INDIVIDUAL DIFFERENCES AND THE LEARNING OF TWO GRAMMATICAL FEATURES WITH TURKISH LEARNERS OF ENGLISH

by

Sebnem Yalcin

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education
University of Toronto

© Copyright by Sebnem Yalcin 2012
This study investigated relationships between individual learner differences and the learning of two English structures that differed in their grammatical difficulty. Using a quasi-experimental design, 66 secondary-level learners of English as a foreign language from three intact classes were provided with four hours of instruction on two L2 structures – one considered relatively easy to learn (i.e., past progressive) and the other relatively difficult to learn (i.e., passive construction). The participants were pretested on their knowledge of both structures and posttested immediately after the instruction. Learners’ progress was measured via written grammaticality judgment tests (GJT) and oral production tasks (OPT). The instruments to measure individual learner differences included a computerized language aptitude test, an L1 metalinguistic awareness test, a motivation questionnaire, a backward digit span test, and a learner retrospection questionnaire.

The results revealed that aptitude and motivation were the two variables that significantly contributed to learners’ gains with respect to the ‘passive’ and that L1 metalinguistic awareness explained significant variation in learners’ gains regarding the ‘past progressive’. These relationships were observed with learners’ performance on the written but not oral measures. A detailed analysis of the aptitude test components revealed that the grammatical inferencing subtask was significantly related to L2 gains on the ‘passive’ – again only with respect to learners’ performance on the written GJT. The results also revealed that learners with different aptitude profiles (i.e., low, medium, high) benefited differently from instruction on the two target features. High aptitude learners performed better than low aptitude learners on the ‘passive’ as
measured by the GJT posttest. With respect to the ‘past progressive’ only learners in the medium aptitude profile group improved significantly on the written GJT.

These findings confirm that language aptitude holds a role in language learning but that there are other factors (i.e., motivation and L1 metalinguistic awareness) that also contribute to L2 progress. These results also provide evidence from a classroom-based study that the grammatical difficulty of what is to be learned is a factor in determining what cognitive abilities L2 learners rely on in their efforts to learn a new language.
Acknowledgements

This thesis has been an outcome of an incredible journey for me. I have been able to work with distinguished scholars and since fall 2007. They have been influential in my thesis in several ways. I also met many great friends during my graduate study. At this point I have the chance to acknowledge them to show my sincerest gratitude.

First and foremost I am grateful to, my supervisor, Dr. Nina Spada who has been my mentor as well as my supervisor for the last five years. Her teaching and research expertise have influenced my academic life and her intellectual and emotional support have shaped my understanding of a great scholar. Her continuous support and encouragement have been extremely important at each step of my graduate life. She has been the ideal scholar who has great knowledge of the field and great enthusiasm for research. I thank her deeply for her support and encouragement since the first days of my PhD study. She always found the ways to bring out the best in me which enabled me to overcome difficulties and frustrations along the way.

I also would like to express my sincerest appreciation to Dr. Alister Cumming who contributed to this thesis with his expertise. He provided detailed and insightful comments and I truly believe that he improved the quality of my thesis with his feedback and suggestions provided on several issues. My sincere appreciation goes to Dr. Jim Cummins who also provided invaluable feedback and insightful comments on my work. His feedback enabled to look at my work from a different perspective. He always supported me with his encouraging comments.

I would like to thank my internal examiners, Dr. Merrill Swain and Dr. Rena Helms-Park. I am grateful for their invaluable comments on my thesis. Their feedback has helped me gain a better understanding of my work. I am really fortunate to be able to get their opinions on my work and to be able to integrate their suggestions to my thesis which I believe improved the quality of my work. I
am also fortunate to have Dr. Rosemary Erlam as my external examiner. I thank her for her genuine and generous comments on my research. Her comments have been extremely motivating for me.

I am also grateful to many other professors and professionals who provided technical and emotional support that contributed to my research. I thank Olesya Falenchuck for her support with the statistical analysis.

I think this is a great opportunity to acknowledge the great friendship I experienced at OISE/UT during those five years. I am lucky that I met Robert Kohls on the first day of my courses in 2007. He provided tremendous support throughout my graduate study particularly during my thesis writing process. He was amazingly generous with his friendship and his support. Thank you, Robert for being a perfect friend and a role model for a graduate student. I was lucky to be a part of a research group where we were engaged in hard but productive work and fruitful discussions. I am indebted to my colleagues, Wataru Suzuki, Yasuyo Tomita, Lorena Jessop, Julie Li-Ju Shiu, Antonella Valeo, and Paul Quinn. I have been very fortunate to meet Stephanie Arnott and Jennifer Shade Wilson. With their friendship, graduate study was more than just courses.

I would like to express my gratitude to my dearest friends, Feride Ozdemir, Zeynep Ondin and Ebru Bag. Feride and Zeynep helped me conduct my pilot study and prepare the materials for my research. They traveled to a different city and worked for my research on their free days. They were intellectually and physically engaged in my research. I have been really fortunate to have them by my side during those difficult preparation days. Their support has been invaluable and I do not have the words the words to thank them. I am also indebted to Ebru Bag, who listened to my thesis stories more than a year. She also helped me score my data and was always ready to discuss my results any time. I also thank Amac Herdagdelen and Nazik Dinctopal-Deniz for their support. They were always ready whenever I called for help.
Dedication

I dedicate this thesis to my parents-Turkan Yalcin, Ismail Yalcin, and my brother Hakan Yalcin who have supported me all the way through this journey even from oceans away.

(Bu tez çok uzaklardan desteklerini her zaman bir nefes kadar yakın hissetğim aileme-annem, babam ve abime ithaf edilmistir).
# Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgements ................................................................................................................ iv

Dedication ............................................................................................................................... vi

Table of Contents .................................................................................................................... vii

List of Tables .......................................................................................................................... x

List of Figures ........................................................................................................................ xii

List of Appendices ................................................................................................................ xiii

Chapter One Introduction ..................................................................................................... 1
  Organization of the Thesis ................................................................................................... 5

Chapter Two Aptitude-related SLA research ...................................................................... 6
  Early Aptitude Research ..................................................................................................... 6
  Skehan’s Model of Aptitude ................................................................................................. 14
  Empirical Work on Aptitude and L2 Instruction ................................................................ 17
    Instructional treatment and aptitude ................................................................................. 17
    Age and aptitude .............................................................................................................. 19
    Noticing, working memory, and aptitude ....................................................................... 21
  Aptitude Complexes and Aptitude-related SLA Research .................................................. 26
    Aptitude-treatment-interaction research ........................................................................ 27
  Summary .............................................................................................................................. 33

Chapter Three Grammatical Difficulty and Aptitude .......................................................... 34
  Definitions of grammatical difficulty in SLA ..................................................................... 34
  Target Features of the Current Study ................................................................................ 41
Grammatical difficulty and aptitude in SLA studies ........................................ 45

Chapter Four Methods .......................................................................................... 49
  Research Design ................................................................................................. 49
  Research Questions ........................................................................................... 49
  Pilot Study ........................................................................................................... 50

Main Study ........................................................................................................... 53
  Participants .......................................................................................................... 53
  Target linguistic structures .................................................................................. 55
  Instructional treatment ........................................................................................ 55
  Instruments ........................................................................................................ 59

Chapter Five Analyses and Results ..................................................................... 73
  Validity and Reliability of the Research Instruments ........................................ 74
    Validity ............................................................................................................. 74
    Reliability ......................................................................................................... 74
  Scoring ................................................................................................................ 75
    Oral production task ........................................................................................ 75
    Grammaticality judgment task ....................................................................... 77
    Aptitude test .................................................................................................... 78
    L1 metalinguistic test ....................................................................................... 79
    Backward digit span test ................................................................................. 79
    Motivation questionnaire ................................................................................ 79
    Retrospection questionnaire .......................................................................... 80

Data Analysis for Research Question 1 (a & b) .................................................... 80
ID variables as predictors for difficult structure .......................................................... 82
ID variables as predictor for easy structure ................................................................. 87
Data Analysis for Research Question 2 (a & b) ............................................................ 90
  Aptitude components as predictors for difficult structure ...................................... 91
  Aptitude components as predictors for easy structure .......................................... 94
  Data Analysis and Results for Research Question 3 .................................................. 98
  Difficult structure: The passive .............................................................................. 98
  Easy structure: The past progressive .................................................................... 100
    Learners’ perception of target structure .............................................................. 103
  Conclusion ............................................................................................................... 107
Chapter Six Discussion and Conclusions .................................................................. 110
  Research Question 1 ............................................................................................ 110
  Research Question 2 ............................................................................................ 117
  Research Question 3 ............................................................................................ 122
  Conclusions and Implications .............................................................................. 126
  Limitations of the Study ....................................................................................... 128
  Directions for Future Research .......................................................................... 129
References ............................................................................................................... 132
List of Tables

Table 1 ....................................................................................................................... 54
Table 2 ....................................................................................................................... 62
Table 3 ....................................................................................................................... 63
Table 4 ....................................................................................................................... 81
Table 5 ....................................................................................................................... 82
Table 6 ....................................................................................................................... 83
Table 7 ....................................................................................................................... 85
Table 8 ....................................................................................................................... 86
Table 9 ....................................................................................................................... 86
Table 10 ..................................................................................................................... 87
Table 11 ..................................................................................................................... 88
Table 12 ..................................................................................................................... 89
Table 13 ..................................................................................................................... 90
Table 14 ..................................................................................................................... 91
Table 15 ..................................................................................................................... 92
Table 16 ..................................................................................................................... 93
Table 17 ..................................................................................................................... 94
Table 18 ..................................................................................................................... 95
Table 19 ..................................................................................................................... 96
Table 20 ..................................................................................................................... 97
Table 21 ..................................................................................................................... 98
Table 22 .................................................................................................................................100
Table 23 .................................................................................................................................100
Table 24 ..................................................................................................................................102
Table 25 ..................................................................................................................................102
List of Figures

Figure 1. SLA processing stages and potential aptitude components ..................................16

Figure 2. Research Design of the Current Study ....................................................................59

Figure 3. LLAMA_D: Sound Recognition Task .................................................................66

Figure 4. LLAMA_E: Sound-Symbol Correspondence Task ..............................................67

Figure 5. LLAMA_F: Grammatical Inferencing Task .........................................................67

Figure 6. LLAMA_B: Vocabulary Learning Task ...............................................................68
List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>Instructional Materials</td>
<td>147</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Past Progressive Grammaticality Judgement Task</td>
<td>154</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Passive Grammaticality Judgement Tat</td>
<td>159</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Past Progressive OPT</td>
<td>163</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Passive OPT</td>
<td>174</td>
</tr>
<tr>
<td>Appendix F</td>
<td>Language aptitude test</td>
<td>179</td>
</tr>
<tr>
<td>Appendix H</td>
<td>Motivation questionnaire</td>
<td>185</td>
</tr>
<tr>
<td>Appendix I</td>
<td>Learner Perception Questionnaire</td>
<td>187</td>
</tr>
<tr>
<td>Appendix J</td>
<td>L1 Metalinguistic Awareness Test</td>
<td>190</td>
</tr>
<tr>
<td>Appendix K</td>
<td>Scoring Scheme for Passive GJT</td>
<td>193</td>
</tr>
<tr>
<td>Appendix L</td>
<td>Scoring Scheme for Past Progressive GJT (Type A Items)</td>
<td>195</td>
</tr>
<tr>
<td>Appendix M</td>
<td>Scoring Scheme for Past Progressive Grammaticality (Type B items)</td>
<td>196</td>
</tr>
</tbody>
</table>
Chapter One
Introduction

This chapter provides an overview of individual differences research within SLA. It begins with a historical overview of aptitude research and continues with a summary of theoretical and empirical research on aptitude and L2 learning. The history of aptitude research starts with the pioneering work of John Carroll and his conceptualization of aptitude as a cognitive ability. The chapter then moves on to a review of more recent conceptualizations of the aptitude construct in line with a current view of L2 learning. Empirical work investigating the relationship between language aptitude and L2 instruction are also briefly introduced in this chapter.

As both a foreign language learner and teacher, I have always been curious about the variation across individual learners’ second language (L2) abilities despite receiving the same type of L2 instruction for the same amount of time, and sharing the same first language background. Similar to other subject areas, such as science and mathematics, there are more or less successful learners in every language classroom. Several individual learner differences are assumed to play roles in that variance among learners. Language aptitude has been identified as one of the individual learner variables that contribute to differences in L2 success. Although language aptitude has been investigated extensively, it has remained somewhat behind the current agenda of SLA research until recently. Individual differences (IDs) research is the domain of work within applied linguistics that investigates questions related to the great variation in rate and ultimate attainment of second (foreign) languages among learners. In order to understand the process of L2 learning better it is important to be able to identify the characteristics of learners that contribute to greater or lesser success. Researchers have identified a number of learner variables that could account for differences in L2 success. These learner variables include age, learning strategies, personality traits, motivation, and foreign language aptitude. Language aptitude has been documented as one of the most consistent predictors of L2
success along with motivation (Dörnyei & Skehan, 2003; Ehrman, 1996; Ehrman & Oxford, 1995). Research has shown that language aptitude is related to L2 learning in classroom contexts (Horwitz, 1987; Wesche, 1981; Hummel, 2009; Kosmos & Sáfár, 2008) as well as natural contexts (Harley & Hart, 2002; Reves, 1983 cited in Dörnyei & Skehan, 2003). The predictive value of language aptitude has also been maintained across different L2 learning conditions such as explicit and implicit learning, and form-focused instruction (FFI) (de Graaff, 1997; Erlam, 2005; Robinson, 1995, 1997, 2002; Sheen, 2006) and across different age groups (DeKeyser, 2000; Harley & Hart, 1997; 2002).

Language aptitude was first demonstrated to be related to L2 learning in the classroom context. Subsequently, it was examined in relation to different L2 learning environments/conditions. Some researchers, for example Krashen (1981) and Reber (1993), claim that language aptitude is only related to explicit (i.e., conscious) learning – the kind of learning that typically takes place via traditional L2 instruction. However, research that has examined the role of aptitude within different L2 instructional contexts has demonstrated that language aptitude plays a role in both explicit and implicit learning conditions (de Graaff, 1997; Erlam, 2005; Robinson, 1997; 2002b). This includes studies that have examined the effects of the interaction between aptitude and different types of corrective feedback on L2 learning (Sheen, 2006).

Empirical research studies have demonstrated that aptitude is related to several aspects of L2 learning. Despite evidence of the predictive value of language aptitude, aptitude is generally under-researched and not well integrated in research on L2 learning and teaching (Dörnyei, 2005; Dörnyei & Skehan, 2003; Sawyer & Ranta, 1998). Until the 1990s, language aptitude was considered to be a stable and monolithic construct. However, recent theoretical work on aptitude influenced by work in cognitive science has resulted in a more dynamic and multicomponential construct, which includes a complex interplay of cognitive abilities (Dörnyei, 2005; Robinson, 2001, 2002; Skehan, 2002). New conceptualizations of aptitude accelerated the process of its integration into SLA research, and also introduced new components of aptitude. For example, working memory (WM) has come to be considered as a more critical component of language aptitude (Miyake & Friedman, 1998, Sawyer & Ranta, 1998; Skehan, 2002). Recent ID
research has included WM as a variable in several empirical studies (Erlam, 2005; Hummel, 2009; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Robinson, 2002b). The findings have shown that WM is related to various aspects of L2 learning such as noticing (Mackey et al., 2002) and awareness (Robinson, 1997; 2002b).

With the changes in the construct of aptitude, SLA researchers see more potential in aptitude research because a more dynamic concept of aptitude has important implications for L2 learning and teaching. That is, aptitude testing in early ID research was used primarily for selecting successful L2 learners for allocating resources (Dörnyei, 2005). However, recent aptitude research aims to use aptitude testing in order to identify aptitude profiles and examine the possibilities of matching learners with compatible instructional approaches. Researchers who support a dynamic conceptualization of aptitude, for example Robinson (2002a) as well as Skehan (2002) propose that if we know more about learners’ aptitude, L2 instruction may be able to better accommodate learners’ needs.

Aptitude-treatment-interaction research has been used quite extensively in educational psychology. This kind of research aims to investigate the potential in matching learner profile and preferences with learning conditions. A seminal work by Wesche (1981) demonstrated that profiling L2 learners and matching them with the right L2 instructional methods based on their ability and preferences led to L2 success whereas aptitude-treatment mismatch resulted in less success. Wesche’s study demonstrated that there is potential in aptitude treatment interaction research. However, this kind of empirical research is rare because it is difficult to investigate in classroom contexts.

Aptitude profiles (Skehan, 1998, 2002) and aptitude complexes research (Robinson, 2001, 2002a) offer a new line of inquiry considered more amenable to empirical investigation (Ortega, 2009). Both Skehan and Robinson argue that learners vary in their cognitive abilities underlying language aptitude. Some learners are hypothesized to be high on some cognitive abilities and low on others. Similarly, some learners can draw on certain abilities on some tasks whereas others rely on other abilities. For example according to Skehan (2002), some learners, based on their aptitude profiles,
could be better at some L2 processing stages (e.g., noticing, or lexicalizing) whereas some may have difficulty in other stages (e.g., controlling). Aptitude research within instructed SLA therefore aims to investigate aptitude profiles in order to provide learners with interventionist techniques at necessary stages in order to boost L2 learning.

The new aptitude framework has resulted in a close examination of aptitude with respect to L2 learning processes (e.g., noticing, awareness or FFI). Research has shown that learners’ aptitude is related to learning gains in foreign language classrooms (Erlam, 2005), noticing the systematic cues in the input (Mackey et. al., 2002), and awareness level (Robinson, 1997, 2002b). As indicated above, aptitude research has also demonstrated that it is related to explicit and implicit learning. Nonetheless many questions in current aptitude-based SLA research remain to be investigated. For example, one question of interest in instructed SLA research has been whether the benefits of instruction depend on the nature of the target feature. Opinions differ over whether easy or difficult structures are better candidates. Krashen (1982) for example argues that only easy structures should be taught. He claims that difficult structures are too complicated to teach. Other researchers like Hulstijn (1995) suggest that difficult structures should be targeted because easy structures are more available to be picked up in the language input. Although a recent meta-analysis of the effects of instruction on ‘easy’ versus ‘hard’ L2 structure revealed no differences (Spada & Tomita, 2010), there is a need for research to investigate this question in relation to ID’s and especially language aptitude. For example, aptitude could be a cognitive ability that boosts learners’ success when learning difficult L2 structures and if so, this could have valuable implications for L2 instruction.

Researchers interested in the interplay between language aptitude, L2 structure, and L2 instruction include Robinson (1997, 2002b) whose studies have examined language aptitude with respect to structure difficulty across different L2 learning conditions. The results indicate that rule difficulty interacts with aptitude. However, this research was carried out in the laboratory setting and with an artificial language. There is a need for similar research in the classroom context, particularly the foreign language classroom where learners share the same first language and have limited exposure to the L2 outside the classroom. An investigation of potential interactions between aptitude and
type of L2 structure (i.e., easy and difficult) in foreign language classrooms should help provide an explanation of why learners who share the same first language background and learning in the same instructional context differ in the progress they make on different target language structures. Such research has the potential to inform curriculum planners and teachers how to best meet learner needs and abilities, especially in EFL contexts where the L2 teaching curriculum is often based primarily on grammatical structures (Fotos, 1998, Nassaji & Fotos, 2004, 2011; Ortega, 2009).

Motivated by current research on aptitude profiles and aptitude complexes, this study examines the role of aptitude and other ID variables in relation to the learning of two target features in English one of which is described as relatively difficult and the other as relatively easy. The study takes place in a foreign language context with a group of Turkish-speaking secondary school students.

**Organization of the Thesis**

The remaining chapters are organized as follows. Chapter Two provides a review of the literature on language aptitude, and discusses theories and research relating to different aspects of L2 learning such as type of instruction, learning context, and age. Chapter Three discusses the second variable in the study: grammatical difficulty. It provides a review of the literature on how SLA research has defined grammatical difficulty based on different linguistic and psycholinguistic theories. Chapter Four addresses the methodology used in the current study and describes the research context, participants, research design, research instruments, and data collection procedures. Chapters Five reports on the results in relation to the three research questions. Chapter Six discusses and interprets the results, identifies limitations of the study, considers the implications of the findings, and provides suggestions for future research.
Chapter Two
Aptitude-related SLA research

This chapter consists of three sections. The first is a review of early aptitude research within SLA and criticism of the aptitude construct. The next is a discussion of empirical work on aptitude and L2 instruction with various measurements of L2 proficiency and aptitude. The third section considers recent conceptualizations of aptitude with a focus on aptitude profiles and aptitude complexes. The chapter concludes with a description of the purpose of this thesis research.

Early Aptitude Research

Aptitude has long been examined as a variable in second language research (Alderson, Clapman, & Steel, 1997; Ehrman, 1996; Ehrman & Oxford, 1995; Erlam, 2005; Gardner, 1988; Harley & Hart, 1997, 2002; Horwitz, 1987; MacLaughlin, 1990; Ranta, 2002; Reves, 1983 cited in Dörnyei & Skehan, 2003; Robinson, 1997, 2002a, b; Wesche, 1981). However, the concept of aptitude has been controversial in many ways. First, some researchers have expressed scepticism about a ‘talent’ that is specific to language learning because aptitude is closely related to general intelligence and hard to measure independently. In addition, language aptitude was initially introduced, especially by Carroll, as a stable characteristic of learners, thereby leaving little room for training. According to Skehan (1998), these two basic concerns about the concept have prevented full integration of foreign language aptitude into SLA research. However, recent research has resulted in an appreciation of Carroll’s original work and an expansion of it into some promising venues for SLA research on aptitude.

Carroll was not the first to introduce the concept of aptitude but his work on the construct and aptitude testing established the parameters in measurement (Dörnyei & Skehan, 2003). Carroll (1982) defined aptitude as “the individual’s capability of achieving a task or program and this capability is made of more or less enduring characteristics of the individual” (p. 84). He remarked that language aptitude did not include motivation or interest and that they were to be evaluated separately. Carroll used
the term foreign language aptitude because his goal was to explore the role of aptitude in the learning of languages other than the mother tongue that were taught in formal settings such as schools or language programs. Carroll’s concept of aptitude was integrated into a learning model: *Model of School Learning*. This model was based on the premise that there were two groups of factors that affect L2 achievement. The first group included *time* and *instructional excellence*. *Time* referred to the length of exposure to language instruction and there was a positive correlation between time provided for L2 learning and L2 achievement. *Instructional excellence* referred to quality of L2 instruction. As can be predicted, there was a positive correlation between instructional excellence and L2 achievement. The second group of factors that affected L2 learning include general intelligence, motivation, and foreign language aptitude. Carroll’s main inquiry was about the nature of foreign language aptitude. He investigated whether it is a ‘general’ unitary ability or a collection of independent, specialized abilities. In his search for possible factors, he examined a number of abilities in the domain of language that correlated with foreign language learning. These abilities included verbal and printed language development factors, lexical factors, reading comprehension, reading speed, reading decoding factors and many more (See Carroll, 1993 for more detail on abilities in the domain of language). Carroll’s factor analytic work on the possible interacting abilities resulted in a definition of aptitude that is a composite of four basic abilities. Carroll defined the four subcomponents as follows (1982, p. 105):

**Phonemic coding ability**: the ability to identify distinct sounds, to form associations between those sounds and symbols representing them, and to retain these associations.

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

**Rote learning ability for foreign language materials**: the ability to learn associations between sounds and meaning rapidly and efficiently, and to retain these associations.

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

Carroll’s Model of School Learning was not fully appreciated in SLA compared to the field of education in general when first introduced but he continued to work on the construct of foreign language aptitude in order to develop a testing instrument. In their
efforts to develop a measurement, Carroll and Sapon first devised a large number of potential predictor tests of foreign language learning and then administered these to learners who were taking language courses. They gathered a great deal of data and were able to examine which aptitude sub-tests successfully predicted L2 achievement. Their work resulted in a four-component model of aptitude, which was published as the Modern Language Aptitude Test (MLAT) in 1959. The test includes four sub-tests that correspond to abilities in the model: phonetic coding ability (test: phonetic script), grammatical sensitivity (test: words in sentences), rote learning ability (test: paired associates), and inductive language learning ability (test: number learning). The MLAT is a paper-and-pencil test battery and it takes 60 minutes to administer. It has been widely used in L2 aptitude measurement studies and it generates a total score which is used to make predictions of language learning success. The composite score can be taken as a monolithic construct, but subcomponents can also be used as separate scores. The MLAT has been the most preferred aptitude measurement in the field showing consistently high correlations with performance in language outcomes in language courses (de Graaff, 1997; Erlam, 2005; Harley & Hart, 1997; 2002; Horwitz, 1987; Hummel, 2009; Kosmos & Sáfár, 2008; Oxford & Ehrman, 1995; Ranta, 2002, Robinson, 1995, 1997; Sáfár & Kosmos, 2008).

Carroll’s work contributed to aptitude research in several ways. First, he defined the aptitude construct. He also proposed a model to explain the interplay of aptitude and several other factors such as instructional excellence and time. Finally he developed a measurement to investigate the aptitude construct. His test has been used extensively in SLA research yet it is not without its limitations and has been criticized on several fronts. For example, its theoretical underpinnings have been questioned by Dörnyei (2005) who posits that the MLAT is a result of a trial and error process that involves administering a great number of intuitively appealing tasks to differentiate good language learners from others (p. 41). The lack of solid theoretical background, according to some researchers, resulted in an unclear relationship between tests and constructs (Sawyer & Ranta, 1998). They argued that the spelling cue task in the test was not conceptually connected to any of the factors proposed. Similarly, inductive language learning ability was not measured
by any of the tasks in the test (Sawyer & Ranta, 1998; Skehan, 1998). Carroll (1982) also admitted that inductive language learning is weakly represented.

Another major criticism concerned the validity of aptitude testing across uninstructed contexts. According to Krashen (1981) inductive ability and grammatical sensitivity in the MLAT operated on conscious language learning, which was, according to him, of less interest compared to acquisition, which led to long term gains. Krashen supported his claims with several research studies which reported significant relations of aptitude and classroom grades (Gardner & Lambert, 1972) or aptitude and self-reports of conscious learning (Bialystok & Fröhlich, 1977). Neufeld (1979) also argued that aptitude testing was arbitrary and unnecessary. He claimed that there was little variation in the rate of first language ability and everyone possessed the same language learning ability. Similar to Krashen’s view, Neufeld proposed that variance in L2 learning was solely due to social factors. Aptitude research conducted in the 1990s provided counter evidence to Krashen’s claims (Harley & Hart, 1997, 2002; Ranta, 2002). Nonetheless there is a lack of empirical research that examines the type of learning that aptitude tests measure.

With more recent work on language aptitude, a different point of view has emerged to address the lack of connection between the MLAT and acquisition-based approaches to L2 learning. Kiss and Nikolov (2005) stated that, “the MLAT fails to take into account the different and equally successful acquisition-based approaches that different learners might apply when learning an L2 (p.205). A similar point was put forward by Robinson (2002b) who urged a revision of available aptitude tests to capture all cognitive abilities operative on different types of L2 learning such as incidental type. Robinson (2002b) also remarked that aptitude tests such as the MLAT failed to capture all cognitive abilities that operate on different types of L2 learning.

Lastly, a practical concern about language aptitude testing, specifically the MLAT, is the language input used in the test. Kiss and Nikolov (2005) argued that the MLAT relied extensively on English and therefore was not suitable for non-native speakers of English. Another issue is the use of natural languages such as Kurdish in the
paired associates" subtest. The test cannot be administered to learners who have prior knowledge of Kurdish. Therefore, prior knowledge of the language of the test is a limitation for the test itself. Some recently developed aptitude tests managed to overcome the language barrier by using artificial or semi-artificial languages (Kiss & Nikolov, 2005; Meara, 2005).

Regardless of these theoretical and practical concerns, the MLAT has been used extensively in empirical research and this has demonstrated that the MLAT is a good predictor of L2 achievement (de Graaff, 1997; Ehrman & Oxford, 1995; Erlam, 2005; Harley & Hart, 1997, 2002; Horwitz, 1987; Hummel, 2009; Ranta, 2002; Robinson, 2002b). Criticism against the MLAT also resulted in attempts to develop alternative aptitude batteries. However, except for one, they have all stayed within the major parameters established by MLAT. The first true alternative was Pimsleur’s Language Aptitude Battery (PLAB) which was developed in response to the underachievement of school children (1966). It included a motivation questionnaire in addition to cognitive tasks. VORD is another test, developed by Parry and Child in 1990 in order to provide an alternative to the MLAT with a language (Turkish) that is different from Western European languages. VORD was designed for students in government language training and Turkish was unlikely to be studied by that group of learners. The Cognitive Ability for Novelty in Acquisition of Language (CANAL-F) aptitude measure was developed by Grigorenko, Sternberg and Ehrman (2000) and was based on a cognitive theory of knowledge acquisition that stresses the importance of coping with novelty in learning. CANAL-F is a different aptitude measurement because, unlike the MLAT and subsequent measures, it was based on the “theory of successful intelligence” by Sternberg (1999). The tasks included learning meanings from neologisms in context, understanding the meaning of passages, sentential inference, and learning language rules. Kiss and Nikolov (2005) developed and validated an alternative aptitude test for young learners. Similar to the MLAT, the test included 4 tasks: phonetic coding ability, language analysis, words in sentences, and vocabulary learning. They used an artificial language called Klingon, in order to control for the effect of prior knowledge of the language input in the tasks.
Most recently, Meara (2005) has developed a computerized aptitude battery which is based on the conceptual underpinnings of the MLAT. The LLAMA aptitude battery consists of a set of exploratory tests based on Carroll’s work, the result of ongoing research at the University of Swansea in Wales. Earlier versions used language samples that were based on Polish and Turkish but the latest version uses stimuli based on a dialect of a language of North Western British Columbia. One of the advantages of the test is that it is fully computerized and presented in a ‘user friendly’ manner (see a detailed view of the test in Chapter 4). However, Meara (personal communication, September 22, 2009) cautions on its use because research to establish the validity and reliability of this aptitude measure is ongoing.

It is widely accepted that aptitude research has not reached its potential within SLA research (Dörnyei & Skehan, 2003; Skehan, 2002; Robinson, 2002a). Skehan (2002) states that “foreign language aptitude has been researched for over thirty years but there’s relatively little empirical work and remarkably little theorizing in comparison to other aspects of applied linguistics” (p.68). There are several reasons for this gap. One major reason is that the notion of aptitude is against the idea of Universalist approaches to L2 learning which aim to address all types of learners across all conditions. The aptitude construct brings along the complexity of individual learner differences into language teaching and learning. A recognition of individual learner differences results in the need to accommodate learner differences in syllabus design and teaching materials. Therefore, aptitude has been either overlooked or neglected by the mainstream SLA research until the 1990s. Another reason is the close association of language aptitude with audio-lingual and grammar-translation methods. That is, aptitude measures were designed and validated with the instructional methods of that time. When the popularity of these instructional methods ended, interest in aptitude virtually disappeared for several years. Another reason for its demise was Carroll’s conceptualization of aptitude as relatively fixed. The assumed stable nature of aptitude prevented it from being integrated into SLA research. Both researchers and practitioners refrained from investing in something that could not be altered. In addition, some SLA researchers, especially Krashen (1982) and Reber (1989, 1991), argued that aptitude was only relevant to ‘learning’ where there was conscious processing of language input. They saw no role for
aptitude in the ‘acquisition’ process, considered to be the process of unconscious learning in natural settings. Even though Krashen’s claims had not yet been tested empirically, they were highly influential. Furthermore, the shift from structure-based approaches to L2 instruction (e.g., the audiolingual method) to communicative approaches led many SLA researchers to be sceptical about the relationship between aptitude and L2 instruction for quite a few years.

