction (Snodgrass & Corwin, 1988) involves adding 0.5 to both the number of hits and the number of false alarms and adding 1 to both the number of signal trials and the number of noise trials, before calculating the hit and false alarm rates (Brown & White, 2005; Hautus, 1995; Macmillan & Creelman, 2005; Rotello et al., 2008; Verde, Macmillan, & Rotello, 2006). The present study adopted the loglinear approach as the method of correction for H or F of 0 or 1.

The two commonly used correction methods are (i) the replacement method involving adjusting only the extreme rates and (ii) the loglinear correction. Adjusting only extreme rates yields biased measures of sensitivity because it does not treat data points equally (Stanislaw & Todorov, 1999). Further, it has been shown that the replacement method is less satisfactory than the loglinear approach (Brown & White, 2005; Hautus, 1995; Verde et al., 2006). Rotello et al. (2008) found that the loglinear correction method was better than the discard method, but it was indistinguishable from the replacement method of correction. However, studies such as Brown & White (2005), Hautus (1995), and Verde et al. (2006) have shown that the loglinear rule resulted in less biased estimates of *d'*.

The present study used the 'one-interval-discrimination-type of CVMT that involved the presentation of one stimulus in a trial. Based on the design of the CVMT in this study, there were a total of 42 Hits and 54 FAs.

Calculating the Statistic d'

First Hit rates (H) and FA rates (F) were calculated as proportions of the total number of Hits and FAs as in the test design, that is, a total of 42 Hits and 54 FAs. The results of calculations revealed a few H and F values of 1 or 0. The loglinear rule was used as a method of correction. Using these corrected H and F values and using the NORM.INV function in excel, the *d'* for each participant was calculated as the difference between the z-transforms of the H and F values, that is, using the formula $d' = z(H) - z(F)$.

The *d* prime scores for five participants were not calculated. Of these, one participant had all his data missing likely due to the computer program not recording the results. Two participants had many non-responses (either because of the delay in responding such that the response time elapsed or because of not responding at all): one had all the first 50 items not responded to (all recorded as time-outs, i.e., as RTs -1990 ms); and one had all the test items except two recorded as time-outs. The *d'* scores for the last two participants were not calculated because the data for these participants had shown that they had the tendency of pressing only one response key for all the test items. One of these was pressing the response key for NEW only, resulting in only correct rejections and misses. Such a participant could not get hits and false alarms as these were the result of responding either correctly to ‘old’ items or incorrectly to ‘new’ items. It was pressing the response key for OLD that resulted in either hits or false alarms. The other participant did just the opposite. The data showed this participant was not just responding OLD for the items but she or he only responded when she or he thought the design that was presented was ‘old’. Consequently, she or he had a few hits and a few false alarms but with many time-outs.

3.6.1.4 The DAT Verbal Reasoning Test (DAT-V)

In the DAT-V, the participants’ answer sheets were marked by the researcher. For each of the 40 problems, there was only one correct answer option. A point was given for a correct choice. Thus, the scores ranged between 0 and 40. All the missing values arose from non-responses. Five items across the three-learner groups had one missing data value each. Three of these were in the university student group. Because the test was paper-and-pencil delivered but at the same time speeded with lots of information on each item, it was not possible to determine the reasons for the non-responses (i.e., whether it was the result of the participants’ not knowing the answer or whether the items were skipped unintentionally). Therefore, to avoid penalising the participants for the non-responses, the scoring proceeded as averages on the total number of items responded to.

3.6.1.5 Llama-B

The dependent measure was the number of correct responses. For every participant, an accuracy score, ranging between 0 and 100, was automatically computed by the Llama-B program. These raw scores were used in the analyses. The only missing value was the score of one primary school participant which was not found.

3.6.1.6 Three-Term Contingency Learning Task (3-Term)

The 3-Term test consisted of four blocks. In each block, a training phase was followed by a test phase. For every participant, an accuracy score with the maximum of 30 in each block was automatically computed by the 3-Term test program. The dependent measure was the number of correct responses. The final score ranged between 0 and 120, and it was computed as the sum of the scores of the four blocks. Only two cases had missing values in block 4 of the test in the primary school group. The computer program might have not saved these scores.

3.6.2 Scoring of language knowledge Tests

As described in the previous sections, the participants completed three language knowledge measures, the EIT, TGJT and UGJT. The following sections present the scoring of these measures. Both the TGJT and UGJT were a Yes/No-response type of tests, requiring the kind of scoring that considers random performance. For several reasons however, the scoring of the TGJT and UGJT was not based on random performance. First, considering the inclusion in the study design of (i) some structures that were not explicitly taught, (ii) the three learner groups from three different levels of educational background, and (iii) most primary and secondary school participants who were not literate enough in the use of a computer (i.e., regarding the TGJT), a very wide range of scores (many of which could be below the chance-level accuracy of .50) was expected as a function of the study design and not of random performance. Second, some proponents of SDT (the scoring that attributes participants' responses to the combination of sensitivity to what is presented and response bias toward saying something) argue that correcting for guessing assumes that either the participant really knows the response (i.e., perfect knowledge) or makes a completely random guess (i.e., absence of knowledge; Huibregtse, Admiraal & Meara, 2002). Such kind of scoring does not consider sophisticated guessing that arises from various degrees of knowing. Sophisticated guessing allows participants not to guess completely at random. Finally, the examination of the responses to the test items revealed no extreme cases preferring only one response type.

3.6.2.1 Elicited Imitation Test (EIT)

The researcher and the research assistant transcribed the recordings of all the participants. As presented below, the researcher used a rubric based on Erlam (2006) consisting of 3 rules to score the participants' responses.

First Criterion: Obligatory Occasion Created – Target Structure Supplied
This criterion described all cases where participants were able to create an obligatory occasion or environment for the target structure and then correctly supplied the target structure irrespective of lexical accuracy. For each of these, the participants scored a point.

Second criterion: Obligatory Occasion Created – Target Structure Not Supplied
This criterion described cases where the obligatory occasion was created but the target structure was not supplied. All these were scored as incorrect responses.

Third Criterion: No Obligatory Occasion Created
This criterion described all cases where no obligatory occasion was created. All these were scored as incorrect responses as well.

Non-responses and missing data were treated differently as described below. The non-responses were conceived of as cases where participants made no attempt to repeat the sentence they had just heard. All such cases were scored as incorrect responses in this task. It was determined that the failure to repeat the sentence could largely be attributed to the participants' failure to store in memory the sentence they had heard. During the trialling of the tests used in this study, the participants had reported that the woman's voice with American English accent that the EIT computer program used was clear enough, suggesting that the EIT program could have minimally influenced non-responses. Further, the participants' computer illiteracy could not have come into play when the participants were completing the task: their only task was to listen and repeat the sentences, requiring no computer response-key pressing.

Missing data values however were instances where the poor sound quality of a participant’s audio recording occurred right at an obligatory occasion. In such cases there were two possibilities: either the participant correctly supplied the target structure which could not be heard due to the poor sound quality of the recording, or the participant did not correctly supply it. Below is an example of a case of missing data.

“Many problems in Malawi are ...”

In the example above, the ellipsis, as used in the transcriptions of the recordings, represented a point at which the sound quality of the recording was not clear enough for the judgment to be made regarding what the participant said. But this is the point where the obligatory occasion to supply a passive verb was created.

The target structure score was the mean of the average score on grammatical items and the average score on ungrammatical items. The averages on each category were thus percentages, with the maximum score of 1.0. All missing values were excluded in the calculation of these averages. This was done to avoid punishing participants for missing data. When responses to all grammatical or ungrammatical items of a target structure were missing, it was impossible to compute a mean score for the category. These are the data points that constituted missing values in this test. The imputation of missing values was followed by the computation of four kinds of scores from grammatical and ungrammatical items mean scores. First, the target structure score was calculated by averaging the grammatical and ungrammatical items mean scores of each target structure. Second, the grammatical sentences composite score was computed as the mean score of all grammatical items mean scores of a target structure. Third, the ungrammatical sentences composite score was computed by averaging all the ungrammatical items mean scores of a target structure. Finally, the EIT composite score was computed as the average score of all the target structure scores.

3.6.2.2 Time Grammaticality Judgment Test (TGJT)

The TGJT computer program automatically recorded the participants’ responses and the RTs it took to respond to each sentence. Both ‘Don’t Know’ responses and non-responses were not scored. It was not determined whether the non-responses occurred as the result of the participants intentionally skipping the question or as the result of the computer program's failure to record the responses. Further, the participants, especially the young learners, were not very computer literate, which might have affected their use of an external mouse in making their choices. Avoiding penalising the participants was the main reason the non-responses were not scored. Therefore, the scoring proceeded as averages on the total number of items responded to. The grammatical sentences and ungrammatical sentences scores were separately calculated as mean scores on the total number of the category items responded to. From these mean scores, the target structure score, the grammatical and ungrammatical sentences composite scores and the TGJT composite score were computed in a similar way as in the EIT above. RTs in the present study were not analysed. The TGJT missing values were conceived in terms of the missing of the grammatical or ungrammatical sentences mean scores of a target structure as the result of excluding 'Don't Know' and non-responses from scoring.

3.6.2.3 Untimed Grammaticality Judgment Test (UGJT)

The UGJT scoring was based on accuracy. Because the participants were not under time pressure, non-responses indicated that the participant did not know the answer. Therefore, all the non-responses were scored as incorrect. For each target grammatical structure, scores were separately calculated as percentages for grammatical and ungrammatical test sentences. From these scores, the target structure score, the grammatical and ungrammatical sentences composite scores and the UGJT composite score were computed in a similar way as in the EIT. There were no missing values in the UGJT scores.

3.6.3 Statistical Analyses

This section presents the statistical analyses employed. The partcipants' scores on all tests were entered onto an excel workbook and then imported into the Statistical Package for the Social Sciences (SPSS) version 22 (IBM Corporation). First, all scores were examined for missing values. The accuracy scores of the EIT, TGJT, UGJT and 3-Term tasks required the computation of overall composite scores for each participant. These composite scores were calculated after replacing missing values in respective test scores.

Hair, Black, Babin, & Anderson (2010) have listed several rules of thumb regarding how to deal with missing data. One of the rules is that variables with as little as 15% missing data are candidates for deletion while higher levels of missing data (e.g., 20% to 30%) should be replaced. Another rule states that any data imputation methods can be applied for missing data under 10%, although the complete case method has been shown to be the least preferred. Several disadvantages are associated with case deletion processes (Hair et al., 2010). First, they tremendously reduce sample sizes and power to find results. Second, they are most affected by any non-random missing data process such that the results may not be generalizable to the population. Lastly, pairwise imputation results in correlations that are “out of range” and inconsistent with other correlations in the correlations matrix because they are all calculated from unique sets of cases. In the present study therefore, missing values below 10% were replaced using group mean imputation. Multiple imputation was used to replace either missing data above 10% or non-random missing data points.

Following the missing value analysis and the subsequent imputation of the replacement values and composite scores, the test scores were graphically and numerically examined for univariate outliers and normality, both for the individual groups and the entire sample as a single group. Outliers, defined as observations with a unique combination of characteristics, typically unusually high or low values, identifiable as distinctly different from the other observations, indicate that scores are not exactly normally distributed. This in turn can seriously distort parametric statistical tests (Field, 2009; Field, 2013; Hair, Black, Babin, & Anderson, 2014; Larsen-Hall, 2010). The inspection of the assumptions of univariate normality was undertaken to inform the decisions about the statistical program to use or the type of inferential statistics to conduct.

Graphically, histograms, Q-Q plots, and boxplots were used to explore data for outliers. For the histograms and Q-Q plots, cases that appeared very different or not approximately falling in the reference line strongly suggested the presence of outliers. Boxplots indicated whether the set of scores had mild and/or significant outliers. Outliers were also examined numerically by using z-scores. In the present study, outliers were determined mainly by the data set's z-scores.[9] Z-scores are standardised values of data whose distribution has a mean of 0 and a standard deviation of 1 (Field, 2009). Converting a raw score (X) into a z-score, involves subtracting the mean of all scores ($\bar{X}$) from the raw score and dividing by the standard deviation of all scores (*s).* Z-scores enable the use of benchmarks that can be applied to any data set regardless of its original mean and standard deviation. For a normally distributed data set, it is expected to have 95% of cases with absolute value less than 1.96, 5% or less with absolute value greater than 1.96, and 1% or less with absolute value greater than 2.58, and no cases above 3.29 (Field, 2009). Based on this, significant outliers in the present study were defined as those with absolute z-score greater than 3.29. Any identified outlier score was replaced by a score that was calculated by converting the outlier score back from the z-score, generally using the mean plus two times (rather than three times) the standard deviation.

The data were then examined for univariate normality. Graphically, histograms, P-P plots and boxplots were inspected (See Field, 2009; Hair et al., 2014; Larson-Hall, 2010). Clustering of scores to right or left as well as the flat peak of the distributions suggested non-normality in histograms while the same was suggested by deviations of data values from the diagonal reference line in P-P plots. Differences in the lengths of boxplots' top and bottom whiskers were also considered as indication for non-normal distribution.

Numeric examination of univariate normality included performing the Shapiro-Wilk test statistic. The Shapiro-Wilk statistic, like the Kolmogorov-Smirnov (K-S), tests the null hypothesis that the data are normally distributed (Field, 2009). A non-significant

---

[9] There have been arguments regarding the accuracy of boxplots in identifying outliers. A boxplot uses the outlier labelling rule (also known as the interquartile range rule) with the multiplier constant of 1.5 as the estimate for determining demarcation criteria for outliers (IBM Support). The length of the box is the interquartile range (IQR). Cases more than three IQR's from the end of the box are labelled as extreme, and they are denoted with an asterisk (*). Values more than 1.5 IQR's but less than 3 IQR's from the end of the box are labelled as outliers, and they are denoted with a circle (o). However, research has demonstrated that the 1.5 multiplier is inaccurate most of the time in identifying outliers (Hoaglin & Iglewicz, 1987).

test (p >.05) suggests the distribution of the sample is not significantly different from a comparable normal distribution. Research suggests the Shapiro–Wilk test has more power to detect differences from normality (Field, 2009). Generally, however, the normality tests are said to have limitations. Research suggests that in large samples small deviations from normality easily lead to significant results, and 'so a significant test doesn't necessarily tell us whether the deviation from normality is enough to bias any statistical procedures that we apply to the data' (Field, 2009, p. 144). Conversely, because these normality tests often do not have enough power to detect violations of the null hypothesis, a non-significant test does not necessarily mean the distribution is normal (Larson-Hall, 2010). Therefore, skewness and kurtosis z-scores were calculated and inspected to further examine normality in the present study. To calculate the z-scores, the following formulae were used: $Z_{skewness}$ = Skewness – 0 / $SE_{skewness}$ and $Z_{kurtosis}$ = Kurtosis – 0 / $SE_{kurtosis}$ (Field, 2009, p. 139). Following Field (2009), non-normality in the present study was defined as skewness or kurtosis z-scores with absolute value greater than 2.58, significant at $p < .01$.

Descriptive statistics were calculated for each test for the individual groups and for the entire sample. For the EIT, TGJT and UGJT scores, the descriptive statistics are reported for target structures scores, grammatical and ungrammatical sentences composite scores as well as for the final overall score computed by averaging all the target structures. Reliability coefficient for internal consistency (Cronbach's α) was calculated for each of the tests except the two Llama subtests. Granena (2013a) points out that the reliability coefficient of .70 is a widely accepted cut-off in the social sciences. In the present study, only SRT had a very low reliability estimate of .40 (see section 4.1.3.2).

Inferential statistics were conducted at three levels. First, for the test scores that did satisfy the assumptions of parametric statistics, one-way between-groups analyses of variance (one-way ANOVAs), with Welch's procedure and a series of planned contrasts, were performed to examine the between-groups multiple comparisons in each of the test scores. Welch's procedure is strongly recommended when the variances of group scores are heterogenous and/or when the group sample sizes are unequal (Field, 2009; Howell, 2012; Larsen-Hall, 2010)[10]. Omega squared (ω^2) effect sizes of the omnibus ANOVA *F*-tests are reported, with the following guidelines: .01 as a small effect; .06 as a medium effect; and .14 as a large effect (Kirk, 1996, cited in Field, 2009, p. 390). The ω^2 is basically thought to be the unbiased estimate of *r,* as it is adjusted to estimate the effect size in the population (Field, 2009). The effect sizes for the contrasts are reported as *r*, calculated from planned orthogonal comparisons' *t*-statistic for unequal variances assumed (see below for the guidelines in interpreting *r*). To examine the between-group comparisons in data with non-normal scores, the Kruskal–Wallis omnibus with the post-hoc Mann-Whitney U tests were performed using the Langtest Web application (Version 1.0, Mizumoto, 2015). The respective *r* effect sizes for statistically significant contrasts are reported for these non-parametric test statistics.

Second, at the level of test instrument validation, the correlational, exploratory and confirmatory factor analyses were conducted to explore the tests' construct validity (i.e., to check for evidence that the tests tapped the types of memory ability or language knowledge as designed). While the correlational analyses were performed to examine the interrelationships between the various test measures, the factor analyses were conducted to explain whether the correlations among multiple observed variables were indeed as the result of one or more assumed underlying explanations or factors (Hair et al., 2014). The exploratory and confirmatory factor analyses were respectively implemented using SPSS version 22.0 (IBM Corp, 2013) and the Equations (EQS) multivariate software version 6.4 (Bentler & Wu, 1989). Non-parametric Spearman Rank Order Correlation (*rho*) is reported for tests' scores with non-normal distributions. Pearson's Correlation Coefficients were computed for normally distributed scores. Following Cohen (1989, p. 79-81), the guidelines for the strength of the relationship between two variables were as follows in absolute value: small effect, *r*=.10 to .29; medium effect, *r*=.30 to .49; and large effect, *r*=.50 to 1.0.[11]

---

[10] Howell further argues that regardless of the sample size, the Games and Howell procedure should be used if the variances are heterogeneous (2012).

[11] Note that the interpretation of effect sizes and the computation of shared variance for Spearman's *rho* are done in the same way as Pearson's *r* (Field, 2009).

Lastly, several analyses were conducted to answer the three research questions of the study. A series of structural equation modelling (SEM) analyses were conducted to investigate the different hypotheses for Research Questions 1 and 2 of the study. The EQS was used to perform all the SEM analyses. Some scores in the present study were non-normal, necessitating the use of the EQS for its capability to handle non-normal data (Bentler, 2006; Narayanan, 2012; Schumacker & Lomax, 2016). The choice to use SEM was necessitated by the need to simultaneously investigate various interrelated dependence relationships among the measured memory ability and language knowledge variables and the latent memory ability construct(s). A latent variable approach allows for more accurate measurement of the constructs of interest (Kaufman et al., 2010). Correlational analyses and a series of Kruskal–Wallis test statistic or one-way ANOVAs, were performed to examine the between-group multiple comparisons in Research Question 3. Finally, all the hypotheses were tested at the conventional .05 or .01 alpha level (2 tailed). Because the significance of *r*, *rho* or the ANOVA test statistic is strongly influenced by the size of sample (Pallant, 2010), attention was paid both to the strength of the relationship or effect and to the amount of shared variance in a correlation or on actual differences in group mean scores in the ANOVAs.

CHAPTER 4. RESULTS

This chapter reports the results of the study. The participants completed six measures of memory abilities and three measures of language knowledge. The first section of the chapter reports the descriptive statistics for the entire sample (i.e., for all the three-group data combined) and for each individual group data since distributional parameters change for each grouping of participants (Carpenter, 2008). Inferential statistics for the correlational, exploratory and confirmatory factor analyses are reported in the second section to validate the test measures. The last section reports findings from a series of SEM, one-way ANOVAs and correlational analyses that were conducted to investigate the three research questions of the study.

4.1 Descriptive Statistics

Descriptive statistics for the memory ability measures are reported first followed by those for the language knowledge measures.

4.1.1 Memory Ability Measures

4.1.2 Declarative Memory Ability Measures.

4.1.2.1 Continuous Visual Memory Test (CVMT).

The CVMT standardised scores had no absolute values greater than 3.29, indicating there were no significant outliers in both the individual group and entire sample scores. The Shapiro-Wilk tests of normality (Table 4.1) indicated that the distributions within all groupings did not differ significantly from comparable normal distributions. Further, neither skewness nor kurtosis absolute z-values were greater than 2.58 in any of the scores. Therefore, all the CVMT scores were considered approximately normally distributed. Table 4.1 reports the resulting descriptive statistics. The one-way between-groups ANOVA test statistic, with Welch’s procedure, indicated mean scores between groups were statistically significantly different (F (2, 64.629) = 6.620, p = .002, w^2 = .110). Planned contrasts indicated significant differences between primary and secondary school groups (t(66.994) = 2.274, p = .026 , r = .268) and between the primary school and university student groups (t(68.338) = 3.615, p = .001, r = .401). There was no statistically significant difference between the secondary school and university student groups (*t*(58.682) = 1.518 , *p* = . 134). The reliability coefficient for internal consistency (Cronbach's α), calculated on the itemisation scores, was .86.

Table 4.1. Descriptive Statistics and Shapiro-Wilk Tests for CVMT D' Scores

			Shapiro-Wilk	
Group (N)	Mean	SD	*D*	*p*
Primary School (41)	1.020	.707	.966	.259
Secondary School (29)	1.365	.562	.974	.675
University Students (32)	1.605	.669	.969	.462
All (102)	1.302	.696	.987	.395

4.1.2.2 The Llama-B

Descriptive statistics for the Llama-B scores are reported in Table 4.2. Only the university group scored within the 25-45 average score range that most people score on the Llama-B test (Meara, 2005). The results of the standardised scores indicated only one case with absolute z-score greater than 3.29 in the entire sample scores. The outlier case was traced to the university student group and replaced by a value calculated using the university students mean score and standard deviation. The Shapiro-Wilk tests of normality (Table 4.2) indicated that the distributions for the primary school and the entire sample scores were significantly different from a bell curve (*D*(41) = .942, *p* = .037 and *D*(102) = .943, *p* = .001, respectively). However, the standardised skewness values indicated only the entire sample scores to be statistically significantly skewed, with the scores piling up onto the left (i.e. to the lower scores) of the distribution. Based on these results, only the entire sample scores were considered as non-normal. The one-way ANOVA test statistic, with Welch's procedure, indicated that the mean scores between the groups were statistically significantly different (*F* (2, 58.345) = 6.170, *p* = .004, w^2 = .105). Planned contrasts indicated a significant difference between the primary school and university student groups (*t*(49.420) = 3.437, *p* = .001, *r* = .439), but there were no significant differences between primary and secondary school groups (*t*(54.443) = 1.759, *p* = .084) and between the secondary school and university student groups (*t*(56.595) = 1.794, *p* = .078). Finally, the Llama-B test reliability estimate was not computed. The test's scores were not itemized data.[12]

Table 4.2. Descriptive Statistics and Shapiro-Wilk Tests for Llama-B Raw Scores

Group (N)	Mean	*SD*	Shapiro-Wilk *D*	Shapiro-Wilk *p*
Primary School (41)	19.3812	10.1359	.942	.037
Secondary School (29)	24.1379	11.8072	.952	.211
University Students (32)	30.5906	16.1276	.954	.189
All (102)	24.2503	13.4734	.943	.000

4.1.2.3 The 3-Term Contingency Learning Test (3-Term)

The 3-Term standardised scores indicated no absolute z-scores greater than 3.29 in any group and entire sample scores, indicating no significant outliers. The Shapiro-Wilk tests of normality (Table 4.3) indicated that the distributions for the primary school and the entire sample scores were significantly different from comparable normal distributions (*D*(41) = .901, *p* = .002 and *D*(102) = .951, *p* = .001, respectively). Both the primary school and entire sample data obtained an absolute skewness z-score greater than 2.58, indicating significant skewness. These two sets of scores were consequently considered non-normal. The resulting descriptive statistics are reported in Table 4.3.

Table 4.3. Descriptive Statistics and Shapiro-Wilk Tests for 3-Term Raw Scores

Group (N)	Mean	*SD*	Shapiro-Wilk *D*	Shapiro-Wilk *p*
Primary School (41)	20.2700	15.169	.901	.002
Secondary School (29)	24.0000	15.655	.958	.288
University Students (32)	38.0600	18.451	.937	.063
All (102)	26.9100	17.981	.951	.001

---

[12] Granena (2013a) proposes a procedure that aims at itemising data for Llama subtests. Future research should aim to do this inorder to calculate the test's reliability.

The Kruskal-Wallis H Test showed that there were significant differences in the 3-Term scores across the three groups ($\chi^2 = 18.601$, *df* = 2, *p* = .001, *r* = .387). The post-hoc Mann-Whitney U, with the Bonferroni adjusted *p*-values, indicated no significant difference between the primary and secondary school groups (*U* = 484.000, *z* = -1.318, *p* = .562, *r* = .158). But there were significant differences between the primary school group and the university student group (*U* = 275.500, *z* = -4.232, *p* = .001, *r* = .495) and between the secondary school group and the university student group (*U* = 274.000, *z* = -2.747, *p* = .562, *r* = .352). The reliability coefficient for internal consistency (Cronbach's α) based on the raw data in the four blocks was .85.

4.1.2.4 The DAT Verbal Reasoning Test (DAT-V)

Descriptive statistics for the DAT-V scores for all participants and for the three learner groups are reported in Table 4.4. The computed standardised DAT-V scores for the individual groups and the entire sample showed no cases with absolute z-score greater than 3.29, indicating absence of significant outliers in all sets of scores. The Shapiro-Wilk tests of normality (Table 4.4) indicated only the entire sample scores distribution as significantly different from a normal distribution (*D*(102) = .971, *p* = .025). However, the skewness and kurtosis *z*-scores for all the sets of scores were not significant. Therefore, all the DAT-V scores were treated as approximately normally distributed.

Table 4.4. Descriptive statistics and Shapiro-Wilk tests for DAT-V average scores

Group (N)	Mean	*SD*	Shapiro-Wilk	
			D	*p*
Primary School (41)	.322	.110	.965	.240
Secondary School (29)	.528	.129	.968	.510
University Students (32)	.670	.100	.976	.676
All (102)	.490	.186	.971	.025

The one-way ANOVA test statistic, with Welch's procedure, indicated the mean scores across groups were statistically significantly different (*F* (2, 61.486) = 98.544, *p* = .001, w^2 = .628). With unequal variance assumed, planned contrasts indicated significant differences in all the three comparisons as follows: between primary and second school groups (*t*(54.188) = 6.962, *p* = .001, *r* = .687); between the primary school and university student groups (*t*(69.257) = 14.047, *p* = .001, *r* = .860); and between the secondary school and university student groups (*t*(52.755) = 4.748, *p* = .001, *r* = .547). The reliability coefficient for internal consistency (Cronbach's α) based on the 40 items was .86.

4.1.3 Procedural memory measures

This section provides descriptive statistics of the two measures of procedural memory ability. The one-way ANOVA group comparisons for these tests are reported in the present study's results section below.

4.1.3.1 Llama-D

The resulting descriptive statistics for Llama-D scores are reported in Table 4.5. The three group mean scores fell within the average score range of 15-35 that most people are reported to have scored (Meara, 2005). The computed standardised Llama-D scores for the individual groups and for the entire sample had indicated no significant outliers. The Shapiro-Wilk tests of normality (Table 4.5) indicated that the distributions for the secondary school (*D*(29) = .897, *p* = .008) and for the entire sample scores (*D*(102) = .960, *p* = .004) were significantly different from comparable normal distributions. But none of the skewness and kurtosis absolute z-values were greater than 2.58 for any sets of the Llama-D scores. Consequently, the scores were considered as approximately normally distributed. Finally, the Llama-D reliability estimate, just like that of Llama-B, was not calculated because the test was not itemised. However, Granena (2016) reported the reliability coefficient for internal consistency (Cronbach's α) of .64 for Llama-D.

Table 4.5. Descriptive Statistics and Shapiro-Wilk Tests for the Llama-D Raw Scores

Group (N)	Mean	*SD*	Shapiro-Wilk	
			D	*p*
Primary School (41)	20.980	11.414	.968	.298
Secondary School (29)	20.690	13.411	.897	.008
University Students (32)	23.590	12.841	.944	.098
All (102)	21.716	12.398	.960	.004

4.1.3.2 Serial Reaction Times (SRT)

Descriptive statistics for the SRT scores for the entire sample and for each of the three learner groups are reported in Table 4.6. The SRT standardised scores indicated the presence of a significant outlier in the primary school data, which was replaced accordingly. The Shapiro-Wilk tests of normality (Table 4.6) indicated significantly asymmetrical distributions in the primary school scores ($D(41) = .930$, $p = .015$) and in the entire sample scores ($D(102) = .973$, $p = .038$). However, the examination of the skewness and kurtosis z-scores indicated no absolute value greater than 2.58, suggesting the individual group scores and the entire sample scores, though a little skewed and kurtotic, were approximately normally distributed.

The Spearman-Brown's split-half reliability was .40 (n =97). This estimate was obtained by itemising the differences between improbable and probable RTs in each of the last six learning blocks where learning had occurred. Though it is a very low estimate as compared to the indices of declarative memory, the reliability index of .40 as reported in this investigation is in the range of the indexes reported in studies of implicit learning. Granena (2016) and Kaufman et al. (2010) reported the split-half reliability of 0.44 and they considered it standard for probabilistic SRT tasks based on the reliability of implicit learning in the literature. Further, A.S. Reber et al. (1991) and Robinson's (1997) replication study reported the split-half reliabilities of 0.51 and 0.52, respectively, using the Spearman–Brown correction.

Table 4.6. Descriptive Statistics and Shapiro-Wilk Tests for the SRT Scores (RTs in Log10 ms)

Group (N)	Mean	*SD*	Shapiro-Wilk	
			D	*p*
Primary School (41)	.0184	.0240	.930	.015
Secondary School (29)	.0058	.0189	.975	.708
University Students (32)	.0127	.0197	.978	.743
All (102)	.0131	.0217	.973	.038

4.1.4 Language Knowledge Measures

This section provides descriptive statistics for the three language knowledge measures: the EIT and TGJT, designed as measures of procedural language knowledge; and the UGJT designed as a measure of declarative language knowledge. Due to the large number of grammatical structures included, only standardised scores were used to examine outliers and normality in the individual target structures. The Shapiro-Wilk tests of normality are therefore reported only for the tests' overall scores. Likewise, only one-way between-groups ANOVAs for the tests' overall scores are reported. Descriptive statistics for the EIT and the TGJT are presented first.

4.1.4.1 The Elicited Imitation Test (EIT)

The resulting descriptive statistics for the EIT percentage scores for the entire sample are reported in Table 4.7. The EIT standardised scores indicated only three outliers in the possessive -*s* and pseudo-cleft scores for the primary school group, which were replaced accordingly. The skewness z-scores of the target structure scores and grammatical and ungrammatical items composite scores indicated that the primary school scores on the structural parallelism, possessive '-*s*', locative inversion, *Wh*-question, and pseudo-cleft were significantly skewed. In the secondary school group and university students, only the locative inversion scores were significantly skewed. The kurtosis z-scores showed that the primary school scores on the pseudo-cleft and the secondary school scores on the locative inversion were also significantly kurtotic, with z-values greater than 2.58. The entire sample skewness and kurtosis z-scores indicated that the participants' scores on the possessive '-*s*' and pseudo-cleft were significantly skewed while the scores on the locative inversion, *Wh*-question and dative alternation were statistically significantly kurtotic. The scores on all these grammatical structures were then considered asymmetrically distributed.

The Shapiro-Wilk tests of normality for the EIT composite scores indicated normal distributions in the secondary school scores (*D*(29) = .946, *p* = .143) and the university students scores (*D*(32) = .947, *p* = .120), but significantly asymmetrical distributions in the primary school scores (*D*(41) = .928, *p* = .012) and in the entire sample scores (*D*(102) = .944, *p* = .001). However, no absolute skewness and kurtosis z-scores were greater than 2.58 in any group and entire sample scores, indicating that the scores were a little skewed and kurtotic but that they did not differ significantly from normality.

Consequently, the EIT composite scores for the groups and for the entire sample were considered normally distributed. The one-way ANOVA test statistic, with Welch's procedure, indicated the mean scores across groups were statistically significantly different (F (2, 61.770) = 41.371, p = .001, w^2 = .387). With unequal variance assumed, planned contrasts indicated significant differences in all the three comparisons as follows: between primary and second school groups (t(60.245) = 2.967, p = .004, r = .357); between the primary school and university student groups (t(69.382) = 8.948, p = .001, r = .732); and between the secondary school and university student groups (t(47.763) = 4.652, p = .001, r = .558). Finally, using the EIT grammatical and ungrammatical sentences mean scores, the reliability coefficients for internal consistency (Cronbach's α) for the EIT composite scores (N = 28), the EIT grammatical sentences composite scores (N = 14) and the EIT ungrammatical sentences composite scores (N = 14) were .88, .86, and .63, respectively.

Table 4.7. Descriptive Statistics for the EIT Group and Entire Sample Scores

Grammatical Structure	All (N=102)		Primary (N=41)		Secondary (N=29)		University (N=32)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Progressive *-ing*	.59	.28	.52	.30	.62	.30	.67	.21
Past *-ed*	.50	.31	.35	.27	.44	.28	.76	.21
Plural *-s*	.50	.29	.37	.30	.47	.26	.71	.20
Adverb placement	.38	.20	.30	.20	.36	.20	.48	.15
Structural parallelism	.27	.28	.16	.27	.25	.26	.44	.25
Third person singular *-s*	.51	.27	.33	.24	.56	.24	.70	.17
Possessive *-s*	.20	.27	.07	.14	.22	.25	.35	.32
Possessive determiner	.49	.31	.34	.29	.57	.32	.63	.22
Definite article	.48	.33	.43	.31	.45	.34	.56	.34
Locative inversion	.20	.24	.13	.19	.10	.22	.39	.22
Wh question	.20	.24	.14	.22	.20	.24	.28	.25
Pseudo-cleft	.16	.26	.04	.09	.18	.26	.31	.32
Dative alternation	.41	.31	.25	.28	.40	.32	.61	.20
Passive	.54	.35	.43	.36	.53	.38	.69	.26
Grammatical composite	.54	.26	.38	.23	.53	.24	.75	.14
Ungrammatical composite	.24	.12	.17	.10	.23	.10	.33	.09
EIT composite	.39	.18	.28	.15	.38	.15	.54	.10

4.1.4.2 Timed Grammaticality Judgment Test (TGJT)

The resulting descriptive statistics for the TGJT percentage scores are in Table 4.8. The standardised scores of grammatical and ungrammatical sentences mean scores for each target structure had indicated eleven absolute z-scores greater than 3.29 across the groups. These were replaced accordingly except for two outliers in the pseudo-cleft ungrammatical scores for the university students which were replaced by the minimum score value for the structure, category and group. For the entire sample, the z-scores had shown five outliers in the grammatical sentences' mean scores across the grammatical structures. These were also replaced accordingly.