Attitudes toward aptitude changed considerably in the 1990s. Several reasons contributed to the change in how researchers began to view language aptitude in SLA. One of the reasons for change in aptitude related to developments in cognitive psychology and educational psychology. Interdisciplinary work provided SLA researchers a fresh point of view. Moreover, aptitude has now been reconceptualised as a composite ability which is more dynamic and potentially trainable. The need to identify alternative cognitive abilities has also been acknowledged and WM has been the first agreed upon aptitude component that should be examined with respect to L2 learning (Miyake & Friedman, 1998; Sawyer & Ranta, 1998). Following the publication of the book by Parry and Stansfield (1990), there was a revival of interest in aptitude which resulted in a substantial body of research investigating the role of individual learner variables in instructed SLA research (Robinson, 1997; 2001; 2002; Skehan, 1986, 1989, 2002). These studies have investigated language aptitude and its relevance to L2 learning processes and stages of L2 development (de Graaff, 1997; DeKeyser, 2000; Harley & Hart, 1997; 2002; Robinson, 1997) within different language learning contexts (Erlam, 2005; Ranta, 2002; Wesche, 1981).

The change discussed above has had important implications for aptitude testing research within SLA research. Since the 1990s, aptitude has been reconceptualised as a multicomponential construct in line with a more process-oriented approach to individual learner variables (Dörnyei, 2005; 2010). This new line of aptitude research also aims to address the dynamic and context-dependent nature of language aptitude (Dörnyei, 2005, 2010). As a result scholars have tried to demystify the cognitive variables that operate under the language aptitude construct. Some researchers proposed that new cognitive variables should be examined with respect to other aptitude components. For example,
Miyake and Friedman (1998) nominated working memory as a key component of language aptitude. According to Dörnyei (2005) “the change in the research objective in aptitude research such as examination of the SLA-specific impact of specific cognitive factors and subprocesses marked the start of Post-Carroll research” (p. 250).

New research perspectives have also led to alternative measurements, which are more related to the instructional contexts of contemporary language teaching methods such as communicative language teaching or form-focused instruction. Theoretical work on the construct of language aptitude has also focused on establishing more connections among some key concepts in current SLA research and aptitude. For example, form-focused instruction (FFI), awareness, and noticing have been studied in relation to language aptitude in recent years (See Skehan, 2002 for a review).

A theoretical restructuring of the aptitude construct as an ID variable has led to new directions regarding where aptitude tests can be useful in SLA research. Originally, the MLAT was used was to identify and select “best” candidates for L2 learning. This selection was aimed at reducing training costs. That is why aptitude was considered anti-democratic for a long time. However, recent approaches to aptitude have introduced new objectives for aptitude testing. For example, program evaluation and tailored instruction have been two important objectives for aptitude testing since the 1990s (Dörnyei, 2005). Aptitude test scores have been used to identify learners’ cognitive strengths and preferences in order to accommodate learner needs by manipulating the quality and quantity of language instruction. This view is consistent with Carroll’s Model of School Learning which underlined the quality of instruction in addition to aptitude. Important work by Wesche (1981) showed that there was great potential in matching learners’ cognitive abilities and preferences with the type of L2 learning.

In line with the directions described above, Skehan (1986, 1989, 1998, 2002) worked both on the construct of aptitude and language aptitude testing. His work can be viewed as pioneering because he attempted to integrate aptitude into various aspects of L2 learning. His work can be viewed in three phases. First, he refuted two major criticisms of language aptitude research based on his own research and that of others.
Then, he updated the construct of aptitude and proposed his own model. Finally he attempted to link aptitude and information processing in one model of language aptitude. Skehan preferred a componential approach to examine language aptitude and he claimed that aptitude components are relevant to both instructed and uninstructed contexts of L2 learning. Skehan (1998, 2002) argued that Carroll’s work on foreign language aptitude was valuable because Carroll proposed an interactive model (i.e., Model of School Learning) that included both instructional variables and IDs in order to explain L2 achievement in foreign language classrooms. Skehan was also interested in other factors that could be related to L2 success and aptitude such as L1 development and family literacy. He investigated those factors in a follow-up study to the Bristol Language Project (Wells, 1981) which had demonstrated that there is variation in the speed of L1 acquisition. In his follow-up study to the Bristol Project, Skehan contacted 100 participants from the original study and measured language aptitude through the grammatical sensitivity and inductive language ability subtests of the MLAT. He was specifically interested in the relation between first language development (measured by several early first language measures), foreign language achievement (measured by standardized achievement tests), and aptitude (measured with MLAT). Although it was almost 12 years later, the results yielded a number of significant correlations between early first language measures and aptitude (Skehan, 1986). Inductive language learning ability correlated with a range of first language measures. The results also showed that aptitude, measured at the age of 13 to 16, was partly predicted by the rate of L1 acquisition, L1 vocabulary knowledge, and family literacy.

The results showed that contrary to Neufeld’s claims (1979), there were more than social factors that contribute to variation in L2 learning. Skehan (1986) was also able to demonstrate that aptitude was related to L2 learning.

**Skehan’s Model of Aptitude**

Findings of his follow-up work on the Bristol Project enabled Skehan to reconceptualise language aptitude and he proposed an aptitude model with three components: auditory ability, linguistic ability, and memory ability (1986, 1989). He
based his model on Carroll’s but combined grammatical sensitivity and inductive language ability into one subcomponent—language analytical ability—because he argued that inductive language learning ability was not reflected in the MLAT. Skehan’s updated concept of aptitude allows for differentiated strengths and weaknesses in aptitude. In his view, people may differ in their cognitive strengths and weaknesses. For example, high grammatical sensitivity does not always suggest or guarantee high memory ability. According to Skehan (2002), foreign language learning abilities were modular and related to general learning mechanisms, and each module was connected to an aptitude component. He proposed three phases of language development in his information processing model: (a) input, (b) central representation, and (c) output. These three phases were related to three aptitude components: auditory processing, language processing, and memory respectively.

Skehan’s most recent work on language aptitude includes his suggestion of possible subcomponents for language aptitude. He also attempted to link potential aptitude components to well-studied SLA processes (e.g., noticing, practice, and automatization). These are indicated in Figure 1 below.
Figure 1. SLA processing stages and potential aptitude components

<table>
<thead>
<tr>
<th>SLA processing Stage</th>
<th>Aptitude Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>noticing</td>
<td>auditory segmentation</td>
</tr>
<tr>
<td></td>
<td>attention management</td>
</tr>
<tr>
<td></td>
<td>working memory</td>
</tr>
<tr>
<td></td>
<td>phonemic coding</td>
</tr>
<tr>
<td>pattern identification</td>
<td>fast analysis/working memory</td>
</tr>
<tr>
<td>extending</td>
<td>inductive language learning ability</td>
</tr>
<tr>
<td>complexifying</td>
<td>grammatical sensitivity</td>
</tr>
<tr>
<td>integrating</td>
<td>inductive language learning ability</td>
</tr>
<tr>
<td>becoming accurate, avoiding error</td>
<td>restructuring capacity</td>
</tr>
<tr>
<td>creating a repertoire, achieving salience</td>
<td>automatization</td>
</tr>
<tr>
<td>automatising rule-based language, achieving fluency</td>
<td>proceduralisation</td>
</tr>
<tr>
<td>lexicalising, dual-coding</td>
<td>memory, chunking, retrieval processes</td>
</tr>
</tbody>
</table>

(Skehan, 2002, p. 90)

As can be seen in Figure 1, Skehan proposes links between SLA processing stages with currently existing aptitude sub-components (e.g., inductive language learning ability, grammatical sensitivity) as well as new components (e.g., working memory and retrieval processes). Empirical work on language aptitude has provided evidence for the predictive value of aptitude in both the classroom and natural contexts. Moreover, Skehan (2002) and Robinson (2002a) argue that aptitude operates in awareness and form-focused instruction that takes place during L2 learning. Therefore, both researchers suggest that there is potential value in research that investigates the relation between current L2 development stages, form-based instructional techniques, and newly introduced aptitude components. Studies in SLA investigating relationships between working memory, awareness and aptitude have confirmed some of these connections (Erlam, 2005; Harley & Hart, 2002; Hummel, 2009; Kosmos & Sáfár, 2008; Mackey, Philp, Egi, Fujii, & Tatsumi 2002; Robinson, 2002b). Skehan’s new aptitude model which aims to connect aptitude components with SLA processes is a compelling one in SLA research (Dörnyei,
Nonetheless the model is still speculative and requires empirical support.

**Empirical Work on Aptitude and L2 Instruction**

In the following section, studies in instructed SLA that have included language aptitude as a variable will be reviewed. The studies will be reviewed with reference to instructional treatment, age, and SLA processes (i.e., noticing, FFI). Some studies fall under more than one category.

**Instructional treatment and aptitude**

Explicit and implicit learning have been the focus of recent discussion in the SLA literature and they are thought to have different learning mechanisms and represent different types of L2 knowledge. Language aptitude has been investigated under both implicit and explicit learning conditions. Reber (1979) argued that aptitude was limited to explicit learning contexts and had no validity in implicit learning. de Graaff was one of the earliest SLA researchers to explore whether language aptitude was relevant to implicit and explicit learning. Contrary to Krashen (1981) and Reber (1979), de Graaff (1997) hypothesized that “language aptitude will affect learning to the same extent in both explicit and implicit instructional conditions, as they both require some kind of focus on form” (p.253). He conducted a laboratory experiment in which fifty-six undergraduate learners studied eXperanto (an artificial language) in order to learn four L2 structures. Learners received two different types of L2 instruction (implicit and explicit) and were tested for L2 proficiency through gap filling, sentence judgement, and vocabulary translation tasks and language aptitude through the *Words in Sentences* and *Paired Associates* subtests of the MLAT. The results showed that learners with high and low aptitude scores performed similarly under both explicit and implicit conditions. That is, aptitude was equally relevant to explicit and implicit learning.

Robinson (1997) investigated whether aptitude affected learning in four different L2 conditions (i.e., incidental, implicit, explicit, and rule-search) and also examined whether aptitude was related to learner awareness developed through the four learning
conditions. He measured language aptitude in 104 non-native speakers of English using the *grammatical sensitivity* and *paired-associates* tests from the MLAT. Language gain was measured through the grammaticality judgement test. He found that measures of aptitude were significantly correlated with learning in three conditions, but not in the incidental condition. Robinson interpreted these results to mean that all instructional conditions were fundamentally similar because participants seemingly used conscious processing strategies that are triggered by the task demands on cognitive resources. He therefore argued that some conditions predisposed learners to rely on memory whereas others required the use of grammatical sensitivity. The studies of de Graaff (1997) and Robinson (1997) demonstrated that language aptitude had predictive validity across different instructional types. The argument that aptitude is only related to L2 learning in formal contexts rather than in natural ‘acquisition-rich’ contexts has also been explored. For example, Ranta (2002) investigated whether language analytical ability was a predictor of L2 proficiency in highly communicatively oriented classrooms where there was virtually no focus on form. This study was important because aptitude, previously thought to be associated with grammar-translation or audio-lingual teaching, had not been investigated in a communicative language teaching setting. In Ranta’s classroom study, francophone children were studying ESL in a 5-month intensive program where the main focus was on the development of interpersonal communication skills through mostly oral activities. L2 proficiency was measured through a variety of language tests including aural vocabulary recognition, listening comprehension, cloze and metalinguistic ability. Ranta preferred to use an L1 metalinguistic task to measure language analytical ability. Results showed that language analytic ability was associated with strong performance for the most successful learners and weak performance of the least successful students. Thus it appears that aptitude is related to L2 success even in a highly communicative classroom context. These results confirmed those of an earlier study (Reves, 1983 cited in Dörnyei & Skehan, 2003) that also explored the interaction of learning context and aptitude in L2 learning. Reves compared two different learning contexts for the same group of Arabic L1 learners of Hebrew and English: a formal classroom environment for learning English and exposure to naturalistic language use for Hebrew learning. The results showed that aptitude tests were the best predictors of language learning success in both contexts.
Another study that provided evidence to support the claim that aptitude was a relevant factor in acquisition-rich contexts was conducted in a bilingual exchange program in Canada. Harley and Hart (2002) pre-tested a group of English speaking secondary school students who spent three months on a bilingual exchange in Quebec. In order to measure language aptitude, the researchers used the language analysis subtest of the PLAB and a memory test from Wechsler’s Memory scale (1972). French language proficiency was measured through several measures including vocabulary, grammar and listening tests. The results showed that analytical language ability was a factor influencing learning in the “acquisition-rich” context of the bilingual exchange. That is, learners’ aptitude scores were related to their performance in language tasks used in the study. A similar study conducted by the same researchers was carried out with French immersion students a few years earlier (Harley & Hart, 1997) but the correlations between language aptitude and L2 proficiency were not as consistent. Nonetheless, the findings demonstrated that aptitude is a predictor of L2 success even in contexts where the primary focus is on content and meaning rather than grammatical form. Therefore, along with de Graaff, Robinson, and Ranta’s findings, the two studies conducted by Harley and Hart (1997; 2002) provided counter-evidence to the claim made by Krashen (1981) and Reber (1979) that language aptitude was only related to successful L2 learning in explicit structure-based instructional contexts.

Age and aptitude

Age has been considered to be an important factor in research on aptitude in L2 learning (DeKeyser, 2000; Munoz, 2007) as well as in SLA research (Abrahamsson & Hyltenstam, 2008). The Fundamental Difference Hypothesis (Bley-Vroman, 1988) states that the cognitive resources and abilities that children and adults depend on when learning a language differ. Children have access to domain-specific ‘innate’ learning mechanisms whereas adults depend on their L1 and general problem-solving capacities to learn a language. Considering such differences among children and adult L2 learners, some SLA researchers assumed that aptitude would be a stronger predictor of successful L2 learning for older learners in classrooms providing traditional form-focused instruction (e.g., Krashen, 1981). Harley and Hart (1997) investigated the relation
between age and aptitude in the French immersion context. They examined whether starting to learn French before or after adolescence would make a difference in the relationship between aptitude and L2 learning. They used several aptitude measures but mainly focused on memory and analytical ability. The results showed that different components of aptitude were associated with the L2 abilities of early and late immersion students. For the early immersion students, measures of memory were reported to be more likely predictors of L2 proficiency whereas analytical language ability was the only significant predictor of L2 proficiency for the older learners. The overall findings revealed that while aptitude was a predictor of L2 success both before and after puberty, it appeared that learners drew upon different abilities at different ages. DeKeyser (2000) also examined aptitude in relation to the effects of the critical period on L2 learning. He investigated Hungarian-speaking immigrants who had arrived in the U.S. at different ages. He used the Hungarian version of the MLAT for aptitude and a grammaticality judgement task for L2 measurement. The results demonstrated that aptitude contributed to L2 learning in adults even in untutored contexts. On the contrary, aptitude played no role in ultimate attainment by child learners. DeKeyser interpreted this to mean that “aptitude, as measured by a test of grammatical sensitivity, is a predictor of ultimate attainment in L2, even after decades of exposure to the language in untutored contexts” (p.518). Another study by Ross, Sasaki and Yoshinaga (2002) investigated the interaction of age and aptitude and they also reported that aptitude measured by the Japanese version of the MLAT was a predictor of successful L2 learning for post-critical period learners in the study.

The above studies were conducted to explore the effects of aptitude across different ages and contexts. The results provide evidence in favour of an aptitude/L2 learning relation in child and adult L2 learners. Both the DeKeyser (2000) and Harley and Hart (1997) studies revealed that aptitude played an important role in L2 learning by children and adults. However, there were distinct differences in the strength of the relationship between aptitude and L2 learning across ages. The role of aptitude was relatively weak in early immersion students and child L2 learners whereas the relation became stronger with age. Moreover, the subcomponents of aptitude that contributed to L2 success differed based on L2 learners’ age. This was evident in the fact that all child
L2 learners in DeKeyser’s study reached native or native-like proficiency regardless of their aptitude but only adult learners with high aptitude scores were able to achieve native-like proficiency.

**Noticing, working memory, and aptitude**

Aptitude has been investigated as one of the important determinants of variance in L2 success; however, the complex and multifaceted nature of language aptitude has prevented researchers from drawing conclusions about its components. Therefore, there are continued efforts to explore alternative constructs that might be a part of the aptitude construct and operate in conjunction with the cognitive skills included in aptitude. Memory is one that is often mentioned in ID research. In the MLAT, memory was included as a rote-learning subtest, but this construct is now considered to be rather outdated and is not consistent with dynamic memory models in current ID research. In line with developments in cognitive psychology, working memory capacity (WMC) has been promoted as a promising component for aptitude-based SLA research (Harrington & Sawyer, 1992; Juffs & Sawyer, 2011; Miyake & Friedman, 1998).

Working memory capacity refers to “the ability to manipulate and store material simultaneously” (Gathercole, 1999, p.1). It plays an important role in all high-level cognition, especially language processing. Baddeley and Hitch (1974) proposed a model of WM and it has since been commonly used in cognitive psychology and psycholinguistic research. It is also the most preferred framework of WM in SLA research. The model is comprised of three components: the central executive, phonological loop, and visuospatial sketchpad. The central executive is the control system of limited attentional capacity and it controls the two other ‘slave’ systems: phonological loop and visuospatial sketchpad. Phonological loop is related to sound and language and is assumed to hold memory traces for a few seconds before they fade, and enables articulatory rehearsal processes similar to subvocal speech. The phonological loop is the most studied part of the WM model in relation to L2 learning. The visuospatial sketchpad, the second part of the ‘slave’ system, provides a continuing record of visual objects (Baddeley, 1999; 2003).
Phonological loop is responsible for the phonological short term memory (PSTM), which, according to the Baddeley and Hitch (1974) model, controls the processing and temporary retention of both familiar and unfamiliar phonological information. It determines the temporary maintenance of acoustic speech-based information. Therefore, it is considered to be an important determinant of variance in L1 processing (Baddeley, 1986; Gathercole & Baddeley, 1993). Phonological loop capacity has been investigated extensively in relation to L1 and L2 vocabulary learning (Cheung, 1996; Gathercole & Baddeley, 1989; Service & Kohonen, 1995). It is also reported to be a good predictor of the ability of children (Service, 1992) and adults (Atkins & Baddeley, 1998) to learn a second language.

WM capacity is also considered to have an important effect on SLA processes such as noticing and attention (Robinson, 2002a; Skehan, 2002). Sawyer and Ranta (2001) stated that “assuming that noticing is crucial to learning, and attention is required for noticing, and attention at any moment is limited by WM capacity, then there must be a close relationship between amount of learning and size of WM” (p. 342). WM capacity has also been reported to be strongly correlated with L2 proficiency (Harrington & Sawyer, 1992) and aspects of L2 comprehension (Daneman & Carpenter, 1980). Furthermore, Miyake and Friedman (1998) claimed that WM can even replace language aptitude in their WM-as-language aptitude proposal. Nevertheless, their hypothesis is a rather strong one and it requires further investigation.

Integration of WM capacity into the aptitude construct has been supported by several SLA researchers (DeKeyser & Juffs, 2003; Robinson, 2002a, b; Skehan, 2002). Skehan (2002) also claims that memory and noticing play an important role in all phases of language development and he places them at the center of his information processing model. According to Skehan the ‘central representation’ phase covers the learners’ progress from the initial stage of zero to the end-state of L2 proficiency and this phase includes learners’ interaction with the language as a system, which includes implicit or explicit processes. The ‘central representation’ is also subject to the effects of external factors that can influence the nature and speed of L2 development. He proposes that language analytical ability plays a major role at this stage. For example, different
capacities that learners draw upon change how input or feedback is attended to or noticed. The last phase is ‘output’ and it is more related to factors that affect L2 fluency. Communication strategies and time-creating devices that learners make use of have been researched as aspects of output. Memory has an important role for output in Skehan’s model because encoding, storage, and retrieval are thought to have important effects in any stage of L2 production (See Skehan 2002 for detail). Skehan also emphasizes that WM capacity is a cognitive ability that can play an important role in the model (See Figure 2 for Skehan’s model).

Of all new aptitude components proposed in recent ID accounts, working memory (WM) capacity has been the most studied in recent aptitude-based SLA research as opposed to rote memory in early research. For example, Mackey et al. (2002) investigated whether there was a relationship between WM capacity and noticing of interactional feedback in the form of recasts. Learners participated in communicative tasks such as picture drawing, picture difference, and story completion and received corrective feedback in the form of recasts on the target feature (i.e., question formation). The recasts were provided during short communicative tasks designed for the study. Phonological short term memory (PSTM) and listening span tests were used to measure WM capacity. The results showed that learners who reported less noticing seemed to have low WM capacities whereas learners with high WM capacities reported more noticing. Participants who obtained high scores on the phonological STM differed in noticing according to their developmental level. Learners with high WM scores showed the greatest progress in the delayed posttest. The researchers interpreted this to mean that there was a relationship between WM and levels of noticing. The findings revealed that learners’ WM capacity had an effect on how much they noticed in the input. Their results supported Robinson (2002) and Skehan’s (2002) proposals about the role of WM. They suggested that among other factors such as grammatical sensitivity, WM was also a factor that determined the amount of noticing. The findings indicated interplay of WM capacity, noticing, and L2 developmental level in L2 success.

Another study that included WM capacity as a possible factor affecting L2 learning was conducted by Kosmos and Sáfár (2008). Researchers examined the
interrelation among PSTM, WM capacity, and L2 proficiency. Participants were 121 secondary students studying in an intensive English program in Hungary. Researchers used a non-word span test to measure PSTM and a backward digit span test to measure WM capacity. Language ability was measured with the Cambridge First Certificate exam at the end of the academic year. Learners attended language classes that integrated a focus on form and meaning. Results demonstrated differential effects of PSTM and WM for beginner and pre-intermediate level students. There was no relationship between PSTM and L2 proficiency for beginner learners but there was a moderate relationship between PSTM and writing and use of English sections on the language test. On the other hand, WM measured by a backward digit span test (DB) was related to scores on the complete language test for beginner level students. Kosmos and Sáfár argued that pre-intermediate level students who achieved a high intermediate level at the end of the term had proceduralized or even automatized grammatical processing but knowledge of vocabulary and retrieval of words was more important for them. Therefore, PSTM instead of WM capacity was found to be related to their success. However, for beginner level learners, WM capacity was considered to aid explicit learning mechanisms. Therefore, WM capacity was related to their performance, not PSTM. In addition, the results also provided evidence that PSTM and WM capacity were different constructs since non-word span and DB were not correlated.

Hummel (2009) also investigated the relationship between PSTM, aptitude and L2 proficiency with 77 native speakers of French who were studying English in a Teaching English as a Second Language program. The learners were described as ‘nonnovice’ language learners. She used three subtests of the MLAT (i.e., spelling cues, words in sentences, and paired associates test) to measure aptitude. In order to measure PSTM, she used a non-word repetition test based on Arabic. English was assessed with the Michigan Test of English Language Proficiency. Correlations showed that PSTM and L2 proficiency were only moderately related, and aptitude and L2 proficiency had a significant but lower correlation. Subsequent multiple regression analyses revealed that PSTM and aptitude explained 29% of the variability in L2 proficiency.
There have been ongoing discussions about the right measurement of memory in the SLA literature following insights from cognitive psychology. Both PSTM and WM tests have been used and they are reported to be two separate constructs but they have been demonstrated to be at play during L2 instruction and several factors such as instruction type and proficiency level affect their role. Following Kosmos and Sáfár (2008) and Hummel (2009), PSTM and WM tests could be taken as promising constructs that can support aptitude research.

Another study that explored the interrelation of WM with other ID variables in L2 learning was conducted by Robinson (2002b). He examined whether WM, along with aptitude and intelligence, was related to three types of L2 learning (i.e., explicit, implicit, and incidental). Robinson used Samoan in order to control for previous knowledge of the target language. Aptitude was measured with the Japanese version of the MLAT and WM was measured with a reading span test. Fifty-five adult learners completed three series-solution problem tasks to learn four different rules in Samoan. The results with respect to WM and aptitude showed that aptitude was correlated with explicit learning and it was only weakly related to incidental learning. However, it was not related to implicit learning. WM had the strongest relationship with incidental learning. Contrary to his earlier work (1997), the findings showed that aptitude was at play in incidental learning and its effects were evident over time, however, the relationship was relatively weak. The first group of findings provided counterevidence for Krashen and Reber’s claim that implicit and incidental learning was robust to individual learner differences such as aptitude. WM and/or aptitude were found to be significantly related to three L2 learning conditions. Robinson argued that available aptitude tests were not tapping incidental learning. He indicated that aptitude tests should be revised based on the renewed conceptions of aptitude in order to examine the role of aptitude across different L2 learning types.

The above review of studies indicates that, in line with Skehan’s proposal and that of others (e.g., Miyake & Friedman, 1998; Sawyer & Ranta, 1998), WM capacity is a promising construct that operates in conjunction with other learner variables such as aptitude and L2 developmental level. A compelling case can be made that future research
in aptitude will include WM as a necessary construct with respect to language aptitude. There is a need for systematic research that would clarify the interrelation of WM, aptitude and L2 learning.

**Aptitude Complexes and Aptitude-related SLA Research**

A new direction in aptitude-related SLA research was introduced with Robinson’s revision of the aptitude complexes theory by Snow (1994). Snow (1994) defined ‘aptitude complexes’ as “hypothesized combinations of aptitude variables that jointly influence learning in some particular situation” (p. 9). Robinson also uses the term ‘aptitude complexes’ and his framework is similar to Skehan’s (1989) in that learners are considered to have different profiles in terms of aptitude. That is, some learners might capitalize on memory abilities and others on linguistic abilities. Robinson (2002a) argues that learners vary in cognitive abilities and any sets of cognitive abilities. Robinson’s Aptitude Complex Hypothesis (ACH) builds on the fact that differences in aptitude complexes result in learner variation in uptake. Robinson (2002a) proposes “a hypothetical structure of ‘aptitude complexes’ for learning processes in SLA: (a) aptitude for focus on form (via recasts), (b) aptitude for implicit learning (via oral input), (c) aptitude for incidental learning (via written content), and (d) aptitude for explicit rule learning. Each aptitude complex includes different ability factors that make up the complex. These include primary cognitive abilities such as pattern recognition and grammatical sensitivity or second order abilities such as noticing the gap and deep semantic processing. Differences in aptitude complexes could then explain why some learners benefit from implicit feedback whereas others benefit from explicit rule learning. Robinson’s proposal approaches ID variables from a situation-oriented framework and addresses combined effects of several cognitive variables rather than modular effects of each variable. His model acknowledges that there is interplay between clusters of learner variables and cognitive demands of L2 instruction; however, this theoretical work on aptitude complexes has not yet been empirically tested.

Skehan (1998, 2002) and Robinson (2002a) both argue that SLA research is far behind in studying the effects of IDs in relation to current issues in SLA research (e.g.,
instructional conditions and stages of L2 learning) compared to other fields such as instructional psychology (Cronbach & Snow, 1977) or psychotherapy (Snow, 1991). Robinson’s work on aptitude complexes and Skehan’s aptitude profiles are important because they establish direct connections between individual learner differences and SLA processes. Indeed, the work by Robinson and Skehan on aptitude has been the trigger for current aptitude-related SLA research. They have worked on the possibility of profiling L2 learners based on aptitude and matching these profiles with interventionist methods of L2 instruction. The need to relate learner IDs to SLA processes such as noticing and FFI has been pointed out by several SLA researchers (Dörnyei & Skehan, 2003; Ortega, 2009; Sawyer & Ranta, 1999; Skehan, 1989). SLA scholars agree on the potential value of such integrated research. Robinson (2002a) argues that when greater connections between IDs and SLA processes are established, a pedagogically-oriented aptitude research agenda could help SLA researchers explore the delivery of optimally effective classroom exposure and practice for L2 learners by matching learner strengths in particular complexes to effective instructional options such as types of pedagogic tasks, and interventionist ‘focus on form’ techniques.

**Aptitude-treatment-interaction research**

One of the earliest studies that can be categorized as pedagogically-oriented aptitude related SLA research was conducted by Wesche (1981). This was a report of a classroom experiment carried out in the Public Service Language Training programs in Canada. Wesche reported on the effectiveness of matching the aptitude profiles (and learning styles) of adult L2 learners with different teaching methodologies. She used a number of aptitude measures including the MLAT and the PLAB to identify learner profiles. Half of the learners were placed into the teaching method that matched their aptitude profiles and the other half were placed into the teaching method that was a mismatch with their aptitude profiles. The results revealed that analytic learners matched with an analytic teaching method did better than analytic learners who were matched with the audio-lingual method (i.e., memory method). Similarly, memory-oriented learners matched with the audio-visual method did better. The results suggested that when aptitude and learner abilities were matched with teaching methods, it was possible to
maximize learner gains. The design of the study is a rare one and to my knowledge there has been no other study that has had the same research design. Although SLA research acknowledges the importance of aptitude-treatment-interaction investigation, there is a lack of empirical research in the field of SLA. There are, however, continuous efforts with similar types of research objectives.

In a more recent study, Erlam (2005) investigated the relationship between aptitude and L2 learning in three different instructional settings with a pre, post and delayed-post test study design. She did not match learner profiles and teaching methods but she was able to examine three types of instructional methods in relation to language aptitude. She conducted the study with 92 high school students in their second year of French instruction. Three instructional groups received three 45-minute instructional sessions on direct object pronouns in French. The first group received deductive instruction, which included form-focused activities following rule explanation. The inductive group received input-based activities that were planned to lead them to hypothesis-testing about direct object pronouns. The last group received structured input and was provided with explicit information, rule explanation, and consciousness-raising activities but no production was included. In order to measure language gains in the target structure, listening and reading comprehension tests and written and oral production tests were used. To measure language aptitude, the researcher used several subtests: the words in sentences subtest from the MLAT, and the sound discrimination test of PLAB. Erlam was also interested in working memory (WM) capacity and the possible relationships between the three different types of instructional treatments. In order to measure WM, she used a phonological short-term memory test (i.e., multisyllabic word test) adapted from Baddeley (1999). First, the findings showed that the words in sentences subtest of the MLAT accounted for 24% of the variance in the written production in the instructional setting. Deductive instruction seemed to minimize the effect of IDs on learning because the L2 gain scores were not related to language aptitude in that context. In the structured input group, learners with greater language analytical ability and WM capacity benefited more from the instructional treatment. Moreover, WM and words and sentences together accounted for 48% of the variance on the written production delayed post test in the structured input group, although this type
of instruction did not engage the learners in any production at any stage. Similarly, learners with better language analytical ability in the inductive group benefited more from the instruction. Results related to the inductive group are similar to Robinson’s (1997) findings, which also showed that the highest correlation between language aptitude components and instruction type was between words and sentences in the implicit condition. The findings are intriguing because the lack of relation between L2 gains and aptitude in the deductive group demonstrates that the role of aptitude could be ruled out because L2 instruction provided the necessary metalinguistic information and there was no role for cognitive ability. However, in the inductive group, learner differences in aptitude made a difference because learners were encouraged to try out hypotheses and therefore drew on their cognitive abilities. Learners’ prior learning experience seems to affect which cognitive abilities are called on during instruction. Erlam argues that deductive instruction was more familiar for the learners who have a similar type of L2 experience in their regular L2 classes. Therefore, deductive instruction was successful in minimizing the role of aptitude by providing everything learners needed. I would argue that the learners in the deductive group were used to receiving explicit instruction piecemeal; therefore, they did not need any other cognitive tools. However, learners in the inductive group had to look for other resources to make hypotheses and look for systematic clues in the input. Therefore, they needed more cognitive resources. Erlam commented that the deductive instruction was more similar to the regular classroom instruction students received in their classrooms. Such familiarity with the teaching method could have minimized aptitude effects. Robinson (1997) also commented that learners in his study were more familiar with rule-learning; therefore learners looked for rules in the inductive learning situation. The results emphasize the role of prior L2 learning experience with respect to aptitude effects in L2 learning.

Sheen (2006) explored the effects of aptitude and anxiety in a quasi-experimental study investigating relationships between corrective feedback and the acquisition of articles in English. She provided four groups of English language learners with four types of corrective feedback. She hypothesized that aptitude and anxiety would mediate the effectiveness of four different types of corrective feedback (i.e., oral and written metalinguistic feedback and oral and written recasts) provided to 177 intermediate level
ESL university students. There were two treatment sessions in which two narratives aimed at eliciting the target structure ‘articles’ were presented. She used several measures to assess the acquisition of the target structure including a dictation test, a narrative writing test, and an error correction test. In order to measure the learner factors that were hypothesized to moderate the effect of corrective feedback, a questionnaire was developed to measure the learners’ anxiety and attitudes towards error correction. She used a multiple-choice test with 14 items, which was designed by Schmitt, Dörnyei, Adophs and Durow (2003), in order to assess aptitude. The test was designed to measure language analytical ability as opposed to overall language aptitude and its components. Learners were given a glossary consisting of words and sentences from an artificial language and they were asked to find the English equivalents of 14 sentences provided in the artificial language in order to figure out the underlying rules. Sheen hypothesized that the effectiveness of the different types of feedback would be moderated by the learners’ language analytical ability. The results revealed that there was a significant correlation between the students’ aptitude for language analysis and both the short-term and long-term gains in all treatment groups except for the oral recast group. Learners’ performance who received oral metalinguistic feedback showed the highest correlation with aptitude, whereas the scores of learners who received oral recasts had almost no correlation with aptitude. The fact that metalinguistic feedback correlated more with language analytical ability makes sense given that the test was developed to measure an ability that is similar to the grammatical sensitivity subcomponent of MLAT.

The findings of the study provided important implications for types of feedback and FFI techniques that can support L2 learners. Oral recasts seemed to be resistant to aptitude effects in this group of learners but feedback types that included metalinguistic information were related to aptitude. The findings were contrary to Erlam’s findings because there was no relation between aptitude and L2 gain in deductive instruction that provided metalinguistic information. Erlam examined the effects of L2 instruction at macro levels. She provided three different instruction types, whereas Sheen’s intervention included the feedback types. In order to interpret the results of these two studies a process-oriented approach to individual learner differences is needed. That is, feedback type and FFI techniques in two studies brought their own task demands and they might
have required a variety of cognitive abilities. Metalinguistic feedback and metalinguistic L2 instruction were found to have different relations with aptitude. Feedback type that includes metalinguistic information seems to be strongly related to language aptitude. Sheen’s study seemed to indicate the active role of aptitude with respect to feedback type. On the contrary, in Erlam’s study, the metalinguistic type of information provided in deductive teaching was found to be unrelated to language aptitude. I would argue that learners’ prior learning experience was the factor that determined the use of cognitive resources in the Erlam study. However, close examination of feedback types in Sheen’s study show that intervention techniques within L2 instruction were closely related to cognitive resources available. Metalinguistic information provided in the recasts was not enough to minimize the effects of aptitude. Another important finding was that oral recasts were not related to aptitude. One might speculate that benefitting from oral recast requires learners to draw on a different cognitive ability, in particular, working memory capacity. However, WM capacity was not included in Sheen’s study; therefore, it is difficult to compare the findings with regard to aptitude clusters.

Although aptitude was a predictor of L2 learning in most of the studies reviewed above, some types of instruction seemed to be resistant to differences in IDs such as the incidental group in Robinson (1997), the deductive group in Erlam (2005) and the oral recast group in Sheen (2005). Some qualities of those L2 learning conditions or FFI techniques seem to make them more appropriate for some learners and therefore, instructional excellence which implies the quality of L2 instruction as suggested by Carroll (1962) also could have contributed to L2 achievement in addition to cognitive abilities such as aptitude. Following this, in some cases, instructional excellence could be more important than language aptitude. Therefore, the effect of aptitude does not seem to depend on the distinction between explicit and implicit type of learning as suggested by Krashen and Reber, but the effect of aptitude may be more related to appropriacy of certain types of L2 instruction. However, the limited number of empirical studies investigating aptitude/instruction interactions has prevented SLA researchers from drawing any firm conclusions about what qualities of instruction relate more or less to aptitude characteristics. There is a need for more research to examine relationships between aptitude and instructional approaches in SLA.
Theoretical work on aptitude complexes and empirical studies have provided some evidence to suggest that matching learner profiles to different L2 learning contexts (e.g., grammar-based) and instruction techniques (e.g., input flood, input processing or metalinguistic instruction) has the potential to provide researchers and practitioners with better opportunities for optimally designing effective L2 classroom curricula and instructional materials (Robinson, 2001; 2002a; Skehan, 2002). This type of research “will hopefully enable SLA researchers to identify the type of instruction which plays to the strengths and preferences of certain profiles in students and will bring the most learning for them” (Ortega, 2009, p. 163-164). As indicated above, aptitude-related SLA research has investigated aptitude in relation to different types of learning (implicit and explicit) or different types of form-focused instruction (corrective feedback, inductive and deductive teaching) and context (natural and instructed settings). However, these studies have only investigated one-way interactions among aptitude and instructional type or context. Other factors in the context and instructional treatment could be moderating the effect of aptitude, such as learner’s prior learning experience, L1 and L2 similarities or differences, and/or the qualities of the target language structure chosen. Several SLA researchers have indicated that interactions among aptitude and other SLA processes are equally important to explore. For example, Robinson (2002, 2005) and Skehan (2002) proposed that different task characteristics, different sets or complexes of cognitive abilities, and affective factors interacted with different subcomponents of aptitude. It should be obvious from the above discussion that aptitude is a relevant predictor of L2 learning across different L2 contexts and across adult and young learners.

Moreover, WM has been shown to be a promising cognitive ability that contributes to L2 achievement. As indicated above, several aspects of SLA research have been examined with regard to aptitude. For example, noticing, recasts, and types of FFI. Other questions relevant to aptitude that could also be examined within the context of SLA research include type of L2 structure. For example, one might expect that the predictive validity of aptitude would differ between easy and difficult L2 structures. Instructed learners might rely more on language aptitude when hypothesizing about the structural regularities in difficult structures but not with easy L2 structures. Several researchers have conducted aptitude related research studies that have included the type
of L2 structure as a variable (DeKeyser, 1995; Robinson, 2002b). DeKeyser (2005) discusses the potential value of two-way or three-way interactions among aptitude, instructional type and target language structure to be acquired. He argues that, “... not much work has addressed the question of the differential impact of factors such as aptitude and motivation on specific elements within morphology and syntax, in other words, on elements characterized by specific types of difficulty” (DeKeyser, 2005, p.16). This is a factor of interest in the present study that is discussed in the next chapter.

**Summary**

Despite widespread acceptance that language aptitude is a cognitive ability that consistently explains variance in L2 learning, there has been a delay in the integration of aptitude research in mainstream SLA research. However, since the 1990s research on language aptitude has taken a process-oriented approach and it has resulted in a more dynamic and multicomponential aptitude construct. This new approach to studying aptitude has also resulted in new venues for aptitude research. Aptitude profiles and aptitude complexes have been introduced as new aptitude models and are based on the premise that learners vary in their strengths in terms of cognitive variables and that different abilities within aptitude complexes operate at different stages of L2 development. The role of aptitude has also been examined with respect to different types of instruction. However, no study to my knowledge has examined whether and how aptitude interacts with different types of L2 structures in the classroom setting.
Chapter Three
Grammatical Difficulty and Aptitude

This chapter discusses different conceptualizations of grammatical difficulty in the SLA literature. It also reviews empirical research that has examined grammatical difficulty in relation to differences in learner aptitude. The chapter concludes with a description of the purpose of this thesis research.

Definitions of grammatical difficulty in SLA

Within the instructed SLA literature, there has been considerable discussion about whether the effectiveness of a particular type of instruction (e.g., implicit/explicit) depends on the type of language structure taught. SLA researchers have put forth different arguments as to whether simple or complex structures should be targeted for L2 instruction. For example, Reber (1993), working in cognitive psychology, argues that if the stimulus domain is simple and the relevant structural properties are made salient, then explicit learning can be effective. Within the field of SLA, Krashen (1982) allows little room for explicit instruction and argues that only “easy” rules are candidates for explicit instruction because “hard” rules can only be acquired. On the other hand, Hulstijn and de Graaff (1994) propose that complex structures are the best candidates for explicit instruction assuming that simple structures are available in the input and are easily recognized and picked up by learners. Moreover, they suggest that the scope and reliability of a feature interact with complexity. Therefore, rules that apply to a large number of cases and hold true for more cases should be taught (Hulstijn, 1995; Hulstijn & de Graaff, 1994). Few studies have directly compared the effects of type of instruction on different target features. However, in a recent meta-analysis, Spada and Tomita (2010) categorized L2 features used as target structures from 41 SLA studies as ‘simple’ or ‘complex’ and examined the effects of explicit and implicit instruction on these features. The results indicate advantages for explicit instruction on both types of language features. However, the authors caution that the findings could be a result of the way in which the target features were classified as ‘simple’ or ‘complex’ as there are several ways in which the definition could have been formulated.
There appear to be four main perspectives on defining grammatical difficulty in the SLA literature: acquisition, linguistic, pedagogical and psycholinguistic perspectives (Collins et al. 2009, Spada & Tomita, 2010). Below I briefly describe each of these.

**Acquisition perspective**

The acquisition perspective makes a distinction between easy and difficult structures based on the acquisition order of specific language features. This approach views late acquired structures as difficult and early acquired structures as easy. For example, based on the results of the morpheme acquisition studies in the 1970s (Bailey, Madden & Krashen, 1974; Dulay & Burt, 1973, 1974), the progressive marker –ing is considered to be an example of an easy structure because it is acquired early contrary to the simple past –ed, which is acquired later and is therefore difficult. Other developmental sequences that have been observed in SLA including negation, question formation, word order and so forth. (Lightbown, 1980; Pienemann, 1989; Ravem, 1973; Schumann, 1979; Wode, 1976) can also be used to specify ease versus difficulty. The acquisition perspective provides a compelling explanation because the definition is based on empirical research on L2 development.

**Linguistic perspective**

Another way to conceptualize grammatical difficulty is based on the linguistic nature of the language feature. This approach includes three viewpoints: (1) L1-L2 differences, (2) markedness, and (3) structural difficulty. Using L1-L2 differences as a criterion for determining difficulty dates back to Contrastive Analysis (Lado, 1957). Proponents of Contrastive Analysis argued that the differences between the learner’s L1 and L2 would result in difficulty, whereas similarities between the L1 and L2 would make it easier to learn. The differences between French and Turkish can be viewed in this way. For example, grammatical gender can cause difficulty for Turkish learners of French because French possesses grammatical gender and Turkish does not. However, using only L1-L2 comparisons as a basis for ease or difficulty is problematic because there is evidence in the SLA literature of common developmental sequences for some target structures across many different L1s (Bardovi-Harlig, 2000, McDonough & Kim,
suggesting that L1/L2 differences or similarities do not consistently predict ease or difficulty of learning.

Another point of view regarding structure difficulty within the linguistic perspective is the notion of markedness. Markedness is based on typological universals or Universal Grammar. This view identifies some language structures as less natural or less frequent than others. These features are referred to as ‘marked’ and markedness can imply difficulty because according to the markedness theory, less marked structures are easier to learn. Two interrelated theories also operate under the Markedness theory: The Noun Phrase Accessibility Hierarchy (Keenan & Comrie, 1977) and Prototype Theory (Rosch, 1977, Rosch & Mervis, 1975) that are used to categorize language structures based on their distinctiveness from contrasting categories or representativeness within a category (Hu, 2002). The hierarchy or categorization in these approaches can also be used to determine qualities of the L2 structure (i.e., grammatical difficulty) because prototypical members are seen as more salient than their counterparts (Rosch, 1977). Based on Markedness theory, relative clause pronoun functioning as a subject (e.g., the footballer who scored the goal has been mobbed) is defined as less marked than a relative pronoun that functions as the object of a preposition (e.g., the house that was moved into was built in the 1920s). Definitions of difficulty have also been based on Givon’s (1995) model of functional markedness. Using this model, Housen, Pierrard, and Van Daele (2005) defined the passive voice in English as complex and French sentence negation as less complex and less marked in their instructed SLA study.

The last component considered within the linguistic perspective is the structural complexity of a language feature. Hulstijn and de Graaff (1994) provide objective criteria to define structural complexity: “the number of (and/or the type) of criteria to be applied in order to arrive at the correct form” (p.103). There is no prediction about the ease of acquisition to be concluded from this definition because researchers note that ease of acquisition depends on several factors such as processing capacity required or transparency of the form-meaning relations in a given structure. According to this approach, and contrary to the acquisition perspective, the simple past tense is identified as easy because it only involves one transformation and English articles could be defined
as comparatively simple. The proposed definition provides a compelling explanation for determining difficulty because the criteria could be applied to all natural and artificial languages in SLA research. Following this, de Graaff (1997) used the same criteria to determine structure difficulty in an artificial language (i.e., eXperanto). Similarly, Spada and Tomita (2010) examined structural difficulty by comparing the number of transformations in their meta-analysis.

Hulstijn and de Graaff’s definition has the advantage of providing objective criteria. However, it lacks support from acquisition and processing data (Collins, 2009; Spada & Tomita, 2010). Furthermore, EFL/ESL teaching experience and acquisition data provides counter evidence to the definition of English articles as simple. Articles have been acknowledged to be extremely difficult to master for many EFL and ESL groups in SLA literature.

**Pedagogical perspective**

The pedagogical perspective approaches difficulty from the point of view of the L2 learner. It focuses on the difficulty that learners experience in dealing with the L2 structure rather than the complexity of the structure itself. It is concerned with the difficulty of the pedagogical explanations provided by EFL teachers or EFL textbooks. There are several ways to assess the pedagogical difficulty of a given target structure such as the amount of metalanguage used to describe the rule (Housen, Pierrard, & Van Daele, 2005) and the number of exceptions to the rule (Hulstijn, 1995). Other ways of defining difficulty from a pedagogical perspective include asking experienced teachers and learners to rate the difficulty of different L2 features (e.g., Robinson, 1996; Shiu 2011; Van Baalen, 1983) and asking learners for rule explanations (Green & Hecht 1992).

One of the reasons for the ambiguity surrounding definitions of grammatical difficulty stems from the interchangeable use of the terms pedagogical rule difficulty and structure difficulty. Some researchers (Robinson, 1996; Van Baalen, 1983) use the term ‘rule’ referring to the pedagogical explanation attached to the grammatical structure
whereas others (Hulstijn & de Graaff, 1994) refer to target structure as a linguistic entity. For the purposes of this study, ‘structure difficulty’ will be used.

Individual learner differences can also determine the difficulty of a given L2 structure. DeKeyser (2005) makes a distinction between objective and subjective difficulty. Subjective difficulty depends on the L2 learner’s perception and experience of what is difficult or easy whereas objective difficulty is based on acquisition or linguistic perspective discussed above. DeKeyser states that subjective difficulty can be influenced by such factors as language learning aptitude and previous L2 learning experience. He further states that, “rule difficulty is an individual issue that can be described as the ratio of the rule’s inherent linguistic complexity to the student’s ability to handle such a rule. What is a rule of moderate difficulty for one student may be easy for a student with more language learning aptitude or language learning experience” (DeKeyser, 2005, p.331). DeKeyser’s view is based on the interplay of linguistic difficulty and learner variables such as language aptitude, prior learning experience and such. The pedagogical approach to difficulty has ecological validity because it takes the agents of the learning into consideration.

**Psycholinguistic perspective**

The final approach in defining grammatical difficulty is referred to as the psycholinguistic perspective. Collins et al. (2009) argue that this approach places input at the center because proponents ascribe a great role for learners’ experience with language input for L2 success. Following this, input frequency and perceptual salience are considered to be two major determinants of difficulty in the psycholinguistic approach. According to N. Ellis (2006a, 2006b) L2 learning is a kind of contingency learning and L2 learners constantly make predictions about possible structures out of what they encounter in the language input. N. Ellis further states that, “Frequency is thus a key determinant of acquisition because rules of language, at the levels of analysis (from phonology, through syntax, to discourse), are structural regularities that emerge from learners’ lifetime analysis of the distributional characteristics of the language learners have to figure out” (2002, p. 144). According to this approach, then, learners’ access to
the target structures in the input is the main determinant of what is easy and difficult (N. Ellis, 2006). However, frequency alone cannot explain why English articles, which are frequently used in the input, are notoriously difficult for many L2 learners.

DeKeyser (2005) proposed three additional factors that contribute to the grammatical difficulty of frequently occurring features in the input: complexity of form, complexity of meaning, and complexity of form-meaning relationships. Complexity of form is defined as “the number of choices involved in picking all the right morphemes and allomorphs to express these meanings and putting them in the right place” (DeKeyser, 2005, p.6). Learners of inflected languages would experience more difficulty than others since there are more choices to be made about morphemes, allomorphs, and their positions. Thus, the past progressive in English could be argued to be an easy structure since it does not involve any allomorphs (Collins et al., 2009). Complexity of meaning is described as novelty or, abstractness, or a combination of both. Novelty could be related to L1/L2 differences. For example, articles, classifiers, grammatical gender, and verbal aspects would be complex for learners whose L1 does not have equivalent forms. In the SLA literature there is ample evidence that these grammatical structures are notoriously difficult to acquire (Ayoun, 2004; Ishida 2004; Leeman, 2003; Butler, 2002; Master, 1997).

When there is no problem of form and meaning, a given structure can still cause difficulty “when the link between form and meaning is not transparent” (DeKeyser, 2005, p. 7). That is, while learners are trying to figure out the form and meaning relationship, three factors are thought to make the situation more difficult: redundancy, optimality, and opacity. When the meaning of a form is expressed by at least two elements in the structure, one of them becomes redundant and, therefore, is difficult to grasp. For example, the simple past in English is indicated with the verb ending and a time phrase (e.g., I walked to school yesterday). Past tense marker –ed becomes redundant since the time phrase yesterday expresses the meaning. Similarly, the alternating presence and absence of certain structures results in optionality and, therefore, leads to difficulty. Another factor that causes a lack of transparency is opacity. Opacity is an issue when different forms stand for the same meaning or the same form stands for a number of
meanings. The third person –s is a perfect example for this case because –s in English can be the third person singular of the verb, the plural of the noun, or the genitive of the noun. To make things more complicated, in each case there are three allomorphs. Learners have to look for cues (i.e., phonological, morphological, or discourse) in order to figure out the relation between the form and the meaning.

Finally, salience has been identified as an important factor in determining whether a given structure is easy or difficult to learn (Doughty & Williams, 1998). Saliency is defined as how easy is it to see, hear, notice, and perceive a structure in the input. Given the fact that attention plays an important role in L2 learning (Robinson, 1995; Schmidt, 2001; Skehan, 1998), perceptual saliency of an L2 structure increases the likelihood of noticing and acquisition. Saliency is closely related to frequency in the input because frequency of a given L2 structure helps learners observe regularities. Perceptual saliency is concerned with the linguistic attributes of the L2 structure and it contributes to how easy it is to recognize the language form (Collins et al., 2009). In a recent meta-analysis to determine the factors contributing to the acquisition order of English morphemes, Goldscheider and DeKeyser (2005) provide convincing evidence that salience is primary. Other factors identified were semantic complexity, morphological regularity, syntactic category, L1 transfer, and frequency. As is evident from the findings, a number of factors contribute to the saliency of L2 forms. There is a complex interplay of the five factors identified; therefore, it is still difficult to account for the fact that L2 learners are exposed to frequent use of English articles in the input but they still remain difficult to acquire. Nonetheless, the salience of some L2 structures seems to make some L2 structures more “available” to be noticed and then acquired. Another example that shows priority of saliency over markedness is provided by Bardovi-Harlig (1987). In her study she provided evidence that preposition stranding was acquired before preposition pied piping although the latter is considered to be the less marked counterpart.

In summary, scholars from different research paradigms have used different criteria for defining L2 grammatical complexity. The fact that, in their meta-analysis, Spada and Tomita (2010) found eight different approaches to describing complex/simple or difficult/easy L2 structures shows that SLA research is far from having common
criteria to define difficulty. Each approach has valid arguments based on linguistic, psycholinguistic and pedagogical perspectives. Thus, one could only define L2 difficulty in relative terms and the operationalization of L2 structure difficulty based on more than one theoretical framework may be more advantageous.

In the following section, two English structures chosen for the purposes of the current study are introduced and defined with respect to structure difficulty.

**Target Features of the Current Study**

The two language features targeted in the present study are the “passive voice” and the “past progressive tense” in English. Following Hulstijn and de Graaff’s (1994) criteria for structural complexity the current study defines the passive voice in English as a difficult L2 structure and the past progressive as an easy target structure. However, as is evident from the discussion above, no single definition is sufficient to define any given L2 structure as difficult or easy. Therefore, in the current study, other criteria to determine the relative difficulty or ease of the target features are used (e.g. L1/L2 differences). Below is a description of how these features have been examined in the SLA literature.

The passive voice has been investigated in a number of SLA studies (Izumi & Lakshmanan, 1998; Ju, 2000; Kuiken & Vedder, 2002; Spada, 2010; Yip, 1994; Williams & Evans, 1998) and it has been reported to be a difficult structure for L2 learners. Several reasons are given for the difficulty of passives. Williams and Evans (1998) argue that the passive voice is not salient owing to the infrequency of the input. Moreover, the formation of the passive is opaque and not straightforward. In terms of formal complexity, the passive requires subject-verb agreement and tense agreement. There are allomorphs of the verb ‘be’ forms (i.e., is, are, was, were) and ‘be’ can be used as a main verb as well. The passive is also reported to be underused in L2 written and oral production (Hinkel, 2004).

When L1 and L2 similarities and differences with respect to the passive voice are considered, the passive voice can be thought of as a difficult structure for Turkish
learners of English due to the syntactic and morphological differences between the two languages and differences in how the passive voice is realized in Turkish. Aksu-Koc and Slobin (1985) report that Turkish L1 learners are very successful in acquiring grammatical morphology in Turkish which is an inflectionally rich language. Turkish morphology is also reported to be remarkably regular and transparent. L2 structures which are not transparent are more likely to be difficult for Turkish learners. The passive voice in English is defined as opaque (i.e., not transparent) for L2 learners. According to DeKeyser (2005) structures are opaque when the relationship between the form and function is difficult to detect. Izumi and Lakshmanan (1998) provide several factors that could contribute to lack of transparency with the passive. The internal argument of the verb appears in the subject position of the passive sentence. The verb morphology also changes in the passive. There is passive auxiliary be and the main verb is realized in its past participle form. In comparison, in order to form passive sentences in Turkish, the suffix –il is added to the verb stem. However, there are some variants of the passive suffix depending on the last letter of the verb stem (i.e., -n, -ln). Lewis (2000) states that the “passive is not used as much as its English equivalent for the sake of elegant variation: e.g., instead of ‘he was rebuked by his father’ a Turk is more likely to say ‘his father rebuked him’ (p. 152)”.

When a transitive verb is passivized, it is usually done for topicalization purposes. The process of passivization makes the direct object of a transitive verb the subject of a passive one. This is similar to how English passive voice works. The examples below are taken from Göksel and Kerslake (2005, p. 149).

**Example 1:**

EV bu yil artik sat-il-acak. (cf. Biz evi bu yil satacagiz.)

house this year finally sell-PASS-FUT

‘The house will finally be sold this year.’

The most remarkable feature of the passive voice in Turkish is the impersonal passive. The addition of a passive suffix to an intransitive verb produces an impersonal passive voice. This type of passive includes no person or group of persons that is
understood to be performing the action denoted by the verb and the closest English equivalent sentences are active sentences with ‘people’, ‘one’ or the impersonal ‘you’ as subject (Göksel & Kerslake, 2005). Consider the following example:

**Example 2:**

Adalara artik denizotobusuyle mi gid-il-ecek?

[Go-PASS-FUT]

Will hovercraft be the way [people] go to the islands? *(This sentence is ungrammatical in English)*

Turkish also allows for double passive voices. That is, when a transitive verb is combined with two passive suffixes, it results in an impersonal passive that refers to a property of a particular entity. However, these constructions are quite marginal. There are some other irregular cases as well but they are not as frequent as impersonal passives (see Goksel & Kerslake, 2005 for details). The passive is realized in all tenses in Turkish similar to English. However, the teaching of the passive voice in the current study included the tenses that the learners had already been taught prior to intervention (i.e., past simple, present simple, present progressive, past progressive, and present perfect).

The “easy” structure chosen for the present study is the past progressive tense in English. SLA studies that have included the past progressive have considered it to have a transparent form-meaning relationship and to be salient in the input (Collins et al. 2009; Révész, 2009; Révész & Han, 2006). Révész (2009) states that the past progressive in English is physically salient as realized via a free morpheme (was/were) and syllabic bound morpheme (-ing). Collins et al. (2009) list several reasons why the past progressive is more available and accessible in the input:

- The progressive occurs across a wide range of common verbs and is less restricted to the prototypical activity class of verbs compared to other tenses. (i.e., simple past)
- The progressive occurs across a range of highly frequent English verbs.

43
All progressive markers appear as intact syllables; therefore, the progressive is more phonetically and perceptually salient.

There are no allomorphs for the –ing marker that would lead to less saliency.

The difference between the simple past and the past progressive is described as grammatical aspect in English. Aspect refers to internal make-up of events compared to tense which locates events or situations in time. *John sang* and *John was singing* show a contrast in aspect, both sentences being in the past tense (Comrie, 1976, 1985 cited in Bardovi-Harlig, 1992). When L1-L2 contrasts are considered, the past progressive tense also has a corresponding structure in Turkish: however, there are differences as to how this tense is realized in Turkish. The primary tense differentiation is between the past and the non-past in Turkish and -(y)DI is the past copula marker and it expresses past tense. An imperfective aspect is expressed by the verbal suffixes -(I) yor, -mAktA, and -(A/l)r (Goksel & Kerslake, 2005). As Turkish morphology allows for the co-occurrence of several suffixes in a fixed order, the past progressive is expressed by adding both the past tense marker -(y)DI and the imperfective marker to the verb stem:

**Example 3:**

James bir bankada calis-yor-du

[WORK  Imperfective  Past Copula]

James was working in a bank

**Example 4:**

Saat ikide calis-iyor-du-m

[WORK- ImP- Past COP- 1st person singular]

I was working at two o’clock
Grammatical difficulty has been investigated in relation to FFI techniques in a number of studies; however, only two studies have investigated the potential interaction between grammatical difficulty, instruction, and aptitude. These studies are reviewed in the following section.

**Grammatical difficulty and aptitude in SLA studies**

Robinson (1997, 2002b) investigated different rule types (i.e., complex and simple) in relation to L2 learning and IDs (i.e., intelligence and aptitude) in two studies. In his first study (1997), he included difficulty of the target language structure as a variable to investigate the effects of explicit and implicit learning in a laboratory setting. He investigated whether aptitude affected learning in four different learning conditions (incidental, implicit, rule-search and explicit) and examined whether aptitude was related to learner awareness developed through the four learning conditions. To determine the difficulty of language structures in his study, Robinson asked fifteen ESL teachers to rate the complexity of the forms that were judged to be ungrammatical in grammaticality judgment tests by ESL learners. This procedure revealed that pseudo clefts of location (e.g., *where Mary and John live is in Chicago not in New York*) were defined as a hard rule and subject-verb inversion in sentences with fronted adverbials of movement was an easy structure (e.g., *into the house John ran/ran John*). Robinson measured the foreign language aptitude in 104 non-native speakers of English using grammatical sensitivity and paired-associates subtests of the MLAT. Participants were exposed to the easy and hard rules in four different learning conditions (i.e., explicit, implicit, rule-search, and incidental) and they completed a grammaticality-judgement task which measured L2 learning. The results revealed significant correlations between performance on easy/hard rules, and memory and grammatical sensitivity in all three types of instruction: implicit, rule-search and instructed learners. However, performance on easy and hard rules was unrelated to aptitude measures in the incidental condition. Robinson interpreted these results to mean that all instructional conditions were fundamentally similar because participants used conscious processing strategies that were triggered by the task demands in those conditions except for the incidental type. However, the tasks presented to learners in those conditions differed in their demands on cognitive resources. Therefore,
some conditions predisposed learners to rely on memory whereas others required the use of grammatical sensitivity. Learning in any condition was reinforced when there was a match between learners’ strengths in memory or grammatical sensitivity. He also stated that superior learning was a result of awareness prompted by aptitude. When learners’ performance on easy and hard rules was compared it was revealed that the performance of instructed learners on hard rules was higher than all participants in other conditions. Robinson concluded that different performance across easy and hard rules in four different conditions was the result of task demands (1997). The study showed that aptitude was a positive contributing factor in three L2 instructional types and it was also related to learners’ awareness level.