Table 4.8. Descriptive Statistics for the TGJT Group and Entire Sample Scores

Grammatical Structure	All (N=102)		Primary (N=41)		Secondary (N=29)		University (N=32)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Progressive -*ing*	.74	.21	.65	.22	.72	.19	.86	.15
Past -*ed*	.70	.21	.63	.20	.62	.20	.85	.15
Plural -*s*	.69	.19	.58	.19	.73	.15	.79	.16
Adverb placement	.72	.19	.67	.21	.68	.17	.80	.15
Structural parallelism	.49	.20	.49	.22	.48	.18	.51	.18
Third person singular -*s*	.69	.22	.59	.23	.69	.21	.83	.14
Possessive -*s*	.55	.30	.37	.25	.53	.29	.80	.19
Possessive determiner	.57	.24	.51	.23	.58	.23	.65	.23
Definite article	.59	.23	.54	.25	.54	.21	.68	.18
Locative inversion	.57	.21	.53	.23	.61	.25	.57	.15
Wh question	.67	.26	.56	.25	.60	.25	.88	.16
Pseudo-cleft	.55	.17	.52	.20	.55	.15	.60	.13
Dative alternation	.69	.21	.59	.24	.67	.16	.82	.14
Passive	.76	.21	.64	.18	.78	.22	.91	.12
Grammatical composite	.69	.13	.65	.14	.70	.11	.74	.11
Ungrammatical composite	.58	.20	.48	.18	.53	.16	.76	.11
TGJT composite	.64	.11	.56	.08	.62	.09	.75	.06

The standardised values of the target structure scores, grammatical and ungrammatical sentences composite scores and the TGJT composite scores indicated one instance of absolute z-scores greater than 3.29 across the groups. In the entire sample, two outliers were identified across the four sets of scores. These were replaced accordingly, using either the group or standard deviation of the respective target structures and sentence categories. The standardised skewness and kurtosis scores for the target structure scores and grammatical and ungrammatical sentences composite scores indicated that the university students' scores on the participle *-ing* and passive structures were significantly skewed. The passive scores were also shown to be significantly kurtotic. With the entire sample, the participants' scores on only the dative alternation were significantly skewed. These sets of scores were considered non-normal.

The Shapiro-Wilk tests of normality for the TGJT composite scores indicated normal distributions in the primary school scores ($D(41) = .981$, $p = .714$), in the secondary school scores ($D(29) = .978$, $p = .781$) and the university students scores ($D(32) = .940$, $p = .077$), but significantly asymmetrical distribution in the entire sample scores ($D(102) = .973$, $p = .036$). However, no absolute skewness and kurtosis z-scores were greater than 2.58 in any group and entire sample scores, indicating the TGJT scores were approximately normally distributed. The one-way ANOVA test statistic, with Welch's procedure, indicated the mean scores across groups were statistically significantly different (F (2, 61.538) = 67.705, $p = .001$, $w^2 = .503$). With unequal variance assumed, planned contrasts indicated significant differences in all the three comparisons as follows: between primary and second school groups ($t(58.216) = 2.773$, $p = .007$, $r = .342$); between the primary school and university student groups ($t(70.069) = 11.268$, $p = .001$, $r = .803$); and between the secondary school and university student groups ($t(47.376) = 6.638$, $p = .001$, $r = .694$). Using the grammatical and ungrammatical sentences mean scores, the reliability coefficients for internal consistency (Cronbach's α) for the TGJT composite scores (N = 28), the TGJT grammatical sentences composite scores (N = 14) and the TGJT ungrammatical sentences composite scores (N = 14) were .74, .68, and .83, respectively.

4.1.4.3 Untimed Grammaticality Judgment Test (UGJT)

The resulting descriptive statistics for the UGJT percentage scores are in Table 4.9. The computed standardised values of grammatical and ungrammatical sentences mean scores for each target structure had indicated five outliers across the three groups. These were replaced accordingly. For the entire sample, the z-scores indicated four outliers across grammatical and ungrammatical sentences mean scores for all target grammatical structures. These were replaced accordingly, using the entire sample mean and standard deviation. The standardised values of the target structure scores, grammatical and ungrammatical sentences composite scores and the UGJT composite scores indicated two cases in the target structure scores with an absolute z-score greater than 3.29 across the groups, which were replaced accordingly. In the entire sample scores, none of the z-scores were greater than 3.29.

Table 4.9. Descriptive Statistics for the UGJT Group and Entire Sample Scores

Grammatical Structure	All (N=102)		Primary (N=41)		Secondary (N=29)		University (N=32)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Progressive *-ing*	.76	.21	.68	.23	.76	.20	.87	.13
Past *-ed*	.69	.20	.55	.16	.71	.15	.86	.16
Plural *-s*	.73	.18	.65	.20	.76	.16	.79	.15
Adverb placement	.70	.18	.63	.20	.72	.17	.78	.14
Structural parallelism	.56	.18	.51	.17	.51	.18	.66	.16
Third person singular *-s*	.69	.21	.56	.18	.75	.19	.81	.18
Possessive *-s*	.65	.25	.47	.20	.71	.22	.85	.12
Possessive determiner	.52	.19	.48	.13	.57	.23	.53	.20
Definite article	.67	.18	.59	.17	.67	.17	.77	.16
Locative inversion	.61	.19	.58	.20	.61	.18	.65	.19
Wh question	.65	.21	.61	.20	.61	.20	.73	.21
Pseudo-cleft	.56	.15	.55	.18	.54	.12	.59	.15
Dative alternation	.78	.14	.70	.15	.80	.12	.86	.11
Passive	.82	.17	.74	.20	.83	.14	.92	.09
Grammatical composite	.70	.13	.68	.14	.68	.14	.75	.10
Ungrammatical composite	.63	.19	.49	.17	.68	.17	.76	.12
UGJT composite	.67	.11	.59	.08	.68	.08	.76	.07

Across the groups, the UGJT's standardised skewness and kurtosis values of the target structure scores and grammatical and ungrammatical sentences composite scores indicated only the university students' scores on the pseudo-cleft structure to be significantly skewed. For the entire sample scores, the skewness and kurtosis z-scores indicated the scores only on the participle *-ing* and passive structures to be significantly skewed and, therefore, not normally distributed.

The Shapiro-Wilk tests of normality for the UGJT composite scores indicated normal distributions in the primary school scores (*D*(41) = .960, *p* = .154), the secondary school scores (*D*(29) = .967, *p* = .470) and the university students scores (*D*(32) = .945, *p* = .105), but significantly asymmetrical distribution in the entire sample scores (*D*(102) = .971, *p* = .024). However, no absolute skewness and kurtosis z-scores were greater than 2.58 in any group and entire sample scores, indicating the UGJT scores were approximately normally distributed. The one-way ANOVA test statistic, with Welch's procedure, indicated the mean scores across groups were statistically significantly different (*F* (2, 63.092) = 47.357, *p* = .001, w^2 = .455). With equal variance assumed, planned contrasts indicated significant differences in all the three comparisons as follows: between primary and second school groups (*t*(99) = 4.708, *p* = .001, *r* = .428); between the primary school and university student groups (*t*(99) = 9.302, *p* = .001, *r* = .683); and between the secondary school and university student groups (*t*(99) = 4.102, *p* = .001, *r* = .381). Using the UGJT grammatical and ungrammatical sentences mean scores, the reliability coefficients for internal consistency (Cronbach's α) for the UGJT composite scores (N = 28), the UGJT grammatical sentences composite scores (N = 14) and the UGJT ungrammatical sentences composite scores (N = 14) were .76, .77, and .87, respectively.

4.2 Validation of Test Instruments

The following sections report inferential statistics that were conducted to provide some evidence for the validity of the memory ability and language knowledge test instruments. For each set of test instruments, evidence for test validity was investigated by first calculating correlation coefficients between test instruments. Following the correlation analysis, Exploratory Factor Analysis (EFA, using SPSS Version 22.0) and Confirmatory Factor Analysis (CFA, using EQS Version 6.4) were conducted to investigate whether the sets of the six memory tests and of the three language knowledge tests would load on two components according to the study design. The entire sample (N = 102) scores were used in this process of test validation. The plural *-s*, was excluded from these analyses and consequently from the whole data set because test sentences for the structure contained noun phrases of two different structures, with some test items having determiners (e.g., ‘Two boys ...’) while others lacking determiners (e.g., ‘Boys ...’). Based on the proposed criterion to determine grammar rule complexity, the plural ‘-*s*’ rule test items could have been differently categorised as complex for those containing a determiner and as simple for those lacking a determiner.

4.2.1 Exploration of Memory Ability Test Instruments

Based on the study design and the review of literature (see Chapters 2 and 3), it was hypothesized that there should be significant correlations between CVMT, Llama-B, 3-Term and DAT-V tests because they measure declarative memory ability, and between SRT and Llama-D tests because they measure procedural memory ability. It was further hypothesised that DAT-V would correlate with SRT and Llama-D since DAT-V was found to have directly predicted implicit learning in Kaufman et al. (2010).

Table 4.10. Spearman's Correlations among the Memory Ability Measures

Measure	Declarative memory				Procedural memory	
	CVMT	Llama-B	3-Term	DAT-V	Llama-D	SRT
CVMT	__					
Llama-B	.14 (.16)	__				
3-Term	.17 (.09)	.44**(.001)	__			
DAT-V	.33**(.001)	.48**(.001)	.41**(.001)	__		
Llama-D	-.07(.49)	.03 (.76)	.13 (.20)	.10(.29)	__	
SRT	.04 (.67)	.03 (.74)	-.05 (.63)	-.02(.86)	-.30**(.001)	__

Note. N = 102; **. Correlation significant at p < .01; CVMT=Continuous Visual Memory Test; 3-Term=Three Term Contingency Learning; DAT-V=Differential Aptitude Test – Verbal section; SRT = Serial Reaction Times. The values in parentheses are *p* values (2-tailed).

The resulting Spearman's *rho* correlations are in Table 4.10. The DAT-V (verbal learning) was positively but moderately and significantly correlated with all the other three declarative memory ability measures. The Llama-B scores were also moderately and significantly related to the 3-Term scores. The lack of significant correlations between the two verbal declarative memory measures (i.e., the Llama-B and 3-Term) and CVMT (a visual non-linguistic test) may indicate that the CVMT indexed to some extent unrelated learning abilities, with presumably the verbal measures indexing domain-specific and the non-linguistic test indexing domain-general learning abilities.

The two procedural memory measures were negatively but moderately and significantly correlated. Overall, the lack of statistically significant correlations between declarative memory and procedural memory measures supports the hypothesised relationship among the two sets of measures. Further, contrary to the results reported in Kaufman et al. (2010), the DAT-V was not significantly related to any of the two procedural memory ability measures.

4.2.2 Exploratory Factor Analysis for memory measures

To define the underlying structure among the memory abilities measures, Exploratory Factor Analysis (EFA) was conducted. It was hypothesized that CVMT, Llama-B, 3-Term and DAT-V tests would load on one factor and SRT and Llama-D would load on a different factor. The analysis began with the examination of the correlation matrix and some initial EFA statistical requirements. The careful inspection of the correlation matrix (see Table 4.10) showed only one-third of the correlations had the absolute value of .30 and above. Several correlations were below the recommended .30 (see Phakiti, 2018b). However, this is not surprising as the two sets of tests are hypothesised to measure two distinct learning abilities. Table 4.11 shows the initial and extraction values of a two-factor solution EFA, with the Principal Axis Factoring (PAF) method. EFA’s initial communalities are based on the variance for each item accounted for by all factors (Osborne, 2014 cited in Phakiti, 2018b). It is recommended that items with extraction values less than .30 be subsequently excluded (Phakiti, 2018b). In the present study, the CVMT and SRT were retained despite each having an extraction value of less than .30.

Table 4.11. Initial and Extraction Values Based On the PAF Method

Communalities		
Measure	Initial	Extraction
CVMT	.126	.103
Llama-B	.333	.463
3-Term	.262	.344
DAT-V	.381	.610
Llama-D	.136	.628
SRT	.090	.133

Extraction Method: Principal Axis Factoring

The Kaiser-Meyer-Olkin (KMO) test of data sampling adequacy and Bartlett's test of sphericity were also conducted. The KMO statistic is required to be larger than .60 (Fabrigar & Wegener, 2012; Field, 2013, cited in Phakiti, 2018b). In this dataset, the KMO was .63. The Bartlett's test provides the statistical significance that the correlation matrix has significant correlations among at least some of the variables (Hair et al., 2014). The requirement is that this statistic should be significant at .05 (Fabrigar & Wegener, 2012; Osborne, 2014, cited in Phakiti, 2018b). For this dataset, the Bartlett's test was significant, χ^2 (15) = 83.900, p = .001. With the PAF extraction method and a two-common factor solution of eigenvalues of above 1, the total variance explained was 56%. PAF is a reasonably robust and a commonly chosen method for factor extraction even in conditions when univariate and/or multivariate normality assumptions have been violated (Phakiti, 2018b).

Table 4.12. Factor Loadings for the Two-Long Memory Ability Measures

Measure	Factor 1	Factor 2
DAT-V	**.769**	.052
Llama-B	**.692**	-.089
3-Term	**.563**	.088
CVMT	**.326**	-.089
Llama-D	-.029	**.798**
SRT	.018	**-.367**

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

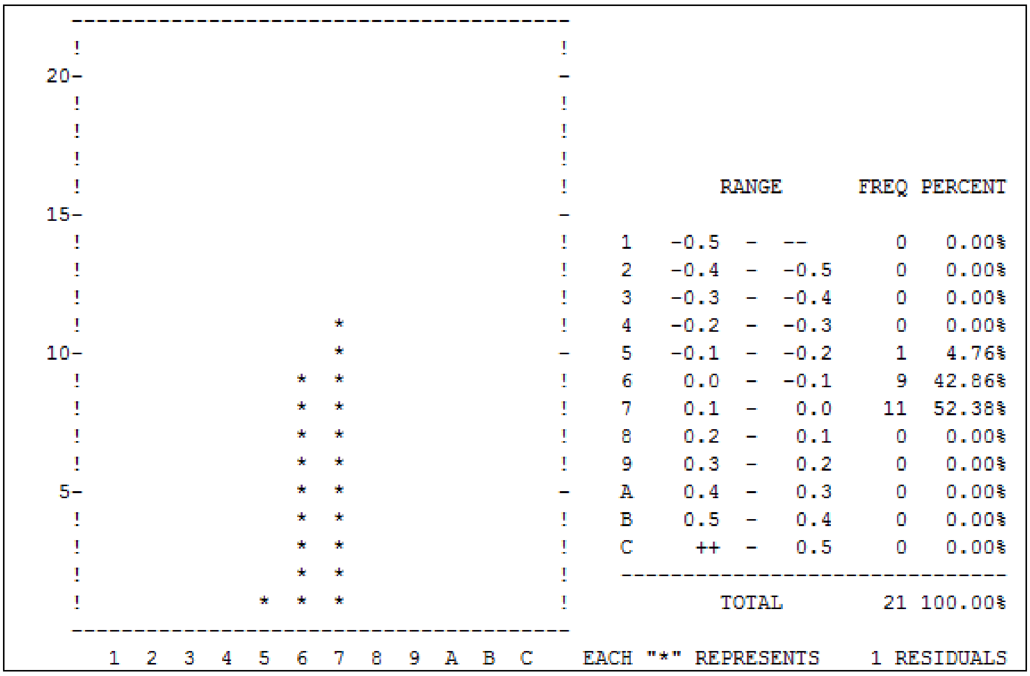
The EFA factor loading results with the PAF extraction and the oblique's Promax rotation methods are presented in Table 4.12. The six tests clearly loaded on the two factors, with DAT-V, Llama-B, CVMT and 3-Term loading on the first factor (consequently named 'declarative memory factor') and SRT and Llama-D on the second factor (named 'procedural memory factor'). However, the SRT test was negatively related to the factor while the Llama-D test was positively related, suggesting that these two tests measured two different aspects of the factor. The Promax rotation method was preferred to orthogonal rotation methods (e.g., Equimax, Varimax and Quartimax). Orthogonal rotation methods assume that factors are unrelated, while oblique rotation methods (e.g., Oblimin and Promax) assume that the factors may be correlated (Phakiti, 2018b). In the present study, some of the proposed models of DP (see Chapter 2) assume that declarative memory interfaces with procedural memory, suggesting that these two memory systems (or factors) may be correlated. Finally, the two factors were weakly and non-significantly related (*r* = .19).

4.2.3 Confirmatory factor analysis for memory ability measures

A Confirmatory Factor Analysis (CFA), using the normal theory maximum likelihood (ML) estimation and the Elliptical non-normal correction, was conducted not only to confirm the manifestation of the factor indicators as identified in the EFA but also to examine how well the specification of the factors in relation to the measured variables matched the actual data. The CFA model assumed the unidimensionality of constructs where each measured variable is hypothesized to relate to, or to be determined by, only a single factor with all cross-loadings fixed at zero. Hair et al. (2014) recommend running CFA models that do not include cross-loadings or within- and between-construct error covariances because significant estimates of these paths raise serious questions about both construct and discriminant validity. The terms *factor, construct* and *latent variable* are used interchangeably. Likewise, the terms *indicator*, and *observed* or *measured variable* are also used interchangeably. Based on the exploratory factor analysis results above, the CVMT, Llama-B, 3-Term and DAT-V tests constituted the declarative memory construct indicators while SRT and Llama-D were the indicators for the procedural memory construct. The covariance matrix used in the analysis is in Table 4.13 in Appendix C.

The evaluation of the model adequacy was based on the inspection of the values of the multivariate normality, standardised residuals, the chi-square statistic and other goodness-of-fit indices. The assumption of the multivariate normality of the observed variables indicated that the observations were independent and identically distributed, with *Mardia's coefficient* (G2, P) of -1.5141 and an associated *z* estimate of -0.7803. It is recommended that *Mardia's coefficient* should not exceed 3 and that the multivariate kurtosis statistic be zero or near zero in CFA and SEM (Phakiti, 2018a). In this dataset, the kurtosis statistic was -0.0222, which was considered as close to zero. The results further showed there were no multivariate outliers (i.e. no participants had two or more extreme scores).

The examination of the standardised residual matrix and the associated average absolute standardized residual and average off-diagonal absolute standardized residual values (in Table 4.14 in Appendix C) showed that all the values centred around zero[13]. It is recommended that these values should be around zero and that their distribution be symmetrical (Byrne, 2006). Any absolute values greater than 2.58 are considered larger (Byrne, 2006). The review of the frequency distribution (Figure 4.1) revealed that most residual values (95.24%) were between -0.1 and 0.1 while the remaining residuals, 4.76% fell between -0.1 and -0.2. The more residuals over the absolute value of 0.10, the less explanatory power the model has (Phakiti, 2018a). Therefore, in this study, the distribution of the residual values was relatively symmetric and quite centred around zero. Overall, these results suggest that the model as whole appeared to be fitting quite well despite the minimal discrepancy in the residuals distribution. Further, the average absolute standardized residual, the average off-diagonal absolute standardized residual and the highest standardised residual values (Table 4.15) all reflected a very good model fit to the data.



```
   ----------------------------------------
   !                                      !
 20-                                      -
   !                                      !
   !                                      !
   !                                      !
   !                                      !            RANGE      FREQ PERCENT
 15-                                      -
   !                                      !    1   -0.5  -  --       0   0.00%
   !                                      !    2   -0.4  -  -0.5     0   0.00%
   !                                      !    3   -0.3  -  -0.4     0   0.00%
   !                    *                 !    4   -0.2  -  -0.3     0   0.00%
 10-                    *                 -    5   -0.1  -  -0.2     1   4.76%
   !                 *  *                 !    6    0.0  -  -0.1     9  42.86%
   !                 *  *                 !    7    0.1  -   0.0    11  52.38%
   !                 *  *                 !    8    0.2  -   0.1     0   0.00%
   !                 *  *                 !    9    0.3  -   0.2     0   0.00%
  5-                 *  *                 -    A    0.4  -   0.3     0   0.00%
   !                 *  *                 !    B    0.5  -   0.4     0   0.00%
   !                 *  *                 !    C     ++  -   0.5     0   0.00%
   !                 *  *                 !    -------------------------------
   !              *  *  *                 !            TOTAL        21 100.00%
   ----------------------------------------
      1  2  3  4  5  6  7  8  9  A  B  C    EACH "*" REPRESENTS    1 RESIDUALS
```

Figure 4.1. Residual histogram and frequency chart for the memory ability dataset

Model fit was evaluated through the examination of goodness-of-fit indexes (e.g. chi-square test, CFI, RMSEA, SRMR, in Table 4.15), parameter estimates and parsimony

---

[13] The standardised residuals are “analogous to Z-scores” and are therefore easier to interpret as compared to unstandardized residual values (Byrne, 2006, p. 94).

for well-fitting model. These statistics are reported in Table 4.15. The independence (or null) chi-square (χ^2) statistic reported for the likelihood ratio test of the Bentler and Bonett (1980) null hypothesis was comparatively high (i.e., $\chi^2_{(15)} = 95.979$). The null model is typically used as a baseline against which alternative models can be compared to evaluate the gain in improved fit (Byrne, 2006). The χ^2 of the alternative model does not only need to be much smaller than that of the null model but also needs its probability value non-significant (Byrne, 2006; Phakiti, 2018a). Byrne (2006) points out that the χ^2 test simultaneously tests the null hypothesis (H_0) postulating that specification of the factor loadings, factor variances-covariances, and error variances for the model under study are valid. Therefore, the probability value associated with χ^2 indicates the likelihood of obtaining a χ^2 value that exceeds the χ^2 value when H_0 is true, suggesting that the higher the probability associated with χ^2, the closer the fit between the hypothesised model and the perfect fit (Bollen, 1989a cited in Byrne, 2006). In the present study, the test of H_0 obtained the χ^2 value of 9.464, with 8 degrees of freedom and a probability of .305, suggesting that the data did fit the hypothesised model quite well.

Parsimony fit indexes (Table 4.15) namely Akaike’s (1987) information criterion (AIC) and Bozdogan’s (1987) consistent version of the AIC (CAIC) were used to assess the number of estimated coefficients required to achieve a specific level of fit so that the model is not overidentified. The two statistics are used to compare two or more models, with the smaller values representing a better fit of the hypothesised model (Byrne, 2006). In comparison to the independence model, the hypothesised two-factor model in the present study had a substantially small AIC and CAIC statistic, representing a better fit of the hypothesised model.

The three relative or incremental fit indexes, namely the Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Comparative Fit Index (CFI) and Incremental Fit Index (IFI), were consistent in indicating that the hypothesised two-factor model in the study represented an adequate fit to the data. The NFI, NNFI and CFI values of .90 are deemed sufficient to accept the tested model (Phakiti, 2018a) whereas values of 1.0 suggest perfect fit. These relative and incremental fit indexes measure the proportionate improvement in fit by comparing a hypothesised model with the independence or null model (Byrne, 2006).

Table 4.15. Goodness-of-Fit Indices for the for the memory ability dataset

Standardised residuals	
Average Absolute Standardized Residual	0.0337
Average Off-Diagonal Absolute Standardized Residual	0.0472
Null model	
Independence Model CHI-SQUARE (df =15)	95.979
Independence AIC	65.979
Independence CAIC	11.605
Hypothesised CFA model	
Model AIC	-6.536
Model CAIC	-35.536
CHI-SQUARE (df = 8)	9.464
P-value for the CHI-SQUARE Statistic	0.305
Fit Indices	
Bentler-Bonett Normed Fit Index	0.901
Bentler-Bonett Non-Normed Fit Index	0.966
Comparative Fit Index (CFI)	0.982
Bollen's (IFI) Fit Index	0.983
Mcdonald's (MFI) Fit Index	0.993
Joreskog-Sorbom's GFI Fit Index	0.971
Joreskog-Sorbom's AGFI Fit Index	0.923
Root Mean-Square Residual (RMR)	5.678
Standardized RMR	0.048
Root Mean-Square Error of Approximation (RMSEA)	0.043

Similarly, the three ‘absolute’ fit indexes, namely the McDonald’s Fit Index (MFI), Goodness-Of-Fit Index (GFI) and the Adjusted Goodness-of-Fit Index (AGFI), had values greater than .90, suggesting that the hypothesised model did fit the data quite well. These indexes are not comparative; instead, they depend only on how well the hypothesised model fits the sample data (Byrne, 2006).

Finally, termed ‘absolute *misfit* indexes’, the Root Mean-Square Residual (RMR), Standardised Root Mean-Square Residual (SRMR), Root mean squared error of approximation (RMSEA) indexes “decrease as goodness-of-fit improves and attain their lower-bound value of zero when the model fits perfectly” (Byrne, 2006, p. 99). The

RMR values are relative to the sizes of the observed variances and covariances, making them difficult to interpret (Byrne, 2006). The standardised residuals range from zero to 1, with a smaller value (e.g., .05) reflecting a well-fitting model (Byrne, 2006). The SRMR and RMSEA for the hypothesised two-factor structure in this study (Table 4.15) were .048 and .043 respectively, suggesting that the hypothesised model represented an adequate fit of the data. The two factors in the model were weakly and non-significantly correlated (*r* = .16). In summary, the evaluation of all the statistics for the hypothesised model fit provided strong evidence to accept the model.

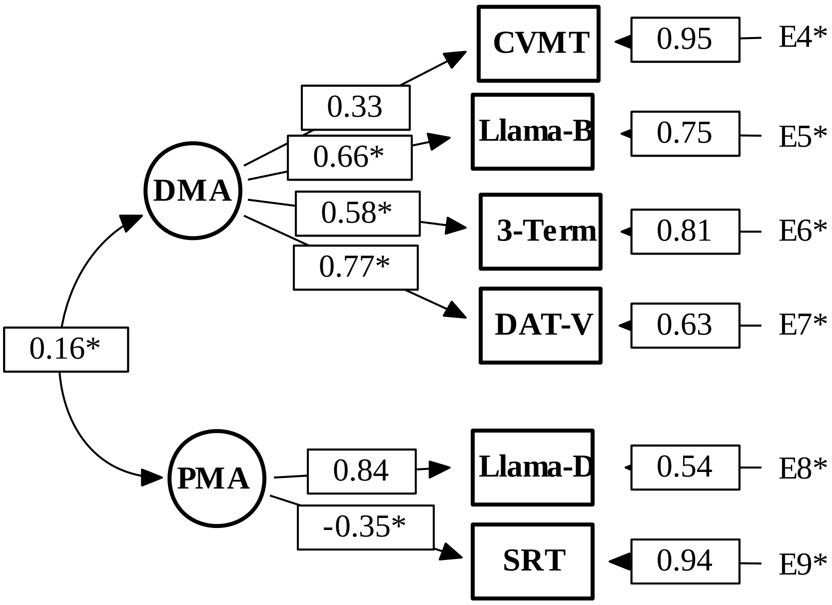


Figure 4.2. Memory abilities' two-factor CFA model with standardised parameter estimates. Note. N = 102; * Estimated parameters; DMA = Declarative memory ability; PMA = Procedural memory ability; CVMT = Continuous Visual Memory Test; 3-Term = Three Term Contingency Test; DAT-V = Differential Aptitude Test – Verbal section; SRT = Serial Reaction Test.

Figure 4.2 presents the CFA results of the two-factor memory ability model, with the standardised parameter estimates. While the indicators for declarative memory ability were positively related to the factor, SRT and Llama-D were, respectively, negatively and positively related to the latent variable. Further, the indicators accounted for .368 (i.e., 37%) and .414 (i.e., 41%) of the total common factor variance for declarative memory and procedural memory ability, respectively[14]. In both cases, the factor

---

[14] The total common factor variance (h^2) is the sum of squared factor loadings divided by the number of variables. $1-h^2$ is then the amount of unexplained variance (see Phakiti, 2018).

accounted for about 60% or above of the proportion of the variance of the related strongest measured variable. Generally, these path estimates are similar to the factor loadings obtained in the EFA above.

The evaluation of the parameter estimates proceeded based on the estimates' statistical significance and their appropriateness or feasibility, that is, whether any estimates fell outside the admissible absolute value range of 0 and 1 (e.g., Byrne, 2006; Phakiti 2018a). The unstandardized and standardised parameter estimates are presented in Table 4.16. The estimates for all the freely estimated parameters, except that of the SRT, were statistically significant at the .05 level and within the admissible range. Finally, the standard error for the SRT was excessively small (i.e., approaching 0). It has been argued that if a standard error approaches zero or if it is excessively large, the test statistic for its related parameter cannot be defined (Bentler, 2005, cited in Byrne, 2006). Standard errors are influenced not only by the units of measurement in variable but also by the magnitude of the parameter estimate itself (Byrne, 2006).

Table 4.16. Statistical significance of parameter estimates for memory ability CFA model

	Unstandardized		Standardised		R^2
Path	DMA	PMA	DMA	PMA	
CVMT	1.000 (.000)		.326		.106
Llama-B	39.363@ (14.293)		.663		.440
3-Term	45.977@ (17.162)		.580		.337
DAT-V	.635@ (.230)		.774		.600
Llama-D		1.000 (.000)		.843	.711
SRT		-.001 (.001)		-.348	.121

Note. DMA = Declarative memory ability factor;
PMA = Procedural memory ability factor; 1.000 Parameter fixed to 1;
@ Statistic significant at the 5% level; Standard errors in parentheses

Non-significant parameters are considered unimportant to a model and should be deleted given adequate sample size (Byrne, 2006). However, research suggests that the statistical significance of parameter estimates is tied not only to sample size but also to model complexity. Simpler models can be tested with smaller samples whereas models with more measured variables require larger samples to produce more information and greater stability (Hair et al., 2014). Larger samples mean less variability and increased stability in the solutions. The following are ways in which models can be complex and therefore requiring larger sample sizes: (i) more constructs that require more parameters to be estimated; (ii) constructs having less than three measured/indicator variables; and (iii) multigroup analyses requiring an adequate sample for each group (Hair et al., 2014). Further, research has shown that larger sample sizes are required as communalities (i.e., as the square of the standardized construct loadings) become smaller (Hair et al., 2014). The problem is exaggerated when models have constructs with only one or two measured or observed variables (Hair et al., 2014). Concluding their discussion on the impact of sample sizes and model complexity, Hair et al., (2014) suggest among others that (i) models containing five or fewer constructs, each with more than three items (observed variables) and with high item communalities (.6 or higher) should have the minimum sample size of 100, and (ii) models with seven or fewer constructs, lower communalities (below .45), and/or multiple under-identified (fewer than three) constructs, must have the minimum sample size of 300. In the present study, the SRT and Llama-D tests were the only two indicators of the procedural memory ability factor, making the model complex. Further, the SRT communality was as low as .14. It was therefore concluded that the SRT nonsignificant parameter estimate in the CFA model was indicative of the small sample size in the present study.

4.2.4 Validating language knowledge test instruments

Pearson correlation coefficients were calculated to investigate whether the EIT, TGJT and UGJT tapped the types of knowledge they were designed for. Based on the test design, it was hypothesized that all three language tests may be correlated because the UGJT may indicate both procedural and declarative language knowledge as grammatical errors can also be assessed based on an unarticulated, procedural knowledge of the rules (Akakura 2009). However, EIT and TGJT would be strongly correlated as the tests are designed to measure procedural knowledge. The UGJT measures declarative knowledge. Consistent with the first hypothesis but contrary to the second, the EIT was even more strongly and significantly correlated with the UGJT (r = .768, p = .001) than was related to the TGJT (r = .685, p = .001). Similarly, the TGJT was more strongly and significantly related to the UGJT (r = .739, p = .001).

Based on the test design, it was hypothesised that in a two-factor solution EFA, the EIT and TGJT would load on one factor whereas UGJT would load on a different factor. Though the correlations (see above) were all above the recommended value of .30, they suggested that the hypothesised two-factor EFA was not attainable. Based on the PAF method, .743, .701, and .806 were the extraction values for the EIT, TGJT and UGJT respectively. Further, the tests of sampling adequacy and sphericity showed that the dataset was factorable, with the KMO statistic of .739 and a significant Bartlett's statistic, $\chi 2$ (3) = 174.198, *p* = .001. However, while a two-factor solution with PAF accounted for as large as 92.7% of the total variance explained, the second factor eigenvalue was substantially low (see Table 4.17), indicating that the two-factor extraction was not plausible. The results of the EFA with the PAF extraction method and oblique's Promax rotation method (Table 4.18) were inconsistent with what was hypothesized. The EIT and UGJT loaded heavily on Factor 1 while the TGJT loaded heavily on Factor 2. The UGJT also loaded quite considerably on Factor 2.

Table 4.17. Total variance explained for language knowledge measures

Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	2.461	82.047	82.047
2	.318	10.606	92.652
3	.220	7.348	100.000

Extraction Method: Principal Axis Factoring; a =When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Motivated in part by the fact that the UGJT loaded quite significantly on both factors as well as by previous research suggesting that L2 learners respond differently to the grammatical and ungrammatical sentences in grammaticality judgment tests (Bialystok, 1979 and Hedgcock, 1993, cited in R. Ellis, 2005a), a decision was made to separately examine the psychometric properties of the grammatical and ungrammatical sentences in the UGJT. The results of the Pearson Product Moment Coefficients showed that UGJT grammatical sentences composite score was moderately and significantly correlated with both the EIT score (*r* = .329, *p* = 001) and the TGJT score (*r* = .308, *p* = 001). The UGJT ungrammatical sentences composite score was strongly and significantly related to both the EIT score (*r* = .620, *p* = 001) and the TGJT score (*r* = .605, *p* = 001). These results were unexpected. Previous research (R. Ellis, 2005a; 2009c) suggests that UGJT ungrammatical sentences score provides the best measure of explicit knowledge. Consequently, one would expect the score to correlate less strongly with the EIT and TGJT, the purported measures of implicit language knowledge. The

EIT and TGJT scores' strong correlation with UGJT ungrammatical sentences score was strong, indicating that the participants may have extensively relied on the same learning abilities when completing the three tasks (see below for what may have accounted for these EIT results). For the TGJT, the fact that the time pressure allocation was for the whole test and not between test items could on some occasions allow participants to reflect on their decisions as they performed the test.

Table 4.18. Factor loadings for language knowledge measures

	Factor	
	1	2
Elicited Imitation Test	.722	.158
Untimed Grammaticality Judgment Test	.530	.399
Timed Grammaticality Judgment Test	.185	.672

Extraction Method: Principal Axis Factoring; Rotation Method: Promax with Kaiser Normalization; a. = Rotation converged in 3 iterations.

Table 4.19. Correlation matrix for the EIT and TGJT and UGJT grammatical and ungrammatical sentences composite scores

Measure	EIT	TGJT Grammatical	TGJT Ungrammatical	UGJT Grammatical	UGJT Ungrammatical
TGJT					
Grammatical	.30**	__			
Ungrammatical	.59**	-.09	__		
UGJT					
Grammatical	.33**	.58**	-.02	__	
Ungrammatical	.62**	.06	.65**	-.18	__

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

A decision was then made to separately examine the psychometric properties of the grammatical and ungrammatical sentences in the TGJT. This decision was based in part on previous research suggesting that L2 learners respond differently to the grammatical and ungrammatical sentences in a grammaticality judgment tests (Bialystok, 1979; Hedgcock, 1993, both cited in Ellis 2005a) as well as on the results from studies such as De Graaff (1997) studies which found 'no relationship between time pressure and accuracy in a timed grammaticality judgment test due to the possibility of explicit knowledge having been accessed in addition to the expected use of implicit knowledge' (Akakura, 2009, p. 29). The Pearson Product Moment Coefficients calculated between the EIT score and the TGJT and UGJT grammatical and ungrammatical sentences composite scores are in Table 4.19.

First, the EIT score was significantly and more strongly correlated with the ungrammatical sentences composite scores than with the grammatical sentences composite scores for the TGJT and UGJT. While the TGJT grammatical sentences composite score was unrelated to the ungrammatical sentences composite scores for the UGJT and TGJT, it was correlated significantly and more strongly with the UGJT grammatical sentences composite score than with the EIT score. In contrast, the TGJT ungrammatical sentences composite score correlated strongly and significantly with both the EIT score and UGJT ungrammatical sentences composite score, but it was unrelated to the UGJT grammatical sentences composite score. Consequently, a decision was made to conduct a second two-factor solution EFA, now with the EIT composite score, the TGJT grammatical sentences and the UGJT ungrammatical sentences composite scores. It was hypothesised that in a two-factor solution EFA, the EIT and TGJT grammatical sentences composite score would heavily load on one factor whereas UGJT ungrammatical sentences composite score would heavily load on a different factor.

Though the statistic of the test of sampling adequacy was lower (i.e., .466) than the recommended .600 and above, the examination of the correlation matrix (Table 4.19), the extracted values of .300 and above, the test of sphericity ($\chi 2$ (3) = 60.738, p = .001), and the first two factors' eigenvalues of about 1 and above showed that the dataset was factorable. The two-factor solution accounted for 89.0% of the total variance explained. With the PAF extraction method and Oblique's Promax rotation method, the UGJT ungrammatical sentences score and EIT loaded more heavily at .902 and .689 on Factor 1 but less heavily on Factor 2 at -.219 and .311, respectively. The grammatical sentences score in the TGJT loaded more heavily on Factor 2 at .599 but less heavily at -.116 on Factor 1. Erlam (2009) points out that an EIT test that measures implicit knowledge should show some stronger relationship with other measures of implicit language knowledge than with measures of explicit language knowledge. Because the EIT loaded together with the ungrammatical sentences composite score for the UGJT (which was inconsistent with the test design) but at the same time it loaded relatively considerably on Factor 2, a decision was made to exclude the EIT from further analyses. As the result of this decision, the grammatical sentences composite score for the TGJT was used as the measure of implicit or procedural language knowledge while the ungrammatical sentences composite score in the UGJT was used as the measure of explicit or declarative language knowledge.

The possible explanation for the behaviour of the EIT in the present study rests on the test's lack of being reconstructive (i.e., its failure in requiring participants to process language stimuli rather than simply to imitate by memorising). Erlam (2006; 2009) argues that the measure of implicit language knowledge must not simply be a measure of rote repetition. Instead, it must be reconstructive. In addition to the delay between presentation of stimuli and repetition, another feature that crucially contributes to the reconstructive nature of the EIT is allowing for participants to focus their attention on meaning rather than on form (Erlam, 2009). The design of the EIT in the present study had excluded this semantic aspect (see Chapter 3), a situation that might have allowed learners access to both explicit and implicit knowledge when they were completing the task. The exclusion of the EIT from further analysis rendered CFA on language knowledge measures impossible because there were now only two that measured the language knowledge variable.

4.3 Present study findings

4.3.1 Memory and language knowledge type

Research Question 1

Do instructed L2 English learners' declarative and procedural memory systems predict the knowledge of grammatical structures, as measured by the tests of implicit and explicit language knowledge?

To answer Research Question 1, zero-order correlations between memory abilities and language knowledge measures were examined. Then, to further assess the association between the two long-term memory abilities and the learners' implicit and explicit knowledge of the thirteen L2 English grammatical structures, SEM analyses that allowed the construction of latent variables were conducted. It was hypothesised that declarative memory would strongly predict the learners' explicit language knowledge while procedural memory would strongly predict implicit knowledge (*Hypothesis 1a*). It was further hypothesised that declarative memory would be strongly related to the learners' declarative knowledge of the grammatical structures which are explicitly taught while the learners' procedural knowledge of grammatical structures which are learned in a more implicit manner would be strongly predicted by the procedural memory system (*Hypothesis 1b*).

4.3.1.1 The association between memory and language knowledge

The results of the calculated Spearman's correlation coefficients between the memory ability and language knowledge measures in Table 4.20 suggest a partial support for the first hypothesis. The declarative memory measures were correlated significantly with the UGJT ungrammatical sentences composite score, with the DAT-V showing the strongest correlation. Further, the implicit language knowledge measure was moderately correlated with the DAT-V but was weakly related to the CVMT. No procedural memory measure was related to either of the language knowledge measures.

Table 4. 20. Spearman's correlations between language knowledge and memory ability measures

	Declarative memory				Procedural memory	
Measure	CVMT	Llama-B	3-Term	DAT-V	Llama-D	SRT
TGJT grammatical	.21*	.16	.10	.39**	.02	.03
	(.036)	(.100)	(.340)	(.000)	(.850)	(.793)
UGJT ungrammatical	.32**	.38**	.32**	.69**	.10	-.13
	(.001)	(.000)	(.001)	(.000)	(.332)	(.211)

**. Correlation is significant at the 0.01 level (2-tailed);
*. Correlation is significant at the 0.05 level (2-tailed);
P-values are in parentheses

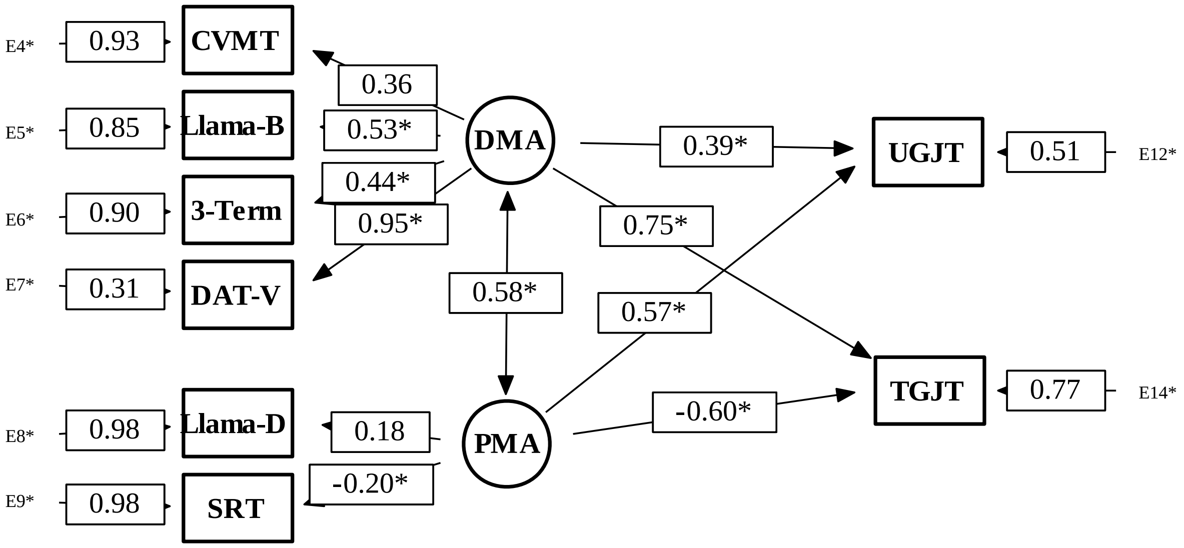


Figure 4.3. Initial SEM model for memory abilities and language knowledge types
Note. * Estimated parameter; DMA = Declarative memory ability; PMA = Procedural memory ability; SRT = Serial Reaction Times; UGJT= UGJT ungrammatical sentences composite score; TGJT = TGJT grammatical sentences composite score

The initial hypothesized SEM model consisted of the measurement model for memory systems and the structural model with the UGJT ungrammatical and TGJT grammatical sentences composite scores as dependent variables. The shared variance of CVMT, Llama-B, DAT-V and 3-Term formed the latent variable representing declarative memory ability. Llama-D and SRT shared variance formed the latent variable representing the procedural memory ability. Using the ML estimation and the heterogeneous geo mean non-normal estimator, the initial model with path loadings is presented in Figure 4.3.

The total number of freely estimated parameters in the model was 19 and the degrees of freedom were 17. Mardia’s Coefficient (G2, P) was −3.6188, with the normalized estimate of -1.4447. The average off-diagonal standardized residual was very small (0.0475) and the RMSEA (0.056), the CFI and GFI indices (0.974 and 0.946, respectively) were within the recommended values, indicating that the hypothesized model explained the observed covariances among the variables quite well. The χ^2 statistic was non-significant suggesting that there was no significant difference between the observed and the model-predicted variances and covariances. As shown in Table 4.21, all the fit indices provided substantial support for the initial hypothesized model.

Table 4.21. Goodness-of-Fit Indices for the initial SEM model

Null model	
Independence Model CHI-SQUARE (df =28)	237.242
Independence AIC	181.242
Independence CAIC	79.743
Hypothesised SEM model	
Model AIC	-11.565
Model CAIC	-73.19
CHI-SQUARE (df =17)	22.435
Probability value for the CHI-SQUARE Statistic	0.16857
The Normal Theory RLS CHI-SQUARE for this ML solution	22.435
Probability Value for the CHI-SQUARE statistic	0.169
Fit Indices	
Bentler-Bonett Normed Fit Index	0.905
Bentler-Bonett Non-Normed Fit Index	0.957
Comparative Fit Index (CFI)	0.974
Bollen's (IFI) Fit Index	0.975
Mcdonald's (MFI) Fit Index	0.974
Joreskog-Sorbom's GFI Fit Index	0.946
Joreskog-Sorbom's AGFI Fit Index	0.886
Root Mean-Square Residual (RMR)	9.717
Standardized RMR	0.068
Root Mean-Square Error of Approximation (RMSEA)	0.056

A decision was made that Llama-D and SRT must each directly predict the language knowledge types in the model. This decision followed in part from the fact that the two tests were found to be differently (i.e., positively and negatively) related to the latent variable, and most importantly from the evident parameter estimates' high instability for both indicators. As shown in the initial model, the Llama-D and SRT shared variance representing procedural memory latent variable had resulted in the significant decrease in the two tests' parameter estimates as compared to those obtained in the CFA, an indication that the parameter estimates were highly unstable. Schoonen (2015) points out that validity and/or reliability issues could be involved if a measured variable does not fit the hypothesized relations quite well. Since the memory abilities-two-factor CFA in the present study (see above) had shown a good model fit and construct validity, then the instability observed in the procedural memory latent variable in this initial SEM model might have been the result of the low reliability found in the SRT test which also had a relatively lower factor loading than that of Llama-D[15]. Alternatively, identification problem of the two-indicator procedural memory measurement model could result in the observed instability of the parameter estimates when regression paths are added in the structural model. The revised model is in Figure 4.4.

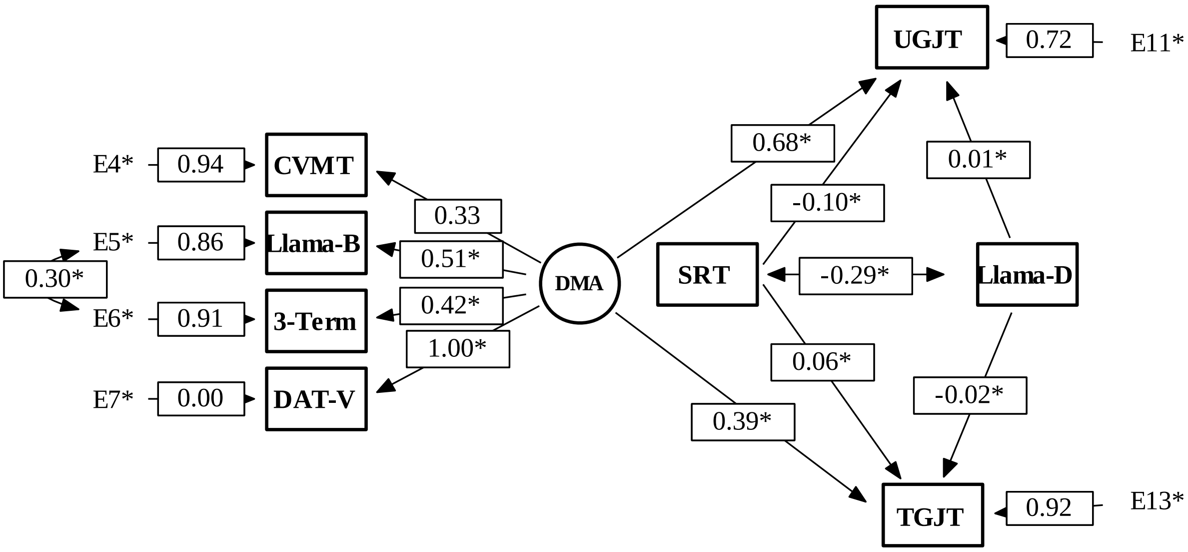


Figure 4.4. Revised SEM model for memory abilities and language knowledge types

Note. * Estimated parameter; DMA = Declarative memory ability; SRT = serial reaction times; UGJT= UGJT ungrammatical sentences composite score; TGJT = TGJT grammatical sentences composite score

The fit indices from the first run of the revised model had shown a poor model fit (see Table 4.22 in the appendices). The examination of the multivariate χ^2 statistic results of the Lagrange Multiplier (LM) test had shown that the error covariances for the Llama-B and 3-Term and for the TGJT grammatical and UGJT ungrammatical sentences composite scores could be freed in the model re-specification to improve the model fit. As reported above, the correlations between the two language knowledge measures was very low and not statistically significant. This was considered as evidence against freeing this error covariance parameter for estimation. The significance of not freeing the parameter for estimation was that discriminant validity between the two language knowledge measures was preserved. Using the ML estimation and the Elliptical non-

---

[15] However, note that the reliability estimate of the Llama-D was not computed (nor was that of Llama-B) because the test was not itemised.

normal estimator, the final analysis with the Llama-B and 3-Term error covariance now estimated showed a substantial improvement in the model fit (Table 4.23).

Table 4.23. Goodness-of-fit indices for memory abilities and language knowledge revised model after re-specification

Independence Model	
Independence Model CHI-SQUARE (df =28)	210.166
Independence AIC	154.166
Independence CAIC	52.666
Hypothesised Model	
Model AIC	-12.577
Model CAIC	-70.577
CHI-SQUARE (df = 16)	19.423
Probability value for the CHI-SQUARE Statistic	0.247
Fit Indices	
Bentler-Bonett Normed Fit Index	0.908
Bentler-Bonett Non-Normed Fit Index	0.967
Comparative Fit Index (CFI)	0.981
Bollen's (IFI) Fit Index	0.982
Mcdonald's (MFI) Fit Index	0.983
Joreskog-Sorbom's GFI Fit Index	0.956
Joreskog-Sorbom's AGFI Fit Index	0.901
Root Mean-Square Residual (RMR)	5.766
Standardized RMR	0.061
Root Mean-Square Error of Approximation (RMSEA)	0.046

The multivariate kurtosis’ Mardia's coefficient (G2,P)was -3.6188 with the normalized estimate of -1.4447.The total number of freely estimated parameters in this re-specified model was 20 and the degrees of freedom were 16.[16] The average off-diagonal standardized residual was still very small (0.0470) and the RMSEA (0.046), the CFI and GFI indices (0.981 and 0.956, respectively) were all within the recommended

---

[16] Note. Model modification and re-specification do not improve multivariate kurtosis statistic (see Phakiti, 2018a). Thus, the Mardia’s Coefficient (G2, P) and the Normalized estimate values are the same for both the initial and the final models.

values, indicating that the model explained the observed variances and covariances among the variables quite well. The χ^2 statistic was not significant, suggesting that there was no significant difference between the observed and model-predicted variances and covariances. As shown in Table 4.23, all the fit indices provided substantial support for the final re-specified SEM model.

Table 4.24. Parameter estimates for the memory and language knowledge model

	Unstandardized			Standardised		
	Llama-D	SRT	DMA	Llama-D	SRT	DMA
UGJT	.000	.892	.570@	.008	-.100	.682
	(.001)	(.654)	(.168)			
TGJT	.000	.364	.214@	-.017	.063	.391
	(-.001)	(.542)	(.077)			

Note. DMA = Declarative memory ability factor; UGJT = UGJT ungrammatical sentences composite score; TGJT = TGJT grammatical sentences composite score; @ Statistic significant at the 5% level; Standard errors in parentheses

The results of the statistical significance of parameter estimates are in Table 4.24. Declarative memory had a positive direct effect on both declarative and procedural language knowledge types, respectively explaining 46% (i.e. $\gamma = .68$; $R^2 = .46$) and 15% (i.e. $\gamma = .39$; $R^2 = .15$) of the variance. The effects of both the SRT and Llama-D on the two language knowledge types were not statistically significant and their parameter estimates were very low, suggesting that these predictors had no effect at all. The DAT-V which significantly correlated with the procedural language knowledge measure had an indirect weak effect on the procedural knowledge (i.e. $\gamma = .39$ [=1.00 x .39]; $R^2 = .15$), suggesting that all the observed direct effect of the declarative memory on the procedural knowledge could only be explained by the DAT-V’s correlation with the TGJT grammatical sentences.

4.3.1.2 Exposure, memory and language knowledge

To investigate the hypothesised double dissociation as the result of the exposure condition where the declarative memory would be strongly related to the learners’ performance on the explicitly taught grammatical structures while the procedural memory to implicitly taught grammatical structures *(Hypothesis 1b),* a pair of scores for each of the language knowledge tests was calculated: first, the set of scores for all grammatical structures where instruction is given (i.e., the 'Explicitly Taught' structures), and second, the set of scores for all grammatical structures where no instruction is provided (i.e., the 'Not Explicitly Taught' structures). Normality was explored by simply checking the distribution of the unstandardized residuals computed from the three group means. The Shapiro-Wilk statistic is reported. The UGJT ungrammatical sentences composite scores for the 'Explicitly Taught' grammatical structures (*N=9, mean* = .60, *SD* = .22; *D*(102) = .994, *p* = .911) were normally distributed. Scores for the 'Not Explicitly Taught' structures (*N=4*, m*ean* = .71, *SD* = .18; *D*(102) = .973, *p* = .033) were significantly non-normal. The TGJT grammatical sentences composite scores for the 'Explicitly Taught' grammatical structures (N=9, m*ean* = .77, *SD* = .13; *D*(102) = .979, *p* = .110) for 'Not Explicitly Taught' grammatical structures (N=4, m*ean* = .51, *SD* = .20; *D*(102) = .982, *p* = .170) were both normally distributed. Table 4.25 presents zero-order correlations among the four sets of scores.

Table 4.25. Spearman's correlations for UGJT ungrammatical and TGJT grammatical sentences composite scores for the two exposure types

	UGJT Ungrammatical		TGJT Grammatical	
	Explicitly taught(N=9)	Not explicitly taught(N=4)	Explicitly taught(N=9)	Not explicitly taught(N=4)
UGJT-Explicitly taught	__			
UGJT-Not explicitly taught	.60** (.000)	__		
TGJT-Explicitly taught	.20* (.045)	.10 (.306)	__	
TGJT-Not explicitly taught	-.00 (.983)	-.24*(0.015)	.40** (.000)	__

**. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed); P-values are in parentheses

The UGJT ungrammatical sentences composite score for the explicitly taught structures was strongly and significantly correlated with the score of the measure for the structures that are not explicitly taught but was weakly correlated with TGJT grammatical sentences composite score for the same exposure type. The TGJT grammatical sentences composite score for the explicitly taught structures was also related moderately to the TGJT grammatical sentences composite score for the implicitly learned structures. These results suggest that the type of exposure had no influence on the two language knowledge types. However, for the structures that are not explicitly taught, the declarative language knowledge measure was negatively but significantly correlated with the procedural language knowledge measure, suggesting that the two language knowledge types may have led learners' performance in opposite direction as a function of the type of exposure.

4.3.1.2.1 Effects of exposure on memory and declarative language knowledge

The zero-order correlations between memory abilities measures and the UGJT ungrammatical sentences composite scores for the explicitly taught structures and for the structures not explicitly taught are presented in Table 4.26.

Table 4.26. Spearman's correlations between memory ability measures and the UGJT ungrammatical sentences composite scores for 'Explicitly Taught' and 'Not Explicitly Taught' structures

Measures		UGJT Ungrammatical	
		Explicitly Taught Structures (N= 9)	Not Explicitly Taught Structures (N= 4)
Declarative Memory	CVMT	.34** (.000)	.12 (.216)
	Llama-B	.38** (.000)	.28** (.004)
	3-Term	.33** (.001)	.13 (.189)
	DAT-V	.71** (.000)	.45** (.000)
Procedural Memory	Llama-D	.09 (.365)	.11 (.290)
	SRT	-.11 (.254)	-.05 (.611)

**. Correlation is significant at the 0.01 level (2-tailed);
P-values are in parentheses

The UGJT ungrammatical sentences composite score for the explicitly taught structures was significantly correlated with all the four markers of declarative memory ability. These results were as expected: both the explicit instruction and declarative measure recruit conscious learning mechanisms/processes. For the structures not explicitly taught, the declarative language knowledge measure was significantly related to two markers of declarative memory ability, an indication that the knowledge of the grammatical structures that are acquired in a more implicitly way may have involved declarative memory.

To assess the role of exposure type on the association between the two long-term memory abilities and the declarative language knowledge, a SEM analysis was conducted. The declarative memory factor and the two procedural memory measures were predictors. The dependent variables consisted of the UGJT ungrammatical sentences composite scores for the ‘Explicitly Taught’ and the ‘Not Explicitly Taught’ structures. The initial structural model did not fit the data well (see the initial model fit indices in Table 4.27 in Appendix C).The LM test results had indicated that re-specifying for estimation the error covariances for the two UGJT grammatical sentences composite scores and for Llama-B and 3-Term had a significant effect in achieving a good model fit, with the incremental univariate LM χ^2 values of 14.866 and 7.040 and the standardised parameter change values of 0.612 and 0.295, respectively. However, the final structural model had only Llama-B and 3-Term error covariance freed for estimation during re-specification. To preserve the discriminant validity of the two language scores (though they were strongly correlated), the Llama-B and 3-Term error covariance was preferred to the error covariance of the language scores.

Figure 4.5 is the final structural model and path loadings with the ML Elliptical estimation. The model’s Mardia's Coefficient (G2, P) was -1.9862 with the Normalized Estimate of -0.7929.

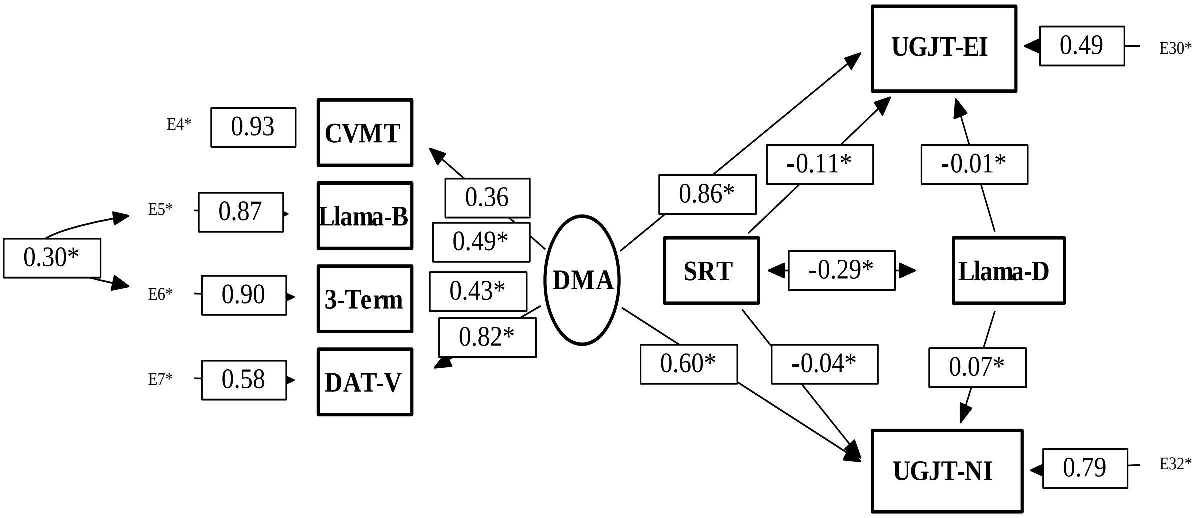


Figure 4.5. Revised SEM model for memory abilities, declarative knowledge and exposure
* Estimated parameters; DMA = Declarative memory ability; SRT = Serial Reaction Times; UGJT-EI = UGJT ungrammatical sentences composite score for explicitly taught structures; UGJT-NI = UGJT ungrammatical sentences composite score for implicitly learned structures

Table 4.28 presents the results of goodness-of-fit for the revised model. The total number of freely estimated parameters in this final model was 20 and the degrees of freedom were 16. The average off-diagonal standardized residual was very small (0.0534) and the RMSEA (0.061), the CFI and GFI indices (0.974 and 0.950, respectively) were all within the recommended values, indicating that the model explained the observed covariances among the variables quite well. The χ^2 statistic was again not statistically significant, indicating that there was no significant difference between the observed and model-predicted variances and covariances. All the fit indices in Table 4.28 provided substantial support for the final model.

Table 4.28. Fit indices for the memory abilities, declarative knowledge and exposure – revised model

Null Model	
Independence Model CHI-SQUARE (df = 28)	259.688
Independence AIC	203.688
Independence CAIC	102.189
Hypothetical Model	
Model AIC	-10.039
Model CAIC	-68.039
CHI-SQUARE (df = 16)	21.961
Probability Value for the CHI-SQUARE Statistic	0.144
Fit Indices	
Bentler-Bonett Normed Fit Index	0.915
Bentler-Bonett Non-Normed Fit Index	0.955
Comparative Fit Index (CFI)	0.974
Bollen's (IFI) Fit Index	0.976
Mcdonald's (MFI) Fit Index	0.971
Joreskog-Sorbom's GFI Fit Index	0.950
Joreskog-Sorbom's AGFI Fit Index	0.887
Root Mean-Square Residual (RMR)	5.767
Standardized RMR	0.059
Root Mean-Square Error of Approximation (RMSEA)	0.061

The parameter estimates with statistical significance are in Table 4.29. Declarative memory was found to have direct, positive significant effect on both exposure conditions, respectively explaining 74% (i.e. $\gamma = .863$; $R^2 = .74$) and 37% (i.e. $\gamma = .605$; $R^2 = .37$) of the variance in the language knowledge scores. The effect was very strong on the explicitly taught structures. The SRT and Llama-D had no effect.

Table 4.29. Parameter estimates for the model for memory abilities, declarative knowledge and exposure type

	Unstandardized			Standardised		
Path	Llama-D	SRT	DMA	Llama-D	SRT	DMA
UGJT Ungrammatical-Explicitly Taught (N=9)	.000 (.001)	-1.137 (.708)	.757@ (.214)	-.011	-.111	.863
UGJT Ungrammatical-Not Explicitly Taught (N=9)	.001 (.001)	-.330 (.741)	.438@ (.134)	.074	-.039	.605

Note. DMA = Declarative memory ability factor; @ Statistic significant at the 5% level; Standard errors in parentheses

4.3.1.2.2 Effects of exposure on memory and procedural language knowledge

Correlational and SEM analyses were conducted to investigate the role of exposure on the association of IDs in the memory abilities and the procedural knowledge. The zero-order correlations between memory ability measures and the TGJT grammatical sentences composite scores for the explicitly taught and not explicitly taught structures are in Table 4.30.

Table 4.30. Spearman’s correlations between procedural knowledge measure in two exposure conditions and memory ability measures

Measure		TGJT Grammatical	
		Explicitly Taught Structures (N = 9)	Not Explicitly Taught Structures (N = 4)
Declarative memory	CVMT	.23* (.022)	.09 (.357)
	Llama-B	.19 (.061)	.13 (.205)
	3-Term	.06 (.523)	.19 (.052)
	DAT-V	.45** (.000)	.18 (.070)
Procedural memory	Llama-D	.00 (.989)	.00 (.964)
	SRT	.15 (.143)	-.12(.224)

**. Correlation is significant at the 0.01 level (2-tailed);
*. Correlation is significant at the 0.05 level (2-tailed);
P-values are in parentheses

The TGJT grammatical sentences composite score for the explicitly taught structures was moderately related to the DAT-V and weakly correlated with the CVMT, suggesting some involvement of declarative memory. This may not be unexpected since the learning environment had involved explicit instruction and the procedural language knowledge test used, the TGJT as used in this study, may have allowed learners some access to explicit knowledge. No statistically significant correlations were found between the procedural memory measures and the two TGJT grammatical sentences composite scores.

In the structural model, the predictors comprised the same four-indicator declarative memory construct and the two procedural memory ability measures. The dependent variables however were the two TGJT grammatical sentences composite scores for the explicitly taught and not explicitly taught structures. The initial model with all error covariances fixed to 0 had not fit the data very well (see the initial model fit indices in Table 4.31 in Appendix C). The LM test results had shown that freeing up the error covariances for the two TGJT grammatical sentences composite scores and for the Llama-B and 3-Term could improve the model fit. Each of these error covariance parameters had shown a considerable difference in the magnitude of the parameter estimate, with the incremental univariate LM χ^2 values of 13.679 and 10.626 and the standardised parameter change values of .377 and .357, respectively. In the final run of the model, the two parameters were then re-specified to be freely estimated. In addition to having been the subset of the TGJT measure, these two language knowledge scores were observed to be moderately correlated (i.e., $r = .40$, $p < .001$, see above). However, the re-specification of the parameter for estimation has implications on the discriminant validity of the constructs for which each measure was an indicator.

The final model with the ML Elliptical estimation is in Figure 4.6. The model's Mardia's Coefficient (G2, P) was -3.0998 with the standardised estimate of -1.2375. The model's goodness-of-fit indices are in Table 4.32. The freely estimated parameters in this final model were 21 with 15 degrees of freedom. The average off-diagonal standardized residual was small (0.0403) and the RMSEA (0.026), the CFI and GFI indices (0.991 and 0.963, respectively) were all within the recommended values, indicating that the model explained the observed covariances among the variables quite well. The χ^2 statistic was non-significant, indicating that there was no significant difference between the observed and model-predicted variances and covariances. All the fit indices (except the Non-Normed Fit index whose value was slightly lower than the recommended .90) provided substantial support for the final model.

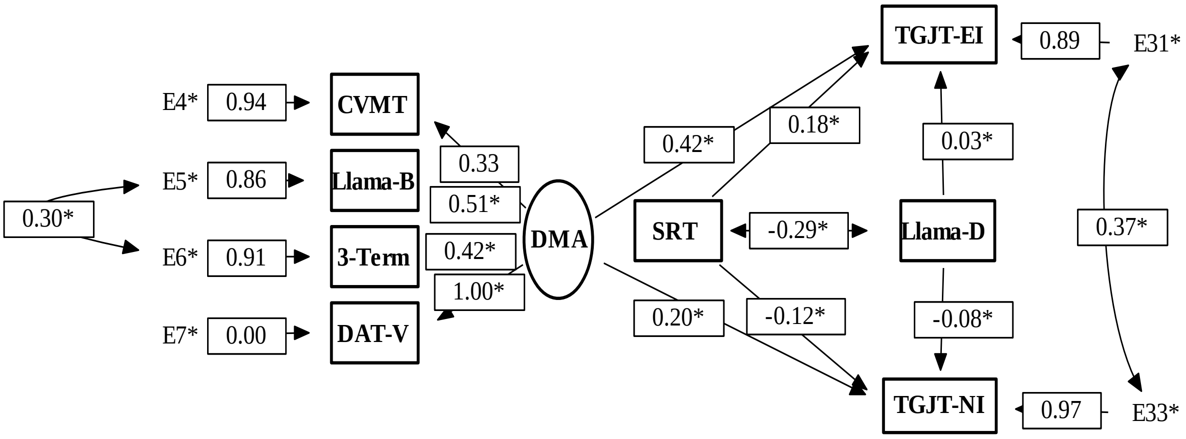


Figure 4.6. Revised SEM model for memory abilities, procedural knowledge and exposure
* Estimated parameters; DMA = Declarative memory ability; SRT = Serial Reaction Times; TGJT-EI = TGJT grammatical sentences composite score for explicitly taught structures; TGJT-NI = TGJT grammatical sentences composite score for implicitly taught structures

Table 4.32. Fit indices for the memory abilities, procedural knowledge and exposure – revised model

Null Model	
Independence Model CHI-SQUARE (df = 28)	150.099
Independence AIC	94.099
Independence CAIC	-7.400
Hypothesised Model	
Model AIC	-13.958
Model CAIC	-68.332
Chi-Square (df = 15)	16.042
Probability Value for the CHI-SQUARE Statistic	0.379
Fit Indices	
Bentler-Bonett Normed Fit Index	0.893
Bentler-Bonett Non-Normed Fit Index	0.984
Comparative Fit Index (CFI)	0.991
Bollen's (IFI) Fit Index	0.992
Mcdonald's (MFI) Fit Index	0.995
Joreskog-Sorbom's GFI Fit Index	0.963
Joreskog-Sorbom's AGFI Fit Index	0.912
Root Mean-Square Residual (RMR)	5.767

Standardized RMR	0.055
Root Mean-Square Error of Approximation (RMSEA)	0.026

The model's parameter estimates are in Table 4.33. The declarative memory had a positive but weak direct effect only on TGJT grammatical composite score for the explicitly taught structures, explaining 17% (i.e. $\gamma = .416$; $R^2 = .17$) of the variance. The SRT and Llama-D had no effects.