In Robinson’s second study (2002b), he examined whether aptitude interacts with rule difficulty across three types of L2 instruction: implicit and explicit, and incidental. He used series solution tasks to measure the learning of Samoan. Three rules were selected that were different in terms of complexity and they were also structurally different from the participants’ native language. Robinson used the Japanese version of the MLAT in order to measure language aptitude. He also measured WM with a reading span test. Several language measures were administered including computerized grammaticality judgement (GJ) tasks, listening GJ tasks, and a guided sentence production test. The results revealed that aptitude and working memory interacted with rule difficulty. That is, performance on one of the target structures (i.e., locative sentences) was positively related to aptitude but not to WM whereas performance on ergative rules was negatively related to aptitude on language tasks. The findings, according to Robinson, demonstrated that incidental learning was affected by the nature of the rule to be learned. Robinson interpreted these results to mean that “…adult incidental L2 learning is invariably affected by the complexity, and the nature of the rule to be learned, and IDs in cognitive abilities interact with rule complexity and rule type during incidental learning” (2002b, p.260). Robinson’s (1997, 2002b) studies demonstrated that, contrary to earlier arguments by Reber et al. (1991) and Krashen (1982), aptitude was related to incidental learning, which is meaning-focused and thus, considered by some researchers to be resistant to individual learner variables. Second, he was able to show that rule complexity interacted with incidental L2 learning. Robinson’s
work remarked the potential value of examining L2 structure difficulty in relation to aptitude complexes or aptitude profiles within SLA research.

A series of laboratory studies that have investigated the role of memory, attention and inductive learning have provided important evidence for the possible interactions reported in Robinson’s studies. For example, Williams (1999) investigated the relationship between memory and inductive learning of morphological rules in a semi-artificial form of Italian. The performance of learners across two types of rules in the target language was measured through a phonological long-term memory test. The results showed that both types of rules were related to memory but form-form types of rules were mostly learned implicitly whereas rules that include form-function mappings were learned via explicit teaching. DeKeyser (2005) interpreted these results as meaning that explicit learning of form-function mappings through explicit teaching was related to grammatical sensitivity whereas implicit learning of semantically redundant form-form rules was related to a function of the ‘memory’ sub-component of language aptitude. Thus, in addition to Robinson (1997, 2002b), Williams’ (1999) study provided evidence that some language structures may be better learned through the help of certain components of language aptitude.

Research that has examined how aptitude interacts with type of L2 structure has taken place in laboratory contexts and in some cases with artificial languages. Therefore, there is a need to further examine the same variables with natural languages in the classroom context. A close examination of language aptitude with regard to type of L2 structure could reveal important information about whether aptitude operates to help L2 learners overcome L2 structure difficulty. If so, aptitude could be viewed as a cognitive tool to support L2 instruction, particularly in EFL learning where there is a major role played by attention to form. Information on learners’ aptitude profiles could be used to play to learners’ strengths in L2 instruction. Difficult L2 structures could be introduced and taught in line with learners’ aptitude profiles. In the case of easy structures, for example, there might be less need for explicit instruction; therefore, more meaning-based activities could be used. Furthermore, learners’ age could also be considered to be a factor. Robinson (2002b) provided evidence for aptitude and rule difficulty interaction in
adult L2 learning. There is a need to examine this relationship with different age groups. Therefore, young adolescents were selected for this study.

To summarize, recent work on aptitude has revealed interaction effects between subcomponents of aptitude and L2 instruction on learning. Furthermore, the nature of L2 structure complexity is considered to be an important variable when investigating relationships between L2 learning and aptitude in instructed SLA. Robinson (2002) proposes that “aptitude complexes” interact with rule type and suggests that the maintenance of learning may draw on different clusters of cognitive abilities and capabilities (i.e., WM, grammatical sensitivity, and inductive language learning ability). DeKeyser (2005) also argues that that nature and degree of difficulty of target language structures interact with the individual learner differences and this interaction has a number of implications for instructional decisions. If there is an interaction between aptitude and the type of L2 structure, then instructional techniques can be offered to help learners make better use of their capabilities. Similarly, Robinson (2002a) states that the “full range of FFI techniques from the least communicatively intrusive ‘input flood’, to the most intrusive, ‘input processing’ or ‘metalinguistic instruction’ can be expected to be differentially related to clusters of cognitive processing and learning abilities, i.e., aptitudes” (p. 131).

The current study aims to examine whether language aptitude is related to two L2 structures defined as easy and difficult in a foreign language curriculum. The main aim of the study is to explore whether language aptitude is a cognitive ability that can boost L2 learning across two types of L2 structures. In the next chapter the research questions and methodology are outlined.
Chapter Four

Methods

This chapter begins by describing the study design and stating the research questions that motivated the study. This is followed by a brief description of the pilot study that was conducted for validation and reliability purposes. Next, a description of the data collection for the main study is described with details of the research site and the participants, the instruments, and data collection procedures.

Research Design

The study employed a pretest-treatment, and posttest design using intact EFL classes. The instructional treatment, one of the independent variables, included four hours of L2 instruction. As indicated in Chapter 3, the two target structures were the passive voice and the past progressive in English. Five individual learner difference (ID) variables were also examined: L2 motivation, L1 metalinguistic awareness, working memory capacity, learners’ perception of L2 structure difficulty, and foreign language aptitude. These specific ID factors were chosen because they were assumed to be closely linked to L2 learning. L2 learning was assessed through two dependent variables: a grammaticality judgment test and an oral production test developed for both target structures.

Research Questions

The study aims to answer the following research questions:

1. To what extent do individual learner variables (i.e., L2 motivation, L1 metalinguistic awareness, working memory capacity, learners’ perception of difficulty, and language aptitude) explain the effectiveness of L2 instruction on selected target structures?
   a) To what extent do IDs explain the effectiveness of L2 instruction on difficult L2 structures?
   b) To what extent do IDs explain the effectiveness of L2 instruction on easy L2 structures?
2. Do specific aptitude components explain the effectiveness of L2 instruction on selected L2 structures?
   a) Do aptitude components explain the effectiveness of L2 instruction on difficult L2 structures?
   b) Do aptitude components explain the effectiveness of L2 instruction on easy L2 structures?

3. What is the effect of difference in aptitude profiles on the learning of selected structures?
   a) Is there an effect of difference in aptitude profiles on the learning of difficult structures?
   b) Is there an effect of difference in aptitude profiles on the learning of easy structures?

Pilot Study

In order to examine whether all materials and instruments were functioning as planned, a pilot study was carried out. This took the form of a free summer English language course that I taught in Turkey. The course was open to school-aged learners in grades 6, 7 and 8 and was completely voluntary. The students (and parents) who expressed interest in participating in the study were all from public schools. This was likely due to the fact that, in Turkey, public and private schools differ in the amount of English language instruction provided. In public schools, learners receive four hours of English per week at the secondary level whereas in private schools, it can be as many as eight hours a week. As a result, there is a substantial difference in the English language proficiency between public and private school students in Turkey.

The language course was organized in three different sessions over a three-month period (June to August, 2010) to reach as large a number of students as possible. Each session included an average of 15 students. As the L2 proficiency of the learners was not the same, it was necessary in some instances to teach several extra hours so that all learners had the prerequisite knowledge to deal with the target structures (i.e., passive voice and past progressive). Learners were pre-tested on two grammaticality judgment
tasks (GJT) and two oral production tasks (OPT) before the target structures were taught. This was followed by four hours of instruction per target structure. The type of instruction learners received was form-focused instruction (FFI), which included attention to form within the context of meaningful practice (Spada, 1997, 2010). The amount of time spent on form and meaning was the same for both structures. Turkish was used when necessary to explain how the structure worked in English but no comparisons were made between the two languages because I wanted the learners to figure out the similarities and differences between two languages for themselves. Several individual learner difference (ID) tests were administered during the instructional sessions but not during the teaching of the two target features. After the instruction, learners completed the posttests. Each group of learners took the delayed posttest two weeks later.

The pilot study served two major purposes in the preparation for the main study. First, it helped to identify the appropriate age group of learners. Second, it explored the ‘teachability’ of the instructional materials. The instructional materials were prepared for EFL learners with at least pre-intermediate proficiency. However, it was difficult to anticipate which grade would be the right target group as there are no benchmarks in language classes in the Turkish educational system. Thus, the materials were taught to a range of learners (i.e., 6th, 7th, and 8th grades) in order to find the right group. It soon became clear that 6th and 7th grade students did not have enough proficiency in English to follow the materials. This was particularly the case with the passive voice. Students in the 8th grade seemed to have a better understanding of the content and were able to follow the classes but there were very few 8th grade students in all sessions, the majority being 6th and 7th grade learners. This led me to conclude that the instructional materials and tests developed for my research were probably more appropriate for 8th grade learners and in particular, 8th grade students in private schools who would have stronger L2 ability.

The last purpose of the pilot study was to test the measurement instruments for reliability and validity. The following instruments were piloted with the learners in three classes:
1. Two GJTs for the passive voice and the past progressive
2. Two OPTs for the passive voice and the past progressive
3. A computerized foreign language aptitude test (LLAMA)
4. A foreign language motivation questionnaire
5. An L1 metalinguistic awareness test
6. A learner profile questionnaire to determine learners’ perception of difficult and easy structures
7. A word test to measure WM capacity

When the reliability of the GJT was examined, moderate levels of internal consistency were observed for the past progressive GJT (Cronbach’s alpha .78, .62, .87 for pre, post and delayed posttest respectively) but low levels for the passive GJT (Cronbach’s alpha .31, .62, .35 respectively). This was anticipated because the learners’ proficiency was insufficient to understand the passive voice. In addition to reliability analyses, item-analyses were conducted on the GJT, and some items were revised in terms of word choice. Learners also experienced considerable difficulty with the oral production task. Very few of these learners had previous experience with oral language measures and they did not produce the targeted structures. This was another reason why I decided to conduct the main study with learners from private schools where they are often required to do oral exams in the L2.

All the measures of individual learner differences worked fairly well except for the word test for working memory capacity. The test was a PSTM test used in previous SLA studies (Erlam, 2005). It was developed to assess WM capacity for processing information and it was a word test of phonological loop (Baddeley, 1999). The original test required participants to orally repeat a list of five words they had read within certain time constraints. A written version of the WM test was used following Erlam (2005). In this version participants were presented with a list of five-syllable words on computer screens. They had 7.5 seconds to view each list, as suggested by Erlam, and then were given a further 20 seconds to write the list down. The test was delivered via powerpoint and students were instructed in Turkish. They didn’t have to touch the screen or the mouse during the test. The test was timed. The five-syllable words were chosen from a
corpus (Sak, Gungor, & Saraclar, 2008) based on their frequency counts in the database. Since there was no available WM test validated for Turkish, I developed one for this study following the criteria used in the development of other WM measures. However, the choice of words was difficult because five-syllable words are usually compounds in Turkish and they are mostly old words that are probably not familiar to the age group of the participants. Students were given one point for each correctly remembered word. Due to time limitations, it was not possible to administer the test twice. Therefore, test-retest reliability could not be conducted. Instead, internal reliability was examined and the results showed that the test had a low reliability. Therefore I decided to use another instrument for WM in the main study – one that would eliminate the effects of the problematic vocabulary items. Thus a backward digit span was used in the main study.

Piloting also resulted in considerable improvements in the language testing instruments, especially the oral production test for the past progressive. While the task succeeded in eliciting the target structure, it did not do so consistently. The storyline was changed and the new version was piloted with four native speakers, three bilingual (English-Turkish and English-Croatian), and one advanced L2 speaker. Their performance confirmed that the revised version more consistently elicited the past progressive and this was used in the main study. The second oral production task targeting the passive voice was developed by Spada et al. (2011). It has been used extensively with other English language learners from different L1 backgrounds.

Main Study

In the following section, details of the main study are provided. The same pre and posttest design was used as in the pilot study once revisions were completed on the instructional materials, language measures, and WM capacity (i.e., backward digit span task). See Figure 2 for an overview of the research design.

Participants

The participants were EFL learners enrolled in the 8th grade at a private secondary school in Turkey. Three intact classes participated in the study and the learners were
reported to be at a pre-intermediate level of English proficiency. For all participants, English was an obligatory subject and they received seven hours of English instruction per week. The participants were homogenous in terms of language and cultural background. All three classes were taught by the same EFL teacher, who agreed to collaborate with the researcher. The curriculum was based on commercially published English textbooks. The textbook they were using was *Up Beat: Pre-intermediate* (Freebairn, Bygrave, & Copage, 2009). Although the general aim of English language teaching at this school was to improve L2 communicative language skills, there was a strong emphasis on language form due to a nation-wide examination (i.e., National Proficiency Assessment test) that contains 17 English grammar questions; teachers have to prepare students for this test. Therefore, it can be said that the pedagogical approach was a combination of communicative language teaching with a heavy focus on language forms.

Background information for the participants is provided in Table 3. In total, there were 37 female and 29 male students. All participants were 13 or 14 years old, and they were native speakers of Turkish. They had received between four and eight years of English instruction prior to the study (*M* = 7.7 years). All students were also studying an additional foreign language (French or German) at school. Most of the students had never visited an English-speaking country except for seven students who had spent between two weeks to two months abroad.

**Table 1**

*Background Information for Participation Three Classes*

<table>
<thead>
<tr>
<th>Class</th>
<th>Age M (SD)</th>
<th>Gender</th>
<th>Mean length of English study M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13.2</td>
<td>13 females 8 males</td>
<td>6.2</td>
</tr>
<tr>
<td>B</td>
<td>13.5</td>
<td>11 females 11 males</td>
<td>6.0</td>
</tr>
<tr>
<td>C</td>
<td>13.5</td>
<td>13 females 10 males</td>
<td>6.7</td>
</tr>
</tbody>
</table>
Target linguistic structures

The main aim of the study was to explore whether and how language aptitude interacts with L2 structure difficulty in L2 instruction in an EFL context. To this end, I included two target structures that are considered to differ in terms of structure difficulty. L2 structure difficulty was defined with respect to structural complexity based on Hulstijn and de Graaff’s (1994) criteria. According to their criteria, a language structure which includes more grammatical operations to arrive at the target form is more difficult compared to others. Using these criteria, the passive voice in English was defined as the difficult structure in the current study. The past progressive tense was defined as relatively easy because it involves fewer operations compared to the passive. Moreover, previous SLA studies have defined the progressive as relatively easy (Collins et al., 2009, Révész, 2009). However, as indicated above, one criterion is not enough to describe any L2 structure as difficult or easy. Therefore, based on my experiences as an EFL teacher and the relevant SLA literature (Collins et al., 2009, Révész, 2009, Spada, 2010) I also took L1-L2 similarities and differences into consideration.

Following Swan (1995) the teaching of the easy structure—the past progressive—focused on the two contexts in which the past progressive is used (p. 395):

a) Indicating that something was in progress around a particular time in the past (prototypical use/context)

* e.g., *I was studying for the exam at ten o’clock last night*

b) Referring to a longer ‘background’ action or situation; the simple past refers to a shorter action or event that happened in the middle of the longer action, or that interrupted it.

* e.g., *while I was running to catch the bus I dropped my books*

Instructional treatment

The instructional treatment was provided by the researcher. I met the English language teacher while I was seeking for a volunteering language teacher. He had been
working at the specific school for five years. He was the English language teacher for all 8th grade students at the school. School administration allowed me to conduct the study with the target groups and they also let me provide the instructional treatment as the language teacher. I replaced the regular language teacher for eight hours during the instructional treatment. I had no familiarity with the language teacher and the school before the study had taken place. I spent some time at the school in order to familiarize myself with the school procedures. I taught each class separately based on their weekly schedule following the same instructional design.

The instructional materials were content-rich and appropriate to the linguistic level of the learners. Consistent with Spada’s (1997, 2010) definition of form-focused instruction, the overall focus was on meaning. The grammar instruction was integrated into communicative practice and activities. Corrective feedback was provided both implicitly and explicitly throughout all activities. The amount of grammar instruction provided for the two target features was the same – both received four hours of instruction.

In the textbook, *Up-beat: Pre-intermediate*, the past progressive was included in the first unit. Therefore, L2 instruction provided for the current study included an overview of the past progressive with more emphasis on *while* and *when* as clauses and did not include detailed information on how to form simple past progressive sentences. The learners were assumed to have covered that in a previous book in the series. The passive, on the other hand, was included in the 7th unit within the course book. When the instructional intervention started, they were about to finish the 6th unit and begin the new unit on the passive voice. Instead of covering the new unit from the course book, the learners were taught the passive voice with the instructional materials that were prepared for this study (See Appendix A).

The lessons for both structures were planned with meaning-oriented activities. Attention to form was provided in the form of 5 to 10 minute activities integrated with the flow of meaning-oriented activities. The explanation provided for both structures was kept concise. The following excerpt shows how the passive voice was introduced:
After working with some sentences about their truth value, the teacher writes the following on the board and introduces:

“Researchers discovered vitamins 100 years ago.”

The teacher asks learners:

*What is the verb? What is the subject of the sentence? What is the object? Who/What is the receiver of the action?*

She underlines the subject, verb and the object on the board and tells them that this sentence is known as active and explains:

*The subject is responsible for the action of the verb.*

Then she writes the following sentence on the board below the first one:

“Vitamins were discovered by researchers 100 years ago.”

The teacher asks learners:

*What is the difference? What is each sentence about? Who does the action? Who/What receives the action?*

The teacher elicits responses and then provides the following explanation, being sure to underline the parts of the sentence that she mentions, e.g., the past participle, as she describes the verb construction of the passive sentence:

*The first sentence is about what researchers do. This sentence is in the active form. The second sentence is about vitamins. The second sentence is in the passive form. In the passive, the verb is made of two parts: the verb “to be”, in this case in the past tense, and the past participle of the verb, which never changes. In the passive sentence, the subject is the person or thing that received the action. It comes before the verb. The person or thing that did the action is called the “agent”.*

In a meaning-based activity that followed the explicit information, the teacher asks learners to work with a partner to find the answers to the true/false questions about the history of medicine inventions by looking at the information on the timeline on a handout (See Appendix A for the handout and the details of activities).

In a similar way, the past progressive is introduced in a meaning-oriented activity, and then learners are provided with a brief explanation about the form and use. While students talk about the questions, the teacher distributes a handout that includes facts about Einstein’s life. She asks learners to work with their partner to answer true or false
and explain their decisions. After the class finishes the true/false activity, she writes the following sentences on the board.

_Einstein immigrated to the U.S. because he was not happy with his life in Germany._

_He was living in Germany in 1900._

She asks learners:

*What is the difference between these two sentences?*

She tries to elicit responses and then provides the following explanation, underlining the parts of the sentence that she mentions (e.g., past form of the verb, auxiliary be)

*The first sentence is in the simple past tense and it indicates that an activity or situation began and ended in the past. The actions can be single actions, or they can be repeated actions.*

*They can take place over a period of time. The second sentence is in the past progressive tense. The past progressive shows that action was in progress at a particular moment or over a period of time in the past. The past progressive emphasized that the action was in progress. In contrast to the simple past, the past progressive does not specify whether the action was completed. The simple past emphasizes completion of the action. In order to form past progressive we use (simple past of to be verb) WAS and WERE and add –ing to the verb stem.*

As shown above, the structures were introduced by providing concise explicit information about their formation and the activities that followed aimed at providing meaningful practice of the L2 structures. The activities were following a theme: “health” for the passive and “history” and “mystery” for the past progressive tense. Meaning-oriented activities for the passive included reading to confirm expectations and discussing the arguments in the texts (e.g., expressing opinions about health practice and so forth). The activities were planned as pair work or whole group discussion. Sentence completion tasks were also used in conjunction with reading tasks.

The theme “history” and “mystery” for the past progressive was used to attract learners’ attention to meaning. Learners were engaged in reading activities to learn about a mystery and then participated in small group discussion to figure out the mystery. There were some short (e.g. five-minute activities) that introduced the structural formation of questions and negation.
Instruments

The study included several measures in order to assess L2 learning outcomes and individual learner differences. I developed or adapted several research instruments to examine these variables. The following section describes the assessment tasks along with the procedures followed during the administration of the testing instruments. As detailed above, all instruments were subjected to piloting.

![Independent Variables Diagram]

**Figure 2. Research Design of the Current Study**

**Assessment tasks for L2 proficiency**

**Grammaticality judgement task**

In order to assess changes in participants’ L2 knowledge of the past progressive and the passive, two GJTs were used. GJTs are frequently used in SLA research and they are assumed to measure L1 and L2 learners’ underlying grammatical competence (Ellis, 1991; Sharwood-Smith, 1988). Many SLA scholars have used grammaticality judgment tasks to examine various L2 structures (Bley-Vroman, Felix, & Ioup, 1988; Ellis, 1991; Gass, 1994; Mandel, 1999; White 1985, 1987). Some of the results have revealed
inconsistencies in learners’ judgments but when designed carefully, the GJT can be a useful instrument to investigate L2 competence (R. Ellis, 1991).

When designing GJTs for both L2 structures, practical considerations about the number of sentences and timing were taken into consideration. Gass and Mackey (2007) recommend between 50 and 60 sentences but they also acknowledge the fact that the longer the test, the more reliable it is. There is a wide range in terms of the number of sentences included in GJTs. There are long GJT tests with more than 200 sentences (Johnson & Newport, 1989) and moderately short ones with 24 sentences (Spada et al., 2011). The grammaticality judgment task in the present study consisted of 40 sentences for the past progressive structure and 40 sentences for the passive voice.

Sixteen different items represented the prototypical context of the past progressive (i.e., indicating something was in progress around a particular time in the past) and 24 items represented the second context where the past progressive is used to describe a longer “background” action. The 16 items representing the prototypical use were adapted from Révész (2007). Following Révész, out of the 16, eight contained activity verbs and eight contained accomplishment verbs, based on Vendler’s (1967) classification of verbs. Half of the items were grammatical and the other half were ungrammatical.

I included the second context for the past progressive in order to determine whether learners had control over all contexts for this specific tense. Twenty-four items of the progressive section were included to measure the second context. Those items included while and when time phrases, and they included a past progressive and a simple past component (e.g., My father was still sleeping when I left for school or While I was going to the yoga class I stopped to buy a magazine). Those 24 items contained 10 activity verbs and 14 telic verbs based on (Bardovi-Harlig, 1987; Collins, et al. 2009). A while clause requires a progressive marker within the same clause and a simple past or a past progressive in the second clause (e.g., While I was watching television last night I heard about the plane crash). Similarly, a when clause usually requires a simple past tense with a past progressive in the second clause (e.g., When I met Tarkan last summer I
was going to Paris). As with the other items, half were grammatical and the other half were ungrammatical.

The error types for the first context of the past progressive included: (a) use of bare progressive in past progressive contexts, and, (b) use of present progressive in the past. For the second context, in addition to the previous error types, ungrammatical items included omission of the past –ed morpheme in the past simple clause. Out of eight ungrammatical items for the first context, four items contained bare progressive errors and four contained present tense errors. For use of the present progressive in the past, three error types were distributed evenly. An example of each error type is provided below:

**Example 1** (bare progressive): *They working at 7 o’clock last night.*

**Example 2** (use of present progressive): *He is playing chess at 10 p.m. last night.*

**Example 3** (missing –ed): *When we arrive the waitress was setting the table.*

For 40 past progressive items there were 10 distracters, both grammatical and ungrammatical. Half of the distracters were included in the GJT for the passive voice. They were also adapted from Révész (2009) and piloted with Turkish learners. The items were comparable in length and syntactic complexity.

Two versions of the test were developed by randomly scrambling the sentence order. The choice of lexical items was determined by the results of the pilot study and in consultation with the English language teacher of the target group. The instructions for the test were the same for all test sessions. Participants were asked to (1) indicate whether each sentence was grammatically correct or incorrect, and (2) to correct the error if they judged the sentence as ungrammatical. Approximately 25 minutes were allotted for the task in order to provide learners with sufficient time. However, it was not a timed task; learners were given an extra three or four minutes if needed but not more than five minutes were allowed in order to gain knowledge about their grammar as opposed to their
formal rule knowledge, following Gass and Mackey (2007). (See Appendix B for past progressive GJT).

The second part of the grammaticality task measured learners’ knowledge of the passive voice. It also included 40 items. The passive items were adapted from Spada et al. (2011). The items included passive statements formed with the present tense (16 items), past tense (16 items) and the present perfect tense (eight items). Sixteen items were grammatical and 24 were ungrammatical. The items included both regular and irregular verbs; however, the number of regular and irregular verbs was controlled in order to have a balanced distribution. Twenty-four items included regular verbs and 16 included irregular verbs. Table 2 provides detailed information about the distribution for passive items. The instructions for the test were the same for all test sessions.

**Table 2**

*Distribution of Regular and Irregular Verbs in Passive GJT*

<table>
<thead>
<tr>
<th>Verb</th>
<th>Grammatical</th>
<th>Ungrammatical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regular Verb</td>
<td>Irregular Verb</td>
</tr>
<tr>
<td>Present Simple Tense</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Past Simple Tense</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Present Perfect Tense</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>TOTAL</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>
### Table 3

**Distribution of Error Types for Passive Sentences**

<table>
<thead>
<tr>
<th>Error Types</th>
<th>Missing AUX</th>
<th>Past Participle Form</th>
<th>-ing instead of PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present Simple Tense</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Past Simple Tense</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Present Perfect Tense</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>8</strong></td>
<td><strong>8</strong></td>
<td><strong>8</strong></td>
</tr>
</tbody>
</table>

The ungrammatical passive items contained three error types: (a) missing auxiliary, (b) past participle form, (c) –ing instead of past participle. The examples of the error types are as follows:

**Example 1** (missing auxiliary): *Several applications received for the job yesterday.*

**Example 2** (past participle form): *The report wasn’t sign by the teacher.*

**Example 3** (-ing instead of past participle): *The winning goal was scoring by Messi.*

Similar to the progressive items, two versions of the test were developed by scrambling the order of the items (See Appendix C for the passive GJT items). Students were provided with a two-page test for each structure. I was available in the class and explained the procedures to the students in their L1. I also provided two examples on the board so that the students fully understood what they were required to do with the sentences. They were also allowed to ask questions in Turkish. The total time given for each grammaticality judgement task session was 20 minutes. Due to time constraints and scheduling problems, the grammaticality judgement task for the past progressive and the passive were administered separately rather than administering the original GJT format which includes both structures. Each test was administered at different times in order to
prevent learners’ fatigue. Each group of learners took the tests in their classroom during class time allotted for the research.

**Oral production task**

In order to measure learners’ control over the target structures in oral production, two measures were used. One of them was designed to elicit the past progressive and the other to elicit the passive. Both oral production tasks had similar designs. They were both picture-cued storytelling tasks delivered one-on-one using Powerpoint and were digitally audiorecorded.

I developed the oral production task for the past progressive. It was contextualized as a hypothetical scenario that placed the test-taker in the role of someone who could confirm the alibi of a person convicted of a crime. The test-taker was presented with five slides that set the background for the story. Then after reading the information together with the researcher, the test-taker was asked to describe the actions and events in the slides, most of which were designed to elicit the past progressive. The slides, which consisted of a series of photographs, were set up in such a way for the learner to contextualize a past event (e.g., *Last Saturday she was shopping*). Moreover, each photograph showed the time it was taken, for example, 11:07 last Saturday. Each picture included a time reference and a verb to be used (At 11:05 she was reading). The task also included non-target items which required a simple past tense, for example, *they bought an ice cream* (See Appendix D for past progressive OPT). The oral production task for the past progressive included the same pictures for both pretest and posttest; however, the names were changed in the story. As indicated earlier, the oral production task was piloted with both native speakers and EFL learners, and the obligatory items were determined based on input from native speakers.

The picture-cued oral production test for the passive voice was developed by Spada et al. (2011). In this task learners were asked to tell a story about a package that was lost in the mail. The learner was first provided with information to contextualize the story in the past tense. After going over the pictures and the verbs to be used to describe them, the learners were prompted with the question: *What happened to the package?* The
target items all included passive sentences in the simple past tense as the prompt was given in the past tense. Similar to the OPT task for the past progressive, there were both target and non-target slides. That is, not all slides required the use of the passive. Some slides to elicit the active were also included to maintain coherence in the story and to contrast with the passive (see Appendix E for a list of target and non-target items).

Following the grammaticality judgement task, learners were pulled out of their regular English language classes for the oral production task. I administered the OPT in a quiet room at the school. I was available with my notebook and a digital recorder. I also placed a microphone in order to record the learners’ voices. I explained that they would be engaged in two speaking activities. I explained the instructions and the storyline in Turkish in order to prevent task difficulty caused by language proficiency level. Each oral production task took around five to six minutes. Due to time restrictions, the oral production tests were administered to only 32 students. Two oral production tests were distributed in a split-block design in order to avoid any task effects.

**Instruments for individual learner differences**

**Language aptitude test**

The LLAMA is an aptitude test designed by Meara (2005). It is a computerized aptitude test based on the theoretical underpinnings of the MLAT. It includes four subcomponents: (1) vocabulary learning task, (2) grammatical inferencing task, (3) sound-symbol correspondence task, and (4) sound recognition task. The LLAMA aptitude tests are delivered through a computer program. They are written in a sparse programming style that minimizes the use of on-screen instructions. Therefore the language of the program itself was not a burden for the test-takers. All subtests follow a similar sequence. The computer records the score for each test-taker in a separate file. The computerized nature of the tests allows the researcher to assess learners’ individual scores precisely. Two minutes are allotted in subtests LLAMA_B-vocabulary learning (See Figure 6) and LLAMA_E-sound-symbol correspondence (See Figure 4) for students to work on the stimuli and provide the answers. LLAMA_B requires learners to study new vocabulary items in an unknown language. They are given two minutes to learn
some vocabulary items and then asked to identify those new items in the testing session. LLAMA_F-grammatical inferencing task presents learners with sentences in an unknown language and requires learners to figure out the rules that operate (See Figure 5). The last task LLAMA_D (sound-recognition) does not include any training but requires learners to recognize speech patterns (See Figure 3). Thus, the total time for all four tests was approximately 45 minutes, including instructions. In order to guarantee that learners would understand the instructions, they were provided in Turkish. I was available with the computer technology teacher during the testing to respond to any questions or concerns about the computers. The program and the procedures were followed exactly as described in the manual. However, a short introduction was prepared using the computer program Flash. A fellow graduate student helped in the preparation of three to four slides for each subtest and I used that material to introduce the test. I used a new set of icons other than those used in LLAMA to avoid a potential practice effect but I used the same interface to train learners.

Figure 3. LLAMA_D: Sound Recognition Task
Figure 4. LLAMA_E: Sound-Symbol Correspondence Task

Figure 5. LLAMA_F: Grammatical Inferencing Task
The aptitude test was administered in the school computer laboratory. Each class was taken to the lab during their computer technologies class time. Each learner had a computer and a pair of earphones. I explained the instructions of the test and also provided screenshots similar to ones used the program. The instructions for the task were provided in Turkish. For two tasks (i.e., LLAMA_F and LLAMA_E) within the test, learners were provided with a sheet of paper to take notes. The aptitude test took approximately 45 minutes. The test was administered to all groups on the same day.

**Working memory capacity: Backward digit span**

In order to measure WM capacity, a backward digit span test was used. The test was adapted from a Hungarian version of a digit span test (Racsmany, Lukacs, Nemeth, & Pleh, 2005). Any model for WM is based on the fact that there is a limited capacity for storing and processing of information and therefore there needs to be a trade-off between storing and processing (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980). WM tests thus differ from simple short-term memory tests. Short-term memory tests include simple memory span tests (i.e., word or digit span) and only require the recall of random
words or numbers whereas WM tests require both storage and processing. The reading span test introduced by Daneman and Carpenter (1980) has been used frequently in SLA studies (Alptekin & Ercetin, 2010; Mackey et al. 2002). The test-taker is asked to read and comprehend a number of sentences and then recall the last word of each sentence. The test-taker is also asked to judge the grammaticality of meaning of the target sentences in order to make sure that they comprehend the task. Efforts to improve WM testing have resulted in alternative measurements. New tests are aimed at lessening the effect of knowledge of language. For instance, Turner and Engle (1989) introduced operation span, which works the same as reading span but uses mathematical operations instead of sentences. In order to lessen the effect of language, backward digit span (DB) tests, extensively used in psychology and neuroscience, are also used in educational testing. DB tests are considered to be complex WM capacity tests. It is accepted in cognitive psychology that the digit backward test and the digit forward (DF) task measure different aspects of memory. DF can be done with simple rote recall whereas DB requires more mental manipulation and visualization of numbers (Lichtenberger & Kaufman, 2009). DB is also included in several intelligence tests such as Weschler IV and WAIS IV. Assessment research in educational psychology has shown that “DB is associated with attention and executive function processes, not the short-term rote-auditory memory processes tapped by DF” (Hale, Hoepner, & Fiorello, 2002, p. 128). This input from psychology research has recently led SLA researchers to use DB as a tool to measure WM capacity. While there are discussions about the choice of the right measurement for different aspects of WM in SLA research (see Miyake & Friedman, 1998 for WM measures) a recent study conducted by Kormos and Sáfár (2008) demonstrated that DB is a valid test of WM capacity and DB is accepted as a promising test of WM capacity in SLA (See Juffs and Sawyer, 2011).

There were 28 items in the backward digit span (BDS) test in the current study. The test was administered individually in Turkish by the researcher in a quiet room. The introduction and first three digit series were administered orally by the researcher while the remaining series were played by a computer. Each number was read out loud or presented at the rate of one per second (See Appendix G for Digit span test).
The test was introduced as a memory game and administered after posttesting with the oral production tests. The researcher met students in a quiet room and had the computer and digital recorders ready. The students were given the instructions and the test in Turkish, and they were digitally recorded during the test. I did not have eye contact with the participants during the test following the manuals for digit span tests (Mather, Wendling, & Woodcock, 2001; Lichtenberger & Kaufman, 2009).

Motivation questionnaire

Learners’ motivation and orientation towards foreign language learning were measured with a questionnaire designed by Dörnyei (1990). This survey is based on a motivation construct that has been reframed for foreign language contexts. Dörnyei (1990) surveyed 134 young adult learners in Hungary. Drawing upon published motivation and attitude scales, his scale consisted of two sections: items that focused on language fields to use L2, and likert scale items concerning intentions, beliefs, values, interests, and attitudes about English language learning. The questionnaire section included 44 likert scale statements, five of which were treated as single-item variables, whereas the rest were used to form 13 subscales. The shortest scale contains two items, the longest contains seven.

The original questionnaire includes 44 items; however, only 36 items were used in the current study since some of the items were not relevant to my target language learners in terms of age and status. The participants were asked to indicate their responses on a five-point scale (Dörnyei, 1990). To help the participants better understand the questionnaire, it was translated into Turkish and administered in Turkish (see Appendix H).

Retrospection questionnaire

In order to obtain information about the participants’ perceptions of the ‘difficulty’ of the two target structures (i.e., DeKeyser’s ‘subjective difficulty’), they were asked to reflect on their learning experience for 10 to 15 minutes by responding to several items in a questionnaire. The first section of the questionnaire focused on the
participants’ biographical information including age, gender, length of English instruction, and prior EFL experience. The second section elicited the participants’ opinions about the difficulty of several L2 structures that were taught in their regular English curriculum. They were asked to rate the difficulty of each of the target structures on a five-point scale (see Appendix I for the questionnaire). Within this section, they were also asked to rate the difficulty of the two target structures on a difficulty scale. To obtain more information about learners’ perceptions of difficulty in relation to the target features, the third section included open-ended questions about the target language structures and asked learners to provide explanations for their choices. For example, participants were asked why they considered certain (i.e., the passive or the past progressive) structures to be easy or difficult. They were also asked to provide a reason for their perceptions, such as lack of L1-L2 correspondence or difficult to memorize. The learners were asked to reflect on their answers and provide explanations for them. The questionnaire was administered in Turkish to ensure that students fully understood the items.

The motivation and retrospection questionnaires were administered at different times during the same week but the procedure was the same. Both questionnaires were two pages in length. I explained the instructions in Turkish and was available during the administration. The students were given 15 to 20 minutes to complete each questionnaire.

**L1 metalinguistic test**

Learners’ metalinguistic knowledge about the L1 was measured with an error detection and correction task. L1 metalinguistic ability refers to learners’ ability to attend to the syntactic and morphological form of sentences (Ranta, 2002). The items included morphological errors in Turkish and the participants were asked to identify the error and to provide a correct form. In addition, learners were asked to provide an explanation for their correction. They were instructed to provide the grammatical rule that they thought was violated. The test included 14 sentences. Seven sentences were grammatical and seven were ungrammatical (see Appendix J for the test). The test was piloted with native speakers of Turkish and language experts.
The learners were given the printed copy of the test, which was in Turkish. It was three pages long. I explained the instructions and was available to answer any questions raised. 20 minutes were given for this test.

In the next chapter a description of the analyses and results are provided.
Chapter Five
Analyses and Results

This chapter reports on the analyses and results of the study. It begins with a description of the procedures used in the validity and reliability testing for all measures. This is followed by a description of the procedures used for scoring the language and learner variable measures. The chapter concludes with a description of the statistical analyses and results in relation to the research questions investigated in this study:

1. To what extent do individual learner variables (i.e., L2 motivation, L1 metalinguistic awareness, working memory, learners’ perception of difficulty, and language aptitude) explain the effectiveness of L2 instruction on selected target structures?
   a) To what extent do IDs explain the effectiveness of L2 instruction on difficult L2 structures?
   b) To what extent do IDs explain the effectiveness of L2 instruction on easy L2 structures?

2. Do specific aptitude components explain the effectiveness of L2 instruction on selected L2 structures?
   a) Do aptitude components explain the effectiveness of L2 instruction on difficult L2 structures?
   b) Do aptitude components explain the effectiveness of L2 instruction on easy L2 structures?

3. What is the effect of difference in aptitude profiles on the learning of selected structures?
   a) Is there an effect of difference in aptitude profiles on the learning of difficult structures?
   b) Is there an effect of difference in aptitude profiles on the learning of easy structures?
Validity and Reliability of the Research Instruments

In the following section, validity and reliability issues related to the ID variables and language measures are reported. As the first research question addresses all the variables under investigation in the study, all of them will be covered in this section.

Validity

The ID variables were measured through questionnaires, a metalinguistic task, a computerized aptitude measure, and a digit span task. The language variables were measured via oral production and grammaticality judgment tasks. To ensure validity of the ID and language tests, expert opinion was sought from thesis committee members. The questionnaires were revised several times before they were piloted, after which they were revised again. The validity of the language measures involved a close examination of the responses provided by the pilot group of learners. Tests were then revised accordingly.

Reliability

The internal consistency of the ID variables and the language tests were obtained using Cronbach’s alpha. The values are reported in relation to each variable in the following sections.

Oral production data

I transcribed all the oral production data. To establish inter-rater reliability, all the oral production data were also transcribed by a second researcher. Reliability checks indicated that there were 13 disagreements (out of 600 target utterances) that would affect scoring for the passive OPT and 15 for the progressive OPT. Disagreements were resolved by the researcher listening to specific items for a third time and making the necessary revisions. Following this, inter-rater reliability estimates indicated high levels for both OPT’s (i.e., the passive and the past progressive). The correlations ranged from .95 to .97.
**Written data**

All GJT tests were scored by the researcher. All the grammaticality tests were scored by another researcher to establish inter-rater reliability. Inter-rater reliability was high; it was .96 for pretest and .98 for the posttest scoring.

**Scoring**

**Oral production task**

**Past progressive**

There were two types of target items for the past progressive: 1) simple past progressive sentences (i.e., She was reading.), and 2) while clause sentences (i.e., while she was reading her friend came). Scoring the first type was conducted following Révész (2007). That is, learners received 3 points for producing the correct progressive in obligatory contexts (e.g., she was reading), 2 points for using the present progressive (e.g., she is reading) and 1 point for bare progressive (e.g., she reading). It was sometimes difficult to draw the line between obligatory contexts and non-obligatory contexts, so 1 point was also given to learners who told the story in the simple past tense and produced correct simple past sentences (e.g., she read a book). Agreement errors were disregarded following Révész (2007).

For the second type of past progressive (e.g., while they were walking she dropped her ice cream), another coding scheme was developed. The maximum score for a correct response was 3 points. That is, if the learner provided a correct use of the past progressive in the while clause and a correct past simple sentence in the second clause, as in the example above, they were given 3 points. In some items where it was also possible to use the past progressive for the second clause learners received 3 points if they provided correct past progressives in both clauses (e.g., while they were swinging, they ate/were eating ice cream or while she was working, she listened/was listening to music). If however, learners produced the bare progressive or the present progressive in the while clause and the correct simple past in the second clause, they got 2 points. Similarly, if they produced the correct progressive in the while clause and made errors on the simple
past, such as producing the present simple sentence or past form of the verb, they also received 2 points (e.g., *while she was talking on the phone, he got bored*). If they failed to produce any progressive marking in the *while* clause but produced a correct simple past or past progressive sentence in the second clause, they were given 1 point (e.g., *while she talking on the phone, he got bored*). Again agreement errors were disregarded (e.g., *while they was sitting at a park*). Finally, if they failed to produce any progressive marking in the *while* clause and failed to produce the correct simple past sentence or past progressive sentence in the next clause, they received no points (e.g., *while she drink her coffee her boyfriend come*). The maximum score for the progressive oral production task was 30 (3 x 10). The reliability coefficient for both pre- and posttest was high (pretest $\alpha= .91$ and posttest $\alpha= .94$).

**The passive**

When scoring learners’ performance on the passive oral production test, I followed the scheme developed by the creators of the task (Spada et al., 2011). If the learners produced a correct passive sentence in an obligatory context, they were given 3 points (e.g., *the package was sent to Canada*). If they failed to produce the auxiliary or the past participle, they received 2 points (e.g., *the package sent to Canada or the package was send to Canada*). As with the progressive oral production test, learners were not penalized if they produced a meaningful and correct active sentence; they received 1 point (e.g., *the post office sent the package*). However, if they made any errors in the active sentence, they received no points. If learners failed to produce the marker of the passive (e.g., *be* auxiliary), they received no points (e.g., *the package sent to Canada*). Likewise if they produced the progressive marking instead of a past participle in the main verb, they also received no points (e.g., *the package was delivering to his house*). Time changes, agreement errors and minor spelling mistakes in the past participle were not penalized. There were nine target items in this task and the maximum score was 27 (3x9). Internal consistency of the passive OPT pre and posttest was within acceptable limits ($\alpha= .74$ and $\alpha= .88$ respectively).
Grammaticality judgment task

Past progressive

The scoring scheme used to measure learners’ performance on the progressive GJT included two categories. The first item type (i.e., she was reading) was scored following Révész (2007). When the learners judged a grammatical past progressive item as grammatical, or supplied the appropriate corrections for ungrammatical items, they were given 3 points. Two points were given when the learners changed grammatical past progressive sentences to the present progressive, changed ungrammatical bare progressive sentences to the correct progressive, or judged ungrammatical present progressive sentences as grammatical. When the learners changed a grammatical past progressive form into the bare progressive, or they judged a bare progressive item as grammatical, they were given 1 point. Learners received a 0 if they judged ungrammatical items as grammatical and made changes to a non-target form different from the ones listed. Agreement errors were disregarded (See Appendix L for detailed scoring procedures and examples).

The scoring scheme was altered slightly for the second type of items with while clauses. These items included an additional type of error: omission of –ed in the second clause. Thus, learners received 2 out of 3 points if they failed to correct an omission of an –ed error in the second clause. The maximum score for the past progressive GJT was 120 (40x3). The reliability coefficient for the past progressive GJT was high (α=.95 for the pretest and α=.96 for the posttest).

The passive

The scoring for the passive GJT was different from the progressive GJT. The maximum score for a passive GJT item was 2 points. When the learners corrected an ungrammatical passive sentence (i.e., missing ‘be’ error, past participle form error or missing past participle error) with correct passive marking (i.e., be auxiliary and past participle of the main verb), they were given 2 points. If they failed to provide the auxiliary or past participle form of the verb while trying to correct an ungrammatical passive sentence, they were given 1 point. If learners judged an ungrammatical passive
sentence with a past participle form error as correct (e.g., *a bracelet is wear around the wrist*), they were given 1 point as the sentence included be AUX. However, if they judged an ungrammatical sentence with a missing AUX or with a progressive marker to be correct (e.g., *apples are picking in the fall*), they received no points because those sentences included no passive markers. If they provide –ing marker instead of past participle, they received 0. Minor participle errors, time change in the correct passive sentence or agreement errors were disregarded. The detailed scoring scheme for the passive is provided in Appendix K. Any correction other than the past participle or ‘be’ auxiliary will be discarded. The maximum score for the passive GJT was 80 (40x2). The reliability coefficient was found to be high (pretest $\alpha= .91$ and posttest $\alpha=.95$).

In assessing development from pretest to posttest, only 40 target items were scored. Responses to distracters were not coded and therefore not included in the analysis.

**Aptitude test**

As indicated earlier, the LLAMA test is delivered on a computer. The scoring was conducted by the computer program. However, the program did not provide individual item scores but rather a total score for each subtest in the test. Therefore I had four subtest scores (grammatical sensitivity, sound-symbol correspondence, sound recognition, and vocabulary learning) and one total aptitude score for each participant. As indicated earlier, there were four tasks in the aptitude test. The vocabulary learning (i.e., LLAMA_B) task included 20 different vocabulary items and the scores range from 0 to 100. The next task was the sound recognition task (i.e., LLAMA_D) and the scores also ranged from 0 to 100. The sound-symbol correspondence task, which presented a set of 22 recorded syllables and the transliteration of those syllables in an unfamiliar alphabet, allowed for 100 as the maximum score. Similarly, the maximum score for the grammatical inferencing task was 100 as a composite score. When the aptitude test scores were used as single scores, the maximum score for the aptitude test was 400.
L1 metalinguistic test

The L1 metalinguistic test and the scoring scheme were developed by the researcher following procedures used by R. Ellis, Loewen, Elder, Erlam, Philip, and Reinders (2010). First, the learners were asked to decide whether each Turkish sentence was grammatical or not. If they judged the sentence as ungrammatical, they were asked to identify and correct the error. They were also asked to explain the rule that was violated. The learners were given 2 points when they judged a grammatical sentence as grammatical. If they judged an ungrammatical sentence as grammatical, they received no points. They also received no points if they failed to correct an ungrammatical sentence even if they judged it to be incorrect. When they corrected the ungrammatical sentence that they judged as ungrammatical but provided no explanation, they were awarded 1 point. If they provided a satisfactory explanation for the rule, with the correction, they received 2 points. There were thirteen items in the test so the maximum score was twenty-six. Reliability of L1 metalinguistic awareness was found to be within acceptable limits (Cronbach’s $\alpha = .77$).

Backward digit span test

The scoring of the digit span test was conducted following the guidelines of several published manuals (Mather, Wendling, & Woodcock, 2001; Lichtenberger & Kaufman, 2009). The total score for the digit backward span test was calculated as the sum of correctly recalled digits. The total score for the digit span task is twenty-eight. When reliability was examined, internal consistency was found to be within acceptable limits (Cronbach’s $\alpha = .76$).

Motivation questionnaire

In scoring the questionnaire responses, negatively worded items were reversed so that the responses could consistently range from 1 (the most negative score) to 5 (the most positive score) throughout the questionnaire. For each student, a total score for each subscale was calculated first. Then the scores were summed for a general motivation
variable. Reliability coefficient of the motivation questionnaire was also high with Cronbach’s alpha at .84.

**Retrospection questionnaire**

As indicated in Chapter 4, the first section of the questionnaire included questions about the participants’ demographic information and prior learning experience. The second section provided learners with a list of nine L2 structures and asked them to rate their difficulty. The responses were scored as 1 “very easy”, 2 “easy”, 3 “difficult” and 4 “very difficult”. The third section included two open-ended questions about learning styles and preferences and the responses were analyzed qualitatively based on the patterns that emerged. This section also included a scale that had to be completed by the participants. The instructions asked the learners to place the two target structures (i.e., past progressive and passive) on the scale based on their perceived difficulty. The responses to this section were scored and analyzed quantitatively.

Lastly, the retrospection questionnaire included a section where the learners were asked to explain the structural properties and rules for the use of the two target structures (i.e., past progressive and passive). The learners’ responses were analyzed qualitatively.

**Data Analysis for Research Question 1 (a & b)**

The quantitative data for the ID variables (i.e., motivation questionnaire, L1 metalinguistic awareness test, digit backward test, learners’ perception of difficulty, and language aptitude) and language variables (i.e., passive GJT and OPT and past progressive GJT and OPT) were analyzed to address the first research question: To what extent do individual learner variables explain the effectiveness of L2 instruction on selected target (i.e., easy and difficult) structures? I used the Statistical Package for Social Sciences (SPSS) for Windows 17.0 to analyze all the quantitative data. I started by computing the descriptive statistics of all the variables used in the study in order to obtain a better understanding of the data and to characterize the sample. I also checked the data for accuracy, missing data, skewedness, and kurtosis. To examine how much variance in the learning of the two target structures could be explained by the proposed ID variables, a multiple regression analysis was carried out. The first step was to check whether the
data met the normality assumptions of the multiple regression analysis. The descriptive analysis revealed that language variables were not normally distributed except for the passive GJT posttest. Thus these data were transformed in order to meet the assumptions of the statistical procedures. Table 4 shows the descriptive statistics for the individual learner variables and Table 5 shows the descriptive statistics for the language variables before transformations were conducted. An examination of the skewedness and kurtosis values and the histograms indicated that the passive OPT and GJT tests, except for the passive GJT posttest, were positively skewed and that the progressive OPT and GJT were negatively skewed. Following Tabachnick and Fidell (1996), positively skewed variables were transformed with the square root calculation of the variables and negatively skewed variables were transformed with the square calculation of the variables. The transformations helped to convert all the variables into normal distributions except for the past progressive GJT data. While several algorithmic functions were used, none of them were able to convert the past progressive GJT data into a normal distribution. This made it necessary to use alternative procedures when analyzing the past progressive GJT data.

When a dependent variable is measured at two points in time, it is possible to use gain scores (i.e., the difference from posttest and pretest scores) instead of posttest scores. Posttests were used as dependent scores in all regression analyses because gain scores are considered to be unreliable and subject to regression to mean effects (Allison, 1990).

**Table 4**

*Descriptive Statistics for Individual Learner Variables*

<table>
<thead>
<tr>
<th>ID Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aptitude</td>
<td>61</td>
<td>204</td>
<td>59.3</td>
<td>60</td>
<td>315</td>
<td>-0.39</td>
<td>-0.40</td>
</tr>
<tr>
<td>L2 motivation</td>
<td>65</td>
<td>129</td>
<td>15.9</td>
<td>88</td>
<td>158</td>
<td>-0.54</td>
<td>-0.49</td>
</tr>
<tr>
<td>WM capacity</td>
<td>36</td>
<td>11</td>
<td>4.47</td>
<td>4</td>
<td>23</td>
<td>0.73</td>
<td>0.68</td>
</tr>
<tr>
<td>L1 metaling</td>
<td>62</td>
<td>14.6</td>
<td>3.4</td>
<td>5</td>
<td>20</td>
<td>-0.34</td>
<td>-0.21</td>
</tr>
<tr>
<td>Students’ rating</td>
<td>60</td>
<td>3.38</td>
<td>1.51</td>
<td>1</td>
<td>5</td>
<td>-0.18</td>
<td>-0.53</td>
</tr>
<tr>
<td>Passive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students’ rating</td>
<td>60</td>
<td>2.05</td>
<td>1.01</td>
<td>1</td>
<td>5</td>
<td>0.7</td>
<td>-0.13</td>
</tr>
<tr>
<td>Past Prog.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* The number of students who took the ID tests varied because not all students were available on the testing days. The digit span test was administered to 36 students due to time limitations. The written ID tests (i.e., motivation and retrospection questionnaire, L1
metalinguistic awareness test) were given to 65, 60, and 62 students respectively. The aptitude test was administered to 61 students).

Table 5

Descriptive Statistics for Language Variables

<table>
<thead>
<tr>
<th>L2 variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min.</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive OPT</td>
<td>30</td>
<td>4.9</td>
<td>4.8</td>
<td>0</td>
<td>17</td>
<td>1.1</td>
<td>0.53</td>
</tr>
<tr>
<td>Pretest</td>
<td>30</td>
<td>4.4</td>
<td>5.3</td>
<td>0</td>
<td>25</td>
<td>2.0</td>
<td>6.6</td>
</tr>
<tr>
<td>Posttest</td>
<td>63</td>
<td>24</td>
<td>15.2</td>
<td>5</td>
<td>78</td>
<td>1.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Passive GJT</td>
<td>62</td>
<td>44.6</td>
<td>20.1</td>
<td>5</td>
<td>80</td>
<td>0.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>Pretest</td>
<td>63</td>
<td>24</td>
<td>15.2</td>
<td>5</td>
<td>78</td>
<td>1.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Posttest</td>
<td>62</td>
<td>44.6</td>
<td>20.1</td>
<td>5</td>
<td>80</td>
<td>0.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>Past Prog.</td>
<td>30</td>
<td>17.6</td>
<td>8.7</td>
<td>0</td>
<td>30</td>
<td>-0.7</td>
<td>-0.3</td>
</tr>
<tr>
<td>OPT pretest</td>
<td>30</td>
<td>19.5</td>
<td>9.0</td>
<td>0</td>
<td>30</td>
<td>-1.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>Past Prog.</td>
<td>62</td>
<td>91.9</td>
<td>28</td>
<td>22</td>
<td>120</td>
<td>-1.0</td>
<td>4.0</td>
</tr>
<tr>
<td>OPT posttest</td>
<td>62</td>
<td>91.9</td>
<td>28</td>
<td>22</td>
<td>120</td>
<td>-1.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Past Prog.</td>
<td>64</td>
<td>101.2</td>
<td>25.6</td>
<td>29</td>
<td>120</td>
<td>-1.3</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Note: 1. Table includes scores are before they were transformed.
2. The number of participants that took the passive and the past progressive GJT was above 60 whereas OPT’s were only administered to 30 students who took both pre and posttests.

Upon completion of the transformations to meet normality assumptions, I followed the necessary steps to conduct a multiple regression.

ID variables as predictors for difficult structure

The first research question asks: To what extent do individual learner variables explain the effectiveness of L2 instruction on selected target structures? In order to investigate this question specifically with regard to difficult L2 structures and with respect to learners’ performance on the passive GJT, a multiple regression was carried out on learners’ posttest GJT scores. To decide which variables were the best predictors for the regression analysis, correlations between the dependent variables and the predictors were examined. Pearson product moment coefficients revealed that passive GJT posttest scores (i.e., the dependent variable) were related to pretest scores ($r = 0.43,$
$p < .001$), to aptitude total scores ($r = 0.40$, $p < .001$) and moderately related to L2 motivation ($r = 0.30$, $p < 0.05$). Scores on the passive GJT test were also related to L1 metalinguistic awareness ($r = 0.34$, $p < .001$). However, aptitude and L1 metalinguistic awareness were also related ($r = 0.43$, $p < .001$). Therefore, in order to prevent multicollinearity, motivation and aptitude were chosen as best predictor variables for the regression analysis (See Table 6 for the correlations among variables in the analysis). Although two other ID variables (i.e., Working memory capacity and student ratings) were initially proposed as variables to be included in such an analysis, given the small sample size it was decided to enter only three variables (i.e. pretest scores, aptitude and motivation) in the regression analysis in order to prevent model overfitting. Overfitting is a problem when there are too many variables relative to the sample size (Tabachnick & Fidell, 1996). An overfitted model has the risk of low predictive power. Moreover, including fewer variables is recommended by Cohen (1990). In order to prevent overfitting and more highly targeted findings, all the regression analyses conducted included three variables chosen as the best predictors in the correlation analyses conducted with the ID variables.

**Table 6**

*Correlations among ID Variables and the Passive GJT Scores*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Passive GJT Pre</td>
<td>1</td>
<td>0.43**</td>
<td>0.37*</td>
<td>0.08</td>
<td>0.15</td>
<td>-0.01</td>
<td>-0.38**</td>
<td>-0.22</td>
</tr>
<tr>
<td>2 Passive GJT Post</td>
<td>1</td>
<td>0.3</td>
<td>0.3*</td>
<td>0.40**</td>
<td>0.34**</td>
<td>-0.29*</td>
<td>-0.39**</td>
<td></td>
</tr>
<tr>
<td>3 WMC</td>
<td>1</td>
<td>0.04</td>
<td>0.27</td>
<td>0.31</td>
<td>-0.45**</td>
<td>-0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 L2 motivation</td>
<td>1</td>
<td>0.04</td>
<td>0.08</td>
<td>-0.021</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Aptitude</td>
<td>1</td>
<td>0.43**</td>
<td>-0.15</td>
<td>-0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 L1 metalinguistic Awareness</td>
<td>1</td>
<td>-0.19</td>
<td>-0.28*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Passive difficulty rating</td>
<td>1</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Progressive difficulty rating</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$, ** $p < .01$, *** $p < .001$
A hierarchical multiple regression analysis was then conducted, using GJT passive posttest scores as the criterion variable and pretest scores, motivation and aptitude as predictor variables for the outcome. Pretest scores were entered into regression analysis in the first step, and then aptitude and motivation were entered together in the next step. Not surprisingly, pretest scores correlated to all corresponding posttests. Therefore, in all hierarchical regression analyses, pretests were entered in the first step. To check their individual contribution, IDs were entered in the next step. This way, pretest scores were controlled and the contribution of individual learner variables was clearly observed.

The results of the regression analysis revealed, not unexpectedly, that pretest scores alone ($\beta = 0.35 \ t [57] = 3.18, \ p < .001$) significantly predicted posttest scores. The Passive GJT pretest scores explained 18% of the variance ($F[1, 57] = 12.62, \ p < .001$). After controlling for pretest scores, aptitude ($\beta = 0.32, \ t [55] = 2.94, \ p < .005$) and motivation ($\beta = 0.25, \ t [55] = 2.3, \ p < .05$) had significant associations with Passive GJT posttest scores. Motivation and aptitude together increased the explained variance to 35% from 18%. Furthermore, when the pretest scores were controlled, 17% of the variability in Passive GJT posttest scores was predicted by both aptitude and motivation. See Table 7 for the regression model. Effect size indices can be calculated several ways with regression. I chose $R^2$ as the effect size index for multiple regression analyses following Trusty, Thompson and Petrocelli (2004). When $R^2$ was examined in the regression model for the passive GJT pretest scores, language aptitude and motivation were observed to have a medium effect size ($R^2 = 0.35$) based on Cohen’s (1988) qualitative descriptors of effect sizes.
Table 7  

*Summary of Multiple Regression Analysis for Individual Difference Variables Predicting Passive GJT Scores*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F_{change}$</th>
<th>Df</th>
<th>$B$</th>
<th>$t$ (within set)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive GJT posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 1: Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive GJT pretest</td>
<td>12.61***</td>
<td>57</td>
<td>0.47</td>
<td>.35***</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>Step 2: ID variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive GJT pretest</td>
<td></td>
<td>57</td>
<td>0.39</td>
<td>.31***</td>
<td></td>
</tr>
<tr>
<td>Aptitude</td>
<td>7.13***</td>
<td>55</td>
<td>0.01</td>
<td>2.3**</td>
<td>0.35</td>
</tr>
<tr>
<td>Motivation</td>
<td>7.13***</td>
<td>55</td>
<td>0.03</td>
<td>2.93*</td>
<td>0.35</td>
</tr>
</tbody>
</table>

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

The second part of the analysis examined learners’ oral performance with regard to the same structure (i.e., the passive). The same steps were taken in carrying out the multiple regression analysis. The potential predictor variables (i.e., individual learner variables) were screened in order to determine the three strongest ones. When the correlations between the criterion variable (i.e., scores on the passive OPT) and potential predictor variables (i.e., ID variables) were examined, no significant correlations were observed among any ID variables except for OPT pretest scores ($r = 0.46$, $p < 0.05$). See Table 8 for correlations among ID variables and the Passive OPT posttest. Therefore, it was decided to follow the same model used for the passive GJT. That is, the Passive OPT pretest scores were entered as the first predictor in the first model. Then motivation and aptitude were entered into the analysis to examine their role in the passive OPT scores controlling for pretest scores.
Table 8

Correlations among ID Variables and the Passive OPT Scores

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Digit backward test</td>
<td>1 0.04</td>
<td>0.27 0.31</td>
<td>-0.45**</td>
<td>-0.32 0.22</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>L2 Motivation</td>
<td>1 0.04</td>
<td>0.08 0.21</td>
<td>0.13 0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Aptitude</td>
<td>1 0.43**</td>
<td>-0.15 -0.22</td>
<td>0.2 0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>L1 metalinguistic awareness</td>
<td>1 -0.19</td>
<td>-0.28* 0.01</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Passive Difficulty Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Progressive Difficulty Rating</td>
<td></td>
<td>0.46**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Passive OPT posttest</td>
<td>1 0.46*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Passive OPT pretest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Contrary to the previous regression analysis, the results with passive OPT scores indicated that only the pretest scores were significantly associated with posttest performance on the OPT (β =0.46 t [28] = 2.74, p < 0.01). Motivation (β = 0.15, t [26] = 0.85, p > 0.05) and aptitude (β = 0.05, t [26] = 0.28, p > 0.05) did not explain any variance even when pretest scores were controlled for. See Table 9 for the summary of regression model tests for the passive OPT scores. Regression model with ID variables and passive OPT scores has relatively low effect size with (R² = .23).