Table 4.33. Parameter estimates for memory abilities, procedural knowledge and exposure type – revised model

	Unstandardized			Standardised		
Path	Llama-D	SRT	DMA	Llama-D	SRT	DMA
TGJT Grammatical-Explicitly Taught (N=9)	.000 (.001)	1.037 (.533)	.229@ (.080)	.029	.177	.416
TGJT Grammatical-Not Explicitly Taught (N=4)	-.001 (.002)	-1.150 (.924)	.178 (.097)	-.076	-.123	.203

Note. DMA = Declarative memory ability factor; @ Statistic significant at the 5% level; Standard errors in parentheses

4.3.2 Memory, Linguistic Difficulty and Language Knowledge

Research Question 2

Does the level of difficulty of grammatical structures have a differential effect on the nature of the relationship between the learners' memory systems and their language knowledge?

To answer Research Question 2, zero-order correlations between IDs in tests of memory abilities and scores on simple and complex grammatical structures were examined. Then the association of the latent declarative memory construct and the two procedural memory abilities with the simple and complex grammatical structure scores was assessed. It was hypothesised that the level of difficulty of grammatical structures would have a differential effect on the nature of the relationship between the learners' memory abilities and their grammar knowledge (*Hypothesis 2a*). Declarative memory ability would be more predictive of simple grammatical structures (especially as measured by the UGJT ungrammatical sentences composite scores) while procedural memory ability would be more predictive of complex structures (especially as measured by TGJT grammatical composite scores; *Hypothesis 2b*). Further, it was expected that the procedural memory ability would predict the learners' performance on complex structures regardless of the exposure condition. However, this relationship would be stronger for the grammatical structures where learning proceeds implicitly than for structures acquired explicitly (*Hypothesis 2c*).

Unstandardized residuals for the UGJT ungrammatical sentences composite scores for simple grammatical structures (*N=4, mean* = .59, *SD* = .23; *D*(102) = .987, *p* = .424) and for complex grammatical structures *(N=9, mean* = .65, *SD* = .20; *D*(102) = .984, *p* = .271) were normal. The TGJT grammatical sentences composite scores for simple grammatical structures (*N=4, mean* = .79, *SD* = .16; *D*(102) = .964, *p* = .007) were non-normal while for the TGJT grammatical sentences composite scores for complex grammatical structures *(N=9, mean* = .65, *SD* = .14 *D*(102) = .983, *p* = .208) were normally distributed.

Table 4. 34. Spearman's correlations between UGJT ungrammatical sentences and TGJT grammatical sentences composite scores on simple and complex structures

	UGJT Ungrammatical		**TGJT Grammatical**	
Measure	Simple	Complex	Simple	Complex
UGJT Ungrammatical				
Simple (N=4)	__			
Complex (N=9)	.763** (.000)	__		
TGJT Grammatical				
Simple (N=4)	-.113 (.260)	.071 (.479)	__	
Complex (N=9)	.059 (.555)	.094 (.349)	.355** (.000)	__

**. Correlation is significant at the 0.01 level (2-tailed); P-values are in parentheses

Zero-order correlations between UGJT ungrammatical sentences and TGJT grammatical sentences composite scores on simple and complex grammatical structures are in Table 4.34. The UGJT ungrammatical sentences composite scores for simple and complex grammatical structures were strongly correlated. Second, the TGJT grammatical sentences composite score for simple structures was moderately related to TGJT grammatical sentences score on complex structures. These results suggest no differential effect of the linguistic difficulty on language knowledge. However, these correlations were very strong in the declarative language knowledge score but moderate in the procedural language knowledge scores. This may indicate that the linguistic difficulty in the present study had differentially affected the two language knowledge types, with the procedural knowledge being affected more than the declarative knowledge.

4.3.2.1 Effects of difficulty on memory abilities and declarative knowledge

The calculated zero-order correlations between memory ability measures and the UGJT ungrammatical sentences composite scores on simple and complex structures are reported in Table 4.35. The declarative memory measures were all significantly related to the UGJT ungrammatical sentences composite score on both simple and complex grammatical structures, suggesting the involvement of the declarative memory at both levels of difficulty. Both Llama-D and SRT were not related to either of the scores.

Table 4.35. Spearman’s correlations between memory abilities measures and UGJT ungrammatical sentences composite scores for simple and complex grammatical structures

Measure		UGJT Ungrammatical	
		Simple Structures (N=4)	Complex Structures (N=9)
Declarative memory	CVMT	.28** (.004)	.29** (.003)
	Llama-B	.35** (.000)	.39** (.000)
	3-Term	.24* (.016)	.35** (.000)
	DAT-V	.56** (.000)	.71** (.000)
Procedural memory	Llama-D	.07 (.461)	.09 (.359)
	SRT	-.04 (0.679)	-.15 (.143)

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
P-values are in parentheses

The structural model comprised the UGJT ungrammatical sentences composite scores for simple and on complex grammatical structures as dependent variables, and the declarative memory latent factor and the two procedural memory measures as predictors. The structural model did not fit the data very well (see Table 4.36 in Appendix C). The LM test indicated re-specifying for estimation the error covariances of the two UGJT ungrammatical sentences composite scores or of Llama-B and 3-Term had considerable effect in achieving good model fit, with the substantial incremental univariate LM χ^2 value of 24.177 and the standardised change value of 1.710 for the language knowledge scores' error covariance. The two UGJT scores were significantly correlated (see above). Freeing the Llama-B and 3-Term error covariance had resulted in a model misfit. Therefore, the final structural model (Figure 4.7) had the language knowledge scores' error covariance re-specified.

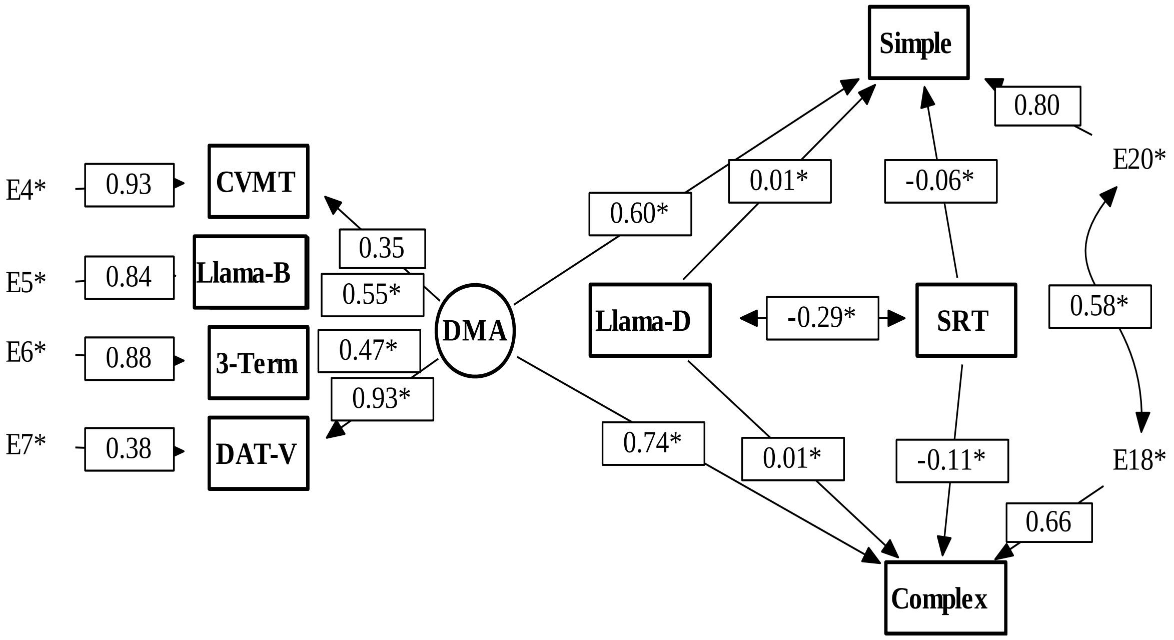


Figure 4.7. Revised SEM model for memory abilities, declarative knowledge and linguistic difficulty
* Estimated parameters; DMA = Declarative memory ability; SRT = Serial Reaction Times; Simple = UGJT ungrammatical sentences composite score on simple grammatical structures; Complex = UGJT ungrammatical sentences composite score on complex grammatical structures

With the ML Elliptical estimation, the Mardia's Coefficient (G2, P) was -3.5435 with the Normalized Estimate of -1.4146 while the freely estimated parameters were 20 and the degrees of freedom were 16. The average off-diagonal standardized residual was very small (0.0458) and the RMSEA (0.026), the CFI and GFI indices (0.996 and 0.961, respectively) were all within the recommended values, indicating that the model explained the observed covariances among the variables quite well. The χ^2 statistic was again not statistically significant, indicating that there was no significant difference between the observed and model-predicted variances and covariances. All the fit indices (Table 4.37) provided substantial support for the final model.

Table 4.37. Fit indices for the memory abilities, declarative knowledge and linguistic difficulty – revised model

Null Model	
Independence Model CHI-SQUARE (df = 28)	311.382
Independence AIC	255.382
Independence CAIC	153.883
Hypothesised Model	
Model AIC	-14.873
Model CAIC	-72.873
CHI-SQUARE (df = 16)	17.127
Probability Value for the CHI-SQUARE Statistic	0.377
Fit Indices	
Bentler-Bonett Normed Fit Index	0.945
Bentler-Bonett Non-Normed Fit Index	0.993
Comparative Fit Index (CFI)	0.996
Bollen's (IFI) Fit Index	0.996
Mcdonald's (MFI) Fit Index	0.994
Joreskog-Sorbom's GFI Fit Index	0.961
Joreskog-Sorbom's AGFI Fit Index	0.912
Root Mean-Square Residual (RMR)	9.464
Standardized RMR	0.058
Root Mean-Square Error of Approximation (RMSEA)	0.026

Table 4.38. Parameter estimates for memory abilities, declarative knowledge and difficulty

	Unstandardised			Standardised		
Path	Llama-D	SRT	DMA	Llama-D	SRT	DMA
UGJT Ungrammatical						
Simple Structures	.000	-.651	.552@	.011	-.062	.602
(N=4)	(.002)	(.874)	(.169)			
Complex Structures	.000	-1.022	.581@	.011	-.114	.738
(N=9)	(.001)	(.650)	(.168)			

@ Statistic significant at the 5% level; Standard errors in parentheses

The parameter estimates are in Table 4.38. The declarative memory had positive, strong to moderate direct effects on both complex and simple grammatical structures, respectively explaining 56% (i.e. $\gamma = .738$; $R^2 = .559$) and 37% (i.e. $\gamma = .602$; $R^2 = .367$) of the variance in the language knowledge measure. The SRT and Llama-D had no significant effect.

4.3.2.2 Effects of difficulty on memory abilities and procedural knowledge

Correlational and SEM analyses were conducted also to assess the role of linguistic difficulty on the association between the learners' IDs in the memory abilities and the procedural language knowledge. The zero-order correlations between memory ability measures and the TGJT grammatical sentences composite scores on simple and complex structures are in Table 4.39.

Table 4.39. Spearman's correlations between memory abilities measures and TGJT grammatical sentences composite scores on simple and complex grammatical structures

Measure		TGJT Grammatical	
		Simple Structures (N=4)	Complex Structures (N=9)
Declarative Memory	CVMT	.07 (.482)	.18 (.078)
	Llama-B	.05 (.611)	.19 (.060)
	3-Term	-.07 (.461)	.15 (.131)
	DAT-V	.30** (.002)	.33** (.001)
Procedural Memory	Llama-D	.02 (.807)	.00 (.966)
	SRT	.04 (.663)	.01 (.936)

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
P-values are in parentheses

Only the DAT-V test, a declarative memory measure, was moderately and significantly related to the TGJT grammatical sentences composites scores for both simple and complex grammatical structures. These results are unsurprising considering evidence suggesting that DAT-V might be directly related to implicit learning (Kaufman, et al., 2010). However, the DAT-V's relation observed with the UGJT ungrammatical sentences composite scores was stronger than that with the TGJT grammatical sentences composite scores. As Kaufman et al. (2010) point out, it is possibly the case that some residual variance in DAT-V is not attributable to declarative, conscious processes but instead accounts for a more specific procedural language learning ability. Both Llama-D and SRT were not related to either of the scores.

The structural model constructed to examine the association of memory, procedural language knowledge and the level of difficulty of structures comprised TGJT grammatical sentences composite scores for simple and for complex grammatical structures as dependent variables, and the declarative memory latent factor and the two procedural memory measures as predictors. The initial model did not fit the data well (see the initial model fit indices in Table 4.40 in Appendix C). The LM statistic indicated re-specifying for estimation the paths joining the simple and complex grammatical structures measures and the Llama-B and 3-Term had each considerable effect in achieving good model fit, with the incremental univariate LM χ^2 values of 10.881 and 11.037 and the standardised parameter change values of 0.335 and 0.361, respectively. The two TGJT scores were significantly correlated. Because freeing the error covariance of the two language scores still has implication on the discriminant validity for the measures, a decision was made to run the model with only the Llama-B and 3-Term error covariance re-specified for estimation. The revised model is presented in Figure 4.8.

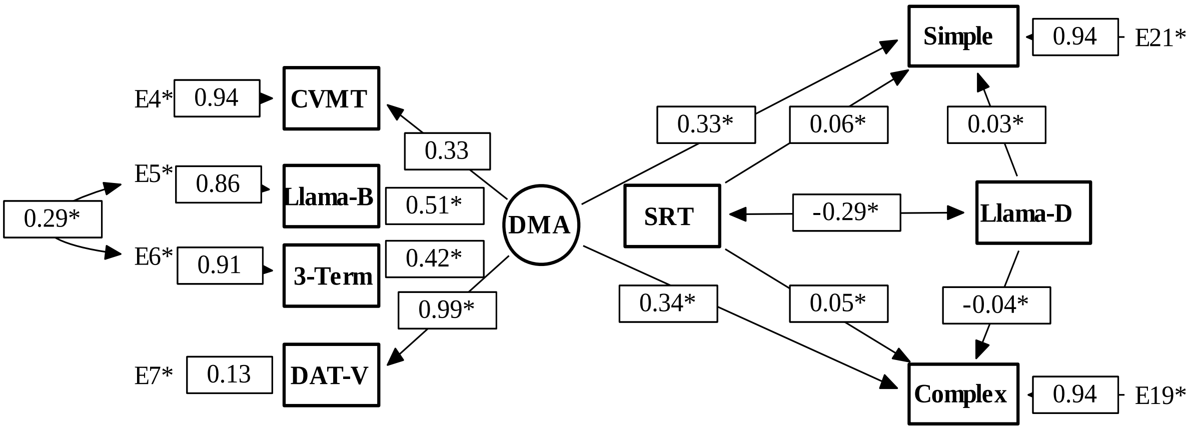


Figure 4.8. Revised SEM model for memory abilities, procedural knowledge and linguistic difficulty
* Estimated parameters; DMA = Declarative memory ability; SRT = Serial Reaction Times; Simple = TGJT ungrammatical sentences composite score on simple grammatical structures; Complex = TGJT ungrammatical sentences composite score on complex grammatical structures

ML Elliptical estimation was used, and overall, the results showed a good fit. The multivariate kurtosis' Mardia's Coefficient (G2, P) was -3.4340 with the Normalized

Estimate of -1.3709. The freely estimated parameters were 20 and the degrees of freedom were 16. The average off-diagonal standardized residual was very small (0.0492) and the RMSEA (0.069), the CFI and GFI indices (0.935 and 0.947, respectively) were all within the recommended values, indicating that the model explained the observed covariances among the variables quite well. The χ^2 statistic was non-significant, an indication that there was no significant difference between the observed and model-predicted variances and covariances. All the fit indices (Table 4. 41) provided substantial support for the final model. The Normed, Non-Normed and AGFI fit indices, however, were slightly lower than the recommended .90.

Table 4.41. Fit indices for the memory abilities, procedural knowledge and linguistic difficulty – revised model

Null Model	
Independence Model CHI-SQUARE (df = 28)	147.959
Independence AIC	91.959
Independence CAIC	-9.540
Hypothetical Model	
Model AIC	-8.231
Model CAIC	-66.231
CHI-SQUARE (df = 16)	23.769
Probability Value for The CHI-SQUARE Statistic	0.095
Fit Indices	
Bentler-Bonett Normed Fit Index	0.839
Bentler-Bonett Non-Normed Fit Index	0.887
Comparative Fit Index (CFI)	0.935
Bollen's (IFI) Fit Index	0.941
Mcdonald's (MFI) Fit Index	0.963
Joreskog-Sorbom's GFI Fit Index	0.947
Joreskog-Sorbom's AGFI Fit Index	0.880
Root Mean-Square Residual (RMR)	5.767
Standardized RMR	0.073
Root Mean-Square Error of Approximation (RMSEA)	0.069

The parameter estimates are in Table 4.42. Both SRT and Llama-D had no effect on both scores. The declarative memory had minimal direct effects on both complex and simple grammatical structures, respectively explaining 12% (i.e. $\gamma = .345$; $R^2 = .119$) and 11% (i.e. $\gamma = .327$; $R^2 = .107$) of the variance in the procedural language knowledge scores. The DAT-V had a positive, but moderate indirect effect on simple structures (i.e. $\gamma = .33$ [=.99 x .33]; $R^2 = .11$) and complex structures (i.e. $\gamma = .34$ [=.99 x .34]; $R^2 = .12$). Overall however, the positive, direct effects of declarative memory on the TGJT grammatical sentences composite scores for the simple and complex grammatical structures were substantially less strong than the effects observed when the language knowledge was declarative.

Table 4.42. Parameter estimates for memory abilities, procedural knowledge and difficulty

	Unstandardized			Standardised		
Path	Llama-D	SRT	DMA	Llama-D	SRT	DMA
TGJT Grammatical						
Simple Structures	.000	.420	.223@	.034	.058	.327
(N=4)	(.001)	(.702)	(.088)			
Complex Structures	-.000	.340	.211@	-.039	.052	.345
(N=9)	(.001)	(.624)	(.081)			

@ Statistic significant at the 5% level; Standard errors in parentheses

4.3.2.3 Effects of difficulty and exposure on memory and knowledge

To examine the hypothesis that procedural memory abilities would be more predictive of the learners' performance on complex structures where learning proceeds implicitly than for structures acquired explicitly, a pair of scores for each language knowledge measure was calculated. One score was calculated for all complex structures which are explicitly taught and another for all complex grammatical structures which are not explicitly taught. Unstandardised residuals for these scores were explored for normality. The UGJT ungrammatical sentences scores for 'Explicitly Taught' complex grammatical structures (*N=6, mean* = .57, *SD* = .23; *D*(102) = .988, *p* = .506) were normal whereas those for 'Not Explicitly Taught' structures (*N=3,mean* = .81, *SD* = .20; *D*(102) = .943, *p* = .001) were not normally distributed. Likewife, the TGJT grammatical sentences scores for complex grammatical structures 'Explicitly Taught' (*N=6, mean* = .75, *SD* = .15; *D*(102) = .984, *p* = .270) were normally distributed while those for the TGJT grammatical sentences scores for complex grammatical structures 'Not Explicitly Taught' (*N=3*, *mean* = .46, *SD* = .22; *D*(102) = .973, *p* = .034).

Table 4.43. Spearman's correlations between UGJT ungrammatical sentences and TGJT grammatical sentences composite scores on complex structures in two exposure conditions

	UGJT Ungrammatical		**TGJT Grammatical**	
Measure	Explicitly Taught (N=6)	Not Explicitly Taught (N=3)	Explicitly Taught (N=6)	Not Explicitly Taught (N=3)
UGJT Ungrammatical				
Explicitly Taught	__			
Not Explicitly Taught	.52** (.000)	__		
TGJT Grammatical				
Explicitly Taught	.18 (.072)	.10 (.305)	__	
Not Explicitly Taught	-.02 (.819)	-.25* (.010)	.39**(.000)	__

**. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed); *P*-values are in parentheses

The resulting zero-order correlations calculated between these scores are in Table 4.43. The two UGJT ungrammatical sentences composite scores on the complex structures, 'Explicitly Taught' and 'Not Explicitly Taught', were strongly and significantly correlated. Likewise, the two TGJT grammatical sentences composite scores for 'Explicitly Taught' and 'Not Explicitly Taught' complex structures were moderately correlated with each other. The UGJT ungrammatical sentences and TGJT grammatical sentences composite scores for 'Not Explicitly Taught' complex structures in the two exposure types were negatively and moderately correlated. Overall, these results suggest that the interaction of type of exposure and linguistic difficulty has no differential effect on the language knowledge types.

4.3.2.3.1 Effects of difficulty and exposure on memory and declarative knowledge

To assess the effects of the interaction of type of exposure and linguistic difficulty on the association of the IDs in the two-long memory systems with declarative language knowledge, a correlational and SEM analyses were conducted with memory measures and the UGJT ungrammatical sentences scores on complex grammatical structures explicitly taught and not explicitly taught. The calculated Spearman's correlation coefficients between these scores are in Table 4.44.

Table 4.44. Spearman's correlations between memory measures and UGJT ungrammatical sentences composite scores on complex structures in two exposure conditions

Measure		UGJT ungrammatical	
		Explicitly Taught (N=6)	Not Explicitly Taught (N=3)
Declarative memory	CVMT	$.33^{**}$ (.001)	.10 (342)
	Llama-B	$.37^{**}$ (.000)	$.23^{*}$ (.022)
	3-Term	$.36^{**}$ (.000)	.12 (.249)
	DAT-V	$.68^{**}$ (.000)	$.51^{**}$ (.000)
Procedural memory	Llama-D	.06 (.521)	.08 (.435)
	SRT	-.14 (.168)	-.08 (.434)

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
P-values are in parentheses

The declarative memory measures were moderately or strongly correlated with the learners' UGJT ungrammatical sentences composite score on complex grammatical structures that are explicitly taught. For the structures not explicitly taught, the Llama-B and DAT-V tests were, respectively, weakly and strongly related to the UGJT ungrammatical sentences composite score on the same structure type. Both Llama-D and SRT were unrelated to either of the scores. Based on the declarative language scores, these results suggest no differential effect of the interplay of linguistic difficulty and exposure type.

The structural model (Figure 4.9) consisted of the UGJT ungrammatical sentences composite scores on complex grammatical structures explicitly and not explicitly taught as dependent variables, and the declarative memory latent factor and the two procedural memory measures as predictors.

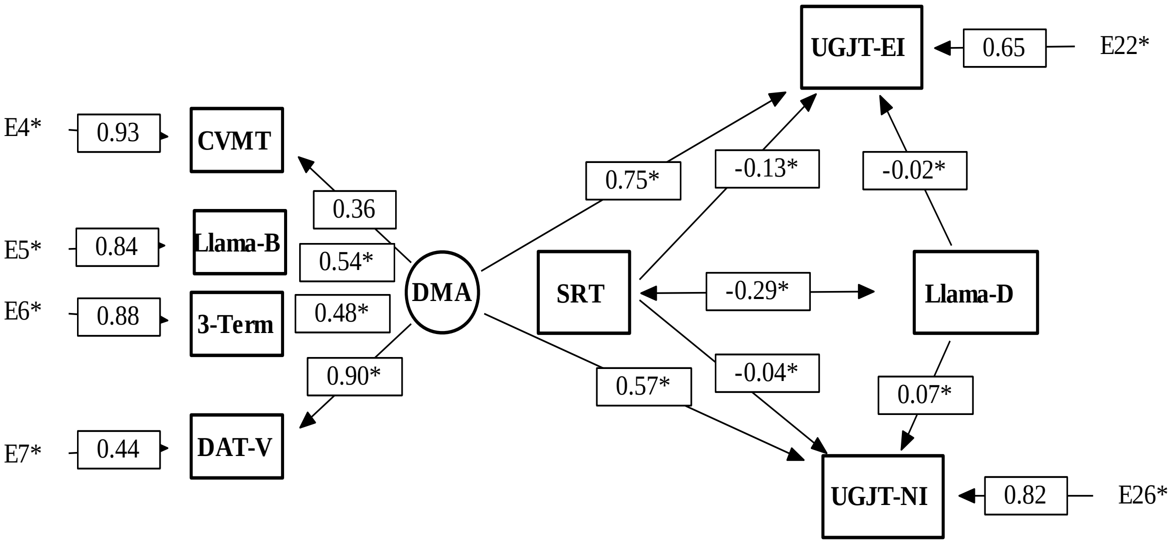


Figure 4.9. SEM model for exposure, memory and declarative language knowledge of complex grammatical structures
* Estimated parameters; DMA = Declarative memory ability; SRT = Serial Reaction Times; UGJT-EI = UGJT ungrammatical sentences composite scores on explicitly taught structures; UGJT-NI = UGJT ungrammatical sentences composite scores on not explicitly taught structures

ML Elliptical estimation was used, and overall, the results showed a very good fit on first run of the model. The multivariate kurtosis' Mardia's Coefficient (G2, P) was -1.7750 with the Normalized Estimate of -0.7086. The freely estimated parameters were 19 and the degrees of freedom were 17. The average off-diagonal standardized residual was very small (0.0574) and the RMSEA (0.066), the CFI and GFI indices (0.966 and 0.944, respectively) were all within the recommended values, indicating that the model explained the observed covariances among the variables quite well. The χ^2 statistic was again not statistically significant, indicating that there was no significant difference between the observed and model-predicted variances and covariances. All the fit indices (Table 4.45) provided substantial support for the model.

The model's parameter estimates are in Table 4.46. Only the declarative memory had strong to moderate direct effects on the scores, explaining 56% (i.e. $\gamma = .75$; $R^2 = .56$) and 32% (i.e. $\gamma = .57$; $R^2 = .32$) of the variance on explicitly and not explicitly taught structures, respectively. The effect was stronger for the explicitly taught structures than was for the implicitly taught structures.

Table 4.45: Fit indices for model for exposure, memory and declarative language knowledge of complex grammatical structures

Null Model	
Independence Model CHI-SQUARE (df = 28)	250.756
Independence AIC	194.756
Independence CAIC	93.256
Hypothesised Model	
Model AIC	-9.421
Model CAIC	-71.045
CHI-SQUARE (df = 17)	24.579
Probability Value for the CHI-SQUARE Statistic	0.105
Fit Indices	
Bentler-Bonett Normed Fit Index	0.902
Bentler-Bonett Non-Normed Fit Index	0.944
Comparative Fit Index (CFI)	0.966
Bollen's (IFI) Fit Index	0.968
Mcdonald's (MFI) Fit Index	0.964
Joreskog-Sorbom's GFI Fit Index	0.944
Joreskog-Sorbom's AGFI Fit Index	0.881
Root Mean-Square Residual (RMR)	9.445
Standardized RMR	0.066
Root Mean-Square Error of Approximation (RMSEA)	0.066

Table 4.46. Parameter estimates for model for exposure, memory and declarative language knowledge of complex grammatical structures

	Unstandardized			Standardised		
Path	Llama-D	SRT	DMA	Llama-D	SRT	DMA
UGJT						
Ungrammatical						
Explicitly Taught	-.000	-.1.352	.688@	-.019	-.129	.750
(N=6)	(.001)	(.773)	(.202)			
Not Explicitly Taught	.000	-.366	.457@	.072	-.040	.570
(N=3)	(.001)	(.799)	(.145)			

@ Statistic significant at the 5% level; Standard errors in parentheses

4.3.2.3.2 Effects of difficulty and exposure on memory and procedural knowledge

Correlational and SEM analyses were also conducted with memory measures and TGJT grammatical sentences scores on complex grammatical structures (not) explicitly taught to assess the effects of the interaction of type of exposure and linguistic difficulty on the association of memory abilities and procedural language knowledge. Spearman's correlations between these scores are in Table 4. 47.

Table 4.47. Spearman's correlations between memory measures and TGJT grammatical sentences scores on complex structures for two exposure types

Measure		TGJT Grammatical	
		Explicitly Taught (N=6)	Not Explicitly Taught (N=3)
Declarative memory	CVMT	.24* (.016)	.04(.696)
	Llama-B	.19 (.051)	.10 (.339)
	3-Term	.13 (.208)	.20* (.045)
	DAT-V	.43** (.000)	.09 (.386)
Procedural memory	Llama-D	.01 (.927)	-.03 (.749)
	SRT	.13 (.197)	-.13 (.187)

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

The CVMT and DAT-V were significantly correlated with the learners' TGJT grammatical sentences composite score on complex grammatical structures which were explicitly taught. The Llama-B score was weakly related to the TGJT grammatical sentences composite score for the explicitly taught grammatical structures. This relationship approached significance level. For complex grammatical structures not explicit taught, only the 3-Term score was weakly and significantly correlated with the TGJT grammatical sentences composite score. Overall, these results suggest that the interplay of linguistic difficulty and exposure had no differential effect on the association of the IDs in declarative memory with the procedural language knowledge.

The structural model consisted of TGJT grammatical sentences composite scores on complex grammatical structures which were explicitly and not explicitly taught as dependent variables. The declarative memory latent factor and the two procedural memory measures were predictors. The model did not fit the data very well (see initial model fit indices in Table 4.48 in Appendix C). The LM statistic indicated re-specification of the error covariances of either the two TGJT scores or the Llama-B and the 3-Term scores had each considerable effect in achieving good model fit, with the incremental univariate LM χ^2 values of 13.727 and 8.909 and the standardised parameter change values of 0.377 and 0.355, respectively. Freeing the error covariance of the Llama-B and 3 -Term resulted in a model that did not fit the data very well. The re-specification of the two TGJT scores error covariance obtained a better model fit, overall. However, this has implications on the discriminant validity of the two test scores in the model. Using the ML Elliptical estimation, the revised model with parameter estimates is in Figure 4.10.

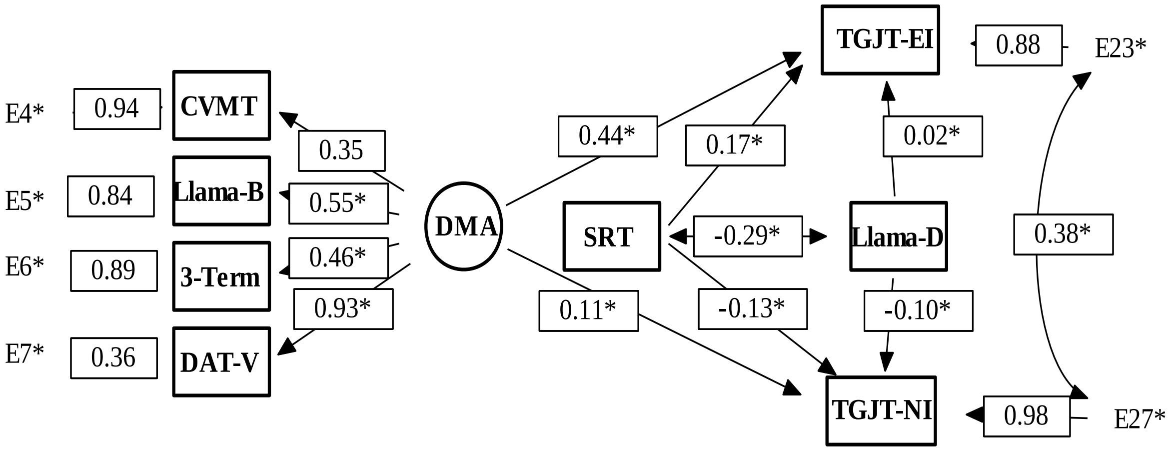


Figure 4.10. SEM model for exposure, memory and procedural language knowledge of complex grammatical structures
* Estimated parameters; DMA = Declarative memory ability; SRT = Serial Reaction Times; TGJT-EI = TGJT grammatical sentences composite scores on explicitly taught complex grammatical structures; TGJT-EI = TGJT grammatical sentences composite scores on complex grammatical structures not explicitly taught

The multivariate kurtosis' Mardia's Coefficient (G2, P) was -2.7104 with the Normalized Estimate of -1.0820. The freely estimated parameters were 20 and the degrees of freedom were 16. The average off-diagonal standardized residual was very small (0.0488) and the RMSEA (0.069), the CFI and GFI indices (0.934 and 0.947, respectively) were all within the recommended values, indicating that the model explained the observed covariances among the variables quite well. The χ^2 statistic was again not statistically significant indicating that there was no significant difference between the observed and model-predicted variances and covariances. All the fit indices (Table 4.49) provided substantial support for the final model. The Normed, Non-Normed and AGFI fit indices were slightly lower than the recommended .90.

Table 4.49: Fit indices for model for exposure, memory and procedural language knowledge of complex grammatical structures

Null Model	
Independence Model CHI-SQUARE (df = 28)	144.176
Independence AIC	88.176
Independence CAIC	-13.323
Hypothesised Model	
Model AIC	-8.389
Model CAIC	-66.389
CHI-SQUARE (df =16)	23.611
Probability Value for the CHI-SQUARE Statistic	0.100
Fit Indices	
Bentler-Bonett Normed Fit Index	0.836
Bentler-Bonett Non-Normed Fit Index	0.885
Comparative Fit Index (CFI)	0.934
Bollen's (IFI) Fit Index	0.941
Mcdonald's (MFI) Fit Index	0.963
Joreskog-Sorbom's GFI Fit Index	0.947
Joreskog-Sorbom's AGFI Fit Index	0.880
Root Mean-Square Residual (RMR)	9.704
Standardized RMR	0.062
Root Mean-Square Error of Approximation (RMSEA)	0.069

The structural parameter estimates are in Table 4.50. Both SRT and Llama-D had no effect on the TGJT scores in both conditions. The declarative memory had a weak direct effect on only complex structures explicitly taught, explaining 19% (i.e. $\gamma = .441$; $R^2 = .19$) of the variance in the TGJT grammatical sentences composite score. The DAT-V had the highest factor loading at .93 of the four declarative memory factor indicators. The measure had a weak indirect effect on the learners' TGJT grammatical sentences composite score for complex grammatical structures taught explicitly, explaining 17% (i.e. $\gamma = .41$ [=.93 x .44]; $R^2 = .17$) of the variance in the score. Overall, these results suggest linguistic difficulty mediated acquisition only when learning proceeded explicitly.