Table 9

Summary of Multiple Regression Analysis for Individual Difference Variables Predicting Passive OPT Scores

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>F change</th>
<th>df</th>
<th>B</th>
<th>t (within set)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive OPT posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1: Control Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive OPT pretest</td>
<td>7.5*</td>
<td>28</td>
<td>0.51</td>
<td>2.7*</td>
<td>0.21</td>
</tr>
<tr>
<td>Step 2: ID variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To investigate the second part of the first research question, which was concerned with learners’ performance on an easy language structure, I examined learners’ knowledge of the past progressive on the GJT in a logistic regression. It was necessary to carry out a logistic regression because the dependent variable (i.e., learners’ performance on the GJT) was negatively skewed and the transformation did not result in a normal distribution. A logistic regression is similar to a multiple regression analysis but it permits the use of categorical variables as predictors or criterion variables (Field, 2009). In order to conduct the logistic regression, I divided the data into two groups equal in numbers: high and low. To decide which individual learner variables to include in the analysis, correlations among the possible ID predictors were examined (see Table 10 for correlations). The results of the nonparametric correlational analyses showed that L1 metalinguistic awareness was the best candidate for a logistic regression. L1 metalinguistic awareness (Kendal’s tau_b= 0.48) and pretest scores (Kendal’s tau_b= 0.53) also appeared to be related to the progressive GJT posttest scores.

Table 10

<table>
<thead>
<tr>
<th>Correlations among Past Progressive GJT and ID Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 WMC</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1 WMC</td>
</tr>
<tr>
<td>2 L2 Motivation</td>
</tr>
<tr>
<td>3 Aptitude</td>
</tr>
<tr>
<td>4 L1 Metalinguistic Awareness</td>
</tr>
<tr>
<td>5 Past progressive categorical</td>
</tr>
<tr>
<td>6 Past Progressive pretest</td>
</tr>
<tr>
<td>7 Past Progressive posttest</td>
</tr>
</tbody>
</table>

*p < .05

**p < .01
Therefore, a logistic regression analysis was conducted to examine whether the membership in the high GJT or low GJT group could be predicted by the pretest and L1 metalinguistic awareness. A test of the full model was statistically significant indicating that the predictors as a set reliably distinguished between high and low GJT posttest scores ($\chi^2 = 29.83, p < .000$ with $df = 2$). Table 11 shows the summary of the logistic regression model. Nagelkerke’s $R^2$ of .54 indicated a moderately strong relationship between prediction and grouping (i.e., high and low GJT scores). Prediction success overall was 77.6% (65.4% for low group and 87.5% for high). However, the results demonstrated that only the pretest scores made a significant contribution to the prediction (Wald’s value= 13.56, $\text{Exp (B)} = 1.08$, $p < 0.01$ with $df = 1$). L1 metalinguistic awareness was not a significant predictor (Wald’s value= 2.78, $p > 0.05$ with $df = 1$) for the past progressive GJT.

Table 11

Summary of Logistic Regression Analysis for Individual Difference Variables Predicting Past Progressive GJT Scores

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$B$</th>
<th>$SE_B$</th>
<th>$e^B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.08*</td>
<td>0.02</td>
<td>1.1</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test</td>
<td>.69*</td>
<td>0.02</td>
<td>1.1</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>0.21</td>
<td>0.12</td>
<td>1.2</td>
</tr>
<tr>
<td>Awareness Constant</td>
<td>-9.56*</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>32.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Df$</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Past Progressive GJT scores (high and low) coded as 1 for high and 0 for low.
A third regression analysis was conducted with the past progressive OPT data. In order to decide which variables were the best predictors, correlations between the dependent variable and possible predictors were examined (See Table 12 for correlations). This procedure revealed that learners’ performance on the progressive OPT posttest was highly related to learners’ pretest scores ($r = 0.74$, $p < .001$) and to L1 metalinguistic awareness ($r = 0.67$, $p < .001$). Thus, OPT pretest scores were entered into the regression analysis in the first step, and L1 metalinguistic awareness was entered in the next step, similar to all previous regression analyses.

**Table 12**

*Correlations among ID Variables and the Past Progressive OPT Scores*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 WMC</td>
<td>1</td>
<td>0.04</td>
<td>0.27</td>
<td>0.31</td>
<td>0.45**</td>
<td>-0.32</td>
<td>0.48*</td>
<td>0.27</td>
</tr>
<tr>
<td>2 L2 Motivation</td>
<td>1</td>
<td>0.04</td>
<td>0.08</td>
<td>-0.21</td>
<td>0.13</td>
<td>0.38*</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>3 Aptitude</td>
<td>1</td>
<td>0.43**</td>
<td>-0.15</td>
<td>-0.22</td>
<td>0.29</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 L1 Metalinguistic</td>
<td>1</td>
<td>-0.19</td>
<td>-0.28*</td>
<td>0.42*</td>
<td>0.67**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Passive Dif.R.</td>
<td>1</td>
<td>0.19</td>
<td>-0.43*</td>
<td>-0.48**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Progressive Dif. R.</td>
<td>1</td>
<td>-0.26</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Progressive pretest</td>
<td>1</td>
<td>0.75**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Progressive pretest</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Dif. R. (student difficulty rating)*

The results revealed that pretest scores ($\beta = 0.74$, $t [28] = 5.93$, $p < .001$) had a significant association with the posttest OPT scores. Progressive OPT pretest scores explained 55% of the variance ($F [1, 28] = 35.21$, $p < .001$). See Table 13 for the regression summary table. After controlling for OPT pretest scores, L1 metalinguistic awareness also indicated a significant association with Progressive OPT posttest scores. Learners’ L1 metalinguistic awareness increased the explained variance from 55% to 71%. Furthermore, when pretest scores were controlled for in the analysis, there was a significant $R$ square change ($\beta = 0.44$, $t [27] = 3.88$, $p < .001$) explained by L1 metalinguistic awareness. $R^2$ of .71 indicates a high effect size but it should be remembered that the regression model with only pretest scores predicting posttest scores also had a high effect size ($R^2=0.55$).
Table 13

Summary of Multiple Regression Analysis for ID Variables Predicting Past Progressive OPT Scores

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F_{change}$</th>
<th>df</th>
<th>$B$ (within set)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Progressive OPT posttest</td>
<td>35.2*</td>
<td>28</td>
<td>0.78</td>
<td>5.9*</td>
</tr>
<tr>
<td><strong>Step 1: Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Progressive OPT pretest</td>
<td>35.2*</td>
<td>28</td>
<td>0.78</td>
<td>5.9*</td>
</tr>
<tr>
<td><strong>Step 2: ID variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Progressive OPT pretest</td>
<td>28</td>
<td>0.59</td>
<td>4.9*</td>
<td></td>
</tr>
<tr>
<td>L1 Metalinguistic awareness</td>
<td>15*</td>
<td>27</td>
<td>42.3</td>
<td>3.9*</td>
</tr>
</tbody>
</table>

* $p < .001$

Results for the first research question demonstrate that two ID variables (i.e., aptitude and motivation) contribute to variance in the passive GJT scores whereas these two variables are not found to be related to the progressive GJT scores. Moreover, no other ID variable could explain any variance in the progressive GJT. With regard to the oral production tasks, L1 metalinguistic awareness explained 16% of variance in progressive OPT scores whereas none of the ID variables could explain any variance in passive OPT scores.

Data Analysis for Research Question 2 (a & b)

The second research question examines the language aptitude variable in greater detail by asking: Do specific aptitude components (i.e. LLAMA_B-vocabulary learning, LLAMA_E-sound-symbol correspondence, LLAMA_F-grammatical inferencing, and LLAMA_D-sound recognition) explain the effectiveness of L2 instruction on selected L2 structures? Aptitude was used as a composite construct in the analysis for RQ 1 (a & b) along with the other ID variables. This research question deals with aptitude separately
from the other ID variables and investigates whether individual aptitude components might make unique contributions to the learning of the two target structures.

As in the sections above, this question is examined in relationship to learners’ performance on the written (i.e., GJT) and oral performance tasks (i.e., OPT) with respect to the difficult and easy target language features. This section begins with a description of the results for learners’ performance on the passive grammaticality test in relation to the subcomponents of aptitude. Next, a description of the results for the oral production tests will be reported. Results for the past progressive GJT and OPT will follow.

**Aptitude components as predictors for difficult structure**

In order to decide which aptitude components might serve as the best predictors for the regression analysis, correlations between the passive GJT and the four aptitude sub-components were examined (See Table 14 for the correlations between the passive GJT and the ID variables). Pearson product-moment coefficients revealed that pretest scores were strongly related to passive GJT posttest scores ($r = 0.43$, $p < .001$). In addition, learners’ performance on the GJT was related to LLAMA_F ($r = 0.40$, $p < .001$) and to LLAMA_E ($r = 0.41$, $p < .001$) (i.e., grammatical inferencing and sound-symbol recognition respectively).

**Table 14**

*Correlations among Aptitude Components and the Passive GJT Scores*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LLAMA_B</td>
<td>1</td>
<td>0.41**</td>
<td>0.35**</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>LLAMA_F</td>
<td>1</td>
<td>0.01</td>
<td>0.44**</td>
<td>0.15</td>
<td>0.41**</td>
</tr>
<tr>
<td>3</td>
<td>LLAMA_D</td>
<td>1</td>
<td>-0.14</td>
<td></td>
<td>-0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td>4</td>
<td>LLAMA_E</td>
<td>1</td>
<td></td>
<td>0.21</td>
<td>0.41**</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Passive GJT pre</td>
<td>1</td>
<td></td>
<td></td>
<td>0.43**</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Passive GJT Post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**p < .01**
A hierarchical multiple regression analysis was next conducted, using the passive GJT posttest scores as criterion variables and the pretest scores, LLAMA_F and LLAMA_E as predictors variables for the passive GJT scores. The pretest scores were entered into the regression analysis in the first step, and then two aptitude subcomponents were entered together in the next step. The results revealed that the pretest scores (β = 0.42, t [57] = 3.65, p < .001) had a significant association with the posttest scores. That is, learners’ performance on the passive GJT pretest explained 18% of the variance (F [1, 57] = 13.11, p < .001) in the posttest. After controlling for pretest only, grammatical inferencing as measured by LLAMA_F (β = .26, t [55] = 2.14, p < .05) had a significant association with passive GJT posttest scores. LLAMA_E (sound-symbol recognition) did not have any significant association with the outcome variable (β = 0.23, t [55] = 1.84, p > 0.05) in the regression model (See Table 15 for regression model). Grammatical inferencing ability (i.e., LLAMA_F) increased the explained variance to 35% from 18%. That is, 17% of the variability in passive GJT posttest scores was predicted by grammatical inferencing when pretest scores were controlled. Regression model has medium effect size with $R^2$=.35.

Table 15

Summary of Multiple Regression Analysis for Aptitude Components Predicting Passive GJT Scores

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F_{change}$</th>
<th>df</th>
<th>$B$</th>
<th>$t$ (within set)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive GJT posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 1: Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive GJT pretest</td>
<td>13.11**</td>
<td>57</td>
<td>0.48</td>
<td>3.6**</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>Step 2: ID variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive GJT pretest</td>
<td></td>
<td>57</td>
<td>0.39</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>LLAMA_F</td>
<td>6.96*</td>
<td>56</td>
<td>0.01</td>
<td>2.14*</td>
<td>0.35</td>
</tr>
<tr>
<td>LLAMA_E</td>
<td>6.96</td>
<td>56</td>
<td>0.02</td>
<td>1.84</td>
<td>0.35</td>
</tr>
</tbody>
</table>

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

A second multiple regression analysis was conducted to examine learners’ performance on the passive OPT scores in relation to the aptitude subcomponents. Again,
aptitude components were screened in order to determine the best predictors. When correlations between the criterion variable and possible predictor variables were checked, no significant correlations were observed among any predictor variables except for the OPT pretest scores. Passive OPT pretest scores are significantly correlated to the passive OPT posttest scores ($r = 0.46, p < 0.05$). Correlations among the predictor and outcome variables are provided in Table 16. The correlations revealed that no aptitude subcomponent is related to the outcome variable but the model used for the earlier was used in order to run a regression analysis. LLAMA_F was the only predictor for the Passive GJT scores; therefore, LLAMA_F was used as predictors for the passive OPT as well. Therefore, the OPT pretest scores were entered as the first predictor in the first model. Next, LLAMA_F (i.e., grammatical inferencing) was entered into the analysis in order to examine whether a similar pattern would exist as it was observed as a possible predictor variable in the correlations.

Table 16

Correlations among Aptitude Components and the Passive OPT Scores

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 LLAMA_B</td>
<td>1</td>
<td>0.41**</td>
<td>0.35**</td>
<td>0.15</td>
<td>0.04</td>
<td>0.34</td>
</tr>
<tr>
<td>2 LLAMA_F</td>
<td>1</td>
<td>0.01</td>
<td>0.44**</td>
<td>0.21</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>3 LLAMA_D</td>
<td>1</td>
<td>-0.14</td>
<td>-0.18</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 LLAMA_E</td>
<td>1</td>
<td>0.31</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Passive OPT post</td>
<td>1</td>
<td>0.46*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Passive OPT pre</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results showed that pretest scores ($\beta = 0.47 \ t [28] = 2.74 \ p < 0.05$) had a significant association with the passive OPT posttest scores and explained 18% of variance in the passive OPT scores. However, the aptitude component of grammatical inferencing as measured by LLAMA_F did not explain any variance when pretest scores were controlled (See Table 17 for the regression model). After controlling for pretest
scores, LLAMA_F ($\beta = -0.03$, $t_{[27]} = -0.18$, $p > 0.05$) had no significant associations with passive OPT posttest scores. Regression model has a medium effect size with $R^2=.31$.

**Table 17**

*Summary of Multiple Regression Analysis for Aptitude Components Predicting Passive OPT Scores*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F_{change}$</th>
<th>df</th>
<th>$B$</th>
<th>$t$ (within set)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive OPT post-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 1: Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive OPT pretest</td>
<td>7.5*</td>
<td>28</td>
<td>0.51</td>
<td>2.7*</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Step 2: ID variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive OPT pretest</td>
<td></td>
<td>28</td>
<td>0.51</td>
<td>2.7*</td>
<td></td>
</tr>
<tr>
<td>LLAMA_F</td>
<td>1.8</td>
<td>26</td>
<td>-0.03</td>
<td>-.18</td>
<td>0.31</td>
</tr>
</tbody>
</table>

*p < .05

**Aptitude components as predictors for easy structure**

To investigate whether a specific aptitude component predicted learners’ knowledge of the past progressive, as measured by the GJT scores, it was necessary to conduct a logistic regression. This was because the past progressive GJT scores were not normally distributed. The logistic regression was conducted following the procedures used for the other regression analyses. The results of the non parametric correlational analyses showed that LLAMA_F, grammatical inferencing (Kendal’s tau_b= 0.19), LLAMA_E, sound-symbol recognition (Kendal’s tau_b:=0.23), and pretest scores (Kendal’s tau_b: 0.6) were related to progressive GJT posttest scores. See Table 18 for nonparametric correlations.
Following the correlation analysis, a logistic regression analysis was conducted to predict the membership of high GJT or low GJT posttest scores with pretest and L1 LLAMA_F and LLAMA_E scores as predictors. The results showed that a test of the full model was statistically significant, indicating that the predictors as a set reliably distinguished between high and low GJT posttest scorers ($\chi^2 = 25.94, p < .000$ with $df = 2$). Nagelkerke’s $R^2$ of .48 indicated a moderately strong relationship between prediction and grouping (See Table 19 for the summary of the logistic regression model). Prediction success overall was 74.1% (61.5% for low group and 84.4% for high). The Wald criterion demonstrated that only pretest scores made a significant contribution to the prediction (Wald’s value $= 13.2$, Exp ($B$) = 1.07 $p < .001$ with $df = 1$). LLAMA_F (Wald’s value $= 0.27$, Exp ($B$) = 0.99 $p > .05$ with $df = 1$) and LLAMA_E (Wald’s value $= 0.83$, Exp ($B$) = 1.01, $p > 0.05$ with $df = 1$) were not significant predictors for the outcome variable past progressive GJT scores.

Table 18

Correlations among Aptitude Components and Past Progressive GJT Scores

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LLAMA_B</td>
<td>1</td>
<td>0.29**</td>
<td>0.23*</td>
<td>0.13</td>
<td>0.08</td>
<td>0.24**</td>
</tr>
<tr>
<td>2</td>
<td>LLAMA_F</td>
<td>1</td>
<td>0.02</td>
<td>0.26**</td>
<td>0.19</td>
<td>0.29**</td>
<td>0.19*</td>
</tr>
<tr>
<td>3</td>
<td>LLAMA_D</td>
<td>1</td>
<td>-0.07</td>
<td>0.03</td>
<td>0.21*</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LLAMA_E</td>
<td>1</td>
<td>0.17</td>
<td>0.21*</td>
<td>0.23*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PP GJT categorical</td>
<td>1</td>
<td>0.52**</td>
<td>0.74**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Prog. GJT Pre</td>
<td>1</td>
<td></td>
<td>0.6**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Prog. GJT Post</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The last multiple regression analysis was conducted with the past progressive OPT data as the outcome variable. In order to decide which variables might be the best predictors for the regression analysis to be conducted on the past progressive OPT posttest, correlations between the dependent variable and aptitude components were checked (See Table 20 for correlations among predictors and the past progressive OPT scores). Pearson product moment coefficients revealed that the progressive OPT posttest (dependent variable) was highly correlated with OPT pretest scores \((r = 0.74, p < .001)\) and also related to LLAMA_B \((r = 0.51, p < .001)\). Following that, OPT pretest scores were entered into regression analysis in the first step, and then LLAMA_B (i.e., vocabulary learning) was entered in the next step.
Table 20

*Correlations among Aptitude Components and the Past Progressive OPT Scores*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 LLAMA_B</td>
<td>1</td>
<td>.41**</td>
<td>.35**</td>
<td>0.15</td>
<td>.37*</td>
<td>.51**</td>
</tr>
<tr>
<td>2 LLAMA_F</td>
<td>1</td>
<td>0.01</td>
<td>.44**</td>
<td>0.15</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>3 LLAMA_D</td>
<td>1</td>
<td>-0.14</td>
<td>0.03</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 LLAMA_E</td>
<td>1</td>
<td>0.07</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Prog. OPT Pre</td>
<td>1</td>
<td>.75**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Prog. OPT Post</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* LLAMA_B is the vocabulary learning task, LLAMA_F refers to grammatical inferencing task, LLAMA_D refers to sound-recognition and LLAMA_E refers to sound-symbol correspondence task in the test.

Results of the multiple regression analysis revealed that the pretest OPT scores ($\beta = 0.74$ $t [28] = 5.93$, $p < .000$) had a significant association with the posttest OPT scores. Table 21 shows the summary of the regression model. Progressive OPT pretest scores explained 55% of the variance ($F[1, 28] = 35.21$, $p < .000$). After controlling for OPT pretest scores, LLAMA_B ($\beta = 0.27$, $t [27] = 2.09$, $p < .05$) had significant associations with progressive OPT posttest scores. LLAMA_B increased the explained variance from 55% to 61%. The results revealed that there is a significant R square change explained by LLAMA_B, vocabulary learning test, when pretest scores are controlled.
The results with respect to the second research question demonstrated that aptitude explained some of the variance in learners’ performance on the language measures. The grammatical sensitivity task explained 17% of the variance in learners’ performance on the passive GJT whereas none of the tasks explained any variance in learners’ performance on the progressive GJT. The vocabulary learning task explained 6% of the variance in learners’ performance on the past progressive OPT. None of the aptitude components explained any variance in the passive OPT scores.

Data Analysis and Results for Research Question 3

Difficult structure: The passive

The third research question asks whether there is an effect of differences in aptitude profiles on the learning of the two target structures. In order to examine the effect of aptitude profiles on L2 learning a 3 (aptitude profile) X 2 (time) mixed design ANOVA with repeated measures was conducted. We began with learners’ performance on the difficult target feature, the passive, on the GJT. Normality assumptions were met with the transformed data. Time was the within group variable with two levels (i.e., pre and posttest) and aptitude profile (i.e., low, medium, and high) was the between group

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F_{change}$</th>
<th>df</th>
<th>$B_{t}$ (within set)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Progressive OPT post-test</td>
<td>35.21**</td>
<td>28</td>
<td>0.78</td>
<td>5.9**</td>
</tr>
<tr>
<td>Past Progressive OPT pretest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2: ID variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Progressive pretest</td>
<td>27</td>
<td>0.68</td>
<td>5.0**</td>
<td></td>
</tr>
<tr>
<td>LLAMA</td>
<td>4.37*</td>
<td>27</td>
<td>4.4</td>
<td>2.09*</td>
</tr>
</tbody>
</table>

*p < .05, **p < .001
variable with three levels. In order to explore the potential effects of aptitude profiles on the learning of the two target features, learners’ scores on the aptitude total score was categorized into three groups: high, medium and low aptitude with equal numbers in each group. I used SPSS to divide learners into three aptitude profile groups with equal numbers. This resulted in the following groupings: low aptitude group with scores between 60-179 (N= 19), medium aptitude group with scores between 180-234 (N=19), and high aptitude group with scores between 235-315 (N=22).

The analysis yielded a significant main effect for time ($F [1, 57] = 75.64, p<.000, \eta^2=0.57$) and all aptitude profile groups improved significantly from pretest to post-test (See Table 22 for ANOVA summary table). A significant interaction between aptitude profiles and time was also observed indicating that learners’ improvement from pretest to post test changed as a function of differences in aptitude profile =($F [2,57]= 3.44, p<.05, \eta^2=0.1$). Post-hoc analyses using Bonferroni comparisons revealed that in the pretest there was no significant difference among the three aptitude groups (low profile, $M= 4.77$; medium profile $M= 4.38$; high profile $M= 4.96$). However, in the posttest there was a significant difference between low aptitude profile learners and high aptitude profile learners (low aptitude profile, $M= 5.85$; medium aptitude profile $M= 6.34$; high aptitude profile $M= 7.35$). The high aptitude learners’ rate of improvement was significantly greater than the low aptitude group. That is, their starting point was similar but after receiving four hours of instruction on the target structure, the high aptitude profile learners improved significantly more than the low aptitude group. When partial eta squared indices for post-hoc analyses are examined, it was observed that low aptitude group comparisons had the lowest effect size ($\eta^2 = 0.13$), medium aptitude comparisons had medium effect size ($\eta^2 = 0.33$), and the highest aptitude profile comparisons had the highest effect size ($\eta^2 = 0.46$). As indicated above effect size for main effect (i.e., time) is relatively high ($\eta^2 = 0.57$) compared to interaction effect with a low effect size ($\eta^2 = 0.10$).
Table 22

*Aptitude Profile Differences and their Effect on Passive GJT*

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aptitude Profile</td>
<td>2</td>
<td>18.83</td>
<td>9.42</td>
<td>2.87</td>
</tr>
<tr>
<td>Error</td>
<td>57</td>
<td>187.25</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>98.15</td>
<td>98.15</td>
<td>75.64**</td>
</tr>
<tr>
<td>timeX Aptitude</td>
<td>2</td>
<td>8.93</td>
<td>4.46</td>
<td>3.44*</td>
</tr>
<tr>
<td>Error</td>
<td>57</td>
<td>187.25</td>
<td>3.28</td>
<td></td>
</tr>
</tbody>
</table>

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

The second analysis was conducted on the oral production data for the passive feature. In order to measure the effects of aptitude profiles on the passive OPT, I conducted a 3 (aptitude profile) X 2 (time) mixed design ANOVA with repeated measures on the passive OPT scores. Again, time was the within group variable and aptitude profiles were the between group variable.

The analysis did not yield a significant main effect for time ($F[1, 27] = 0.37, p > 0.05, \eta^2 = 0.01$). That is, the three aptitude profile groups (i.e. high, medium and low) did not improve significantly from pretest to posttest (See Table 23 for ANOVA summary). There was also no significant interaction for time and aptitude profile ($F[1, 27] = 0.69, p > .05, \eta^2 = 0.07$).

Table 23

*Aptitude Profile Differences and their Effect on Passive OPT*

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aptitude Profile</td>
<td>2</td>
<td>2.54</td>
<td>1.27</td>
<td>0.72</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>47.42</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Time X Aptitude</td>
<td>2</td>
<td>0.39</td>
<td>0.09</td>
<td>1.09</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>17.18</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

*Easy structure: The past progressive*

The third research question also examines whether there is an effect of differences in aptitude profiles on the learning of the easy structure. A mixed design ANOVA with
repeated measures was also planned, however as indicated above, the scores on the past progressive GJT were not normally distributed and the transformations did not resolve this. Therefore, the statistical analysis used in this instance was a general estimating equation (GEE), which is considered to be a statistical analysis similar to a repeated measure ANOVA, but does not require variables to be normally distributed. GEE\(^1\) requires the restructuring of the data. The negatively skewed data for the Past progressive GJT were restructured for the analysis. To do this I selected the options for a Gamma distribution model with identity link but the Gamma distribution is appropriate for positively skewed (not negatively skewed) variables, I reversed the GJT data to reflect “mastery” and “non-mastery” of the target feature. The analysis is conducted with the new form of the data. In the current study, past progressive GJT scores were restructured into a categorical variable (i.e., mastery and non-mastery). Therefore, the findings in this analysis were interpreted with reference to mastery and non-mastery of the target structure.

The findings demonstrated a main effect for time ($\chi^2 = 24.47$, $p < .001$ with $df = 1$). More importantly, there was an interaction between time and aptitude categories ($\chi^2 = 11.68$, $p < .003$ with $df = 2$). That is, the medium aptitude profile group improved significantly with regard to mastery of the past progressive GJT after receiving L2 instruction on the target structure but low and high aptitude profile learners did not improve with respect to mastery after receiving L2 instruction on the target structure. When pairwise contrasts with the Sidak test\(^2\) were checked, there was no significant improvement from the pretest to posttest with the low (sidak significance= 0.16) and high (sidak significance= 0.58) aptitude profile learners. However, the medium aptitude group improved significantly to the mastery after the instruction (sidak significance= .001).

---

\(^1\) The GEE approach is a statistical analysis used for longitudinal and correlated response data analysis (Stokes, 1999). GEE is based on quasilikelihood of inferencing compared to likelihood of inferencing in parametric analyses (Qin & Lawless, 1994).

\(^2\) Sidak is a post hoc test that includes a correction similar to Bonferonni correction which is considered to be less conservative and selected when there is a risk of loss of power associated with Bonferonni correction (Field, 2009).
Thus, the GEE analysis revealed that only medium aptitude profile learners improved with regard to progressive GJT after receiving L2 instruction.

In order to measure the effects of aptitude profiles on the past progressive OPT, I conducted a 3 (aptitude profile) X 2 (time) mixed design ANOVA with repeated measures. The analysis did not yield a significant main effect for time ($F [1, 27] = 0.04, p > 0.05, \eta^2 = .001$). Also, the three aptitude profile groups did not improve significantly from pretest to posttest. Furthermore, there was no significant interaction effects for time and aptitude profiles ($F [1, 27] = 1.09, p > 0.05, \eta^2 = 0.07$). See Table 24 below for a summary of the ANOVA analysis.

### Table 24

**Effects of Aptitude Profile on Progressive OPT**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aptitude Profile</td>
<td>2</td>
<td>2.54</td>
<td>1.27</td>
<td>0.72</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>47.42</td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Time X Aptitude</td>
<td>2</td>
<td>1.39</td>
<td>0.69</td>
<td>1.09</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>17.18</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

### Table 25

**Aptitude Profiles on Language Tests**

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Passive GJT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low aptitude</td>
<td>23.3</td>
<td>36.1</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>Medium aptitude</td>
<td>20.1</td>
<td>42.5</td>
<td>5.8</td>
<td>4.1</td>
</tr>
<tr>
<td>High Aptitude</td>
<td>27.1</td>
<td>55.5</td>
<td>5.1</td>
<td>8.1</td>
</tr>
<tr>
<td><strong>Progressive GJT</strong></td>
<td>72.1</td>
<td>82.7</td>
<td>12.6</td>
<td>13</td>
</tr>
<tr>
<td>Low aptitude</td>
<td>91.6</td>
<td>107.4</td>
<td>18.2</td>
<td>20.2</td>
</tr>
<tr>
<td>High Aptitude</td>
<td>109.3</td>
<td>112.1</td>
<td>20</td>
<td>22.4</td>
</tr>
</tbody>
</table>
The findings with respect to the third research question reveal that all aptitude profile groups improved significantly from pretest to posttest when learning the passive voice (See Table 25). With regard to past progressive learning, however, only the medium aptitude group improved significantly from pretest to posttest. However, learners were significantly different on past progressive GJT pretest with respect to mastery. The significant difference between two groups was also observed on the posttest. Furthermore, with respect to oral tasks none of the aptitude profiles improved significantly from pretest to posttest.

**Learners’ perception of target structure**

*Qualitative and quantitative data analysis*

Qualitative data were collected to gain more in-depth information about the learners’ perceptions of the target features that might help explain their performance on the oral and written language measures. In a retrospection questionnaire, learners were asked to provide a rule and/or description of how the past progressive and passive voice work in English. They were also asked to indicate whether their first language had helped in their learning of the two target structures (See Appendix I for the retrospection questionnaire). The responses were analyzed with a special focus on responses from the 10 high aptitude learners and the 10 low aptitude learners based on their LLAMA total scores. I was also interested in discovering whether there were differences in how high and low aptitude profile learners explained the grammatical structures examined in the current study. High and low aptitude groups were defined based on the total aptitude scores. This categorization does not include a medium group as in the quantitative analysis for the same research questions but the learners in the high and the low group correspond to the ones in the quantitative analyses.

In order to explore how the learners perceived the two target structures in terms of difficulty I asked them to rate the difficulty of the two target structures on a scale from 1 (easiest) to 5 (hardest). The questionnaire was administered after the learners had studied
both structures for four hours each. Therefore, I assume that they had the necessary information to comment on both structures. The learners’ responses showed that the passive voice was considered to be a more difficult structure to learn when compared with the past progressive (passive difficulty $M= 3.38$, $SD= 1.1$, past progressive difficulty $M= 2.05$, $SD= 1.01$). A paired-samples t-test indicated that this difference was statistically significant $t (59) = -7.45$, $p < .001$.

The questionnaire included open-ended questions which asked learners to comment on the difficulty of the selected L2 structures. The majority of the learners noted that the rules of the past progressive were easy to memorize, but the rules for the passive were more complex and that it was also more difficult to differentiate among the multiple uses of the passive voice. It was evident that there were several reasons that contributed to difficulty of the passive voice. For example, some learners reported that they had difficulty in deciding which tense to use in the passive voice. Some emphasized that it is not always the passive rules that create the difficulty but the rules of many tenses that are applied when using it.

As indicated above, learners’ rule explanations for both structures were analyzed in order to see whether there was any difference in how high and low aptitude learners articulated the rules of the target structures. I was interested in seeing whether high aptitude learners would provide more or better metalinguistic explanations. I also examined whether there was any difference in how the learners’ first language helped them when learning two target structures. Below are some selected responses from low aptitude learners:

**Participant 1** indicated that passive was only a little more difficult than the past progressive. She thought that the L1 did not help learn to the past progressive or the passive. She was not able to provide an adequate explanation for the past progressive: *an action that started at a certain point in time and ended at a certain time but took a long time*. However, she did not provide an explanation for the passive voice.

**Participant 2** rated passive and progressive as equal structures in terms of difficulty and she indicated that Turkish helped her learn both structures. She provided a
mechanical explanation for the description of the past progressive rule: *there is* –ing suffix and *was/were*. For the passive she gives the following explanation: *was/were gets an –ed*.

**Participant 3** considered the past progressive to be more difficult than the passive voice. She also indicated Turkish helped when learning both structures. She described the past progressive as: *a past event that started in the past and it has effects in the present.* She provided the following explanation for using the passive voice: *it happened in the past and ended in the past*. She also articulated how the passive voice is formed: *verb+ to be+ ed.*

**Participant 4** rated the past progressive as more difficult than the passive and she provided no response for the role of Turkish in learning the two structures. She provided the following explanation for the past progressive: *it is used for actions that started in the past and continues now*. She did not provide an explanation for the passive voice.

**Participant 5** also rated the progressive as more difficult and provided no response regarding L1 support and rule explanation.

The following are selected responses from the high aptitude group:

**Participant 6** indicated that the past progressive was easier than the passive voice. He noted that Turkish lessons on the passive voice helped him understand the passive better. For the explanation of the past progressive he provided the following: *things that took a certain time in the past. The helping verbs are was/were. The main verb gets –ing suffix.* For the passive rule explanation he provided the following: *sentences in which the verbs do not answer the who question. The subject is not important; the important thing is that the action is done.*

**Participant 7** also indicated that the progressive was an easy structure compared to the passive. She also reported that the Turkish classes helped her when learning the passive voice but it had no effect on the past progressive. When asked for the rule explanation for the past progressive she provided the following: *for the actions that took some time in the past and the structure is was/were + Ving.* For the passive voice, she provided the following explanation: *it doesn’t specify the time but changes based on time/tense. Objects become the subject of the sentence. An action that is done by something is expressed.*
Participant 8 also rated the passive voice as more difficult than the past progressive. He remarked that the L1 did not help for the past progressive and did not mention L1 support for the passive. The past progressive rule explanation that he provided: *was/were and –ing are used. It is used for tenses that happened in the past and took some time.* For the passive explanation, he provided the following: *it is used in the sentences where the doer is not important.*

Participant 9 indicated that both the passive voice and the past progressive are easy structures. He explained that past progressive is used for actions that started at a certain time in the past. The subject is followed by the past helping verb and the main verb is followed by the progressive form. For the passive voice, he indicated that it is the equivalent of the “edilgen yapısı” in Turkish. There is no object in such sentences. So-called subject is followed by the helping verb and the main verb takes the past participle.

Participant 10 also indicated that both structures were easy. She added that past progressive was a familiar structure. The passive, however, was a new one. She also thought that the passive is less frequently used in daily life and the examples were not familiar, making it difficult. When explaining the effect of L1, she said: passive voice is “edilgen yapısı” in Turkish and we have just covered it in Turkish classes so it hasn’t helped yet. Progressive is “şimdiki zamanın hikayesi” in Turkish and this piece of information helped me differentiate past simple from the progressive. She describes the past progressive as follows: it is used for actions which started in the past and it is not clear whether it is over or not. It is realized with the past form of to be and the verb and –ing suffix. In the next question, she describes the passive: passive is used when we want to emphasize the object. It is realized with to be verb and the V3.

The responses from the low aptitude learners show a pattern: there is a lack of consensus about the difficulty of the target structures. Some of the learners in this group identify the past progressive as more difficult than the passive voice and they are not making use of the L1 in learning the target structures. Moreover, they provided less developed explanations for the rule/structure explanations compared with the high aptitude learners.

The learner responses to the open-ended items in the retrospection questionnaire confirmed the results of the quantitative analysis. First, there is a substantial difference in
responses from the low aptitude learners compared with the high aptitude learners. Although most of the low aptitude learners considered the passive to be more difficult, they seemed to underestimate the magnitude of the difference. For example, they indicated that the passive was only slightly more difficult than the progressive. They were also not so sure about the role of their knowledge of the passive voice in Turkish. Another difference among the low and high aptitude learners was the lack of an adequate rule or explanation of the structural properties for the target structures. Most of the low aptitude profile learners could not provide an explanation for the passive. Some were confused about the use of the past progressive.

On the contrary the high aptitude profile group seemed to agree on the difficulty of the passive voice. All of the learners in the high aptitude group agreed that the passive voice was more difficult than the past progressive except for two learners who considered both structures to be equally easy. The learners in this group were able to articulate the rule or the use of the passive. Some of them provided detailed information about how their L1 helped them learn. It was evident that they were making use of the Turkish lessons they had received on the passive voice in their L1 as well.

**Conclusion**

This chapter reported the findings from the data analyses with respect to the three research questions examined in the study. Quantitative data analyses were used to investigate all three research questions while some qualitative data analyses were integrated with respect to the third research question.

The results with regard to the first research question revealed that different ID variables contributed to learners’ performance on the passive and progressive tasks. That is, language aptitude and motivation contributed to explain some meaningful variance in the passive GJT whereas the same ID variables had no relation to the past progressive GJT. Moreover, no ID variable included in the current study could explain any meaningful variance in the past progressive GJT. There are also differences between two target structures with respect to the oral production tasks. L1 metalinguistic awareness contributed to differences in learners’ performance on the progressive OPT but no ID
variable explained any variance in learners’ performance on the passive oral production task.

The results for the second question indicated that different components of aptitude operated in the learning of the difficult and easy L2 structures. The grammatical sensitivity task explained the significant variance in learners’ performance on the passive GJT but no aptitude component explained any differences in learners’ performance on the past progressive GJT. With regard to oral performance, no aptitude component explained any meaningful variance in learners’ performance on the passive OPT but the memory task in the aptitude test explained some meaningful variance in learners’ performance on the past progressive OPT.

The results with respect to the third research question showed that the low, medium, and high aptitude groups improved on the GJT for the difficult feature (i.e. passive) from the pretest to posttest after receiving four hours of L2 instruction. Moreover, the high aptitude learners were observed to be better than low aptitude learners on the posttest. However, only the medium aptitude learners improved significantly from the pretest to posttest on the GJT for the easy structure (i.e. past progressive). Low and high aptitude learners did not improve significantly from pretest to posttest after four hours of instruction. However, it was observed that high aptitude learners were significantly better than low aptitude learners on the passive GJT posttest.

Learners’ performance in their use of both structures on the oral production task showed that none of the groups improved significantly from pretest to posttest. It should be remembered that the scores on the past progressive OPT were rather high because learners previously studied the structure. Arguably, learners’ level of proficiency with respect to the past progressive prior to L2 instruction prevented a substantial level of improvement with respect to the easy structure.

The findings from the questionnaire to measure learners’ perceptions of the target features showed that high aptitude learners provided better metalinguistic explanations for the two structures. Moreover, high aptitude learners demonstrated a better
understanding of the rules of the target structures and they were able to articulate the rules about the target structures in an efficient manner. High aptitude learners were also more aware of the L1 language support when learning L2 structures.
Chapter Six
Discussion and Conclusions

This chapter discusses the results related to the three research questions investigated. That section is followed by a discussion of the implications of the findings, the limitations of the study and suggestions for future research.

Research Question 1

The results showed that aptitude and L2 motivation explained 17% of the variance on the passive GJT posttest scores but did not explain any meaningful variance on the passive OPT posttest scores. The findings for the written task (i.e., GJT) are consistent with previous research which has documented aptitude as the most consistent predictor of L2 success (Dörnyei & Skehan, 2003; Oxford & Ehrman, 1995). In a review of ID research, Dörnyei and Skehan (2003) conclude that “aptitude could be a central construct especially when L2 instruction includes focus on form” (p. 601). The findings with respect to the passive voice appear to support Dörnyei and Skehan’s conclusion and confirm results from previous research conducted in foreign language classrooms (Erlam, 2005; Kiss & Nikolov, 2005; Grigorenko, Sternberg, &Ehrman, 2000; Sáfár & Kosmos, 2008) as well as experimental studies conducted in different L2 learning contexts (de Graaff, 1997; Ehrman & Oxford, 1995; Horwitz, 1987; Ranta, 2002; Reves cited in Skehan, 1989; Robinson, 1997).

In the current study, learners participated in form-focused activities and received concise explicit grammatical explanation that was integrated in meaning-based activities. However, similar to the learners in most foreign classrooms, learners in the present study were more familiar with rule explanation, mechanical drills and other features typical of more traditional structure-based teaching. Therefore, in order to gain insights into the present findings, they are best compared with studies examining IDs in foreign language classrooms. One such study is, Erlam (2005) in which she investigated the interaction of aptitude with the learning of one target structure (i.e., direct object pronouns in French) and three types of L2 instructional methods in a French foreign language classroom. The
deductive instruction included form-focused activities following rule explanation. Learners were provided with correction and corrective feedback and a chart of pronouns was provided for learners to consult. The inductive group received input-based activities that were planned to lead them to hypothesis-testing about direct object pronouns. The last group received structured input and was provided with explicit information, rule explanation, and consciousness-raising activities but all activities were input-based and no production was required. Erlam considered direct object pronouns to be difficult for English learners of French due to several morphosemantic distinctions such as gender, person, and number (See Erlam, 2003a for details). This makes the designation of the target structure in her study similar to the designation of the passive voice in the present study. No relation between learner aptitude and L2 gains were observed in the ‘deductive’ group but there were significant relationships between aptitude and L2 gains in the ‘inductive’ group (i.e., scores in the MLAT correlated significantly with written production in the immediate posttest) and the ‘structured input’ group (i.e., scores in the MLAT correlated significantly with written production in both immediate and delayed L2 gains). Interestingly, the results of Erlam’s study showed that deductive instruction, the type which she defined as most similar to the learners’ regular foreign language instruction, did not interact with differences in learners’ aptitude to produce variation in learning outcomes. However, in the ‘inductive’ group where learners took an active role in hypothesis testing, aptitude played a positive role in enabling learners to benefit from the instruction without rule explanation. Similarly, aptitude played a positive role in the ‘structured input’ group where learners did not produce the target structures during the instruction yet high aptitude learners produced the target structure more frequently on written tests. While the instruction provided in the current study is not directly comparable to the instructional treatments in Erlam (2005), it is most similar to ‘structured input’ because learners in Erlam’s ‘structured input’ group received explicit information and rule explanation about the target structures as the students did in the current study. Erlam’s learners also worked on input-based activities that encouraged them to process spoken and written input and they were not required to produce the target structure. Although there was no restriction on the production of the target structures in
the current study; there was less emphasis on oral production compared to written production and the learners were exposed to primarily input activities.

Erlam interpreted her findings with respect to two major theories in aptitude research: Carroll’s (1962) Model of School Learning and Skehan’s model of aptitude (1989). According to Erlam, the findings with regard to deductive instruction confirm Skehan’s (1989) claim that less structured input allows a greater role for language aptitude. The fact that deductive instruction appears to have been resistant to the effects of language aptitude is also consistent with Carroll’s Model of School Learning which argues that language aptitude is only one of the factors that contribute to L2 success. The fact that learning gains in deductive learning were significantly higher than the two other groups enabled Erlam to speculate that learners in the deductive group were more motivated and participated more because the L2 instruction fit their expectations and preferences. Furthermore because the deductive instruction provided the necessary information to L2 learners in an explicit and structured manner, learners did not have to call upon language aptitude to figure out the language regularities. According to Erlam (2005) the motivation that came out of the match between learner preferences and L2 instruction resulted in minimizing the effects of aptitude.

Erlam’s study provides compelling evidence that some features of L2 instruction can minimize the effects of language aptitude. One type of L2 instruction was unrelated to the effects of aptitude. Therefore, I would argue that the type of instruction is an important factor in determining the role of ID variables in L2 learning. In the current study, language aptitude is related to L2 learning of difficult structures during instruction that combined attention to form within meaning-based instruction. Additionally, unlike the deductive group in Erlam’s study, explicit information and rules were not provided in as systematic and detailed manner. Learners in the current study were given opportunities to test their hypotheses in communicative practice. Skehan (1989) argued that greater aptitude would lead to success in dealing with less structured material. Erlam’s study supports Skehan’s argument showing that aptitude plays a more significant role in the ‘inductive’ and ‘structured’ input groups. In the current study, all learners were provided with meaningful input and less structured concise explicit information integrated into
meaningful practice, which left room for aptitude effects with regard to the difficult L2 structure.

Another important finding in Erlam’s study relates to learners’ performance on the oral production tasks. These tasks required learners to produce the target structure in a limited time with a focus on meaning. Oral production gains were found to relate to aptitude only with the inductive group. However, the relation was negative. That is, learners with greater language analytical ability were less likely to produce the target structure in oral communication. In the present study there was no significant relation between aptitude and oral production gains. Therefore, I would speculate that high levels of aptitude were important for written tasks but not with respect to the difficult structure (i.e., passive voice) on the oral production task.

Results with regard to past progressive GJT scores were different from what was observed with the passive GJT scores. As will be recalled language aptitude was not related to past progressive GJT scores but L1 metalinguistic awareness emerged as a related ID variable. Although L1 metalinguistic awareness was observed to correlate with performance on the past progressive GJT, it did not contribute to explain any significant variance in the GJT posttest scores. However, as indicated earlier, the past progressive scores were highly skewed due to the learner’s previous knowledge of the target structure. Therefore, it is possible that the power of statistical analysis was lessened due to the nature of the data, making it difficult to capture the role of L1 metalinguistic awareness in the regression model.

To better understand the past progressive findings it may be more helpful to examine the oral performance results as measured by the OPT. The results of the regression analysis revealed that L1 metalinguistic awareness explained a meaningful variance in the OPT posttest scores. I would argue that L1 metalinguistic awareness emerged as a predictor variable in learners’ performance on the past progressive OPT because of similarities between Turkish and English. As noted earlier, in Turkish the past progressive tense is much more frequent in daily use compared to the passive voice and its meaning and use in Turkish cover exactly the same meaning and use in English.
Therefore, this similarity could have led learners to use their L1 metalinguistic awareness as a resource. Familiarity with the target structure is another factor that could have led them to refer to their L1 knowledge. It should be recalled that metalinguistic awareness and aptitude scores have been reported to be correlated (Alderson et al., 1997; Ranta, 2002). As a result, Ranta (2002) considers metalinguistic awareness and aptitude as two sides of the same coin. Moreover, they both fall into ‘language analytical’ ability in Skehan’s terms. Thus it is likely that in the present study, learners’ previous knowledge of the past progressive enabled them to draw upon their L1 metalinguistic ability as a cognitive resource reflecting on L1-L2 similarities between the two structures.

The findings related to the learners’ oral and written performance on the passive voice compared with those for the past progressive revealed different patterns. Dörnyei (2010) reports that “correlations between aptitude and L2 attainment indices are often as high as 0.50 and meaningful correlations with motivation have usually been within the range of 0.30-0.40” (p.248). Language aptitude was found to be a variable that explained 12% of the variation in the passive. Admittedly, this represents a small portion of the variance. Nonetheless it is a significant finding and none of the other ID variables accounted for as much of the variance. Motivation also made a significant contribution to the passive GJT scores (5%). Thus, the combination of motivation and aptitude accounted for 17% of the variation on the passive written test (i.e., GJT). This is consistent with the socio-educational model of SLA developed by Gardner (1985; 2001) which represents a modular and stable view of ID variables. The finding that language aptitude and motivation were not related to oral and written production for the easy structure (i.e. past progressive) are not consistent with the model. A more process-oriented perspective on individual learner differences might help to explain this. Since the 1990s, ID variables, particularly motivation, have been reconceptualized and researchers have given more consideration to the learning context (Dörnyei, 2000, 2005, 2010). According to Dörnyei (2010), a situated approach to IDs acknowledges that an individual’s motivation is not stable and it shows constant fluctuation in various degrees. The findings in relation to the past progressive might be better explained by a situated approach to ID variables confirming Dörnyei’s (2010) argument that “ID variables display a considerable amount of variation from time to time and from situation to situation” (p.252). This argument

114
raises the question as to what made the difference across these two learning situations. One might speculate that the differences in the L2 target structures (e.g., difficulty) and different learner experiences with the target structures played a role. I would argue that differences in the nature of the target structures and learners’ previous knowledge about the past progressive contributed to the variation in the learning outcomes. Even though the two learning situations were similar, learners’ responses to the learning experience were also shaped by their previous knowledge and task demands. The passive was a novel structure for them, they were engaged in the instructional activities, and they benefitted from the instruction. The instruction provided was motivating for them containing meaningful activities and concise explicit information. However, the same type of instruction on the past progressive was not motivating because they had already been taught the structure. Therefore, the L2 instruction provided on the past progressive might not have been sufficiently challenging and may not have matched learners’ expectations about L2 instruction on this language features. Nonetheless, the nature of the L2 feature (i.e., the past progressive) played a role in determining what cognitive abilities were to be called upon. Learners probably made use of their L1 metalinguistic knowledge on their past progressive performance since it is relatively easier to recognize similarities between L1 and L2 similarities or differences when you have at least some knowledge of the target structure. Some researchers have argued that metalinguistic awareness is closely related to language analytical ability and therefore to language aptitude (Ranta 2002; Skehan, 1989). Following these arguments, I would argue that based on the similarities in L1-L2 on the specific target structure, learners relied on their metalinguistic awareness rather than language aptitude. So this change between aptitude and L1 metalinguistic awareness could be interpreted as a slight change in the cognitive resources used across the two language features being learned.

Looking at her findings, Erlam (2005) suggests that the issue of motivation is also an important factor in her study especially for the learners in the deductive group. Erlam argues that the deductive instruction was motivating for learners and minimized the effect of aptitude. The findings of the current study are consistent with her suggestion. Despite the fact that the instruction was the same for both structures in the current study, aptitude and motivation are only related to the learning of the passive voice. Similar to the
deductive instruction in Erlam’s study that was compatible with learners’ previous experience and preferences, I would argue that the instruction on the passive voice in the present study suited learners’ needs best and thus motivation emerged as a second variable accounting for learners’ performance in addition to language aptitude. Contrary to Erlam’s findings where motivation was a stronger predictor than aptitude, L2 instruction on the passive voice did not minimize the effect of language aptitude. That is, while motivation was related to learning, it did not rule out the effects of aptitude in the current study. I would argue that the nature of the L2 structure is what distinguishes these results. The passive voice was both difficult and novel for the learners in the present study; therefore, they benefited from higher levels of aptitude to learn the new structure. On the other hand, the learners in Erlam’s study were reported to have some familiarity with the target structure even though it was also considered to be difficult.

Learning the passive voice for learners in this study was challenging and probably required learners to look for patterns and rules in the input. Although they were given some information about the target structure, it was concise and perhaps not enough. As a result, they relied on their language aptitude ability. I assume that the novelty and difficulty of the passive voice also contributed to learners’ motivation to learn the target structure.

The results regarding the first research question indicate that different ID variables contributed to the explained variance in learners’ performance on the two target structures. These results partially support Krashen’s (1981) claim that aptitude is only correlated with success in “learning” contexts where formal accuracy and metalinguistic information are the focus of language instruction. This is because the L2 instruction provided in this study included attention to both form and meaning. However, because the attention to form was embedded within communicative practice, the instruction provided in this study is also similar to that provided in other studies where the role of aptitude in L2 learning has been investigated in more communicative classrooms (Harley & Hart, 1997; Horwitz, 1987; Oxford & Ehrman, 1995; Sáfár & Kosmos, 2008). In fact, the instruction provided in the current study seems to stand somewhere between the deductive learning in Erlam’s (2005) study and the communicative instruction in Ranta’s
In this way, the findings could also be taken as counterevidence to Krashen’s claims. I would argue, however, that the present findings simply show that aptitude also has predictive value in classrooms where a combination of form and meaning-based instruction is provided.

Research Question 2

The second research question examines the unique contributions of aptitude sub-components to the learning of the two target structures. This research question was examined through four aptitude component scores (i.e., grammatical sensitivity, sound-symbol correspondence, sound recognition, and vocabulary learning) and the language measures. The first part of this question focuses on the passive target feature. It will be recalled that the regression analysis indicated that only grammatical inferencing ability (i.e., LLAMA_F) made a significant contribution to the passive GJT scores in addition to pretest scores. It explained an additional 17% of the variance. The results also indicated that grammatical inferencing ability as measured by the LLAMA_F test did not contribute to learners’ performance on the passive OPT. With regard to the past progressive and the aptitude sub-components, the logistic regression revealed that the grammatical inferencing subcomponent did not explain any variance in past progressive GJT scores. However, memory measured via the vocabulary learning subtest explained 6% of variance in the past progressive OPT scores. LLAMA_B test contributed significantly to the past progressive OPT scores even though the contribution is relatively small.

These two aptitude components (i.e., grammatical inferencing and memory) are comparable to the two most extensively used aptitude tasks in the MLAT that have been investigated in earlier ID research (i.e., words in sentences and paired associates). These tests have been reported to be related to several language skills including written production (Erlam, 2005), vocabulary, and listening comprehension (Harley & Hart, 1997, 2002). LLAMA_F corresponds to the Words in Sentences in the MLAT. The word in sentences in the MLAT was developed to measure grammatical sensitivity of language learners. Grammatical sensitivity refers to “the ability to recognize the grammatical
functions of words (or other linguistic entities) in sentence structures” (Carroll, 1981, p. 105). The findings for the passive GJT indicated that that when learners were judging the grammaticality of sentences on the written task they relied more on their grammatical sensitivity than the three other aptitude subcomponents measured by the LLAMA. The same aptitude subcomponent did not explain any variance in the past progressive GJT scores. This result is important considering that the two structures are considered to differ in terms of their difficulty. When learners were exposed to a novel structure in the instruction learners had to rely on their grammatical inferencing ability in order to determine the grammaticality of the target sentences. The instruction probably pushed learners to look for regularities and rules in the input. Learners were provided with some explicit information but it is likely that learners continued to formulate hypotheses and look for clues in the input to confirm such hypothesis.

Different aptitude components were used as a cognitive resource when learners performed on the past progressive tasks. Rote memory ability, as measured by the vocabulary subtask (i.e., LLAMA_B), appears to have been more useful for learners to use the past progressive with respect to oral performance. Those findings with regard to written and oral performance on both structures partially confirm Skehan’s aptitude profile (2002). As indicated in Chapter 2, different aptitude components come into play at different stages of his dynamic model. Noticing is the first stage where learners are introduced to a new structure and learners direct their attention to some aspects of the language system. Pattern identification follows this stage and learners are expected to make generalizations based on the language input available to them. Lexicalizing is the last stage where the learner is capable of using the language domain without error. The learner is then able to produce a rule-based creative version of the language domain. According to his model, grammatical sensitivity comes into play at the beginning of the second stage (i.e., pattern identification) and memory is necessary in the last stage (i.e., lexicalizing). Skehan notes that the model is only a proposal and that it has not been tested empirically, but the findings of the current study with respect to the passive GJT and OPT are consistent with Skehan’s model. That is, when learners encountered a novel/difficult structure (i.e., the passive) they worked through “pattern identification” using their language analytical ability. A close examination of the GJT scores showed
that learners were able to identify the errors and correct them suggesting that after going through the pattern identification stage they were able to make generalizations based on the regularities in the input they had received. However, because the instruction only lasted for four hours, they probably did not have enough time to automatize this knowledge. Therefore, they were not successful in extending their understanding of the passive to produce it spontaneously in oral production. Automatizing and lexicalising are the last two steps in Skehan’s model. Arguably, learners in this study needed more time to reach and move through these stages with respect to the passive voice.

Learners’ performance on the past progressive is also consistent with Skehan’s model of aptitude. Memory is the aptitude component that contributed most to learners’ performance on the past progressive OPT and memory also comes into play in the later stages in Skehan’s model. It will be recalled that learners’ performance on both the written and oral tasks for the past progressive was high particularly on the GJT which produced a ‘ceiling effect’. Nonetheless, a detailed examination of learners’ performance on the OPT suggests that learners were at later stages of processing (i.e., automatizing or lexicalizing) with respect to the past progressive. According to Skehan’s model, learners at this stage are able to choose freely, and appropriately, between lexicalized representation of an interlanguage form, and its rule-based creative version (p. 89). It is evident that in the oral performance, learners were producing the target structure frequently. They were also able to switch between when and while clauses in the past progressive. This is in line with rule-based creative use suggested by Skehan.

The findings above offer support for Skehan’s proposal that different aspects of language aptitude operate during different stages of L2 learning. The findings demonstrate that different aptitude components contribute to L2 learning across two L2 structure types in an EFL context. Although Skehan’s proposal was developed with adult L2 learners in mind, the current study provides support for it with young learners in an EFL context. The present findings also present support for Skehan’s attempts to link different components of aptitude to nine stages in language acquisition. However, because Skehan’s model of information processing and work on aptitude-SLA stage links
have not been tested empirically my speculations will require further empirical investigation.

Another noteworthy finding of the study is related to the type of learning that is measured by aptitude tests. There is limited information about the type of knowledge that is measured in aptitude tests (Dörnyei, 2010). Krashen (1981) argued that aptitude tests measure only “learned” knowledge but his ideas were not tested empirically. The MLAT has been used extensively in SLA research and is considered to be a valid measure of language aptitude (de Graaff, 1997; Harley & Hart, 1997; 2002; Oxford & Ehrman, 1996; Ranta, 2002; Robinson, 2002; Wesche, 1981). Recently however, efforts have been made to develop new aptitude tests. One example is Meara’s (2005) computer-delivered LLAMA aptitude test; however, we do not yet have enough proof of its validity. The only up-to-date study to have investigated the components of LLAMA tests is a factor analysis conducted by Grenane (2010) in which she reported that the three LLAMA subtests (i.e., grammatical sensitivity, sound-symbol correspondence, and vocabulary learning) loaded onto one factor characterized as explicit learning rather than implicit learning and the fourth component—sound recognition—loaded onto implicit learning. The findings of the current study are mostly in line with Grenane’s results. Sound recognition was not significantly related to other aptitude tasks except for one (i.e., vocabulary learning) in the current study but grammatical sensitivity, vocabulary learning and sound-symbol correspondence tasks were all interconnected. Therefore they seem to behave similarly. Grammatical sensitivity and vocabulary learning are the two aptitude components that are related to oral or written performance in the L2 tasks in the present study. Furthermore, a closer examination of aptitude scores with respect to other ID variables provides support for this argument. All the language aptitude components, except for sound recognition, are significantly correlated to L1 metalinguistic awareness scores. Other researchers have also provided evidence for a close relationship between grammatical sensitivity and metalinguistic knowledge (Alderson, Clapham, & Steel, 1997; Ranta, 2002). As indicated above, Ranta (2002) also claims that language analytical ability, which includes grammatical sensitivity and inductive language learning ability, is an overlapping term with metalinguistic ability. It is important to note that LLAMA_D (i.e., sound recognition), which, according to Grenena’s findings, measures implicit learning, is not
related to any of the language outcomes or to the L1 metalinguistic task used in this study. Thus, it seems reasonable to argue that the aptitude test used in the present study measures explicit learning, which is the type of learning that learners engage in during their regular EFL instruction. However, a detailed factor analysis was not conducted on the LLAMA scores in the current study; therefore, suggestions about the type of knowledge measured by it and the subtests require further empirical support.

Another noteworthy finding of the present study is related to the aptitude and working memory capacity (WMC) test. It has been proposed that WMC is significantly related to language aptitude (Miyake & Friedman, 1998, Sawyer & Ranta, 2001). Miyake and Friedman (1998) even argued that WMC could be taken as aptitude and that this was particularly the case with adult learners. Yoshimura (2000) also demonstrated that language aptitude as measured with the language aptitude battery for Japanese is significantly related to working memory measured with a reading span test in adult learners. However the current study provides no support for the close relation between WMC and aptitude as none of the aptitude components are significantly related to WMC. Even though this study, conducted with young adolescents, provides no counterevidence to Miyake and Friedman’s argument, it also provides no evidence in support of such a connection in the L2 instruction of this age group. The findings with regard to the WMC could be due to the instrument preferred in the current study. The WMC test used in the current study was chosen mostly due for practical reasons. Although backward digit span is accepted as a valid measure of WMC in SLA research (see Juffs & Harrington, 2010 for a review) some researchers, for example, Tagarelli, Borges-Mota, and Rebuschat (2011) argue the opposite. Therefore, one might argue that lack of relation between WMC and any other ID variable especially aptitude could be due to the WMC test. Another possibility is the fact that there could be a threshold level for the WMC across young adolescents; therefore, it is difficult to tap any variance across learners. Learners in the current study were all 8th grade learners and they are all 13 year old. Therefore, they could have been rather similar in terms of working memory capacity.
Research Question 3

The last research question explored whether there is a difference in the effect of language aptitude profiles on the learning of the two target features. Below, the first section addresses the quantitative results with respect to the difficult structure (i.e., the passive) followed by the easy structure (i.e., the past progressive). Qualitative analyses of learners’ knowledge about the target structures are discussed at the end.

It will be recalled that there were three aptitude profile groups (i.e., low, medium, and high) in the final analysis. The findings indicated that all aptitude groups improved significantly on the written GJT for the passive voice after receiving four hours of instruction. However, a close examination of the results showed that learners in the different aptitude profile groups benefitted differentially from the instruction. That is, high aptitude learners performed better than low aptitude learners on the more difficult feature (i.e., the passive voice) as measured by the GJT posttest. It should be recalled that the three aptitude groups started out at the same level with respect to performance on the passive voice, but after receiving four hours of instruction, the high aptitude group improved significantly more than the low aptitude group. This is consistent with Carroll’s (1990) views about aptitude and his argument that “the different levels of ability on aptitude tests correspond to increasing difficulties and longer times to learn whatever has to be learned in acquiring a foreign language” (p. 26). While four hours of instruction might be insufficient to adequately teach the passive voice, it seems that the high aptitude learners made the most of it and developed enough understanding of the L2 structure to apply it to new sentences. It is important to note that three groups, including the medium and low aptitude learners, also benefited from the instruction, but the low aptitude learners did not learn as much as the high aptitude learners. This finding is important in two ways. First, it provides additional evidence to support the claim that aptitude is a predictor of L2 learning particularly with respect to learners’ performance on a written GJT measuring the knowledge of a difficult language feature. Second, it indicates that low levels of aptitude do not mean failure for learners – they also improved and became more successful using the target feature.
The results related to aptitude profiles and improvement over time with respect to learners’ ability to produce the passive on the OPT differed considerably. There was no significant improvement from pretest to posttest regardless of high, low or medium aptitude ability. This could be due to the limited time available for instruction and for practice. That is, four hours of instruction might not have been enough for EFL learners to study a novel target structure and to proceduralize that knowledge. On the other hand, four hours of instruction appears to have been enough for learners to improve their knowledge of the passive feature as measured by the written GJT. Furthermore, the qualitative data provide compelling evidence that the learners had acquired some declarative knowledge of the structure. Therefore, it seems reasonable to conclude that learners possessed explicit knowledge of the passive voice. However, learners’ performance on the OPT suggests that more effort and time to develop learners’ oral production abilities is warranted. This is discussed in more detail below in the pedagogical implications section.