Table 4.50. Structural parameter estimates for model for exposure, memory and procedural language knowledge of complex grammatical structures

	Unstandardized			Standardised		
Path	Llama-D	SRT	DMA	Llama-D	SRT	DMA
TGJT Grammatical						
Explicitly Taught	.000	1.143	.266@	.020	.170	.441
(N=6)	(.001)	(.626)	(.092)			
Not Explicitly Taught	-.002	-1.284	.094	-.099	-.128	.106
(N=3)	(.002)	(1.010)	(.094)			

@ Statistic significant at the 5% level; Standard errors in parentheses

4.3.3 Exposure Length, Memory and Language Knowledge

Research Question 3

Do L2 instructed learners from different age groups and educational levels show different patterns in (a) their language knowledge, and (b) their declarative and procedural memory systems.

Research Question 3 was largely exploratory. However, one specific prediction was made that learners would differ in their language knowledge. It was hypothesised that as a function of the length of exposure, the university and secondary school learners would perform better on both declarative and procedural language knowledge measures than the primary school learners would (*Hypothesis 3a*).

4.3.3.1 Between-group differences in declarative language knowledge

To investigate learners' group differences in declarative language knowledge, a one-way between-group ANOVA was conducted on the UGJT ungrammatical sentences composite scores. To assess whether linguistic difficulty and exposure condition mediated the group differences, a series of one-way ANOVAs were performed on the UGJT ungrammatical sentences composite scores for structures explicitly taught and those not explicitly taught and for complex and simple structures. The descriptive statistics and univariate normality tests for the UGJT ungrammatical sentences composite scores for each learner group are reported in section 4.1.2.3. Table 4.51 presents descriptive statistics for the four UGJT ungrammatical sentences composite scores based on linguistic difficulty and exposure condition. For all the scores, the university students outperformed the other groups. The primary school had the lowest mean scores. The examination of the computed standardised skewness and kurtosis values of these four UGJT scores showed they were all normally distributed across the three groups.

Table 4.51. Descriptive statistics for four UGJT ungrammatical sentences composite scores based on linguistic difficulty and exposure condition

Measure	All		Primary		Secondary		University		Reliability
Group	(N = 102)		(N = 41)		(N = 29)		(N = 32)		estimate
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	(N)
UGJT Ungrammatical									
Explicitly Taught	.60	.22	.43	.17	.66	.20	.75	.14	.85(9)
Not Explicitly Taught	.71	.18	.63	.22	.73	.15	.79	.12	.58(4)
Complex Structures	.65	.20	.50	.17	.72	.16	.77	.12	.81(9)
Simple Structures	.59	.23	.47	.21	.58	.21	.74	.16	.64(4)

M = Mean; *SD* = Standard Deviation; N = number of items

With Welch's procedure, the results of one-way ANOVA omnibus tests on the five UGJT scores are in Table 4.52. There were statistically significant differences between the three learner groups across all the five scores. The omega squared (ω^2) values (Table 4.52) show that the group differences were all very strong except for structures that are not explicitly taught.

Table 4.52. ANOVA omnibus *F*-Test with Welch procedure on sets of UGJT ungrammatical sentences composite scores

UGJT Ungrammatical	*F*-Statistic[a]	df1	df2	*p*	w^2
Overall Score	32.431	2	62.277	.000	.36
Explicitly Taught	39.228	2	61.471	.000	.40
Not Explicitly Taught	7.189	2	64.021	.002	.11
Complex Structures	31.787	2	63.241	.000	.38
Simple Structures	20.696	2	62.724	.000	.25

a. Asymptotically F distributed.

The planned orthogonal comparisons (Table 4.53), reported with unequal variance assumed, indicated significant differences in all the comparisons except those between the secondary school group and university students on the scores for structures not explicitly taught and for the complex structures. These results suggest that the learners' educational level after primary school had no differential effect when a structure was complex or implicitly acquired.

Table 4.53. Planned contrasts on sets of UGJT ungrammatical sentences composite scores (*p*-values in bold non-significant)

UGJT Ungrammatical	Comparison		*t*	*df*	*p*	*r*
Overall Score	Primary	Secondary	4.541	61.449	.000	.50
	Primary	University	8.087	69.689	.000	.70
	Secondary	University	2.332	49.277	.024	.32
Explicitly Taught	Primary	Secondary	4.929	54.495	.000	.56
	Primary	University	8.815	70.796	.000	.72
	Secondary	University	2.172	49.805	.035	.29
Not Explicitly Taught	Primary	Secondary	2.071	67.940	.042	.24
	Primary	University	3.771	64.441	.000	.43
	Secondary	University	1.734	53.747	**.089**	.23
Complex Structures	Primary	Secondary	5.493	63.996	.000	.57
	Primary	University	7.876	69.981	.000	.69
	Secondary	University	1.451	52.157	**.153**	.20
Simple Structures	Primary	Secondary	2.199	60.393	.032	.27
	Primary	University	6.381	70.939	.000	.60
	Secondary	University	3.394	51.849	.001	.43

4.3.3.2 Between-group differences in procedural language knowledge

To investigate learners' group differences in procedural language knowledge, a one-way between-group ANOVA was conducted on the TGJT grammatical sentences composite scores. To assess whether linguistic difficulty and exposure condition mediated the group differences in the TGJT grammatical sentences composite scores, a series of one-way ANOVAs and the non-parametric Kruskal-Wallis H tests were carried out on the scores for the explicitly taught, not explicitly taught and for complex and simple structures. The descriptive statistics and univariate normality tests for the

TGJT grammatical sentences composite scores for each learner group are reported in section 4.1.2.2. In Table 4.54 are descriptive statistics for the four TGJT grammatical sentences composite scores based on linguistic difficulty and exposure condition. Across the four sets of scores, the primary school had the lowest mean scores. In the university students scores, the skewness and kurtosis absolute z-scores for the TGJT grammatical sentences composite scores for the explicitly taught and for simple grammatical structures were above 2.58, an indication that these two sets of scores were non-normal.

Table 4.54. Descriptive statistics for four TGJT grammatical sentences composite scores based on linguistic difficulty and exposure condition

Measure Group	Primary (N = 41)		Secondary (N = 29)		University (N = 32)		All (N = 102)		Reliability Estimate
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	(N)
TGJT Grammatical									
Explicitly Taught	.72	.13	.77	.11	.83	.11	.77	.13	.68(9)
Not Explicitly Taught	.47	.24	.55	.17	.54	.16	.51	.20	.57(4)
Complex Structures	.61	.15	.65	.13	.71	.12	.65	.14	.63(9)
Simple Structures	.73	.18	.84	.15	.81	.12	.79	.16	.48(4)

M = Mean; *SD* = Standard Deviation; N = number of items

With Welch's procedure, the results of one-way ANOVA omnibus tests indicated no significant differences between the three groups in the TGJT grammatical sentences composite scores for the structures not explicitly taught (*F* (2, 65.533), 1.562, $p = .217$, $w^2 = .02$) but there were statistically significant differences in the TGJT grammatical sentences composite scores (i.e., 'overall score'; *F* (2, 64.899), 5.667, $p = .005$, $w^2 = .09$). With equal variances assumed, the planned orthogonal comparisons in the TGJT grammatical sentences composite scores (overall score) indicated significant differences between the primary school group and the secondary school group ($t(99) = 2.046$, $p = .043$, $r = .201$) and between the primary school and university student groups ($t(99) = 3.384$, $p = .001$, $r = .322$), but no differences between the secondary school group and the university students ($t(99) = 1.177$, $p = .242$, $r = .117$).

In the TGJT grammatical sentences composite scores for the explicitly taught structures, the Kruskal-Wallis H statistic indicated significant group differences (χ^2 =

15.648, $df = 2$, $p = .001$, $r = .351$). The post-hoc Mann-Whitney U, with the Bonferroni adjusted p-values, indicated a significant difference between the primary school and university students scores ($U = 309.000$, $z = -3.870$, $p = .001$, $r = .453$). But there were no significant differences between the primary school group and secondary school group ($U = 454.000$, $z = -1.678$, $p = .280$, $r = .201$) and between the secondary school group and the university students ($U = 310.500$, $z = -2.225$, $p = .078$, $r = .285$).

When data were analysed by linguistic difficulty, the one-way ANOVA omnibus test showed statistically significant differences in the TGJT grammatical sentences composite scores on complex structures (F (2, 64.596), 5.530, $p = .006$, $w^2 = .08$). The planned orthogonal comparisons with equal variances assumed indicated significant differences between the primary school group and the university students scores ($t(99) = 3.258$, $p = .002$, $r = .311$), but there were no significant differences between the primary school and secondary school groups ($t(99) = 1.191$, $p = .236$, $r = .119$) and between secondary school and university student groups ($t(99) = 1.870$, $p = .064$, $r = .183$).

Further, the Kruskal-Wallis H Test showed that there were significant differences in the TGJT grammatical sentences composite scores on simple structures across the three groups ($\chi^2 = 7.1879$, $df = 2$, $p = .027$, $r = .218$). The post-hoc Mann-Whitney U, with the Bonferroni adjusted p-values, indicated a significant difference between the primary and secondary school groups ($U = 391.000$, $z = -2.447$, $p = .043$, $r = .292$). But there were no significant differences between the primary school group and the university students ($U = 489.000$, $z = -1.881$, $p = .180$, $r = .220$) and between the secondary school group and the university students scores ($U = 399.500$, $z = -0.946$, $p = .1.000$, $r = .121$). These results were surprising since one could expect to find no differences between primary and secondary school learners. A decision was made to examine these results further by analysing the data by linguistic difficulty and exposure type.

Table 4.55 presents descriptive statistics for the four TGJT grammatical sentences composite scores based on interaction of linguistic difficulty and exposure condition. The range of group mean scores was smaller in explicitly taught simple grammatical structures than in not-explicitly taught simple structures. With Welch’s procedure, one-way ANOVA omnibus test indicated no significant differences between the three groups in the TGJT grammatical sentences scores for explicitly taught simple structures

(*F* (2, 63.110), 1.858, *p* = .165, w^2 = .02) but there were statistically significant differences in the TGJT grammatical sentences scores for not-explicitly taught simple structures (*F* (2, 64.471), 4.357, *p* = .017, w^2 = *.07*). These results suggest that the observed non-significantly different simple structures mean scores between primary school learners and university students resulted from the difference in mean scores for not-explicitly taught simple structures between the two groups.

Table 4.55. Descriptive statistics for four TGJT grammatical sentences composite scores based on interaction of linguistic difficulty and exposure condition.

Group	Primary (N = 41)		Secondary (N = 29)		University (N = 32)		All (N = 102)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Simple								
Explicitly Taught (N=3)	.79	.15	.85	.15	.84	.14	.82	.15
Not Explicitly Taught (N=1)	.57	.37	.79	.26	.73	.22	.68	.31
Complex								
Explicitly Taught (N=6)	.69	.16	.73	.12	.83	.12	.75	.15
Not Explicitly Taught (N=3)	.43	.25	.47	.21	.48	.18	.46	.22

M = Mean; *SD* = Standard Deviation; N = Number of items

4.3.3.3 Between-Group Differences in Memory Abilities

To investigate whether the learners differed in their abilities in the two-long memory systems, a between-group ANOVA was conducted with declarative memory ability (DMA) factor, SRT and Llama-D scores. The DMA factor score estimates were computed for each participant using the regression method. The regression method is commonly used to create factor score estimates, and it involves summing the regression coefficients of items influenced by a common factor (Phakiti, 2018b). Descriptive statistics for the DMA factor scores were as follows: Primary School, N = 41, Mean = -0.677, and *SD* = 0.5612); Secondary School, N = 29, Mean = 0.072, *SD* = 0.6487; and the University students, N = 32, Mean = 0.802, and *SD* = 0.6264. The examination of skewness and kurtosis z-scores indicated the DMA group scores were all approximately normally distributed. Similar results were obtained using unstandardized residuals (*D*(102) = .975, *p* = .063). The Llama-D and SRT descriptive statistics and normality tests are presented in section 4.1.1.2.

The one-way ANOVA omnibus tests using the Welch procedure showed that there were no significant group differences in the Llama-D scores F (2, 60.742), 0.507, $p = .605$, $w^2 = -.009$) and SRT scores F (2, 65.026), 2.998, $p = .057$, $w^2 = .063$) but there were statistically significant differences in the DMA scores F (2, 60.953), 55.112, $p = .001$, $w^2 = .508$). The planned orthogonal comparisons, with equal variances assumed, showed that there were significant differences between all the three comparisons as follows: between the primary school group and the secondary school group, t(99) = 5.082, $p = .001$, $r = .455$; between the primary school group and the university student group, t(99) = 10.323, $p = .001$, $r = .720$; and between the secondary school group and the university students, t(99) = 4.688, $p = .001$, $r = .426$. These results provide support to the hypothesis that variability is the characteristic of declarative, explicit, and intentional processes.

This chapter has reported the results of the present study with reference to each of the three research questions. The following chapter summarises and discusses these findings within the available theoretical and empirical L2 acquisition research.

CHAPTER 5. DISCUSSION

This chapter presents a discussion on the interpretation of the findings of the present study regarding the role of the declarative and procedural memory systems and linguistic difficulty in the acquisition of L2 English in an instructed context. In the following sections, each of the three research questions and related hypotheses and findings are revisited and discussed in terms of current L2 acquisition research on the role of long-term memory systems, linguistic difficulty and learning conditions as internal and external factors in L2 acquisition.

5.1 Memory, exposure and language knowledge type

The first research question asked whether instructed L2 English learners' declarative and procedural memory systems would predict grammatical knowledge, as measured by the tests of implicit and explicit language knowledge. Further, by differentiating the targeted grammatical structures into those that are explicitly taught and those that are not explicitly taught (see section 3.4), this research question examined whether the predictive effects of the declarative and procedural memory systems would vary according to exposure condition. It was hypothesised (1a) that the declarative memory would strongly predict the learners' declarative language knowledge while the procedural memory would strongly predict procedural knowledge (*Hypothesis 1a*); and that declarative memory would be strongly related to learners' declarative knowledge of the grammatical structures which are explicitly taught while learners' procedural knowledge of grammatical structures learned in a more implicit manner would be strongly predicted by the procedural memory system (*Hypothesis 1b*).

To address this question, initial correlational analyses and then SEM analyses were conducted on the three independent variables (i.e., the declarative memory factor and the two procedural memory measures, the SRT and Llama-D tests) first with the overall UGJT ungrammatical and TGJT grammatical sentences composite scores and then with the same language knowledge test scores but now based on whether the grammatical structures were explicitly taught or not.

5.1.1 The association between memory and language knowledge.

To examine *Hypothesis 1a* regarding the predictive effect of the memory systems on the language knowledge regardless of the exposure condition and the level of linguistic difficulty (i.e., overall score), the dependent variables were the UGJT ungrammatical sentences composite/overall score (as the declarative knowledge measure) and the TGJT grammatical sentences composite score (as the procedural knowledge measure). The language knowledge composite scores were calculated from all the 13 grammatical structures scores regardless of the level of linguistic difficulty, but with respect to sentence categories (i.e., grammatical vs. ungrammatical). *Hypothesis 1a* was partially supported. Only the declarative memory system had a strong positive direct effect ($R^2 = .46$) on declarative language knowledge and a weak positive effect ($R^2 = .15$) on procedural language knowledge (Figure 4.4 and Table 4.24). If the non-interface position between the memory systems and the language knowledge types is not assumed, that is, if Ullman's DP model is assumed, then these results would suggest that the declarative memory system was responsible for learners' declarative language knowledge and, to some extent, for procedural language knowledge.

However, it appears assuming the dissociation between the declarative and procedural memory systems may best explain the observed predictive nature of the declarative memory system in the present study. First, while the strong predictive effects of declarative memory on declarative knowledge are expected, it is possible that the use of the TGJT as the procedural language knowledge measure might have allowed the learners some access to explicit language knowledge as they completed the task (see R. Ellis, 2005a; 2006; 2009c; R. Ellis & Loewen, 2007). This access to declarative language knowledge may in turn explain why the declarative memory system weakly predicted the learners' performance in the TGJT. Another possible explanation for these findings lies in the effects of the DAT-V test. Although, the DAT-V test did not correlate with any of the procedural memory measures, it was the only declarative memory measure that was moderately and significantly correlated with the TGJT grammatical sentences composite score (Table 4.20). The SEM results (Figure 4.4) indicated that the declarative memory's predictive effect on the TGJT score was largely accounted for by the indirect effect of the DAT-V test. The Llama-B test had the second highest declarative memory factor loading but its indirect effect on procedural language knowledge was very low ($R^2 = .03$). These results may be unsurprising considering that the DAT-V test was found to be predictive of implicit learning in Kaufman et al. (2010). However, this interpretation does not appear reasonable in the absence of significant correlations between either the DAT-V test or the TGJT grammatical sentences composite score and any of the procedural memory measures.

Explaining the declarative memory systems’ predictive effects on procedural language knowledge, in terms of the TGJT, as allowing access to explicit knowledge, or the DAT-V test indexing implicit learning, has one vital implication: that is, the declarative and procedural memory systems appeared to be dissociable, with the declarative memory system predicting only declarative language knowledge but not the procedural language knowledge. That no role was found for the procedural memory system may be explained as the consequence of how the grammatical structures were acquired. Generally, these grammatical structures are explicitly taught to students which may result in the development of only analysed, conscious, declarative language knowledge if we assume the non-interface position of the two types of learning processes. Thus, these results appear to support Paradis’s DP model (2004; 2009) and Reber’s implicit learning theory (A. S. Reber 1989; 1993; A. S. Reber et al., 1991) which do not just assume the non-interface position between the declarative and procedural language knowledge types (discussed in detail in section 2.1) but also claim the one-to-one relationship of the distinction between declarative/procedural memory and explicit/implicit knowledge. That is, what is acquired explicitly remains as verbalisable, conscious knowledge.

These findings where only the declarative memory system (but not the procedural memory system) was predictive of explicit language knowledge are not consistent with those of Kaufman et al. (2010) in which L2 French and L2 German scores were significantly correlated with implicit learning after controlling for psychometric intelligence, working memory, explicit associative learning, and processing speed. In Kaufman et al.’s study, explicit learning was significantly correlated with L1 English and L2 French scores but not with L2 German scores. However, the partial correlation between the language scores and explicit learning were not reported[17]. Further, though the GCSE language exams involve both on-and-off-line test measures (i.e., coursework

---

[17] It is important to note that the language scores were also correlated with general intelligence, working memory as well as processing speed.

and written, listening, speaking, and reading examinations), which can allow test-takers to rely on implicit and explicit knowledge types, Kaufman et al. did not report the exact context under which the two target L2s were learned (i.e., whether implicitly or explicitly). Further, there were no separate measures for declarative and procedural language knowledge.

However, the finding in the present study where only the declarative memory system was predictive of the declarative language knowledge does partially support the results reported in Pretz et al. (2010; 2014) as well as those in Kidd and Kirjavainen (2011) and Lum and Kidd (2012). In Pretz et al. (2010; 2014), procedural memory measures (SRT and AGL) were not correlated with ACT English scores (see Pretz et al., 2014 corrigendum article) obtained from undergraduates with a mean age of 19.29. The lack of correlation between the procedural memory measures and ACT English language scores was, however, expected because the ACT scores were obtained from multiple-choice tests that measure cognitive abilities such as critical thinking, reasoning, and problem solving in each of the four ACT subject areas. These tasks should allow test-takers to heavily rely on their explicit knowledge. There were no direct measures of declarative memory and language knowledge in this study, making the comparison with the current study not that informative.

Kidd and Kirjavainen (2011) and Lum and Kidd (2012), on the other hand, found no direct role for declarative and procedural memory on the grammar acquisition of L1 Finnish and L1 English with younger learners of approximately five years of age. However, the results in both studies indicated a direct and significant relationship between the declarative memory system and vocabulary acquisition. In L1 acquisition, the procedural memory and declarative systems are posited to respectively underly the learning of arbitrary language (i.e., the mental lexicon as well as irregular grammatical rules) and rule-based language (i.e., complex mental grammar). Therefore, the results in Kidd and Kirjavainen (2011) and Lum and Kidd (2012) seem surprising as the procedural memory system would be expected to strongly correlate with the learner's L1 rule-based language knowledge scores especially at age 5. However, the lack of the relationship between the procedural memory tests and grammar-rule knowledge in both Kidd and Kirjavainen (2011) and Lum and Kidd (2012) might have arisen from the fact that there were no separate measures for declarative and procedural language knowledge types.

5.1.2 Effects of exposure type on memory and language knowledge.

To examine *Hypothesis 1b* regarding the modulating effects of exposure condition on the association of the memory systems and knowledge regardless of the level of linguistic difficulty of the targeted grammatical structures, the UGJT ungrammatical and TGJT grammatical sentences composite scores for the “explicitly taught”, on one hand, and for the “not explicitly taught” structures, on the other hand, were used as dependent variables in the SEM analyses. *Hypothesis 1b* was partially supported. These results are discussed in detail below.

5.1.2.1 Effects of exposure type on memory and declarative knowledge

As hypothesised, declarative memory strongly predicted the learners’ declarative language knowledge of grammatical structures which are explicitly taught with a large effect size ($R^2 = .74$). However, the predictive effect of the declarative memory on the learners’ declarative knowledge of grammatical structures that are not explicitly taught was moderate ($R^2 =.37$; see Figure 4.5 and Table 4.29). Procedural memory as measured by the SRT and Llama-D tests was not predictive of the UGJT ungrammatical sentences composite scores for explicitly and not explicitly taught grammatical structures. The observed predictive medium effect of declarative memory on the implicitly acquired grammatical structures in the present study may perhaps be explained as the function of either the language knowledge measure used or the study design. The use of the UGJT might have encouraged the learners to reflect on rules and then use their declarative language knowledge when making the responses. Further, the distinction between “explicitly taught” and “not explicitly taught” was based on the evaluation of the curriculums and the researcher’s experience as a student and a teacher. Because the ex post facto study design was adopted, it is possible that, for some grammatical structures categorised as not explicitly taught, some explicit instruction was in fact provided to some learners at some points during learning.

5.1.2.2 Effects of exposure type on memory and procedural knowledge.

As measured by procedural language knowledge test (Table 4.33 and Figure 4.6), procedural memory had no predictive effects on the TGJT grammatical sentences composite scores for the “explicitly taught” and “not explicitly taught” grammatical structures. Declarative memory, however, had a positive but weak direct effect (R^2 = .17) only on TGJT grammatical sentences composite score for the “explicitly taught” grammatical structures. This finding might not be surprising; the TGJT might have allowed learners some minimal access to explicit knowledge only for those grammatical structures they were explicitly taught. In fact, the TGJT grammatical sentences composite score for the grammatical structures “explicitly taught” was positively but weakly correlated with the UGJT ungrammatical sentences composite score for structures acquired in a similar way (r = .20, p = .05; Table 4.25).

These results have an important implication. As measured by the procedural language knowledge test, it was expected to find a strong predictive effect for the procedural memory system on the learners’ procedural language knowledge of the grammatical structures considered to be learned in a more implicit manner. The lack of this predictive effect may suggest that the exposure condition was not differential. Alternatively, because the GJTs encourage reflection on the rules, the TGJT grammatical sentences scores might have not necessarily indexed procedural learning processes. Further, it is possible that the predictive power for both memory systems could be negatively affected by the lack of the construct/discriminant validity of the two TGJT grammatical sentences composite scores used as dependent variables in the SEM model. The problem of the lack of model fit in the initial SEM model resulted in a decision to include the error covariance of the two language knowledge scores for estimation. Freeing the error covariance for estimation might have diminished discriminant validity of the two language knowledge measures. The descriptive statistics of the two TGJT grammatical sentences composite scores had shown a wider mean difference with a significant *t*-test statistic and a very large effect size ($t(101)$ = 13.461, p = .0001, r = .80). But despite this large difference in the two means, the two TGJT scores were moderately correlated (r = .40, Table 4.25), which might suggest that the construct validity issue may in fact arise as the result of the two scores coming from the same measure/task and therefore being considerably correlated.

The findings of the effects of the exposure type on the relationship between language knowledge and memory in the present study do not appear to support those reported in Faretta-Stutanberg and Morgan-Short (2018) and Tagarelli et al. (2016). Faretta-Stutanberg and Morgan-Short (2018) found no predictive role for the declarative and procedural memory systems in the “at-home” group of learners. The “at-home” setting is a traditional classroom context, “the natural context that ‘explicit’ training conditions are generally designed to reflect” (Faretta-Stutanberg & Morgan-Short, 2018, p. 71). Because explicit learning processes are expected in the study “at-home” context, strong predictive effects for declarative memory on learning are expected. The procedural memory system (as well as WM) predicted behavioural gains in learners in the “study-abroad” or immersion setting condition, “the natural context that ‘implicit’ and ‘incidental’ training conditions are generally designed to reflect” (Faretta-Stutanberg & Morgan-Short, 2018, p. 71). However, Faretta-Stutanberg and Morgan-Short’s (2018) study results need to be interpreted with caution. First, in addition to the limited control over the L2 exposure and use in the two conditions, the sample sizes of the study were very small: 17 for the “at-home” group, and 13 for the “study-abroad” group. Second, there was only one language knowledge measure, the UGJT. Previous research (R. Ellis, 2005a; 2009c; R. Ellis & Loewen, 2007; Gutiérrez, 2013; Zhang, 2015; see section 2.1 for a discussion) suggest that though the UGJT is weighted in favour of declarative language knowledge, its grammatical and ungrammatical test items index different language knowledge types. Using Signal Detection Theory (SDT) to score the UGJT, Faretta-Stutanberg and Morgan-Short (2018) calculated *d'* scores where both ungrammatical and grammatical sentences raw scores are used in the computation (see section 3.7.1.3).

Similar results to Faretta-Stutanberg and Morgan-Short (2018) are reported in Tagarelli et al. (2016). Tagarelli et al. in their investigation of the relationship between IDs in cognitive abilities, exposure conditions, and linguistic complexity in L2 learners using a semi-artificial language, found that correlations between cognitive abilities (i.e., procedural memory and WM) and learning outcomes accounted for little of the variance in overall performance. WM and procedural memory, hypothesised to be more important in explicit and implicit conditions respectively, did not correlate with a language knowledge measure in the instructed group. The procedural memory system, however, was found to be strongly but negatively correlated with the UGJT (i.e., $r = -.586$, $p = .003$; Tagarelli et al., 2016, p. 308) in the incidental group. Like Faretta-Stutanberg & Morgan-Short (2018), Tagarelli et al. (2016) only used the UGJT as the L2 language knowledge measure. Furthermore, there was no direct test measure for the declarative memory system in the study.

The findings in the present study do, however, appear to partially support results reported in such experimental studies as Hamrick (2015), Morgan-Short et al. (2014), Carpenter (2008) and Yalçin and Spada (2016). In these studies, declarative memory appears to play a role if L2 learners are explicitly trained and/or if they remain at lower levels of proficiency. For instance, Carpenter (2008), examining the contributions to L2 acquisition of exposure type, linguistic difficulty and IDs in declarative, procedural and WM memory, found the declarative memory system predictive of accuracy on all grammatical structures (regardless of linguistic difficulty) at both low and advanced proficiency for the explicitly trained group. She concluded that declarative memory "remained important for L2 acquisition throughout learning in the explicit condition" (p. 289). For the implicitly trained group, Carpenter found that the declarative memory system was predictive only at low proficiency levels.

Hamrick (2015) and Morgan-Short et al. (2014) report similar results to those obtained in the implicitly trained group in the Carpenter (2008) study. Though the L2 learners in Hamrick (2015) and Morgan-Short et al. (2014) were trained only implicitly/incidentally, declarative memory measures positively correlated with, or predicted, syntactic development at early stages of acquisition while the procedural memory system positively correlated with, or predicted, language knowledge development at later stages of acquisition. These similar results in the two studies were obtained under slightly different conditions: a more explicit condition in Morgan-Short et al. (2014), as the test instruction at the beginning of training made participants aware that they had to learn; and an implicit condition in Hamrick (2015), as participants in the study were not sensitized to language learning.

Yalçın and Spada (2016) report results like those in Carpenter (2008), where declarative processes appeared to support the learning of explicitly taught grammatical structures. Yalçın and Spada (2016) found that the passive UGJT scores were predicted by Llama-F (grammatical inferencing) while the progressive Oral Production Test (OPT) scores were predicted by Llama-B (rote [associative] memory). These results were interpreted as lending support to claims (i) that different components of aptitude contribute to the learning of difficult (passive) and easy (progressive) L2 structures in different ways and (ii) that different components of aptitude may be involved at different stages of language acquisition as proposed in Skehan's (2002) aptitude profile model. However, Yalçın and Spada's (2016) results also show that while the other three Llama subtests

(Llama-B, Llama-E and Llama-F, subtests that index explicit learning processes) positively and moderately correlated with each other, Llama-D (a subtest that indexes implicit learning processes; see Chapter 2 for a discussion of these) did not correlate with any of the Llama subsets (see Table 1 in Yalçın & Spada, 2016, p. 251). Further, Llama-D did not correlate with any grammatical structure scores most likely because these structures were learned explicitly. Thus, the results in Yalçın and Spada (2016) appear to indicate that explicit instruction given to the learners modulated the way the declarative and procedural memory systems were engaged in the learning process.

In summary, the results regarding the modulating effects of the exposure type on the association of the memory systems and language knowledge partially support the differential nature of exposure conditions. As measured by both the declarative and procedural language knowledge tests, the procedural memory system did not predict the learners' knowledge of grammatical structures, either explicitly taught or not. The declarative memory system strongly predicted the declarative knowledge of the explicitly taught grammatical structures. These results are not surprising because, as pointed out above, the mode of learning these grammatical structures was predominantly explicit for these learners. These findings, therefore, support Paradis's DP model (e.g., 2004; 2009) and Reber's implicit learning theory (e.g., 1989; 1993) claims that explicit learning results in verbalisable, conscious knowledge (probably regardless of the level of proficiency, as found in Carpenter, 2008), without interfacing with procedural learning processes. However, the present study's design problem regarding the criterion to distinguish exposure conditions makes the interpretation of the results, especially those pertaining to "not explicitly taught" grammatical structures difficult. Furthermore, because GJTs encourage reflection on the rules, then both language knowledge measures were weighed in favour of explicit cognitive processes. Results from a well-controlled implicit condition or an immersion context with comparable learners and more robust measures of procedural language knowledge would shed more light on these findings.

5.2 Memory, difficulty, exposure and knowledge type

Research question 2 asked whether the level of difficulty of grammatical structures has a differential effect on the nature of the relationship between the learners' memory systems and their grammar knowledge. This question was addressed using a series of correlational and SEM analyses. Predictive measures remained the declarative memory factor and the two procedural memory tests, the probabilistic SRT and Llama-D test. The outcome measures were language knowledge scores for simple and/or complex grammatical structures.

A. S. Reber (1989; 1993) claims that while the declarative/procedural memory and declarative/procedural knowledge distinctions are in a one-to-one relationship, procedural learning is the process whereby a complex, rule-governed knowledge base is acquired. Thus, *Hypothesis 2a* predicted that the level of difficulty of grammatical structures would have a differential effect on the nature of the relationship between the learners' memory systems and their language knowledge, with (*Hypothesis 2b)* the declarative memory system strongly predicting simple grammatical structures (especially as measured by the UGJT ungrammatical sentences scores) and the procedural memory system strongly predicting complex structures (especially as measured by TGJT grammatical sentences scores). It was further hypothesised (*Hypothesis 2c)* that the procedural memory system would be predictive of the learners' performance on complex structures regardless of the exposure condition, with a stronger effect expected for those grammatical structures where learning proceeded implicitly than for structures acquired explicitly.

5.2.1 Memory, difficulty and language knowledge.

To examine *Hypotheses 2a* and *b,* the dependent variables were the UGJT ungrammatical sentences scores (as the declarative knowledge measures) and the TGJT grammatical sentences scores (as the procedural knowledge measures) calculated for simple and complex grammatical structures regardless of exposure condition. *Hypotheses 2a* and *b* were partially supported.

As measured by declarative language knowledge test (the UGJT), the correlational analysis showed that the declarative memory measures were all significantly related to the UGJT ungrammatical sentences scores for both simple and complex grammatical structures (Table 4.35), suggesting the involvement of the declarative memory at both levels of difficulty. These correlations, however, were all slightly higher for learners' performance on the complex grammatical structures.

Contrary to the predicted results, a SEM analysis showed the procedural memory system, as measured by the SRT and Llama-D tests, unpredictive of the learning of either types of grammatical structures. The declarative memory, however, had strong, positive direct effects on complex structures (R^2 = .56) and moderate direct effects on simple grammatical structures (R^2 = .37; Figure 4.7 and Table 4.38). Descriptive statistics indicated that the learners were much better at judging ungrammatical complex structures as ungrammatical (mean = .79, *SD* = .16) than they were at judging the incorrectness of the ungrammatical simple structures (mean = .65, *SD* = .14). At first glance, this appears unexpected (as per *Hypothesis 2b*). One explanation for the unexpected results where declarative memory strongly predicted the UGJT score for complex but not for simple grammatical structures would be the lack of construct validity of the two language knowledge scores since their error covariances parameter in the SEM model was estimated for purposes of model fit (Figure 4.7). However, the most probable explanation for why the declarative memory system predicted the declarative knowledge for complex structures better than for simple structures is that the learners learned both types of structures explicitly, but they had to use their declarative memory system more intensively to learn the complex ones. On the contrary, the procedural memory system failed to predict the knowledge of either types of grammatical structures because the learners did not use it to learn the structures.

As measured by the procedural language knowledge test, the hypothesis that the procedural memory system would strongly predict complex structures was not supported (Figure 4.8 and Table 4.42). Procedural memory, as measured by the SRT and Llama-D tests, had no effect on either simple or complex grammatical structures. It was instead the declarative memory system that had small, direct positive effects on both complex structures (R^2 = .12) and simple grammatical structures (R^2 = .11). However, these small effects of the declarative memory system on the TGJT grammatical sentences scores might largely be accounted for either by the TGJT allowing learners some access to explicit knowledge or by the indirect effects of the DAT-V test, a test that some research has found to index implicit learning abilities. The DAT-V test was the only declarative memory test that correlated moderately with the TGJT grammatical sentences scores for simple (r = 30) and complex grammatical structures (r = .33).

From these results, no clear conclusions can be made regarding whether the level of difficulty of grammatical structures modulated language learning processes. Potentially, the procedural memory system was not predictive of either knowledge types for either type of grammatical structures because the grammatical structures were predominantly learned explicitly. Further, no conclusions can be made regarding Reber's (1989; 1993; A. S. Reber et al., 1991) claim that procedural learning is the process where a complex, rule-governed knowledge base is acquired. This is because the memory system was not used in the learning of the structures. Therefore, these results need to be interpreted with caution. The adoption of the ex post facto design meant there was no control over the amount of exposure to individual grammatical structures targeted. The unequal amounts of input may explain these results. For instance, it is possible that learners had received extensive instruction on complex grammatical structures as compared to simple structures. To examine the effects of complexity and of exposure to input on the acquisition, the outcome scores were differentiated based on whether the complex structures were explicitly taught or not (see section 3.4). This aspect of the research was addressed by *Hypothesis 2b* and the results are described below. These results formed the basis for interpreting RQ2 with the aim of avoiding the confounding variable of (the amount of) exposure to input.

5.2.2 Memory, difficulty, exposure and language knowledge

To examine *Hypothesis 2c,* the dependent variables were the UGJT ungrammatical sentences scores and the TGJT grammatical sentences scores for explicitly taught and not explicitly taught complex structures. As pointed out above, it was expected that the procedural memory system would be predictive of the learners' performance on complex structures regardless of the exposure condition, with a stronger effect expected for those grammatical structures not explicitly taught than for structures taught explicitly. This hypothesis was not supported.

As measured by the TGJT grammatical sentences scores for explicitly taught and not explicitly taught complex structures, a SEM analysis revealed that the procedural memory system had no predictive effect on either score (Figure 4.10 and Table 4.50). Instead, the declarative memory system weakly predicted only the TGJT grammatical sentences score for explicitly taught complex structures (R^2 = .17). Since the TGJT might have allowed learners some access to explicit knowledge, especially for structures for which explicit instruction was used, then some predictive effect of declarative memory on the learners' language performance is expected. As measured by the UGJT ungrammatical sentences scores for explicitly taught and not explicitly taught complex structures, a SEM analysis revealed that the procedural memory system had no predictive effect on either score (Figure 4.9 and Table 4.46). Instead, the declarative memory significantly and strongly predicted the UGJT ungrammatical sentences score for explicitly taught complex structures (R^2 = .56) but it moderately predicted the UGJT ungrammatical sentences score for not explicitly taught complex structures (R^2 = .32).

In summary, the interaction of the level of linguistic difficulty and exposure type did not modulate procedural learning processes in the present study. The procedural memory system had no predictive effects on the learners' procedural language knowledge of explicitly taught and not explicitly taught complex structures. On the contrary, the interaction of the level of linguistic difficulty and exposure type modulated declarative learning processes. The explicitly taught complex structures were learned better as declarative knowledge than not explicitly taught structures. Though these results are contrary to the prediction, they are not surprising. Because grammar learning is largely explicit for these learners, the procedural memory's lack of predictive effects for not explicitly taught complex structures (especially as measured by the procedural language knowledge test) may simply indicate the learners' lack of procedural knowledge of the structures due to inadequate "naturalistic" exposure to the structures.

The present study's results regarding the modulating role of the level of linguistic difficulty or its interplay with exposure type on declarative and procedural learning processes do not appear to support those of Granena (2013b; 2014) and Tagarelli et al. (2016). Granena (2013b), examining the interplay of procedural memory, exposure type and morphosyntactic L2 attainment in 18-year-old Chinese learners of Spanish largely naturalistically, found that rule type (agreement and non-agreement) differentially modulated the procedural learning processes. Procedural memory was found to play a role in the learning of only agreement grammatical structures, both in late learners who started learning the L2 after age 16 and early childhood learners who started learning the L2 between ages 3 and 6. However, Granena's (2013b) results do not clearly address the effects of linguistic complexity. First, though the non-agreement grammatical structures (which comprised structures that make essential contributions to meaning i.e., subjunctive mood, perfective/imperfective aspect, and Spanish passive) might have been more cognitively taxing than agreement structures (which consisted of noun–adjective gender, subject–verb, and noun–adjective number agreement), both structural types were considered complex and difficult for L2 learners. Without the inclusion and examination of both linguistically simple and complex grammatical structures, it is difficult to arrive at meaningful conclusions as regards the role of grammatical feature complexity. Second, because Granena's (2013b) main aim was to investigate whether procedural learning is an ability with meaningful IDs, the study did not include measures of declarative memory which would shed more light on the modulating effects on the targeted grammatical structures, especially in light of the fact that the Llama-D test, a procedural memory measure, had correlated significantly with both the overall score of the Metalinguistic Knowledge Test (MKT) and the MKT's grammatical and ungrammatical test item scores which index explicit language knowledge. However, in contrast to the present study where the learning condition was largely explicit, Granena (2013b) investigated the role of the procedural memory system on L2 learning in an immersion context where input is vast. This may explain the differences in the predictive effects of the procedural memory system.

Like Granena (2013b), Tagarelli et al. (2016) provide partial support of Reber's claim that procedural learning is the process where complex, rule-governed knowledge is acquired. Though Tagarelli et al. (2016) found that linguistically simple rules were learned better in both explicit and implicit conditions, and that WM did predict the learning of simple structures, they found the predictive effects of the procedural memory system only on the most linguistically complex grammatical structure in the incidentally trained group. The UGJT was the only language knowledge measure that was used in Tagarelli et al.'s (2016) study.

The results of the present study also do not support those reported in Ettlinger et al. (2014) with only incidentally trained learners on a semi-artificial grammar. Ettlinger et al. (2014) found that declarative memory was only associated with the learning of a complex grammatical feature while the procedural memory system was associated with the learning of both simple and complex grammatical structures. Thus, in Ettlinger et al.'s (2014) study, the simple rule was best learned as procedural knowledge. However, the learners in the study were told at the beginning of the study about learning a new language, which may have led to conscious learning of the target grammatical structures.

The present study findings lend support to previous studies such as Granena (2014), Yalçın and Spada (2016) and Robinson (1997) which have reported results that do not support Reber’s (1989; 1993; A. S. Reber et al. 1991) claim that procedural learning is the process where a complex, rule-governed knowledge base is acquired. However, the designs (learning conditions and/or measurements) of these studies, like the present study, appear have been weighted in favour of explicit learning processes. Granena (2014), investigating the role of cognitive abilities (as measured by Llama subtests) in the acquisition of L2 Spanish agreement and non-agreement structures studied in her (2013b) study, found a statistically significant association only between explicit learning abilities and the learners’ declarative knowledge of agreement grammatical structures. These learners had acquired the L2 Spanish in an immersion context.

Yalçın and Spada (2016) report results indicating that the level of linguistic difficulty does not modulate procedural learning processes. The explicitly instructed passive (a linguistically complex grammatical structure) and progressive tense (a linguistically simple grammatical structure) were found to have been predicted by the declarative learning processes. Llama-D, the test that indexes implicit learning, did not predict learning. Similar results are reported in Robinson (1997) who found strong and significant correlations between the declarative memory system (as measured by the MLAT5 Paired Associates) and scores for both linguistically simple and complex grammatical structures in the explicitly trained conditions.

Finally, Carpenter (2008) reports results that partially support those of the present study. Examining the contributions of exposure type, linguistic difficulty and IDs in declarative, procedural and WM memory to the acquisition of an artificial language at advanced proficiency, Carpenter (2008) found that linguistic difficulty of three grammatical structures: morphosyntax, indexed by agreement (noun-determiner and noun determiner-adjective); syntax, indexed by phrase structure (S-V-O); and the lexical aspects of grammar, indexed by verb argument (transitivity) did not modulate the explicit learning processes. The declarative memory system predicted the learning of all the three grammatical structures. For the implicit condition, the procedural memory system only predicted agreement and phrase structures.

5.3 Exposure length, memory and language knowledge

Research Question 3 was largely exploratory and asked whether the L2-instructed learners from different age groups and educational levels showed different patterns in (a) their language knowledge, and (b) their declarative and procedural memory abilities. To address this question, a series of one-way between-group ANOVA were conducted on the declarative and procedural language knowledge measures (i.e., UGJT ungrammatical and TGJT grammatical sentences scores) and on the memory systems measures. Though the research question was exploratory, one prediction was made regarding the learners' declarative and procedural language knowledge as described below. *Hypothesis 3a*: It was hypothesised that as a function of the length of exposure, the university and secondary school learners would perform better on both declarative and procedural language knowledge measures than the primary school learners. No predictions were made regarding group differences on the memory measures.

5.3.1 Between-group differences in language knowledge.

Learner-group differences in declarative and procedural language knowledge were examined for the overall score for the knowledge measures. To examine the modulating effects of exposure condition and the level of linguistic difficulty on learner-group differences in language knowledge, the language knowledge scores were further split based on exposure type and linguistic difficulty. *Hypothesis 3a* that as a function of the length of exposure, the secondary school and university students would perform better than primary school learners on both declarative and procedural language knowledge measures, was supported. Descriptive statistics for UGJT ungrammatical and TGJT grammatical sentences composite scores as well as for UGJT ungrammatical and TGJT grammatical sentences scores for explicitly and not explicitly taught, and for linguistically simple and complex structures, all showed that secondary school and university students outperformed primary school learners (Tables 4.51 and 4.54). Further, for all sets of UGJT ungrammatical sentences scores, secondary school students performed slightly lower than university students. For the sets of TGJT grammatical sentences, secondary school learners had slightly higher scores on simple structures and on structures not explicitly taught than university students, indicating that the two groups might have not differed much in their procedural language knowledge of these structures.

A series of one-way between-group ANOVA omnibus tests showed statistically significant differences between the three learner groups across all the five sets of UGJT ungrammatical sentences scores. The group differences were all large (i.e., as in Table 4.52 indicating the Omega squared (ω^2) effect sizes of above .10).[18] The results of follow-up planned orthogonal comparisons showed some important performance differences between the three learner groups (Table 4.53). First, the planned comparisons indicated significant group differences in all the comparisons except those between the secondary school group and university students on scores for structures not explicitly taught and for the complex structures. These results suggest that the learners' educational level after primary school had no differential effect when a structure was complex or likely to be implicitly acquired. Second, the effect sizes for the group differences between primary school and secondary school learners and between primary and university students were all large ($r \geq .50$) for all sets of scores except those on structures not explicitly taught. These results suggest that primary school learners differed widely in their declarative knowledge from secondary school and university students, with secondary school and university students having much stronger declarative language knowledge. This difference is unsurprising because both secondary school and university students had been exposed to much more (explicit) language input than primary school learners and were more able to learn explicitly because of maturational changes. Third, for all the three comparisons, moderate-to-weak effect sizes were observed only on structures that were not explicitly taught, suggesting that learners did not differ much relative to other structural types.

For the sets of TGJT grammatical sentences scores, a series of one-way between-group ANOVA omnibus tests showed varying results. Several group differences on the sets of TGJT scores were found. First, on the TGJT grammatical sentences scores (overall score), the primary school group differed from the secondary school group ($r = .20$) and from university student group ($r = .32$). These effect sizes were small as compared to those observed in the comparisons of the learners' declarative knowledge, suggesting

---

[18] The guidelines for interpreting Omega squared effect sizes are as follows: .01 as a small effect; .06 as a medium effect; and .14 as a large effect

that the group differences for the overall score on the procedural language measure were not that pronounced.

Further, exposure type was found to have modulated performance on the TGJT grammatical sentences. First, the primary school group differed significantly from the university student group on the TGJT grammatical sentences scores for explicitly taught structures (*r* = .45). No differences were observed between the primary school and secondary school groups and between the secondary school and university student groups. On structures not explicitly taught, no differences were found between learner groups. This is an interesting finding. Because these comparisons were between three-learner groups at different educational levels, finding no group differences in these structures may suggest that implicit learning processes are indeed stable with minimal IDs as claimed by Reber's (A. S. Reber et al., 1991). However, these results be interpreted with caution because all learner groups scored around .50 (Table 4.54), suggesting most learners' responses, irrespective of group, were merely by chance without being confident.

Similarly, linguistic complexity appeared to have modulated learners' scores on the TGJT grammatical sentences. It was the primary school group that differed significantly from the university student group on the TGJT grammatical sentences scores for complex structures (*r* = .31). No differences were observed between the primary school and secondary school groups and between the secondary school and university student groups. These results appear unsurprising. Younger learners' (i.e., primary school learners) performance is expected to differ from that of university students due to differences in the exposure length. However, the lack of significant differences in the other two comparisons may arise because these structures were complex and that the exposure lengths between the primary and secondary school groups and between secondary school and university students may not be significant enough to modulating learning.

On the TGJT grammatical sentences scores for simple structures, a significant difference was found only between the primary and secondary school groups (*r* = .29). There were no significant differences between the primary school group and the university students and between the secondary school group and the university students scores. That primary school learners' performance was not significantly different from that of university students but significantly different from that of the secondary school group is surprising. It would be expected that one finds no differences for all comparisons or to find differences in the primary school and university students' comparison as the result of the big difference in the exposure time. To get a clear picture of the effects of linguistic complexity, group comparisons for simple structures were analysed further by exposure type. It was observed that the three learner groups differed significantly only in simple structures that are not explicitly taught, suggesting that the difference in the overall simple structures' mean scores observed between primary and secondary school groups resulted from the difference in the groups' mean scores for not-explicitly taught simple structures. Thus, linguistic difficulty or complexity interacted with exposure type to modulate the proceduralized knowledge of simple structures. One possible interpretation would that the explicit learning of simple structures either influenced or interfaced with implicit learning processes as proposed in the versions of the weak interface position (N. C. Ellis, 1994; 2005; 2008; R. Ellis, 1993; Schmidt & Frota, 1986; Sharwood Smith 1981) or as proposed in the strong interface position (DeKeyser, 2007; 2003; 1998; Sharwood Smith, 1981. The most probable explanation for these results, especially considering that the learning environment is generally explicit, would be that the effects of explicit teaching and the use of the TGJT (as it encourages reflection) may have enabled the learners to access, to some extent, their explicit knowledge when judging those simple structures that are explicitly taught.

5.3.2 Between-group differences in memory abilities.

To address the question of learner-group differences in declarative and procedural memory, a series of between-group ANOVAs were conducted on the declarative memory measure (i.e., the declarative memory ability factor, namely the DMA scores) and on the procedural memory measures, the SRT and Llama-D test scores. As detailed in Chapter 4, the DMA factor score estimates were computed for each participant from the four declarative memory measures: the CVMT, Llama-B, 3-Term and DAT-V.

The results of a series of one-way ANOVA omnibus tests showed no significant group differences in the Llama-D scores (w^2 = -.009) and in the SRT scores (w^2 = .063) but there were statistically significant group differences in the DMA scores (w^2 = .508). All the three comparisons in the DMA scores were significantly different with moderate to large effect sizes as follows: between primary and secondary school groups (*r* = .46); between primary school group and university students (*r* = .72); and between secondary school group and university students (*r* = .43). Descriptive statistics for the DMA factor scores showed that university students had the highest mean score (Mean = .80) while the primary school learners had the lowest (Mean = -.68).

These results appear to provide support to Reber's (1989; 1993; A. S. Reber et al. 1991) hypothesis that variability is the characteristic of the explicit, intentional, conscious learning processes. In the present study, the three-learner groups varied in terms of not only their educational level but also their ages, with the primary school learners as the youngest group (mean age = 14.1, SD 2. 05) and the university students as the oldest group (mean age = 23.9, SD 3.14). Despite these differences in ages and education levels, the learners showed no significant variations in their performance on measures that index implicit learning abilities.

To the knowledge of the researcher for the present study, no previous studies in L2 acquisition research have attempted to examine the declarative and procedural memory systems among learner groups of different ages and educational levels. A great deal of research has indirectly investigated the question of IDs in the procedural memory system by looking at whether the procedural memory was predictive of learning or not. This research has found procedural memory predictive of, or strongly related to, the scores of outcome variables (e.g., Carpenter, 2008; Ettlinger et al., 2014; Granena 2013b; Hamrick, 2015; Kaufman et al., 2010; Kidd, 2012; Pretz et al., 2010; 2014; Tagarelli et al., 2016). This has been viewed as evidence that the memory system is an ability with meaningful IDs. A. S. Reber et al.'s (1991) is the only previous study, to the knowledge of the researcher of the present study, that has directly examined learners' variability in implicit and explicit learning processes. A. S. Reber et al. designed both the explicit and implicit learning tasks in such a way that they were both of equivalent difficulty. All the 20 subjects were run through the two learning tasks such that the scores on both tasks were taken from the same sample. Using a simple heterogeneity of variance test, A. S. Reber et al. (1991) found that the distribution of scores from the explicit task had a "dramatically greater variance than that from the implicit" (p. 893), suggesting higher individual variability in the explicit learning task than in the implicit learning task. These results are supported by those found in the present study. The present study, however, differs from A. S. Reber et al.'s (1991) study in the use of ANOVA as a statistical procedure for analysing the data and in not controlling for the difficulty of the declarative and procedural learning tasks. Thus, the two studies are not directly comparable.

This chapter has summarised the present study’s results and discussed them with reference to each of the research questions. The discussion of the findings has also been considered in relation to the relevant previous studies in L2 acquisition research. What follows is the conclusion chapter.

CHAPTER 6. CONCLUSION

In Malawi, a small African country, proficiency in L2 English determines one's livelihood, educational opportunities, and socio-political participation. However, because learners invariably come into contact with English at school, it is only through schooling experiences that students can develop English proficiency. In recent years, there has been an unending country-wide outcry regarding learners' poor achievements in L2 English. The goal of the present study was to explore the source of the learners' L2 English knowledge in Malawi by examining the representation of the their L2 knowledge systems as a function of the learning environment. Therefore, the general aim of the study was to investigate the role of the two most important long-term memory systems (namely, declarative and procedural memory) and linguistic difficulty in the acquisition of L2 English in an instructed context in Malawi.

The study adopted an ex post facto design with 103 L2 English learners from primary school, secondary school and university levels. At the time of data collection, all the participants had learned English as their L2 for at least 7 years, largely through explicit instruction in the government's national school programme. A total of 14 English grammatical structures were targeted. However, scores for 13 grammatical structures were used for analysis (see section 4.2).

Positing domain-general processes in L2 acquisition, the study was couched within the cognitive theoretical approaches which hold that the declarative and procedural memory systems are the two most important long-term memory systems in the brain and they sustain a range of functions and domains including language acquisition (e.g., Lovibond & Shanks, 2002; Mitchell et al., 2009; Whittlesea & Price, 2001; McLaren et al., 2014; P. J. Reber et al., 1996; Squire, 1992; 2004; Squire & Zola, 1996; Ullman, 2001a; 2001b; 2015; Ullman et al., 1997; M. Paradis, 2009; Saville-Troike & Barto, 2017; A. S. Reber, 1989; 1993; A. S. Reber et al., 1991). The two memory systems are affected by both internal and external factors. Neurological research has shown that individuals differ in terms of the strength of the declarative and procedural memory systems (Carpenter, 2008; Granena, 2013; Kaufman et al., 2010). As pointed out in section 2.2.2, whether individual differences (IDs) exist in the procedural memory system is still a contested issue. External factors such as learning condition and linguistic difficulty have also being found to mediate the involvement of the two long-term memory systems. As discussed in the literature review in Chapter 2, three models explain how the declarative and procedural memory systems are implicated in language acquisition: Ullman's DP Model (Ullman, 2001a; 2001b; 2004; 2015; Ullman et al. 1997); Paradis's DP Model (Paradis 1994; 2004; 2009) and Reber's Implicit Learning Theory (A. S. Reber 1989; 1993; A. S. Reber et al., 1991). Reber's implicit learning theory claims that procedural memory is not an ability and that the memory system is implicated more especially when the stimulus environment is complex. The present study conceptualised both the declarative and procedural memory systems as abilities which vary according to individuals.

To investigate the role of linguistic complexity, the targeted 14 grammatical structures in the present study were categorised as simple or complex structures based on the levels of their linguistic complexity. In the absence of clear criteria for determining the extent of linguistic complexity of grammatical features, the present study drew on the core aspects of Pienemann's PT (e.g., Pienemann, 2005; 2015; Pienemann et al., 2005; Pienemann & Keßler, 2012) to propose a criterion that focuses on the intrinsic structural complexity of grammatical features to objectively categorise the features into simple and complex (see section 3.4). The point of feature unification in the constituent structure was used to distinguish linguistically simple functors from linguistically complex functors. As conceptualised in PT, mapping and linguistic linearity were the two linguistic aspects that were used to distinguish syntactically simple structures from complex ones. To allow the investigation of the role of exposure type, the grammatical structures were further differentiated according to whether they were "explicitly taught" or not (see section 3.4.). Document analysis of primary and secondary school syllabuses as well as the present study's researcher's experience were used to determine those structures explicitly taught and those that were not.

Evidence from the correlational and exploratory factor analyses shown that TGJT grammatical sentences and UGJT ungrammatical sentences provided good measures for procedural and declarative language knowledge types, respectively (see section 4.2). For the memory systems measures, evidence for test validity from correlational, exploratory and confirmatory factor analyses, shown that the shared variance of the CVMT, DAT-V, Llama-B and 3-Term tests formed the declarative memory construct whereas the shared variance for the SRT and Llama-D formed the procedural memory construct (see section 4.2). To answer the present study's research questions, a series of

Structural Equation Modelling (SEMs), ANOVAs and correlational analyses were conducted, with the declarative memory factor, SRT and Llama-D as predictors (see Chapter 4 for the rationale for the SRT and for Llama-D to directly predict outcome scores) and TGJT grammatical and UGJT ungrammatical sentences scores as dependent variables.

This concluding chapter presents, in section 6.1, a summary of the key findings of the study. The study’s theoretical, methodological and pedagogical implications are presented in section 6.2. Limitations and directions for future research are presented in sections 6.3. Section 6.4 concludes the chapter.

6.1 Summary of key findings

Research Question 1 asked whether instructed L2 English learners’ declarative and procedural memory systems predicted grammar knowledge, as measured by tests of declarative and procedural language knowledge. The results suggest dissociation of declarative and procedural learning processes, with explicit instruction resulting in only declarative learning processes with no possibility of interfacing with the procedural learning processes. The results showed that the procedural memory system was not predictive of the learners’ language knowledge, possibly because the targeted grammatical structures are largely explicitly taught and learned. Only the declarative memory system predicted the learners’ declarative knowledge of the grammatical structures. When data were analysed by exposure type, still no predictive role was found for the procedural memory system. As hypothesised, the declarative memory system very strongly predicted only the declarative language knowledge scores for grammatical structures that are explicitly taught. Due to the fact that the grammatical structures are generally explicitly taught and that the language knowledge measures (the GJTs) used encourage reflection, it was unsurprising for the declarative memory system (a) to have weakly predicted learners’ procedural knowledge of all structures irrespective of the exposure type, (b) to have weakly predicted learners’ procedural knowledge of grammatical structures explicitly taught, and (c) to have moderately predicted learners’ declarative language knowledge scores for grammatical structures which are not explicitly taught.

Research Question 2 asked whether the level of difficulty of grammatical structures or its interplay with exposure type had a differential effect on the nature of the relationship between the learners' memory systems and their knowledge of grammatical structures. Contrary to A. S. Reber's (1989; 1993) claim that implicit learning is the process whereby a complex, rule-governed knowledge base is acquired, the results of the present study showed that linguistic difficulty indeed modulated learning but only for the declarative learning processes. Declarative memory predicted the learners' declarative knowledge of complex grammatical structures better than simple structures. However, as pointed out above, these results need to be interpreted with caution because the role of linguistic difficulty must have masked how the structures are learned. Because both complex and simple grammatical structures were largely learned explicitly, the learners must have used their declarative memory intensively when learning complex structures. The interaction of the level of linguistic difficulty and exposure type modulated declarative learning processes only. The declarative memory system predicted the learners' declarative knowledge of the explicitly taught complex structures better than the structures that were not explicitly taught. Because the grammatical structures were largely learned explicitly, the learners must have very little opportunity to learn the structures naturalistically. Due to inadequate naturalistic exposure to the structures, procedural memory was not used for learning. This explains why the memory system did not predict the learning of either type of complex grammatical structures.

Research Question 3 explored whether the L2-instructed learners from different age groups and educational levels showed different patterns in (a) their language knowledge, and (b) their declarative and procedural memory abilities. To explore the effects of linguistic complexity and exposure type on language knowledge, further group comparisons were conducted on declarative and procedural language knowledge scores for simple and complex structures as well as for explicitly and not explicitly taught grammatical structures. Overall, large effect sizes were observed in almost all comparisons between primary school learners and the other two groups of learners, an indication that the younger learners widely differed in their declarative knowledge from the older secondary school and university students. Primary school learners performed significantly lower on all five sets of declarative language knowledge scores than did secondary school and university students. That secondary school and university students differed significantly on all the declarative knowledge scores, except those of complex grammatical structures and grammatical structures not explicitly taught, appeared to suggest that complex and not explicitly taught structures rely on similar (implicit) learning processes as the result of lengthy exposure.

The results further showed that the three learner groups did not differ much on their overall procedural language knowledge, as indicated by relatively smaller effect sizes, than those found for declarative language knowledge. Both exposure type and linguistic difficulty had modulated learning. One key finding was that while primary school learners' performance on procedural knowledge for explicitly taught grammatical structures was significantly different from that of university students (with a medium effect size), no group differences were found in learners' procedural language knowledge for grammatical structures that were not explicitly taught. This finding, as pointed out above, lends support to the claim that implicit learning processes are characteristically stable with minimal IDs.

Another key finding of the present study comes from the results of learner-group comparisons in declarative and procedural memory systems. No group differences were found on procedural memory measures, but all the three groups performed significantly differently from each other on declarative memory measures. These results were taken as supporting A. S. Reber's (1989; 1993; A. S. Reber et al., 1991) claim that variability is the characteristic of the explicit, intentional, conscious learning processes.

To put it more succinctly, the principal findings of the study suggest that (1) the procedural memory system played no role in the learning of grammatical structures, suggesting the learners' knowledge was represented only as declarative knowledge; (2) linguistic complexity and exposure type appeared to modulate only the declarative learning processes; and (3) no group differences were found in the procedural memory system. The following section discusses the implications of these results for L2 acquisition theory, methodology and pedagogy.

6.2 Implications

6.2.1 Pedagogical.

This section addresses the practical or pedagogical implications of the present study. First, in a more general way, prior experience and knowledge of how to promote the development of L2 declarative and procedural knowledge can go a long way in assisting a language teacher to become more effective in the delivery of language lessons. For instance, based on the results of the present study where only the declarative memory system strongly predicted the declarative knowledge of the grammatical structures which are generally explicitly taught, the teacher would be more effective in developing learners' declarative knowledge by explicitly teaching those structures. Further, at the level of curriculum and material design, language teachers, educators or language programme designers and examiners can carefully choose materials and task types (either those involving explicit or implicit learning processes) based on the goals of the language learning lessons. For language learning tasks that aim at developing learners' proficiency, the best task type would be that requiring the learners to employ their implicit learning mechanisms.

Another important pedagogical implication is that explicit instruction draws on only the declarative learning processes. The results of the present study showed only the declarative memory system strongly predicting the declarative language knowledge of the explicitly taught grammatical structures. The lack of predictive effects for the procedural memory system may imply that the learners lacked procedural learning processes possibly because the instruction was explicit. As argued by R. Ellis (2005b), the primary goal of L2 learning for many learners such as those who participated in the present study is the development of implicit linguistic knowledge which is required for authentic and effective communication. Authentic and effective communication require a high level of language abilities such as advanced-level fluency and comprehensibility, and appropriately organising the discoursal delivery (Robinson, 2005a). The researcher of the present study, who, as a student and an L2 English teacher, has gone through the same education system as the participants in the study, suggests based on the present study's results, that the communicative problems that L2 English learners in this education system encounter may be linked to the fact that they lack procedural knowledge of L2 English. It is therefore proposed that the adoption of task-based teaching methods that should encourage the procedural learning of grammatical structures would develop the learners' procedural language knowledge (or their linguistic competence).

It is important at this point to point out that the lack of significant association between procedural memory measures (i.e., the SRT and Llama-D tests) and the procedural language knowledge measure (i.e., the TGJT grammatical sentences score) in the present study might have arisen either from the low reliability of the memory measures or from the fact that the language knowledge test did not actually measure the learners' procedural memory.[19] Consistent with research on the construct validity of language knowledge types which shows that the measures of implicit and explicit knowledge are not pure measures (R. Ellis, 2008 cited in Akakura, 2009), Vafaee et al. (2017) found that the different types of GJTs measure different levels of explicit knowledge. Future investigations using oral production tests and measures of brain activation during language processing are required to gain further insights about the role of the memory systems in L2 development.

6.2.2 Theoretical implication.

This study contributes to the question of whether language learning is a domain-specific skill or not. Linguistic theoretical (or nativist) approaches assume that language learning is a domain-specific skill involving the acquisition of highly abstract linguistic principles and constraints called Universal Grammar (UG, e.g., Bley-Vroman, 1989; Boeckx, 2006; Chomsky, 1995; Eubank, 1993; 1994; Schwartz & Sprouse, 1996; Vainikka & Young-Scholten, 1994). The findings of this study suggest that domain-general cognitive abilities, particularly the declarative memory system, play important roles in L2 grammatical acquisition. The declarative memory system, likewise the procedural memory system, supports learning in other domains including L1 (see Chapter 2), suggesting no fundamental differences between L1 and L2 acquisition. As is well known, a wide range of factors including cognitive, linguistic, affective, and sociocultural factors play a significant role in L2 acquisition. However, the most probable explanation for the involvement of these other factors in L2 acquisition could be that their engagement is mediated by domain-general processes such as those supported by the two long-term memory systems (Morgan-short et al., 2014). Because no role was found in the present study for the procedural memory system, future research should investigate the predictive nature of the memory system in similar learning environments to the present study using other test measures such as event-related potentials (ERPs) to measure brain activation during processing.

---

[19] Note that while the SRT and Llama-D were correlated, the TGJT grammatical sentences composite score correlated with two declarative memory measures (see Table 4.20)

The second obvious theoretical implication of this study concerns the declarative/procedural learning/knowledge and declarative/procedural memory distinction. The results of this study suggest that this distinction may indeed be a one-to-one relationship, and that conscious, declarative processes do not interface with unconscious, procedural processes. Thus, the effect of explicit instruction benefits only the declarative learning processes. This is the view as proposed in Paradis' (2004; 2009) DP model, A. S. Reber's (1989; 1993; A.S. Reber et al., 1991) evolutionary theory of implicit learning and in the non-interface perspectives of language knowledge. Consistent with these theoretical perspectives, only the declarative memory system played a role in the acquisition of the grammatical structures in the present study. The lack of the predictive effects of the procedural memory system may therefore be the result of the exposure condition that does not advantage procedural learning processes. Further, even in the face of the long periods of explicit training and practice in the present study (i.e., considering that the learners were exposed to many hours of training at primary school, secondary school, and university), the procedural memory remained unpredictive of learning. These findings may additionally imply that a procedural process does not depend on declarative learning process as espoused in skill acquisition research (but see Beaunieux et al., 2006 for contrary findings). Even at high proficiency, explicitly trained learners may not have the reliably automatic syntactic processing capacity found in L1 and subserved by the procedural memory (Morgan-Short et al., 2012). Thus, only exposure conditions in which explicit learning and/or knowledge are minimised lead to dependence on procedural memory processing.

Another theoretical contribution of this study relates to the degree of IDs in the two long-term memory systems. The results of the present study suggest that procedural learning processes display tighter distributions in a population when compared with declarative systems. The results showed that the three learner groups varied significantly in their declarative memory system but not in their procedural memory. However, as pointed out above, these results were based on group comparisons and not on the individual learner comparison of their declarative and procedural memory capacities. Future research should not only examine the degree of variance of IDs in the two memory systems using a simple heterogeneity of variance test (as in A. S. Reber et al., 1991) but also employ measures of other cognitive abilities such as processing speed and personality (intuition and impulsivity) which in Kaufman et al. (2010) correlated with implicit learning.

The study findings also contribute to the debate about whether declarative and procedural learning processes in L2 acquisition are supported by stimulus-specific or domain-general learning mechanisms. The measure of declarative memory system, which was found predictive of the L2 English grammatical structures in the present study, consisted of two strands of indicators: one, of a visual non-linguistic indicator (i.e., the CVMT test); and the other, of three verbal indicators (i.e., the Llama-B, 3-Term and DAT-V tests). That the visual non-linguistic CVMT test not only correlated weakly and moderately with the TGJT grammatical and UGJT ungrammatical sentences scores (Table 4.20), but also loaded on the same factor with the three linguistic measures of the declarative memory, suggests a partial contribution of a domain-general mechanism to the explicit learning of an L2. However, since there was only one non-linguistic measure of declarative memory in the present study, future research, taking a solid methodological approach, should examine this question further using several non-linguistic and linguistic measures of the declarative and procedural memory systems on solid methodological ground.

The present study also contributes to our understanding of L2 acquisition, irrespective of learners’ L1 background. The L2 English learners in the present study were native speakers of Bantu family languages widely spoken in central and southern parts of Africa, with grammars that bear no similarity to English. While there has been considerable research examining the role of domain-general processes in L2 acquisition, the findings from studies like the present one, with learners with L1 backgrounds not explored yet, may further our understanding of the role of cognitive abilities, linguistic difficulty and learning conditions in L2 acquisition.

6.2.3 Methodological.

The study also has several methodological implications. The first is that the triangulation of the measures allowed the exploration of construct validity. There have been calls to examine the validity of constructs such as declarative and procedural memory and declarative and procedural language knowledge (e.g., Buffington & Morgan-Short, 2018; Granena, 2013a; Vafaee et al., 2017). In the present study, correlational analyses, EFA and CFA showed that the SRT and the Llama-D were indictors for the procedural memory system while the DAT-V, CVMT, 3-Term and Llama-B tests were indicators for the declarative memory construct. While CVMT and Llama-B tests are increasingly becoming common measures of the declarative memory in L2 acquisition research that seeks to account for the role of cognitive abilities, the use of DAT-V and 3-Term tests is uncommon in this research. EFA and CFA showed that both the DAT-V and 3-Term loaded strongly on the declarative memory factor ($r >$ .50). Further, of all the measures of the declarative memory, the DAT-V test was the strongest indicator of the factor ($r >$ .70) but also highly correlated with the UGJT ungrammatical sentences composite score (Table 4.20). Future research should explore further the use of these measures of the declarative memory system.[20] For the language knowledge measures, correlational analyses and EFA provided further evidence that scores on the TGJT grammatical sentences and the UGJT ungrammatical sentences are indicators of separate aspects of language knowledge. Further, the analyses showed that EIT that does not include a semantic aspect in its design is indeed not a valid measure of implicit language knowledge.

The second methodological implication is the significance of the use of CFA and SEM as statistical tools for test measures' validation and for data analysis. First, the use of CFA and SEM in the present study led to a critical decision to use each of the two procedural memory measures (i.e., the SRT and Llama-D tests) to directly predict the outcome variables. In addition to the fact that the two measures were negatively correlated, the factor loadings in the SEM model (Figure 4.3) showed high variability in the scores of the two tests, revealing lower factor loadings than those identified in the measurement model, the CFA (Figure 4.2). Such unstable indicators could have an impact not only on further SEM analyses but also on the factor score computation. It is very unlikely that other statistical analyses would have revealed this instability in the two test scores. Second, the use of SEM allowed for the simultaneous investigation of various interrelated dependence relationships among variables, in which inferences were made either between the predicting factor and the outcome variables or between factor indicators and outcome scores.

---

[20] The drawback with the use of the DAT-V and the 3-Term tests is that they are both language dependent.

Lastly, by adopting the ex post facto design and investigating the learners' L2 knowledge systems in a more natural setting of classroom instruction rather than in a controlled, experimental setting, the present study maximised the ecological validity of the findings in L2 acquisition research.

6.3 Limitations and future research

The present study took a solid methodological approach by exploring learners' knowledge of a wide range of L2 grammatical structures (i.e., 13 in total were used in the analyses) and by employing sophisticated analyses (namely EFA, CFA and SEM) that allowed the examination of the underlying construct design features and various interrelated dependence relationships of test measures. However, like many others, this study is not without limitations. This section presents these limitations and suggestions for future research.

First, the most obvious limitation is the inherent limitations of the ex-post facto design that the present study adopted. By adopting such a design, the study's results were predicted based on an action, namely the explicit learning of grammar rules that had already occurred. Thus, there was no control over this independent variable; there was not only no control over some instances of implicit instruction but also over the amount of explicit input for individual grammatical structures. This study's findings therefore need to be interpreted with caution as there could be more than one possibility or cause for the effects seen in the results. Experimental investigations with novel structures of natural language grammars are required.

The present study was also limited in its operationalisation of the declarative and procedural language knowledge types. Each of the language knowledge types was operationalised as learners' judgements on only one measure: judgements of the UGJT ungrammatical sentences as the measure of declarative language knowledge, and judgements on the TGJT grammatical sentences as the measure of procedural language knowledge. Gaining further insights into the association of memory systems and language knowledge, would require the inclusion of measures of declarative and procedural language knowledge types such as the Metalinguistic Knowledge Test (MKT; a declarative knowledge measure) and Oral Production Tests (OPTs; procedural knowledge measures). In the present study, the EIT, designed as a procedural language knowledge measure but not implemented as used in Erlam (2006, see sections 3.5 and 4.2), was excluded from analysis because the examination of construct validity of the measure showed it had allowed learners' reliance on declarative knowledge. Suzuki and DeKeyser (2015) also found that the MKT, a declarative knowledge measure, significantly predicted the EIT but not a Word Monitoring Test (WMT), suggesting that the EIT draws on explicit type of linguistic knowledge.

Further, the TGJT grammatical sentences score in the present study might not have necessarily indexed learners' procedural language learning processes. In addition to the inherent design feature of the GJTs which directs learners to focus on form when completing the tasks, the TGJT in the present study was implemented differently from that in R. Ellis' (2005a; 2009c) studies. In the present study, the TGJT did not require a timed response to each item, which might have allowed some participants to occasionally reflect on their responses. Further, recent empirical evidence appears to suggest that the un-speeded and speeded GJTs and EIT appear to measure different levels of explicit language knowledge, namely automatised explicit knowledge and less automatised explicit knowledge (Suzuki, 2017; Suzuki & DeKeyser, 2015; 2017; Vafaee et al., 2017). Suzukia & DeKeyser (2015) further point out that it is difficult to determine if indeed the TGJT and/or the EIT index(es) implicit learning abilities in studies without independent measures, for example, of implicit/explicit learning aptitude. In the present study, the TGJT grammatical sentences score was associated with two of the measures of the declarative memory system but was not associated with either of the procedural memory measures (see Table 4.20), suggesting that the knowledge measure might have not necessarily tapped the learners' procedural language knowledge as per design. Instead, learners appeared to have accessed their declarative knowledge. However, no clear conclusions on this issue can be reached from the results of the present study. Including other measures of procedural language knowledge (e.g., the OPT) or of neuroimaging tests would be a valuable addition to verify these results. In addition, as described below, the procedural memory measures in the present study seem to have lacked power and construct validity.

Another limitation of this study is the observed lack of power and construct validity in the procedural memory system measures. The battery of only two procedural tests, the SRT and Llama-D, was not adequate and resulted in factor identification problems in the measurement model (see section 4.3.1). Further, the SRT had a very low internal consistency (Spearman-Brown’s split-half = .40; N = 97) while the Llama-D test’s reliability was not calculated because scoring was not itemised (see section 4.1.3). Perhaps more reliable measures might have shown some role of procedural learning in the L2 English acquisition. The issue of low reliability in the SRT test is briefly addressed in Section 4.1.3, and studies such as Kaufman et al. (2010) and Granena (2013b) found the SRT predictive of outcome scores despite the test’s low reliability estimates of around .44. A. S. Reber et al. (1991) argue that in contrast to explicit tasks, implicit tasks have a restricted score range, which results in relatively small IDs and low reliability estimates. The interplay between the nature of the SRT task and the type of the participants may also explain the low SRT reliability in the present study. While six other tests were also computer-administered, the SRT was the only test in the present study that was scored based on reaction times. Thus, speed in responding was the aspect of utmost importance in the test. However, most primary and secondary school participants were not literate enough in the use of a computer. Therefore, their response times to stimuli might have delayed because of their lack of familiarity with computers.

Contrary to the expected positive correlation, the SRT and Llama-D were negatively related (see Table 4.10), suggesting the lack of construct validity of the tests. In Granena’s (2012; 2016) studies, the SRT (with the version as in the present study) and Llama-D were positively related, with a weak correlation in the (2016) study (*r* = .20). In both Granena’s studies, the SRT’s RTs were not log10-transformed as is the case in the present study. However, one important finding in Granena’s studies (e.g., 2013a; 2013b; 2016) is that the Llama-D and SRT tests do not seem to measure the exact same underlying learning skills despite these tasks having consistently shown a lack of significant correlations with measures that tap into learners’ analytical/explicit processes.

A distinction between the SRT and Llama-D tests can be made on several dimensions. First, the SRT test is non-verbal and visual and involves sequence learning while Llama-D is verbal and auditory and involves phonetic awareness. Second, following Seger (1994, cited in Kaufman et al., 2010), the SRT constitutes a motor-based form of implicit learning while Llama-D can be considered as representing the judgement-based paradigm of implicit learning. Seger further points out that implicit learning paradigms differ in the ratio of explicit to implicit processes required for successful performance on the tasks. Third, it is possible that SRT and Llama-D tests measure different aspects of implicit learning as regards information processing (Granena, personal communication, January 2020). Granena observed that, [given] a task, greater activity during encoding (i.e., more learning) may be negatively associated with subsequent retention/retrieval (i.e., less memory). In other words, a greater capacity to retrieve what is learned may be linked to a lesser capacity to encode (i.e., learning less but retaining more). It could be the case therefore that the SRT test may be measuring the encoding component (i.e., implicit learning) while the Llama-D test may be measuring the retention/retrieval component (implicit memory).

It is therefore possible that the different contributions of domain-general and domain-specific, of motor-based and judgement-based implicit learning mechanisms, and of encoding and retrieval processes, coupled with the confounding group variable, may have accounted for the SRT and Llama-D observed negative association in the present study. In fact, Kaufman et al. (2010, p. 337) cite research (e.g., Gebauer & Mackintosh, 2007) that has “shown that various implicit learning paradigms do not correlate well with each other”. A recent study has also found no convergent validity among procedural memory measures (Buffington & Morgan-Short, 2018). Future research should examine these results and claims by means of CFA and SEM, with several other tests of procedural memory such as the WPT (Foerde et al., 2006; Knowlton, et al., 1994), the AGL (A.S. Reber, 1967; 1989; A. S. Reber et al., 1991) and the TOL (Kaller et al., 2011; Kaller et al., 2012).

The present study’s operationalisation of linguistic difficulty and the exposure conditions of the targeted L2 English grammatical structures constitutes another limitation of the present research. Half of the 14 targeted grammatical structures in the present study were functors and the other half were syntactic structures. Because some functors (e.g., articles, classifiers, and grammatical gender) are difficult to learn and are strongly resistant to instructional treatments (DeKeyser, 2005) while some syntactic structures are easier to learn (e.g., the canonical SVO order), no previously tested criterion for distinguishing linguistic complexity (e.g., transformational rules or T-units) would apply to all of these cases. Further, transformational rules and other subjective criteria such as cognitive abilities and pedagogical judgments have considerable drawbacks (see section 2.3). Therefore, a novel criterion was proposed and used in the present study to determine simple and complex grammatical structures (see section 3.4).

The criterion was based on aspects of Pienemann's Processability Theory (PT, e.g., Pienemann, 1998a; 1998b; 2015; Pienemann et al., 2005; Pienemann & Keßler, 2012; Pienemann & Lenzing, 2015). To distinguish the complexity of functors, the point of feature unification in the constituent structure was used. For syntactic structures, mapping and linguistic linearity as conceptualised in the PT were used to distinguish between simple and complex syntactic structures. However, the problem with this criterion is that it has not been tested before. For exposure conditions, the grammatical structures were categorised as either explicitly taught or not based on the document analysis of syllabuses and the present study's researcher's experience (see section 3.4). It would be more insightful if teachers and students were interviewed as well on the structures that are taught.

The study was also limited by the confounding variable of group. The study adopted a cross-sectional design with three age groups of learners. However, most of the analyses were based on the data of all three age groups with the assumption that the learners, irrespective of their age differences, learned the tasks they were asked to perform in similar ways. However, it is possible that some of the unexpected results in the present study may have to do with the different groups of learners included. Given the small individual group sample sizes, it was not possible to conduct SEM analyses to examine whether declarative and procedural memory correlated differently with different age groups. Future research should attempt to include large samples of learners from the same L2 backgrounds but of different age groups. The research with such methodological approach should shed more light on the role of the declarative and procedural memory systems in L2 acquisition.

Finally, another obvious limitation of this study is the lack of a comparison or control group. Results from a comparable group of learners of L2 English in immersion settings, for instance, should have provided an indication of the magnitude of the effect of exposure type. Such results would comprehensively ascertain the role of explicit instruction in the present study. Future research should aim to replicate and extend the findings of the present study with a sample of a control group of L1 or L2 English learners in immersion settings.

6.4 Conclusion

The present study demonstrates the role of declarative and procedural memory, the two long-domain-general memory systems, in the acquisition of L2 English grammatical structures in an instructed classroom environment. Only the declarative memory system was found to be predictive of learning. No role was found for the procedural memory system, which appears to suggest that the learners lacked procedural processing of their language knowledge which is relevant for effective and authentic communication. Linguistic complexity was found to have modulated only the declarative learning processes possibly because grammar is explicitly taught. The findings make a valuable theoretical contribution to L2 acquisition research on the role of the declarative and procedural memory systems, exposure type and linguistic complexity. Further, the present study has important pedagogical implications for L2 learners, language teachers, and educators or language programme designers.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in second language acquisition, 30*(4), 481-509.

Akakura, M. A. (2009). Effect of explicit instruction on implicit and explicit second language knowledge: An empirical study on English article acquisition (Doctoral dissertation, the University of Auckland).

Alexiadou, A. (2004). On the development of possessive determiners. In E. Fub, & C. Trips (Eds.), *Diachronic clues to synchronic grammar* (pp. 31-58). Amsterdam: John Benjamins Publishing Company.

Alexiadou, A. (2005). Possessors and (in) definiteness. *Lingua*, *115*(6), 787-819.

Alexiadou, A., & Wilder, C. (1998). Introduction. In C. Wilder (Ed.), *Possessors, predicates and movement in the determiner phrase* (pp. 1-16). Amsterdam: John Benjamins Publishing Company.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind.* New York: Lawrence Erlbaum Associates.

Anderson, J. R. (1976). *Language, memory and thought*. Hillsdale, NJ: Lawrence Erlbaum.

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in second language acquisition, 37*(2), 185-196.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International journal of psychological research, 3*(2), 12-28.

Beaunieux, H., Hubert, V., Witkowski, T., Pitel, A., Rossi, S., Danion, J., . . . Eustache, F. (2006). Which processes are involved in cognitive procedural learning? *Memory, 14*(5), 521-539. DOI: 10.1080/09658210500477766

Bell, P. K. (2015). Explicit and implicit learning: Exploring their simultaneity and immediate effectiveness. *Applied linguistics, 1*(22), 1-22.

Bendixen, A., & Andersen, S. K. (2013). Measuring target detection performance in paradigms with high event rates. *Clinical neurophysiology, 124*(5), 928-940.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bentler, P. M., & Wu, E. J. (1989). *EQS 6.4 for windows* [computer software]. Encino, CA: Multivariate Software, Inc.:

Bialystok, E. (1978). A theoretical model of second language learning. *Language learning, 28*(1), 69-83.

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis* (1st Ed.). Mahwah, NY: Erlbaum. DOI: 10.4324/9781410601667

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

Boeckx, C. (2006). *Linguistic minimalism: Origins, concepts, methods and aims*. Oxford: Oxford University Press.

Boeckx, C. (2007). Pseudoclefts: A fully derivational account. In J. Bayer, T. Bhattacharya & H. M. T. Babu (Eds.), *Linguistic theory and south Asian languages essays in honour of K. A. Jayaseelan* (pp. 29-40). Amsterdam: John Benjamins Publishing Company.

Bohnacker, U. (1997). Determiner phrases and the debate on functional categories in early child language. *Language acquisition, 6*(1), 49-90.

Bowles, A. R. (2003). Implicit and explicit learning of artificial grammar structures and natural language morphology: Implications for second language learning (Doctoral dissertation, University of Colorado).

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in second language acquisition, 33*(2), 247-271.

Bresnan, J. (2001). *Lexical functional syntax*. Malden, MA: Blackwell.

Bresnan, J., & Kanerva, J. M. (1989). Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic inquiry, 20*(1), 1-50.

Brill-Schuetz, K., & Morgan-Short, K. (2014). The role of procedural memory in adult second language acquisition. *Proceedings of the annual meeting of the cognitive science society, 36*(36), 259-265.

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & cognition, 41*(2), 281-296.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied psycholinguistics, 27*(2), 185-209.

Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional effects and individual differences in L2 morphology learning. *Language learning, 67*(1), 171-207.

Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior research methods, 37*(3), 436-449.

Buffington, & Morgan-Short. (2018). Construct validity of procedural memory tasks used in adult-learned language. In C. Kalish, M. Rau, J. Zhu & T. T. Rogers (Eds.), *Proceedings of the 40th annual conference of cognitive science society* (pp. 1420-1425). Wisconsin: Cognitive Science Society.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins Publishing Company.

Buyl, A., & Housen, A. (2015). Developmental stages in receptive grammar acquisition: A processability theory account. *Second language research, 31*(4), 523-550.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd Ed.). New York: Routledge.

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

Carrol, J. B., & Sapon, S. M. (1959). *Modern language aptitude test: MLAT*. New York: Psychological Corporation.

Carroll, J. B. (1958). A factor analysis of two foreign language aptiude batteries. *The journal of general psychology, 59*(1), 3-19.

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

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd Ed.). Boston: Heinle & Heinle Publishers.

Chimpololo. (2010). Assessing pupils' language proficiency in a bilingual education programme: Insights from two primary schools in Malawi. *The Buckingham journal of language and linguistics, 3*(1), 1-10.

Chiuye, G. M. (2005). The promise and perils of the Malawi School Certificate of Education Examinations: An investigation of cognitive structures of questions on English papers (doctoral dissertation, the University of Akron).

Chiyembekezo, R., Kondowe, W., & Ngwira, F. F. (2019). Teaching English speaking skills to L2 students using role-play method: What do teachers in Malawi say? *Researchers world-journal of arts, science & commerce, 10*(1), 12-20.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Cohen, J. (1989). *Statistical power analysis for the behavioral sciences* (2nd Ed.). New York: Lawrence Erlbaum Associates.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The modern language journal, 93*(3), 336-353.

Collins, P. C. (1991). *Cleft and pseudo-cleft constructions in English*. London: Routledge.

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*(3), 356-371.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in cognitive sciences, 5*(12), 539-546.

Dalrymple, M. (2001). *Lexical-functional grammar*. Wiley Online Library. Retrieved from users.ox.ac.uk › ~cpgl0015 › lfg

De Graaff, R. (1997). The experanto experiment: Effects of explicit instruction on second language acquisition. *Studies in second language acquisition, 19*(2), 249-276.

DeKeyser, R. M. (1995). Learning second language grammar rules. *Studies in second language acquisition, 17*(03), 379-410.

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: Cambridge University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(4), 499-533.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 312-348). Oxford, UK: Blackwell.

DeKeyser, R. M. (2005). What makes learning second language grammar difficult? A review of issues. *Language learning, 55*(S1), 1-25.

DeKeyser, R. M. (2007). Skill acquisition theory. In B. V. J. Williams (Ed.), *Theories in second language acquisition: An introduction* (pp. 97-114). Mahwah, NJ: Erlbaum.

DeKeyser, R. M. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in second language acquisition, 38*(2), 353-363.

Denhovska, N., Serratrice, L., & Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: The role of frequency and working memory. *Language learning*, *66*(1), 159-190.

Dornyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah N.J: Lawrence Erlbaum Associates.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA review, 19*(1), 42-68.

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). Oxford: Blackwell Publishing Ltd.

Doughty, C. J., & Long, M. H. (2003). The scope of inquiry and goals of SLA. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 3-15). Oxford: Blackwell Publishing Ltd.

Eichenbaum, H. (2002). *The cognitive neuroscience of memory: An introduction.* New York: Oxford University Press.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press.

Elder, C., & Manwaring, D. (2004). The relationship between metalinguistic knowledge and learning outcomes among undergraduate students of Chinese. *Language awareness, 13*(3), 145-162.

Ellis, N. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). Amsterdam: John Benjamins Publishing Company.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in second language acquisition, 27*(2), 305-352.

Ellis, N. C. (2008). Implicit and explicit knowledge about language. In J. Cenoz, & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., pp. 1878-1890). New York: Springer Science.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL quarterly, 27*(1), 91-113.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in second language acquisition, 24*(2), 223-236.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language learning, 54*(2), 227-275.

Ellis, R. (2005a). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*, *27*(2), 141-172.

Ellis, R. (2005b). Principles of instructed language learning. *System, 33*(2), 209-224.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied linguistics, 27*(3), 431-463.

Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-26). Bristol: Multilingual Matters.

Ellis, R. (2009b). Investigating learning difficulty in terms of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 143-166). Bristol: Multilingual Matters.

Ellis, R. (2009c). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31-64). Bristol: Multilingual Matters.

Ellis, R. (2015a). Form-focused instruction and the measurement of implicit and explicit L2 knowledge. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 417-441). Amsterdam: John Benjamins Publishing Company.

Ellis, R. (2015b). Researching acquisition sequences: Idealization and de-idealization in SLA. *Language learning, 65*(1), 181-209.

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in second language acquisition, 29*(1), 119-126.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition, 28*(2), 339-368.

Ellis, R., & Roever, C. (2018). The measurement of implicit and explicit knowledge. *The language learning journal,* 1-16.

Epstein, S. D., Flynn, S., Martohardjono, G., Flynn, S., Martohardjono, G., & O’Neil, W. (1998). The strong continuity hypothesis: Some evidence concerning functional categories in adult L2 acquisition. In F. S. Matohardjono, & W. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 67-77). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and brain sciences, 19*(4), 677-714.

Erçetin, G., & Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied psycholinguistics, 34*(4), 727-753.

Erlam, R. (2003a). The effects of deductive and inductive instruction on the acquisition of direct object pronouns in French as a second language. *The modern language journal, 87*(2), 242-260.

Erlam, R. (2003b). Form-focused instruction in L2 French (Doctoral dissertation, University of Auckland).

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language teaching research, 9*(2), 147-171.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied linguistics, 27*(3), 464-491.

Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. 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. 65-93). Bristol: Multilingual Matters Bristol.

Ettlinger, M., Bradlow, A. R., & Wong, P. C. (2014). Variability in the learning of complex morphophonology. *Applied psycholinguistics, 35*(04), 807-831.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. (2015). The relationship between artificial and second language learning. *Cognitive science,*1-26.

Eubank, L. (1993). On the transfer of parametric values in L2 development. *Language acquisition, 3*(3), 183-208.

Eubank, L. (1994). On the transfer of parametric values in L2 development. *Language acquisition, 3*(3), 183-208.

Evans, J. S. B. (2003). In two minds: Dual-process accounts of reasoning. *Trends in cognitive sciences, 7*(10), 454-459.

Faretta-Stutenberg, M. (2014). Individual differences in context: A neurolinguistic investigation of working memory and L2 development (doctoral dissertation, University of Illinois at Chicago).

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second language research, 34*(1), 67-101.

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., . . . Elger, C. E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science, 285*(5433), 1582-1585.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles: SAGE Publications Ltd.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Ed.). Los Angeles: SAGE Publications Ltd.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation.* Boston: Allyn & Bacon.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the national academy of sciences, 103*(31), 11778-11783.

Gass, S. M., & Mackey, A. (2012). *The Routledge handbook of second language acquisition.* London: Routledge Taylor & Francis Group.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of experimental psychology: learning, memory, and cognition, 33*(1), 34.

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in second language acquisition, 37*(2), 269-297.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language learning, 51*(1), 1-50.

Granena, G. (2012). Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment (Doctoral dissertation, University of Maryland).

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-130). Amsterdam: John Benjamins Publishing Company.

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language learning, 63*(4), 665-703.

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied linguistics, 35*(4), 483-503.

Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied psycholinguistics, 37*(3), 577-600.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second language research, 29*(3), 311-343.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics* (1988 reprint edition Ed.). Los Altos Hills, CA: Peninsula Publ.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and individual differences, 38*, 44-53.

Grigornko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *The modern language journal, 84*(3), 390-405.

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in second language acquisition, 35*(3), 423-449.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). New York: Pearson.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis: Pearson new international edition* (7th Ed.). Edinburgh: Pearson Education Limited.

Hamrick, P. (2013). Development of conscious knowledge during early incidental learning of L2 syntax (Doctoral dissertation, Georgetown University).

Hamrick, P. (2014a). Recognition memory for novel syntactic structures. *Canadian journal of experimental psychology/Revue Canadienne de psychologie expérimentale, 68*(1), 2-7.

Hamrick, P. (2014b). A role for chunk formation in statistical learning of second language syntax. *Language learning, 64*(2), 247-278.

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and individual differences, 44*, 9-15.

Hamrick, P., Lum, J. A., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences, 115*(7), 1487-1492.

Han, Y. (1996). L2 learners' explicit knowledge of verb complement structures and its relationships to L2 implicit knowledge (Doctoral dissertation, Temple University). Retrieved from https://search.proquest.com/docview/304268322.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language teaching research, 2*(1), 1-23.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in second language acquisition, 19*(3), 379-400.

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

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior research methods, instruments & computers, 27*(1), 46-51.

Helke, M. (1970). The grammar of English reflexives (Doctoral dissertation, Massachusetts Institute of Technology).

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association, 82*(400), 1147-1149.

Hooker, C. A. (1994). Idealisation, naturalism, and rationality: Some lessons from minimal rationality. *Synthese, 99*(2), 181-231.

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in second language acquisition, 38*(2), 163-175.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA.* Amsterdam: John Benjamins Publishing Company.

Howard Jr, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and aging, 12*(4), 634-656.

Howell, D. C. (2012). *Statistical methods for psychology*. Australia: Wadsworth Cengage Learning.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language testing, 19*(3), 227-245.

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). Oxford, UK: Blackwell Publishing.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in second language acquisition, 27*(2), 129-140.

Hulstijn, J. H., & Graaff, R. d. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA review, 11*, 97-112.

IBM Corp. (2013). *SPSS statistics: Version 22.0* [computer software]. Armonk, NY: IBM Corp.

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London Task: Psychometric properties of a structurally balanced problem set. *Psychological assessment, 24*(1), 46-53.

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural brain research, 216*(1), 1-8.

Kamwendo, G. H. (2003). Is Malawi guilty of spoiling the queen's language? *English today, 19*(02), 30-33.

Kamwendo, G. H. (2016). The new language of instruction policy in Malawi: A house standing on a shaky foundation. *International review of education, 62*(2), 221-228.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321-340.

Kayambazinthu, E. (1998). The language planning situation in Malawi. *Journal of multilingual and multicultural development, 19*(5), 369-439.

Keating, P. (2005). D-prime (signal detection) analysis. Retrieved from URL http://www.linguistics.ucla.edu/faciliti/facilities/statistics/dprime.htm

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language learning, 58*(4), 703-746.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental psychology, 48*(1), 171.

Kidd, E., & Kirjavainen, M. (2011). Investigating the contribution of procedural and declarative memory to the acquisition of past tense morphology: Evidence from Finnish. *Language and cognitive processes, 26*(4-6), 794-829.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & memory, 1*(2), 106-120.

Krashen, S. D. (1981). *Second language acquisition and second language learning* (1st Ed.). Oxford: Oxford University Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition* (1st Ed.). Oxford: Pergamon.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Oxford: Pergamon.

Krashen, S. D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Kunje, D., Selemani-Meke, E., & Ogawa, K. (2009). An investigation of the relationship between school and pupil characteristics and achievement at the basic education level in Malawi. *Journal of International Cooperation in Education, 12*(1), 33-49.

LaBrozzi, R. M. (2009). Processing of lexical and morphological cues in a study abroad context (doctoral dissertation, the Pennsylvania State University, University Park).

LaBrozzi, R. M. (2012). The role of study abroad and inhibitory control on processing redundant cues. In K. Greeslin, & M. Diaz-Campos (Eds.), *Selected proceedings of the 14th hispanic linguistics symposium* (pp. 228-241). Somerville, MA: Cascadilla Proceedings Project.

Langacker, R. W. (1995). Possession and possessive constructions. In J. R. Taylor, & R. E. MacLaury (Eds.), *Language and the cognitive construal of the world* (82nd ed., pp. 51-79). Berlin: Mouton de Gruyter.

Larrabee, G. J., Trahan, D. E., & Curtiss, G. (1992). Construct validity of the continuous visual memory test. *Archives of clinical neuropsychology: The official journal of the national academy of neuropsychologists, 7*(5), 395-405. DOI: 0887-6177(92)90152-D [pii]

Larsen-Hall, J. (2010). *A guide to doing statistics in second language research*. New York: Rutledge.

Leow, R. P. (2015). Implicit learning in SLA: Of processes and products. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 47-66). Amsterdam: John Benjamins Publishing Company.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in second language acquisition, 38*(4), 801-842.

Li, S. (2017). The effects of cognitive aptitudes on the process and product of L2 interaction. In L. Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 41-70). Amstardam: John Benjamins Publishing Company.

Li, S. (2018). Language aptitude. In A. Burns, & J. Richards (Eds.), *The Cambridge guide to learning English as a second language* (pp. 63-72). Cambridge: Cambridge University Press.

Li, S. (2019). Six decades of language aptitude research: A comprehensive and critical review. In Z. E. Wen, P. Skehan, A. Biedron, S. Li & R. L. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research and practice* (pp. 78-96). New York: Routledge.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., . . . Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language learning, 63*(3), 530-566.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic bulletin & review, 21*(4), 861-883.

Long, M. H. (2007). Age differences and sensitive periods controversy in SLA. In M. H. Long (Ed.), *Problems in SLA* (pp. 43-74). Mahwah, New Jersey: Laurence Erlbaum Associates, Publishers.

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The modern language journal, 82*(3), 357-371.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of experimental psychology: Animal behavior processes*, *28*(1), 3.

Lum, J., & Kidd, E. (2012). An examination of the associations among multiple memory systems, past tense, and vocabulary in typically developing 5-year-old children. *Journal of speech, language, and hearing research, 55*(4), 989-1006.

Lum, J., Kidd, E., Davis, S., & Conti-Ramsden, G. (2010). Longitudinal study of declarative and procedural memory in primary school-aged children. *Australian journal of psychology, 62*(3), 139-148.

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

Macmillan, & Creelman. (2005). *Detection theory: A user's guide* (2nd Ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Maganga, J., Mwale, L., Mapondera, A., & Saka, T. (2010). *Learning achievement of standards 3 and 7 learners in Malawi.* Zomba: Malawi Institute of Education.

Manns, J. R., & Squire, L. R. (2001). Perceptual learning, awareness, and the hippocampus. *Hippocampus, 11*(6), 776-782.

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*(3), 379-413.

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

McLaren, I. P., Forrest, C., McLaren, R. P., Jones, F. W., Aitken, M., & Mackintosh, N. J. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology of learning and memory, 108,* 185-195.

Meara, P. (2005). *LLAMA language aptitude tests: The manual*. Swansea, UK: Lognostics.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für sprachwissenschaft, 28*(1), 5-34.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and brain sciences, 32*(2), 183-198.

Miyake, A., & Friedman, N. P. (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, New Jersey: Lawrence Erlbaum Associates, Inc.

Mizrachi, A., Padilla, O., & Susuwele-Banda, W. (2010). *Active-learning pedagogies as a reform initiative: The case of Malawi.* Washington, DC: American Institute for Research.

Mizumoto, A. (2015). Langtest (version 1.0) [web application] [computer software], http://langtest.jp/

Mmela, E. (2006). Implementing integrated literacy approaches in an English classroom in Malawi (Doctoral dissertation, Virginia Polytechnic Institute and State University).

Malawi Institute of Education. (2004a). Syllabus for literacy and languages – English Standard 7. In Malawi Institute of Education, *Malawi primary school syllabuses Standard* 7 (pp. 54-104). Domasi: Malawi Institute of Education.

Malawi Institute of Education. (2004b). Syllabus for literacy and languages – English Standard 8. In Malawi Institute of Education, *Malawi primary school syllabuses Standard 8* (pp. 51-118). Domasi: Malawi Institute of Education.

Malawi Institute of Education. (2013a). *Republic of Malawi syllabus for English Forms 1 and 2*. Domasi: Malawi Institute of Education.

Malawi Institute of Education. (2013b). *Republic of Malawi syllabus for English Forms 3 and 4*. Domasi: Malawi Institute of Education.

MoEST. (2015). *National education standards: Primary and secondary education*. Lilongwe: MoEST.

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

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

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

Morgan-Short, K., & Ullman, M. T. (2012). The neurocognition of second language. In Gass, Susan M and Mackey, Alison (Ed.), *The Routledge handbook of second language acquisition* (pp. 282-299). New York: Routledge.

Narayanan, A. (2012). A review of eight software packages for structural equation modeling. *The American statistician, 66*(2), 129-138.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology, 19*(1), 1-32.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning, 50*(3), 417-528.

Ouellet, M., Beauchamp, M. H., Owen, A. M., & Doyon, J. (2004). Acquiring a cognitive skill with a new repeating version of the Tower of London Task. *Canadian journal of experimental psychology/Revue Canadienne de psychologie expérimentale, 58*(4), 272.

Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th Ed.). Berkshire UK: McGraw-Hill Education.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic approaches to bilingualism, 1*(3), 213-237.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implication for bilingualism and second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). London: Academic Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amstardam: John Benjamins Publishing Commpany.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins Publishing.

Phakiti, A. (2018a). Confirmatory factor analysis and structural equation modeling. In A. Phakiti, P. De Costa, L. Plonsky & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 459-500). London: Springer Nature Limited.

Phakiti, A. (2018b). Exploratory factor analysis. In A. Phakiti, P. De Costa, L. Plonsky & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 423-457). London: Springer Nature Limited.

Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied linguistics, 10*(1), 52-79.

Pienemann, M. (1998a). Developmental dynamics in L1 and L2 acquisition: Processability theory and generative entrenchment. *Bilingualism: Language and cognition*, *1*(01), 1-20.

Pienemann, M. (1998b). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins Publishing Company.

Pienemann, M. (2005). Discussing PT. In M. Pienemann (Ed.), *Cross-linguistic aspects of processability theory* (pp. 61-83). Amsterdam: John Benjamins Publishing Company.

Pienemann, M. (2015). An outline of processability theory and its relationship to other approaches to SLA. *Language learning, 65*(1), 123-151.

Pienemann, M., Di Biase, B., & Kawaguchi, S. (2005). Extending processability theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of processability theory* (pp. 199-251). Amsterdam: John Benjamins Publishing Company.

Pienemann, M., & Keßler, J. (2012). Processability theory. In S. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 228-247). New York: Routledge.

Pienemann, M., & Lenzing, A. (2015). Processability theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 159-179). New York: Routledge Taylor & Francis.

Poletiek, F. H. (2002). Implicit learning of a recursive rule in an artificial grammar. *Acta psychologica*, *111*(3), 323-335.

Pothos, E. M. (2005). Expectations about stimulus structure in implicit learning. *Memory & cognition*, *33*(1), 171-181.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological bulletin, 133*(2), 227-244.

Pothos, E. (2010). An entropy model for artificial grammar learning. *Frontiers in psychology*, *1*(16), 1-12.

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.

Pretz, J. E., Totz, K. S., & Kaufman, S. B. (2014). The effects of mood, cognitive style, and cognitive ability on implicit learning: Corrigendum. *Learning and individual differences*, *20*(3), (2010) 215–219.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., & Crystal, D. (1985). *A comprehensive grammar of the English language*. London: Longman.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, *6*(6), 855-863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General, 118*(3), 219.

Reber, A. S. (1992). The cognitive unconscious: An evolutionary perspective. *Consciousness and cognition, 1*(2), 93-133.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.

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*(5), 888.

Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral neuroscience, 110*(5), 861.

Rebuschat, P. (2009). Implicit learning of natural language syntax (Doctoral dissertation, University of Cambridge).

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language learning*, *63*(3), 595-626. DOI: 10.1111/lang.12010

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in second language acquisition, 37*(2), 299-334.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar: Handbook in generative syntax* (pp. 281-337). Dordrecht: Kluwer.

Rizzi, L., & Shlonsky, U. (2006). Satisfying the subject criterion by a non-subject: English locative inversion and heavy NP shift. In Frascarelli (Ed.), *Phases of interpretation* (pp. 341-361). Berlin: Mouton de Gruyter.

Robinson, P. (2003). Attention and memory during SLA. In L. C. Daughty, & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631-678). Oxford: Blackwell.

Robinson, P. (1995). Attention, memory, and the “noticing” hypothesis. *Language learning, 45*(2), 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*(01), 27-67.

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

Robinson, P. (2002a). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam: John Benjamins Publishing Company.

Robinson, P. (2002b). Learning conditions, aptitude complexes, and SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113-133). Amsterdam: John Benjamins Publishing Company.

Robinson, P. (2005a). Aptitude and second language acquisition. *Annual review of applied linguistics, 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*(2), 235-268.

Roehr, K. (2006). Metalinguistic knowledge in L2 task performance: A verbal protocol analysis. *Language awareness, 15*(3), 180-198.

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

Rosenbach, A. (2004). The English *s*-genitive: A case of degrammaticalization? In O. Fischer, M. Norde & H. Perridon (Eds.), *Up and down the cline – The nature of grammaticalization* (pp. 73-96). Amsterdam: John Benjamins Publishing Company.

Ross, S., Yoshinaga, N., & Sasaki, M. (2002). Aptitude-exposure interaction effects on *Wh*-movement violation detection by pre-and-post-critical period. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 266-299). Philadelphia: John Benjamins Publishing Company.

Rotello, C. M., Masson, M. E., & Verde, M. F. (2008). Type I error rates and power analyses for single-point sensitivity measures. *Perception & psychophysics, 70*(2), 389-401.

Safir, K. (2004). *The syntax of anaphora*. Oxford: Oxford University Press.

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). London: Oxford University Press.

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

Saville-Troike, M., & Barto, K. (2017). *Introducing second language acquisition* (3rd Ed.). London: Cambridge University Press.

Schacter, D. L., & Wagner, A. D. (1999). Remembrance of things past. *Science, 285*(5433), 1503-1504.

Schiff, R., & Katan, P. (2014). Does complexity matter? Meta-analysis of learner performance in artificial grammar tasks. *Frontiers in psychology, 5*(1084), 1-10.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics, 11*(2), 129-158.

Schmidt, R. W., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Schoonen, R. (2015). Structural equation modeling in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 213-242). New York: Routledge.

Schoorlemmer, M. (1998). Possessors, articles, and definiteness. In A. Alexiadou, & C. Wilder (Eds.), *Possessors, predicates, and movement in the determiner phrase* (pp. 55-86). Amsterdam: John Benjamins Publishing Company.

Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (4th Ed.). New York: Routledge.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological research, 61*(3), 175-190.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research, 12*(1), 40-72.

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

Selinker, L., & Gass, S. (2008). *Second language acquisition: An introductory course*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied linguistics, 2*(2), 159-168.

Silva, L. H. R., & Roehr-Brackin, K. (2016). Perceived learning difficulty and actual performance: Explicit and implicit knowledge of L2 English grammar points among instructed adult learners. *Studies in second language acquisition, 38*(2), 317-340.

Singleton, D. M., & Ryan, L. (2004). *Language acquisition: The age factor* (2nd Ed.). New York: Multilingual Matters Ltd.

Skehan, P. (1991). Individual differences in second language learning. *Studies in second language acquisition, 13*(2), 275-298.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-94). Philadelphia: John Benjamins Publishing Company.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of experimental psychology: General, 117*(1), 34.

Spada, N., Shiu, J. L., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language learning, 65*(3), 723-751. DOI:10.1111/lang.12129

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A Meta-Analysis. *Language learning, 60*(2), 263-308.

Squire, L. R. (1992). Declarative and non-declarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience, 4*(3), 232-243.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of learning and memory, 82*(3), 171-177.

Squire, L. R., & Knowlton, B. J. (1994). Memory, hippocampus, and brain systems. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 825-837). Cambridge: MIT Press.

Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. Gazzaninga (Ed.), *The new cognitive neurosciences* (pp. 756-776). Cambridge: MIT Press.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu.rev.neurosci, 27,* 279-306.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the national academy of sciences, 93*(24), 13515-13522.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers, 31*(1), 137-149.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In K. Frankish, & J. S. B. Evans (Eds.), *In two minds: Dual processes and beyond* (pp. 55-88). Oxford: Oxford University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford: Oxford University Press.

Strong, C. H., & Donders, J. (2008). Validity of the continuous visual memory test (CVMT) after traumatic brain injury. *Journal of clinical and experimental neuropsychology, 30*(8), 885-891.

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

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied psycholinguistics, 38*(5), 1229-1261.

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language learning, 65*(4), 860-895.

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language learning, 67*(4), 747-790.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th Ed.). Boston, MA: Allyn & Bacon.

Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2015). Working memory, learning conditions and the acquisition of L2 syntax. In Z. E. Wen, M. B. Mota & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 224-247). Bristol: Multilingual Matters.

Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. *Proceedings of the annual meeting of the cognitive science society, 33*(33), 2061-2066.

Tagarelli, K. M., Ruiz, S., Vega, J. L. M., & Rebuschat, P. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in second language acquisition, 38*(2), 293-316.

The psychological corporation. (1995). *DAT for selection-verbal reasoning (European adaptation)*. Orlando, FL: Harcourt Brace and Company.

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

Trahan, D. E., & Larrabee, G. J. (1988). *Continuous visual memory test: Professional manual*. Odessa, FL.: Psychological Assessment Resources.

Trahan, D. E., Larrabee, G. J., & Quintana, J. W. (1990). Visual recognition memory in normal adults and patients with unilateral vascular lesions. *Journal of clinical and experimental neuropsychology, 12*(6), 857-872.

Tulving, E. (2000). Introduction to memory. In M. Gazzaninga (Ed.), *The new cognitive neurosciences* (pp. 727-732). Cambridge: MIT Press.

Ullman, M. T. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of psycholinguistic research, 30*(1), 37-69.

Ullman, M. T. (2001b). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition, 4*(2), 105-122.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1-2), 231-270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language 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.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 135-158). New York: Routledge Taylor & Francis Group.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge and use. In G. Hickok, & S. A. Small (Eds.), *The neurobiology o f language* (pp. 953-968). San Diego, CA: Academic Press.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of cognitive neuroscience, 9*(2), 266-276.

Unterrainer, J. M., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive brain research, 17*(3), 675-683.

Urry, K., Burns, N. R., & Baetu, I. (2015). Accuracy-based measures provide a better measure of sequence learning than reaction time-based measures. *Frontiers in psychology, 6*, 1158.

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in second language acquisition, 39*(1), 59-95.

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X’-theory: Evidence from Korean and Turkish adults learning German. *Language acquisition studies in generative grammar, 31*(4), 71-89.

Van den Bos, E., & Poletiek, F. H. (2015). Learning simple and complex artificial grammars in the presence of a semantic reference field: Effects on performance and awareness. *Frontiers in psychology, 6* (158), 1-7.

Vasterling, J. J., Duke, L. M., Brailey, K., Constans, J. I., Allain Jr, A. N., & Sutker, P. B. (2002). Attention, learning, and memory performances and intellectual resources in Vietnam veterans: PTSD and no disorder comparisons. *Neuropsychology, 16*(1), 5.

Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of *d* ', A_z, and *A*'. *Perception & psychophysics, 68*(4), 643-654.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wall, M. (2012). How to analyse reaction time (RT) data: Part 1. Computing for psychologists. Retrieved from https://goo.gl/2XBYFr

Whelan, R. (2008). Effective analysis of reaction time data. *The psychological record, 58*(3), 475-482.

White, L. (2003). *Second language acquisition and universal grammar*. London: Cambridge University Press.

White, L. (2015). Linguistic theory, universal grammar, and second language acquisition. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition* (pp. 34-53). New York: Routledge.

Whittlesea, B. W., & Price, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & cognition*, *29*(2), 234-246.

Williams, B. A., & Pearlberg, S. L. (2006). Learning of three-term contingencies correlates with raven scores, but not with measures of cognitive processing. *Intelligence*, *34*(2), 177-191.

Williams, J. N. (2012). Working memory and SLA. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 445-459). New York: Routledge.

Winke, P. (2005). Individual differences in adult Chinese second language acquisition: The relationships among aptitude, memory and strategies for learning (Doctoral dissertation, Georgetown University).

Yalçın, Ş, & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom-based study. *Studies in second language acquisition, 38*(2), 239-263.

Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PloS one, 7*(8), e42993.

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in second language acquisition, 37*(3), 457-486.

Ziori, E., & Pothos, E. (2015). Artificial grammar learning: An introduction to key issues and debates. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 247-274). Amsterdam: John Benjamins Publishing Company.

APPENDIX A: BACKGROUND QUESTIONNAIRE/ BIODATA FORM

PERSONAL BACKGROUND QUESTIONNAIRE

With this questionnaire, I would like to get an impression of the personal background and English language use of primary school, secondary school and university learners in Malawi. It consists of 29 items. Please note that not all items may apply to you personally. Should you think that a certain item does not apply to you (for example when you are asked about the language use of your children and you don't have any children), you may cross out the number in front of that question and move on to the next. It is important that you answer these questions on your own, because I am interested in *your* language use. If you don't understand a certain question, please do not hesitate to ask. There are no right or wrong answers!

1) **What is your name?** ....
2) What is your date of birth? ……………………………………………………
3) Are you: ☐ Male ☐ Female ?
4) Where were you born? Village/Town: ....District: ……..Country: ....
5) What nationality do you have? ☐ Malawian ☐ Other, namely ....
6) Have you ever lived in a country other than Malawi for a longer period (that is, more than 6 months)? ☐ no ☐ yes, in: (town)………(country)………………
7) If you have indicated that you have ever lived in a country other than Malawi, did you attend any English classes in an educational environment, like a school or some similar institution: ☐ no ☐ yes, less than 1 month ☐ yes, less than 3 months
   ☐ yes, less than 6 months ☐ yes, less than 1 year ☐ yes, more than 1 year
8) What language(s) did you acquire before starting school? (i)............... (ii)............
9) What language or languages have you learned at school? (i)................(ii)..............
10) What language or languages did you learn outside of an educational environment (so outside of school or work)? (i)................... (ii)....................
    (iii).............................
11) Have you repeated any school year (i.e. class)? ☐ no ☐ yes, namely ....
12) Do you ever go to church? ☐ no, never ☐ yes, sometimes ☐ yes, regularly
13) If you have indicated you go to church, could you please indicate in which language the services are held? ☐ English ☐ Chichewa ☐ Other, namely ....
14) In general, how would you rate your English language proficiency before school?
    ☐ none ☐ very bad ☐ bad ☐ sufficient ☐ good ☐ very good

15) In general, how would you rate your English language proficiency at present?

☐ none ☐ very bad ☐ bad ☐ sufficient ☐ good ☐ very good

16) How often do you speak English?

☐ rarely ☐ few times a year ☐ monthly ☐ weekly ☐ daily

17) What is your current marital status?

☐ married ☐ separated/divorced ☐ widow/widower ☐ with partner ☐ single

18) With what language(s) was your (ex) partner brought up?

☐ English ☐ Chichewa ☐ Other, namely: ....

19) What language or languages do you mostly use when talking to your (ex) partner?

☐ only English ☐ only Chichewa ☐ only other, namely.........................

☐ English, Chichewa and other, but mostly English

☐ English, Chichewa and other, without preference

☐ English, Chichewa and other, but mostly Chichewa and/or other

☐ Chichewa and other, without preference

☐ Chichewa and other, but mostly Chichewa

☐ Chichewa and other, but mostly other

20) What language or languages does your (ex)partner mostly use when talking to you?

☐ only English ☐ only Chichewa ☐ only other, namely........................

☐ English, Chichewa and other, but mostly English

☐ English, Chichewa and other, without preference

☐ English, Chichewa and other, but mostly Chichewa and/or other

☐ Chichewa and other, without preference

☐ Chichewa and other, but mostly Chichewa

☐ Chichewa and other, but mostly other

21) Do you have children? ☐ no ☐ yes, number: ....

22) What language or languages do you mostly use when talking to your children?

☐ only English ☐ only Chichewa ☐ only other, namely........................

☐ English, Chichewa and other, but mostly English

☐ English, Chichewa and other, without preference

☐ English, Chichewa and other, but mostly Chichewa and/or other

☐ Chichewa and other, without preference

☐ Chichewa and other, but mostly Chichewa

☐ Chichewa and other, but mostly other

23) What language or languages do your children mostly use when talking to you?

☐ only English ☐ only Chichewa ☐ only other, namely.......................

☐ English, Chichewa and other, but mostly English

☐ English, Chichewa and other, without preference

☐ English, Chichewa and other, but mostly Chichewa and/or other

☐ Chichewa and other, without preference

☐ Chichewa and other, but mostly Chichewa

☐ Chichewa and other, but mostly other

24) Could you, in the following table, please indicate to what extent you use English in the domains provided? You may simply tick the box.

I speak English	all the time	frequently	sometimes	rarely	very rarely
With relatives					
With friends					
At school					
At work					
In church					
Other					

25) Do you ever listen to English songs? ☐ yes ☐ no

26) Do you ever watch English television programmes?

☐ yes ☐ no ☐ I would love to, but I can't get them

27) Do you ever listen to English radio programmes?

☐ yes ☐ no ☐ I would love to, but I can't get them

28) Do you ever read English newspapers, novels or magazines? ☐ yes ☐ no

29) You have come to the end of this questionnaire. Is there anything you would like to add? This can be anything from language-related comments to remarks about the questionnaire or research itself.

…………………………………………………………………………………………

……...……………………………………………………………………………………

……

Thank you for your participation

Contacts: Mr. G. Tchesa, gtch098@aucklanduni.ac.nz;
0999180720/+64223917805
Mr. Wongani Nyondo, 0884322677

APPENDIX B: TEST MATERIALS

Grammaticality Judgment Test Sentences

Participle *-ing*

Grammatical

1. Alice and Chatha were making noise when you entered.
2. The light went out while Takondwa was reading.
3. We were watching TV all night.

Ungrammatical

4. While Chisisi was cook, the phone rang.
5. When Alice saw him, he was played chess.
6. She is drive to work at the moment.

Past *-ed*

Grammatical

1. Peter travelled to Lilongwe two weeks ago.
2. Jane studied French last year.
3. Alice baked Jane a cake last night.

Ungrammatical

4. John live abroad five years ago.
5. Alice completed her assignment and print it out
6. Chisisi fetch the dog some water yesterday evening.

Plural *–s* (without determiner 2 (1gr-1ungr); with determiner 4 (2gr-2ungr))\

Grammatical

1. Many students come to school late nowadays.
2. Girls normally do better in school than boys do.
3. Thirty apples in the basket are rotten.

Ungrammatical

4. Alice sold a few old coins and stamp.
5. River are not as big as lakes are.
6. Several country in this region were still very poor.

Adverb placement

Grammatical

1. Chatha completely rejected this proposal.
2. Chisisi drives her boat slowly.
3. Peter has quietly asked me to leave the house.

Ungrammatical

4. That girl eats greedily anything since the accident.
5. Chatha writes very well English.
6. Alice baked John carefully a cake.

Structural parallelism

Grammatical

1. Not only do I hate chili, but I also hate beans.
2. Students must either write a report or read another book.
3. Jerry is neither rich nor famous.

Ungrammatical

4. Tamara neither enjoys drinking nor singing.
5. Either Chisisi got the perfect score in English or in Mathematics.
6. Alice cooked not only meat for Jane but also for Chatha.

3rd person singular –*s*

Grammatical

1. Peter doesn’t like vegetables.
2. Most intelligent students plan their compositions.
3. My dog chases that black cat.

Ungrammatical

4. Takondwa live in a small house in Matawale.
5. The employees enjoys the new building.
6. One student don’t understand these assignments.

Possessive –*s*

Grammatical

1. The car's front seat lacks a belt.
2. The lions' usual source of water has dried up.
3. John’s bed looks very small.

Ungrammatical

4. Alice is still living in his rich uncle house.
5. That boy plans on stealing Peter car.
6. The girls books are on the table.

Possessive determiner *her/his*

Grammatical

1. Mary always stays at home because her uncle is very strict.
2. Maria schools here while her brother schools at Mnsanjama.
3. Charles plays a guitar. His girlfriend plays a flute.

Ungrammatical

4. The boy’s name is James and her sister is Maria.
5. The young man was crying while her mother was watching TV.
6. She is Mercy. His brother is Peter.

Definite article *the*

Grammatical

1. I have a cat. The cat is black
2. The phone on my desk belongs to Ken.
3. The moon is very bright tonight.

Ungrammatical

4. They had the very good time at the party
5. A man who lives next door is Chinese.
6. A sun rises in the east.

Locative Inversion

Grammatical

1. Through the forest went Jane.
2. On Christmas day Chatha phoned.
3. Round the track Alice raced.

Ungrammatical

4. On the boat Alice was.
5. On Saturday night danced Takondwa.
6. In the evening studied Jane.

Wh-question

Grammatical

1. What does Jennifer love?
2. What did the manager leave on the table?
3. What do Mable and Alice sing at karaoke every Sunday night?

Ungrammatical

4. What Chisisi and Chatha bought?
5. Travellers take what to their destinations?
6. Where daddy is going?

Pseudo-cleft

Grammatical

1. Where the car was was in the carpark not in the road.
2. Where the apples were was in the basket.
3. Where Mr. Zimba teaches is at the university.

Ungrammatical

4. Where is Alice is in the country not in the city.
5. Where cooked Jane is in the kitchen not in the bathroom.
6. Where Carl writes are at a desk not on the floor.

Dative alternation

Grammatical

1. Jennifer cooked Peter dinner.
2. Some students reported the problem to the police.
3. My father bought ten pens for my brother.

Ungrammatical

4. The teacher described the students the picture.
5. That driver collected John the money.
6. Alice donate the museum a painting.

Passive

Grammatical

1. The entire house was painted by Tom.
2. These cars are made in Japan.
3. The letter is written by Alice.

Ungrammatical

4. Dinner cooks by Charles nowadays.
5. My bags take into the house by Jennifer.
6. The video posted on Facebook by Alex.

UGJT sample test sentences

Sentence 1: I am believing your story.

Correct Incorrect

a. This sentence is grammatically correct / incorrect. ☐ ☐
(Tick one box.)

b. How certain are you of the judgement you just made? ☐ %

not certain totally
at all certain
0 -----------------------------------100

c. The main way I decided was: (Tick one box.) Rule Feel
A. I used a rule. (by rule) ☐ ☐
B. It just sounded right/wrong. (by feel)

Sentence 2: My father is a doctor.

Correct Incorrect

a. This sentence is grammatically correct / incorrect. ☐ ☐
(Tick one box.)

b. How certain are you of the judgement you just made? ☐ %

c. The main way I decided was: (Tick one box.)
A. I used a rule. (by rule) Rule Feel
B. It just sounded right/wrong. (by feel) ☐ ☐

Sentence 3: My teacher made me to work hard.

Correct Incorrect

a. This sentence is grammatically correct / incorrect. ☐ ☐
(Tick one box.)

b. How certain are you of the judgement you just made? ☐ %

c. The main way I decided was: (Tick one box.)
A. I used a rule. (by rule) Rule Feel
B. It just sounded right/wrong. (by feel) ☐ ☐

Sentence 4: He is marrying a girl who she lives next door. Correct Incorrect

a. This sentence is: ☐ ☐

b. How certain are you? ☐ %

Rule Feel

c. The main way I decided was: ☐ ☐

Sentence 5: Moon was shining brightly**.** Correct Incorrect

a. This sentence is: ☐ ☐

b. How certain are you? ☐ %

c. The main way I decided was: Rule Feel

☐ ☐

Sentence 6: My father gave me some advices. Correct
Incorrect

a. This sentence is: ☐ ☐

b. How certain are you? ☐ %

c. The main way I decided was: Rule Feel

☐ ☐

If you have any questions about how to do this test, please ask one of the supervisors now. You can take as much time as you need for each item.
Once you have completed an item, please DO NOT go back and change your answer.

Elicited Oral Imitation Test

Practice Sentences

1 We drove to the park.
2 I'll call her tomorrow night.
3 You can buy meat at the butcher shop.
4 My brother just bought a plot of land.
5 Sometimes they take their dog for a walk.
6 We're going to play football this afternoon.

Test Sentences

Participle -*ing*

Grammatical

Every family is eating supper now.

Ungrammatical

I was come from school yesterday when it started to rain.

My teacher is now meet all the students in the class.

Past -*ed*

Grammatical

Malawians elected a new president.

Ungrammatical

A long time ago people walk on foot.

When man invented the motor car, life change for everyone.

Plural –*s* (without determiner 2 (1gr-1ungr); with determiner 4 (2gr-2ungr))\

Grammatical

Girls normally do better in school than boys do.

Ungrammatical

River are not as big as lakes are.

Young people like cigarettes and fast car.

Adverb placement

Grammatical

Our schools always provide good education.

Ungrammatical

All students in Malawi write very well English.

Young people visit often clubs and drink a lot.

Structural parallelism

Grammatical

Malawi is neither rich nor famous.

Ungrammatical

Students neither enjoy writing nor singing.

Many people either hate vegetables or beans.

3rd person singular *–s*

Grammatical

An intelligent student plans her compositions.

Ungrammatical

All teachers enjoys their teaching job.

Everyone love comic books and read them

Possessive *–s*

Grammatical

The lions' source of food is meat.

Ungrammatical

The pig head lacks horns.

My sister ambition is to become a teacher.

Possessive determiner *her/his*

Grammatical

A good father spends time with his children.

Ungrammatical

A good husband respects her wife.

Ethel Kamwendo is famous, and his brother is famous too.

Definite article *the*

Grammatical

We have a lake. The lake has beautiful beaches.

Ungrammatical

A hotel on Mount Zomba is very beautiful

A sun rises in the east.

Locative Inversion

Grammatical

Round the track athletes race.

Ungrammatical

On Christmas day drink many people.

In the milk fats are.

Wh-question

Grammatical

What do children love? Playing, right?

Ungrammatical

What Malawi bought from Zambia last year? Maize and fuel, right?

Travellers take what to their destinations? Buses and trains, right?

Pseudo-cleft

Grammatical

Where Zomba district is is in the South.

Ungrammatical

Where cook many people is in the kitchen not in the bathroom.

Where students write better are at a desk not on the floor.

Dative alternation

Grammatical

Good teachers describe every picture to students.

Ungrammatical

Class teachers must collect students all the fees.

People should report the police stolen money

Passive

Grammatical

All the cars in Malawi are made in Japan.

Ungrammatical

Most photos post on Facebook by young women.

Many problems in Malawi created by ourselves.

APPENDIX C: STUDY RESULTS TABLES

Table 4.13. Covariance matrix for the memory ability CFA model

Measure	CVMT	Llama-B	3-Term	DAT-V	Llama-D
CVMT	0.485				
Llama-B	1.333	181.532			
3-Term	1.641	107.604	323.309		
DAT-V	0.043	1.272	1.402	0.035	
Llama-D	-0.547	1.463	34.559	0.334	153.71
SRT	-0.001	0.004	-0.015	0.000	-0.079

Note. N = 102

Table 4.14. Standardized residual matrix for the memory ability CFA model

Measure	CVMT	Llama-B	3-Term	DAT-V	Llama-D
CVMT	0.000				
Llama-B	-0.074	0.000			
3-Term	-0.058	0.059	0.000		
DAT-V	0.079	-0.006	-0.031	0.000	
Llama-D	-0.108	-0.083	0.075	0.038	0.000
SRT	-0.017	0.050	-0.004	-0.026	-0.000

Note. N = 102

Table 4.22. Fit indices for the memory abilities and language knowledge types first run of the revised model

Multivariate Kurtosis	
Mardia's Coefficient (G2, P)	-3.6188
Normalized Estimate	-1.4447
Standardised Residuals	
Average Absolute Residual	2.5759
Average Off-Diagonal Absolute Residual	3.3118
Null Model	
Independence Model CHI-SQUARE (df = 28)	210.166
Independence AIC	154.166
Independence CAIC	52.666
Hypothesised Model	
Model AIC	1.69
Model CAIC	-63.56
CHI-SQUARE (df = 18)	37.69
Probability Value for the CHI-SQUARE Statistic	0.00426
Fit Indices	
Bentler-Bonett Normed Fit Index	0.821
Bentler-Bonett Non-Normed Fit Index	0.832
Comparative Fit Index (CFI)	0.892
Bollen's (IFI) Fit Index	0.898
Mcdonald's (MFI) Fit Index	0.908
Joreskog-Sorbom's GFI Fit Index	0.918
Joreskog-Sorbom's AGFI Fit Index	0.836
Root Mean-Square Residual (RMR)	10.753
Standardized RMR	0.087
Root Mean-Square Error of Approximation (RMSEA)	0.104

Table 4.27. Fit indices for the memory abilities, declarative knowledge and exposure – initial model

Multivariate Kurtosis	
Mardia's Coefficient (G2, P)	-1.9862
Normalized Estimate	-0.7929
Standardised Residuals	
Average Absolute Residual	2.3958
Average Off-Diagonal Absolute Residual	3.0803
Null Model	
Independence Model CHI-SQUARE (df = 28)	259.688
Independence AIC	203.688
Independence CAIC	102.189
Hypothesised Model	
Model AIC	-3.047
Model CAIC	-64.672
CHI-SQUARE (df = 17)	30.953
Probability value for the CHI-SQUARE Statistic	0.02024
Fit Indices	
Bentler-Bonett Normed Fit Index	0.881
Bentler-Bonett Non-Normed Fit Index	0.901
Comparative Fit Index (CFI)	0.940
Bollen's (IFI) Fit Index	0.943
Mcdonald's (MFI) Fit Index	0.934
Joreskog-Sorbom's GFI Fit Index	0.930
Joreskog-Sorbom's AGFI Fit Index	0.853
Root Mean-Square Residual (RMR)	9.707
Standardized RMR	0.068
Root Mean-Square Error of Approximation (RMSEA)	0.090

Table 4.31. Fit indices for the memory abilities, procedural knowledge and exposure – initial model

Multivariate Kurtosis	
Mardia's Coefficient (G2, P)	-3.0998
Normalized Estimate	-1.2375
Standardised Residuals	
Average Absolute Standardized Residual	0.0481
Average Off-Diagonal Absolute Standardized Residual	0.0616
Independence Model	
Independence Model CHI-SQUARE (df = 28)	150.099
Independence AIC	94.099
Independence CAIC	-7.400
Hypothesised Model	
Model AIC	3.387
Model CAIC	-58.237
Chi-Square (df = 17)	37.387
Probability Value for the CHI-SQUARE Statistic	0.00298
Fit Indices	
Bentler-Bonett Normed Fit Index	0.751
Bentler-Bonett Non-Normed Fit Index	0.725
Comparative Fit Index (CFI)	0.833
Bollen's (IFI) Fit Index	0.847
Mcdonald's (MFI) Fit Index	0.905
Joreskog-Sorbom's GFI Fit Index	0.918
Joreskog-Sorbom's AGFI Fit Index	0.827
Root Mean-Square Residual (RMR)	9.085
Standardized RMR	0.082
Root Mean-Square Error of Approximation (RMSEA)	0.109

Table 4.36. Fit indices for the memory abilities, declarative knowledge and difficulty – initial model

Multivariate Kurtosis	
Mardia's Coefficient (G2, P)	-3.5435
Normalized Estimate	-1.4146
Standardised Residuals	
Average Absolute Standardized Residual	0.0575
Average Off-Diagonal Absolute Standardized Residual	0.0733
Independence Model	
Independence Model CHI-SQUARE (df = 28)	311.382
Independence AIC	255.382
Independence CAIC	153.883
Hypothesised Model	
Model AIC	12.965
Model CAIC	-52.285
CHI-SQUARE (df = 18)	48.965
Probability Value for the CHI-SQUARE Statistic	0.00011
Fit Indices	
Bentler-Bonett Normed Fit Index	0.843
Bentler-Bonett Non-Normed Fit Index	0.830
Comparative Fit Index (CFI)	0.891
Bollen's (IFI) Fit Index	0.894
Mcdonald's (MFI) Fit Index	0.859
Joreskog-Sorbom's GFI Fit Index	0.896
Joreskog-Sorbom's AGFI Fit Index	0.792
Root Mean-Square Residual (RMR)	11.670
Standardized RMR	0.090
Root Mean-Square Error of Approximation (RMSEA)	0.131

Table 4.40. Fit indices for the memory abilities, procedural knowledge and difficulty – initial model

Multivariate Kurtosis	
Mardia's Coefficient (G2, P)	-3.434
Normalized Estimate	-1.3709
Standardised Residuals	
Average Absolute Residual	2.2805
Average Off-Diagonal Absolute Residual	2.9321
Independence Model CHI-SQUARE (df = 28)	147.959
Independence AIC	91.959
Independence CAIC	-9.540
Model AIC	7.154
Model CAIC	-58.095
CHI-SQUARE (Df = 18)	43.154
Probability Value for the CHI-SQUARE Statistic	0.00076
Fit Indices	
Bentler-Bonett Normed Fit Index	0.708
Bentler-Bonett Non-Normed Fit Index	0.674
Comparative Fit Index (CFI)	0.790
Bollen's (IFI) Fit Index	0.806
Mcdonald's (MFI) Fit Index	0.884
Joreskog-Sorbom's GFI Fit Index	0.907
Joreskog-Sorbom's AGFI Fit Index	0.814
Root Mean-Square Residual (RMR)	9.183
Standardized RMR	0.094
Root Mean-Square Error of Approximation (RMSEA)	0.118

Table 4.48. Fit indices for exposure, memory and procedural language knowledge of complex grammatical structures – initial model

Multivariate Kurtosis	
Mardia's Coefficient (G2, P)	-2.7104
Normalized Estimate	-1.082
Standardised Residuals	
Average Absolute Standardized Residual	0.0482
Average Off-Diagonal Absolute Standardized Residual	0.0617
Null Model	
Independence Model CHI-SQUARE (df = 28)	144.176
Independence AIC	88.176
Independence CAIC	-13.323
Hypothesised Model	
Model AIC	0.353
Model CAIC	-61.271
CHI-SQUARE (df = 17	34.353
Probability Value for the CHI-SQUARE Statistic	0.008
Fit Indices	
Bentler-Bonett Normed Fit Index	0.762
Bentler-Bonett Non-Normed Fit Index	0.754
Comparative Fit Index (CFI)	0.851
Bollen's (IFI) Fit Index	0.864
Mcdonald's (MFI) Fit Index	0.918
Joreskog-Sorbom's GFI Fit Index	0.924
Joreskog-Sorbom's AGFI Fit Index	0.839
Root Mean-Square Residual (RMR)	8.461
Standardized RMR	0.078
Root Mean-Square Error ff Approximation (RMSEA)	0.101

APPENDIX D: SAMPLE OF MALAWI’S EDUCATION SYSTEM SYLLABUS WITH REQUIRED INFORMATION IN ALL COLUMNS

Term: 1

Core element: Structure and use of language

Primary outcome: The learner will be able to understand how sounds, words and grammar can be used to create and interpret texts

Assessment standard	Success criteria	Theme/Topic	Suggested activities	Teaching, learning and assessment methodologies	Suggested teaching and learning resources
We will know this when learners are able to:	Learners must be able to:				
Demonstrate an understanding of various question forms in oral and written texts	• **Formulate** various question forms in oral and written texts	• Question form	• Discussing various question forms in oral and written texts, e.g.: "Would it be alright if … " • Asking and answering various forms of questions, including those of the question tag	• Question and answer • Group discussions • Explanations • Peer observations and assessment • Teacher observations • Pair and group work	• Wall charts with various question forms • Recorded texts • Pictures
Demonstrate an understanding of language form and	• **Describe** uses or functions of	• Language structure and grammar	• Identifying various parts of speech in different types of texts	• Question and answer • Group discussions • Explanations	• Wall charts with various question forms

structure in oral and written narratives, descriptions, reports, and argumentative texts	phrases and clauses. • **Identify** verb tenses in sentences	• Tenses	• Identifying verb tenses in oral and written texts, e.g.: Present Perfect Tense or Continuous Tense	• Peer observations and assessment • Teacher observations • Pair and group work	• Recorded texts • Pictures
Show an awareness and correct use language in oral and written texts	• **Use** language in oral and written texts	• Awareness of language use	• Completing gaps using various sentences • Filling gaps in sentences by using correct structure or language	• Brainstorming • Discussion • Pair work • Group work • Explanation	• Books • Learners • Charts • Cards • Pictures
		• Awareness of language use	• Discussing sentences with correct structure and language	• Questions and answer • Explanations • Demonstrations • Pair and group work • Peer observations and assessment • Teacher observation	• Wall charts • Pictures • Recorded sentences or passages • Narratives

(From Primary School Year 8 syllabus, MEI, 2004b, pp. 69-71)