The results of learners’ performance on the easy structure, the past progressive, pattern differently. Only the medium aptitude profile learners improved significantly from pretest to posttest on the GJT. What this suggests is that the low aptitude learners did not benefit from L2 instruction or they did not add anything new to their repertoire of the past progressive because they had already possessed a high level of knowledge about it before the instruction. Similarly, high aptitude learners did not benefit from the instruction probably because there was less room for improvement. It should be recalled that when only posttest scores are compared the high aptitude learners were significantly better than the low aptitude learners. Yet the medium aptitude group was found to be the only group that improved. It will be recalled from the findings of RQ 1 that language aptitude was not correlated to performance on the past progressive and that metalinguistic awareness emerged as a factor. In accordance with Sawyer and Ranta’s (2001) claim, I would speculate that L2 instruction provided for the past progressive seemed to be below the necessary threshold level to trigger the language effects of aptitude in learning. Therefore, learners did not have to draw on language aptitude skills; instead they relied
on their available metalinguistic knowledge. However, this speculation warrants further empirical investigation.

Similar to the results for the passive OPT, there was no significant improvement over time on the past progressive OPT for any of the aptitude profile groups. There was a different pattern for the progressive OPT, however, because learners had relatively high scores on the pretest ($M = 17.9$ out of 30). Similar to pretest scores in the progressive GJT, learners also performed well on the past progressive OPT. Therefore there was less room for improvement. The OPT is considered to be a more difficult task in comparison with the GJT, but the results show that learners were successful on both measures. This suggests that the differences between learners’ performance on the oral production tasks for the passive and past progressive is more related to the difficulty of the target feature than to aptitude profiles.

The qualitative data analysis revealed that there are differences in how high and low aptitude profile learners diverge in their perceptions of difficulty and their articulation of the structural properties of the two target features. Qualitative data gathered through student questionnaires provided further support for the effects of aptitude profiles on L2 learning as measured by the GJT. The responses to the questions that asked for descriptions of the rule/structural characteristics of the two target structures showed that there were differences between how high and low aptitude profile learners articulated the rules/structures of the target features. High aptitude students were much more aware of the structure and better at articulating the rules for and the use of the passive and the past progressive.

As indicated in Chapter 4, coincidentally the learners in the current study had recently studied the Turkish equivalent of the passive voice in their Turkish classes. When I consulted with the Turkish language teacher, I found out that he provided similar types of explanations for the introductory passive voice. However, I did not include any Turkish-English comparisons in the instructional intervention. Qualitative data supports the finding that high aptitude learners were able to make use of the metalinguistic knowledge about the passive in Turkish. They indicated that Turkish helped them
understand the passive voice better. On the other hand, all but two low aptitude learners did not provide satisfactory explanations on the same questions that asked for rule/structure explanations. They typically provided limited information/explanation for the questions. Furthermore, most of them were not aware of the help from Turkish because they either said it did not help learn the new structures or they did not respond to the question.

Another difference between the responses from the high and low aptitude groups was that the high aptitude group considered the passive more difficult than the past progressive, whereas only a few low aptitude learners indicated that they found the passive more difficult. High aptitude learners were able to explain the difference between the passive and the progressive, both in terms of complexity and use, noting that the passive was the more difficult structure to produce. Their behaviors are in line with what Ehrman (1996) suggests about behaviors of high aptitude learners in her study of the relation between language aptitude and personality traits (the MLAT):

*Those who are open to new material can tolerate contradictions, establish hypothesis to be tested, focus on meaning, and find ways to link the new with the previous knowledge structures seem to have an advantage in managing the complex demands of language and culture learning. The weakest students appear to be overwhelmed by the chaos they encounter; the strongest meet it head on, may even embrace it to a degree* (p. 209).

The responses that the low aptitude learners provided in the current study seemed to be at the surface level without much evaluation and/or reflection about the L2 structures and their use. On the other hand, participant 10, who is from the high aptitude group, indicated that she considers both structures to be easy but admits that the passive is not frequent in daily Turkish; therefore, it could be considered to be more difficult. This statement suggests that high aptitude learners in the current study are more aware of the L2 structures and the learning entailed in such structures. They are also more capable of reflecting on their previous experiences in learning. I would argue that the above stated differences between high and low aptitude learners’ perceptions are consistent with the
quantitative results. There are substantial differences between high and low aptitude learners in how they perceive difficulty and how they are aware of the L2 process. These differences are presumably reflected in their learning experience. When learners are more aware of the possible difficulties and available cognitive resources such as L1 knowledge, they are possibly better equipped for language learning experience.

The findings with regard to the last research question show that the nature of the target structure is a factor that contributes to the variance in L2 gains. The nature of an L2 structure, namely the difficulty, interacts with the aptitude profiles within this target group of learners. Differences in aptitude did not prevent learners from learning the difficult structure in a relatively short time. However, there was a significant difference in the rate of improvement between high aptitude and low aptitude learners. As for the easy structure, another factor, the previous knowledge of the L2 structure, seems to be intervening. That is, aptitude profiles do not seem to make any difference in learners’ level of improvement because only medium profile learners improved significantly after receiving L2 instruction. As the learners were familiar with the easy structure, it is difficult to tap the effects of aptitude profiles with respect to gains in the easy structure.

Qualitative analysis also revealed some difference between high and low aptitude learners with regard to their perceptions of L2 structure difficulty and rule definition. High aptitude learners seem to be more aware of the nature of L2 structures and they provide better rule explanations compared to low aptitude learners who seem to underestimate the difficulty of the L2 structures and who seem to be less aware of rules attached to the target structures.

Conclusions and Implications

The findings of the current study support previous research that has revealed a positive role for aptitude in L2 learning. The findings also indicate that language aptitude is related to L2 learning in primarily meaning-based language classrooms in which attention to form was integrated into the lessons. However, the study also demonstrates a complex interplay of several ID variables. It acknowledges that language aptitude plays a role in language learning but there are other factors (i.e., motivation and L1
metalinguistic awareness) that also contribute to learning. The nature of the L2 structure, structure difficulty in particular, affects which cognitive abilities L2 learners rely on.

The results also show that a lack of aptitude does not debilitate learners. This confirms R. Ellis’ (2009) argument that aptitude enhances the rate and ease of learning but it is not a prerequisite for L2 acquisition. Learners are expected to overcome a lack of aptitude with the help of effective L2 instruction and the right amount of time allotted for learning. Moreover, the quality of instruction may be more important than aptitude abilities. The study further reveals that when learners are introduced to a new and relatively difficult structure, aptitude is a cognitive ability that may help them to learn the target structure in less time and in more depth.

The study has several pedagogical implications for L2 teachers in general and EFL instructors teaching in this research context (i.e., Turkey) in particular. The results have also shown that individual differences in language aptitude did not prevent learners from learning a difficult structure in four hours of L2 instruction. However, learner gains were limited to written production of the target structure; therefore, the findings suggest that high grammatical sensitivity contributes to grammatical accuracy in written production but not to accuracy in oral production. EFL teachers acknowledge that written grammatical accuracy is considered to be more important in EFL curriculum in Turkey. Thus, learners are less familiar with oral communicative tasks. Furthermore, there are few (if any) opportunities for learners of English in Turkey to participate in communicative interactions in the target language. Therefore, it is important that EFL teachers and curriculum planners aim to develop better oral production skills in L2 learners.

If teachers have access to information about individual learner differences, classroom activities could be planned based on learner strengths or weaknesses. Aptitude tests used to be difficult to administer in classrooms but continuous attempts have resulted in more practical and affordable tests. Therefore, teachers might consider administering such tests to learn more about their learner profiles. If the majority of learners in the language classroom lack strong aptitude skills, language teachers might
consider including more explicit information about the language system underlying the target language domain or they might consider modifying their instructional approach in other ways. In particular, teachers may choose from various form-focused techniques within the explicit or implicit continuum (Doughty & Williams, 2003). Matching learners’ strengths with different FFI techniques could boost learner development, as suggested by Kiss and Nikolov (2005).

Another practical implication would be the use of L1 metalinguistic awareness in support of L2 instruction. The findings showed that L1 metalinguistic awareness can operate as an effective cognitive tool in L2 learning especially with L2 structures that are similar in the L1. Using first language knowledge as a resource is not favoured in language teaching due to old teaching methods that overused L1 translation. However, the current study indicates that L1 metalinguistic knowledge is already at play during L2 learning. Therefore, language teachers could collaborate with native language teachers to gain some insights to support learners when there is a difficulty encountered due to L1-L2 difference or similarity. It is important that in EFL contexts such as in this study, learners share the same L1, and this could be used as a valuable tool for supporting the L2 learning of certain structures as suggested by several researchers in SLA (Ahmar, Lightbown, & Spada, 2010; Cook, 2001; Cummins, 2000).

**Limitations of the Study**

The findings of this study must be interpreted in relation to limitations in the research design and research instruments used. The study was conducted in an EFL context with participants who were grade eight students (i.e., around fourteen years old) at a private school. Therefore, the findings of the study may not be applicable to other L2 learner populations. The lack of a control group also prevents generalizibility. Second, the current study targeted two L2 structures and focused the instruction and measurements only on those two structures. The findings cannot be generalized to other L2 grammatical structures.

A further limitation is related to the oral production tasks used in the current study. Oral production is difficult to elicit from L2 learners. The learners were not very
familiar with oral production tasks delivered one-on-one with an English speaking adult. Although efforts were made to make them feel comfortable, learners’ performance on the two oral production tasks seemed to be greatly affected by the unfamiliarity of the task compared to the GJT which is a more familiar task/test for this group of learners.

The oral production task for the passive voice did not elicit the passive from most of the participants. It might be possible that the learners did not get the idea of ‘the passive’ and they did not understand what was expected of them. The passive OPT could be revised by (1) modeling some passive sentences at the introductory script or (3) using by phrases on the slides. Another possible explanation for the limited production with regard to the passive OPT could be the nature of the target structure. As indicated earlier, the passive voice is used less frequently in oral production (Hinkel, 2004). Therefore, a different target structure which is also considered to be difficult but used more frequently might elicit more production. Another option would be using different elicitation tasks. For example, an elicited imitation task might be an alternative test to measure the spontaneous use of the target structures. Lastly, another aptitude test that enables the researcher to measure reliability would help in future studies.

The last limitation is related to the length of instruction. The participating school and the EFL teacher were quite generous in terms of time and resources. They allowed me to enter the classrooms and teach the target structures for eight class hours in total. In addition, they let me administer whole group and individual tests. However, one of the disadvantages of classroom research is, limited time for instruction and this is still a concern in the present study. It is likely that four hours of instruction was not enough to observe the effects on L2 learning in this study particularly with regard to learners’ oral production. Thus, a study which included more time to study and practice the structures would have been better to observe the long-term effects of L2 instruction.

**Directions for Future Research**

To my knowledge, the current study is the first empirical attempt to investigate individual learner differences with respect to grammatical difficulty in a classroom context. As generalizability of the findings is not possible, there is a need for replication
studies in a variety of learning contexts such as EFL, ESL and bilingual settings with a variety of L1 backgrounds targeting different L2 structures. In order to apply the findings to a larger population, replication studies should address the methodological weaknesses of the current study. For example, further studies could choose both target structures as novel structures in order to tap the effect of aptitude with regard to grammatical difficulty.

More dynamic conceptualizations of aptitude constructs in SLA (Dörnyei, 2010, Robinson, 2002; Skehan, 2002) have prompted the development of new aptitude tests (Grigorenko, Sternberg, & Ehrman, 2000; Kiss & Nikolov, 2005). Therefore, replication studies could be conducted with alternative aptitude tests that are validated and checked for reliability.

In addition to replication studies, twin studies, such as Hulstijn (1997), that are conducted in the classroom and laboratory would be interesting to explore the effects of individual learner differences on L2 learning. The laboratory setting offers the control and isolation of variables which is impossible to achieve in classroom research. Thus, the outcomes can be more reliably attributed to the independent variable that is under investigation. On the other hand, while laboratory studies provide such advantages, classroom-based research provides greater ecological validity (Spada, 2005). Learners’ previous knowledge of the past progressive prevented me from drawing clear conclusions about the L2 structure difficulty. Another valuable investigation would be to explore similar research questions using alternative methods.

In the currently study qualitative methods were not used except for the learner retrospection questionnaire. More detailed information about the learners’ L2 development and their perceptions about their progress could shed light on many aspects of the current study. For example, learners’ perception about the oral production tasks could be gathered through individual debriefs and interviews. Stimulated recalls could be administered to explore learners’ thinking during the aptitude test. Given current discussions about whether aptitude test are measures of explicit or implicit learning, learners’ input about what they were thinking about and aware of while completing an aptitude test could provide useful information about the cognitive skills involved in
aptitude testing. Stimulated recalls used to explore learners’ thinking during written and oral tasks could also provide valuable information about how learners experience those tasks in terms of the type of L2 knowledge they access in order to complete them (Shui, 2011).

In conclusion, this study of the interaction between language aptitude, L2 grammatical difficulty, L2 instruction and learning has demonstrated that aptitude is not a prerequisite for language learning. Nonetheless, it is an important cognitive ability that affects the rate of L2 learning particularly when learners are dealing with difficult structures. The study acknowledges that there is a complex interplay of individual learner differences in the case of L2 learning. One or two individual learner factors are not sufficient to explain the variance in L2 success, which is why a number of individual learner variables were investigated in this study. The findings have revealed that individual ID variables contributed differently to L2 learning. Moreover, the nature of the L2 structures played an important role in determining the influence of ID variables. Difficulty of the target structure has been shown to be a factor that affects the role of aptitude on L2 learning in addition to other variables. More empirical work on the construct of aptitude and improvements in aptitude testing should provide useful information for L2 researchers, L2 teachers, and L2 learners across a variety of contexts and will also allow for more detailed studies of the interaction between aptitude and L2 structure difficulty.
References


instructed second language acquisition (pp. 235-269). Brusells: Mouton De Gruyter.


Ranta, L. (2002). The role of learners’ language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159-179). Amsterdam: John Benjamins


Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp.139-155). New York; Cambridge University Press.


Appendix A
Instructional Materials

Past Progressive Teaching Materials

<table>
<thead>
<tr>
<th>Lesson 1</th>
<th>45 MIN</th>
<th>1. HANDOUT 1: T/F and don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Albert</td>
<td></td>
<td>2. HANDOUT 2: Reading Albert Einstein</td>
</tr>
<tr>
<td>Einstein)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tell the students that you are going to be discussing important events and people in history. Ask the learners to identify 5 people who they consider to be among the most influential in world history and to ask their partner’s ideas as well. Introduce Albert Einstein and indicate that the class will be learning interesting facts about Einstein and his life.

While participants talk, the teacher distributes **Handout 1 (T, F & Don’t know)**. Ask learners to work with their partner to try to answer “True or False” and explain why they think so.

When the class has finished the T/F activity, write the following sentences on the board.

_Einstein immigrated to the U.S. because he was not happy with his life in Germany._

_He was living in Germany in 1900._

Ask learners:

_What is the difference between these two sentences?_

Elicit responses and then provide the following explanation, being sure to underline the parts of the sentence that you mention, e.g. past form of the verb, auxiliary be.

_Tell the learners that the first sentence is in the simple past tense and it indicates that an activity or situation began and ended in the past. The actions can be single actions, or they can be repeated actions. They can take place over a period of time._

_The second sentence is in the past progressive tense. The past progressive shows what was in progress at a particular moment or over a period of time in the past. The past progressive emphasized that the action was in progress. In contrast to the simple past, the past progressive does not specify whether the action was completed. The simple past emphasizes completion of the action. In order to form past progressive we use (simple past of to be verb) WAS and WERE and add –ing to the verb stem._
Distribute **Handout 2 (Reading passage)**. Ask learners to read the passage about Einstein and find out whether they were right or wrong in their guesses. Help learners with unknown words. Ask learners about what they learned from the text. Ask some comprehension questions and check their understanding.

- e.g: *Was Einstein successful at school? Where was he born? When did he graduate? What was he doing between 1902-1909? What did he achieve in 1905?*

**(Form, 5 minutes)** Take up whether the 13 statements are true or false as a class. As learners to practice the past progressive by reading out the information in the text that provides the answer. As you write the past progressive sentences on the board guide their attention to the form of the sentence. For example, for the first sentence “*In 1889, Einstein was studying in the Swiss Polytechnic Institute.*” explain that

>This sentence emphasizes that the situation was in progress during the year 1889.

<table>
<thead>
<tr>
<th>Lesson: 2</th>
<th>45 min.</th>
<th>Handout 3:Past Simple vs Progressive Handout 4: Paragraph completion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Question formation and Negative Sentences</td>
<td></td>
</tr>
</tbody>
</table>

Ask students to try a few more sentences in **Handout 3** and check their answers when they are finished.

Tell the students that

We use past progressive for actions in progress (not necessarily completed) in the past.

*E.g:* *I was explaining it to my friend last night. (Perhaps I didn’t finish)*

The past progressive is often describes a scene. It gives background information for stories. The simple past gives the action of the story.

*E.g.: The sun was shining. We were swimming. (background description)*  
*Suddenly, a strange ship appeared. We swam back to the sea. (action)*

Distribute **Handout 4** and ask them to work in pairs. When they are finished, ask learners to share their stories.

Provide corrective feedback on errors.

**Question formation in Past progressive**
Explain the students that in order to form questions with past progressive tense we follow the same rules as in other rules. For Yes/No questions we use Was/were at the beginning of the sentence and subject is in between the auxiliary was/were and main verb.

Was/were + Subject + Verb –ing?

E.g: A: Were you sleeping at 11.00 last night? I called you but you didn’t answer.
   B: Yes, I was.

In order to form WH- questions, we start with the WH- word and the auxiliary (was/were), subject and the main verb-ing follow them.

WH-

What
Where + Was/Were + Subject+ Main Verb-ing?
Who

E.g: A: What were you doing last night?
   B: I was helping helping my little sister with her math homework.

Tell the learners to

A. Work in pairs. Take turns asking and answering questions about times in your lives. Use ideas from Columns A and B or use your own ideas.

Example: A: what were you thinking about five minutes ago?
   B: My lunch.

A __________________ B

Where/live? five minutes ago
What/think about? last night at midnight
What/do? Last year?
Who/teach you? Three years ago
Who/live with? At 3 o’clock this morning.

B. Write the most interesting answer and share is with the class.

ASK THE STUDENTS TO BRING THEIR FAVOURITE PHOTO to share with the class.
(5 min.) Start by asking about several individual students’ health. Introduce some vocabulary related to health and ask students to contribute with the items related to health. This can be a spider diagram that you use for reviewing vocabulary related to health.

Surgery, doctor, surgeon, cancer, flu/cold, kidney etc…

(10 min.) While participants talk, distribute Handout 1. Ask learners to work with their partner to try to answer “True or False” and give an explanation for why they think so for as many of the 13 statements as they can.

1. Surgery was invented in the 1800’s.
2. The first X-rays were taken using radium.
3. Florence Nightingale introduced modern nursing methods
4. Vitamins were discovered in the early 1900’s.
...

(10 min.) When the class has finished the T/F activity, write the following active sentence on the board.

“Researchers discovered vitamins 100 years ago.”

Ask learners:

What is the verb? What is the subject of the sentence? What is the object? Who/What is the receiver of the action?

Underline the subject, verb and the object on the board. Tell them that this sentence is known as active and explain:

The subject is responsible for the action of the verb.

Now write the following sentence on the board below the first one:

“Vitamins were discovered by researchers 100 years ago.”

Ask learners:

What is the difference? What is each sentence about? Who does the action? Who/What receives the action?
Elicit responses and then provide the following explanation, being sure to underline the parts of the sentence that you mention, e.g. the past participle, as you describe the verb construction of the passive sentence:

*The first sentence is about what researchers do. This sentence is in the active form. The second sentence is about vitamins. The second sentence is in the passive form. In the passive, the verb is made of two parts: the verb “to be”, in this case in the past tense, and the past participle of the verb, which never changes. In the passive sentence, the subject is the person or thing that received the action. It comes before the verb. The person or thing that did the action is called the “agent”.*

(10 min.) Distribute Handout 2. Ask learners to work with a partner to find the answers to the T/F questions by looking at the information on the timeline in the handout. Do #4 “Vitamins were discovered in the early 1900’s.” as an example by locating the response on the timeline under “Vitamins...” and noting that it’s true. Vitamins were discovered in the early 1900s.

(10 min.) Take up whether the 13 statements are true or false as a class. Then, for each passive statement (i.e., 1, 2, 4, 8, 12), ask learners to practice the passive by reading out the information on the timeline that provides the answer along with the relevant statement. As you write all the passive statements on the board, guide their attention to the form of the sentence. For example, for the first sentence “Surgery was invented in the 1800s” explain that

Sometimes the agent is identified and written after the verb, using the word “by”, for example, “by researchers”. At other times, it is not mentioned; for instance, when it is unimportant.

Explain to the students:
*In this sentence, the agent is not mentioned because it is not important. In this general statement, we assume a doctor invented it, but who did it is not important. Ask students for examples.*

*This explanation could include L1 explanation*
Lesson 2
Activity 2 | 45 min. | Handout 3: Med: It’s not same (article)
Handout 4: Med: It’s not same (comp. Qs)

(10 min.) Tell learners they will be learning more about how health care differs around the world. Distribute Handout 3 and ask learners to read the article. When they are done, check briefly for vocabulary questions.

(10 min.) Then ask them to read the article again and identify the number of passive sentences. Provide the answer (i.e., there are 16 uses of the passive). Break the classes into groups of three. Assign each group two examples of the passive that they must prepare to explain. After five minutes, choose two groups to provide one explanation each. Ask them to explain how they can identify the sentences as passive, for example what the components of the verb are.

*This part could include L1 explanation by the students.

(10 min.) Distribute Handout 4. Ask learners to answer the questions on the handout on their own and then check their answers with a partner. Correct the answers as a class. Provide corrective feedback and encourage learners to answer in full sentences.

(10 min.) Explain that the passive can be formed with many tenses. Display Poster 1, and distribute Poster 1 handouts. Show the class that the past participles of many verbs are listed on the back of the poster. With the class, go through the examples of how different tenses are used with passives.

(10 min.) Distribute Handout 6 and ask the learners to complete it. Remind them that more than one tense may be possible. Take up as a class, eliciting answers and providing corrective feedback.
Medical Practices - throughout the ages (Handout 1)

1. Surgery was invented in the 1800s.
2. The first X-rays were taken using radium.
3. Florence Nightingale introduced modern nursing methods.
4. Vitamins were discovered in the early 1900’s.
5. The first transplant of an artificial heart took place in the 1950’s.
6. The first vaccination was for polio.
7. Freud developed the psychoanalytic method of treating mental illness.
8. Radium was first used to diagnose and treat cancer in the 1890’s.
9. Penicillin was the first antibiotic drug.
10. The first successful organ transplant was a liver.
11. Doctors first studied human anatomy in the 1800’s.
12. Anaesthesia was first used in the 1900’s.
13. Cryosurgery uses extreme heat during surgery.

Appendix B
Past Progressive Grammaticality Judgement Task

NAME & SURNAME______________________ Date:

Decide whether the following sentences are CORRECT or INCORRECT. Circle only ONE answer for each answer. If you circled INCORRECT, correct the mistake by crossing out and writing the correct form. You can choose not sure option if you have no idea.

1. The class was having a break when I got there.
   CORRECT/INCORRECT/NOT SURE

2. She is reading a book at 4 yesterday afternoon.
   CORRECT/INCORRECT/NOT SURE

3. She was baking a cake at 7 yesterday morning.
   CORRECT/INCORRECT/NOT SURE

4. He is playing chess at 10 p.m. last night.
   CORRECT/INCORRECT/NOT SURE

5. I lost some money while I was buying a cup of coffee yesterday.
   CORRECT/INCORRECT/NOT SURE

6. While we were riding in the village, we discover many beautiful places.
   CORRECT/INCORRECT/NOT SURE

7. Look! It snowing.
   CORRECT/INCORRECT/NOT SURE

8. Kaan was living abroad when he met his wife.
   CORRECT/INCORRECT/NOT SURE
9. While I studying in the library I met an old friend.

CORRECT/INCORRECT/NOT SURE

10. I lost many photos while I was changing apartments last winter.

CORRECT/INCORRECT/NOT SURE

11. While I was swimming yesterday morning, I saw an octopus.

CORRECT/INCORRECT/NOT SURE

12. They were studying late last night.

CORRECT/INCORRECT/NOT SURE

13. My father was still sleeping when I left for school this morning.

CORRECT/INCORRECT/NOT SURE

14. They are writing a letter at 6 p.m. yesterday evening.

CORRECT/INCORRECT/NOT SURE

15. They working at 7 o’clock last night.

CORRECT/INCORRECT/NOT SURE

16. While I am walking in the street, I saw an accident.

CORRECT/INCORRECT/NOT SURE

17. While I watching television last night, I heard about the plane crash.

CORRECT/INCORRECT/NOT SURE

18. My friend was having dinner when I phone her last night.

CORRECT/INCORRECT/NOT SURE

19. He driving to work at 7 a.m. yesterday.

CORRECT/INCORRECT/NOT SURE
20. At 5 p.m. yesterday I jogging on the beach.  
CORRECT/INCORRECT/NOT SURE

21. They were flying to London at 7.30 yesterday morning.  
CORRECT/INCORRECT/NOT SURE

22. Selim was drawing a picture when I entered the room.  
CORRECT/INCORRECT/NOT SURE

23. When I met Tarkan last summer, I was going to Paris.  
CORRECT/INCORRECT/NOT SURE

24. When we arrive at the restaurant, the waitress was setting the table.  
CORRECT/INCORRECT/NOT SURE

25. She was painting the house at 11 last night.  
CORRECT/INCORRECT/NOT SURE

26. I drove 10 km while it was snowing heavily.  
CORRECT/INCORRECT/NOT SURE

27. I hurt my knee while I was dancing on Saturday.  
CORRECT/INCORRECT/NOT SURE

28. At 7 o’clock she was singing.  
CORRECT/INCORRECT/NOT SURE

29. While my mum is telling a story, my little brother fell asleep.  
CORRECT/INCORRECT/NOT SURE

30. I am running the 3rd km when I crashed into a stranger.  
CORRECT/INCORRECT/NOT SURE
31. At 9 a.m. last Tuesday they are sitting in a restaurant.
   CORRECT/INCORRECT/NOT SURE

32. She eating a sandwich at 9 o’clock last night.
   CORRECT/INCORRECT/NOT SURE

33. I haven’t seen her for many years.
   CORRECT/INCORRECT/NOT SURE

34. She was talking on the phone at 4 a.m. in the morning.
   CORRECT/INCORRECT/NOT SURE

35. While I was waiting at the bus stop it started raining.
   CORRECT/INCORRECT/NOT SURE

36. He is riding a horse at the moment.
   CORRECT/INCORRECT/NOT SURE

37. He was lying in bed at 8 o’clock last night.
   CORRECT/INCORRECT/NOT SURE

38. At 4 p.m. yesterday I was walking to school.
   CORRECT/INCORRECT/NOT SURE

39. Mozart die while he was composing the Requim.
   CORRECT/INCORRECT/NOT SURE

40. She make lunch for me yesterday.
   CORRECT/INCORRECT/NOT SURE

41. While they are eating in a café, the waiter dropped the dishes.
   CORRECT/INCORRECT/NOT SURE
42. While I was going to the yoga class I stopped to buy a magazine.

CORRECT/INCORRECT/NOT SURE

43. My grandfather died last summer.

CORRECT/INCORRECT/NOT SURE

44. I making a pasta when I burnt my hand.

CORRECT/INCORRECT/NOT SURE

45. I learnt different types of coffee while I was working in a coffee shop.

CORRECT/INCORRECT/NOT SURE
Appendix C
Passive Grammaticality Judgement Tak

Name & Surname______________________ Date:

Decide whether the following sentences are CORRECT or INCORRECT. Circle only ONE answer for each answer. If you circled INCORRECT, correct the mistake by crossing out and writing the correct form. You can choose not sure option if you have no idea.

1. Some of teacher’s words misunderstood.
   CORRECT/INCORRECT/NOT SURE

2. Milk is sold in bags in Canada supermarkets.
   CORRECT/INCORRECT/NOT SURE

3. Smoking has prohibited recently.
   CORRECT/INCORRECT/NOT SURE

4. The water in the swimming pool is changed every week.
   CORRECT/INCORRECT/NOT SURE

5. Tea is growing in the Black sea region.
   CORRECT/INCORRECT/NOT SURE

6. She travelled to London last spring.
   CORRECT/INCORRECT/NOT SURE

7. Facebook is using by many young people.
   CORRECT/INCORRECT/NOT SURE

8. Many houses destroyed by the earthquake.
   CORRECT/INCORRECT/NOT SURE

9. Stamps are purchased at the post office.
   CORRECT/INCORRECT/NOT SURE

10. The road has been fixing twice this year.
    CORRECT/INCORRECT/NOT SURE
11. The pyramids were built in 2500 B.C.
   CORRECT/INCORRECT/NOT SURE

12. That window has been broken three times this month.
   CORRECT/INCORRECT/NOT SURE

13. Othello was written by Shakespeare.
   CORRECT/INCORRECT/NOT SURE

14. Special food is preparing for the holidays.
   CORRECT/INCORRECT/NOT SURE

15. Experiments done on rats.
   CORRECT/INCORRECT/NOT SURE

16. Fast food has forbidden in school canteens.
   CORRECT/INCORRECT/NOT SURE

17. Cars are repairing by mechanics.
   CORRECT/INCORRECT/NOT SURE

18. My room painted blue last weekend.
   CORRECT/INCORRECT/NOT SURE

19. Several trees were planted last summer.
   CORRECT/INCORRECT/NOT SURE

20. The yoga classes have been taught by Susan since 2009.
   CORRECT/INCORRECT/NOT SURE

21. The winning goal was scoring by Arda.
   CORRECT/INCORRECT/NOT SURE

22. She lose her credit card last winter.
   CORRECT/INCORRECT/NOT SURE

23. The dishes were washed by my little brother.
   CORRECT/INCORRECT/NOT SURE

24. The homework is collect every day.
   CORRECT/INCORRECT/NOT SURE
25. Our office was moving to Montreal last winter.
   CORRECT/INCORRECT/NOT SURE

26. Millions of dollars have been borrowed by big companies.
   CORRECT/INCORRECT/NOT SURE

27. A bracelet is wear around the wrist.
   CORRECT/INCORRECT/NOT SURE

28. The X-ray taken in the emergency room.
   CORRECT/INCORRECT/NOT SURE

29. Apples are picked in the fall.
   CORRECT/INCORRECT/NOT SURE

30. Aspirin taken for a headache.
   CORRECT/INCORRECT/NOT SURE

31. The decision is made by the government.
   CORRECT/INCORRECT/NOT SURE

32. Tea is prefer over coffee in Turkey.
   CORRECT/INCORRECT/NOT SURE

33. Listen! He singing a song now.
   CORRECT/INCORRECT/NOT SURE

34. The report wasn’t sign by the teacher.
   CORRECT/INCORRECT/NOT SURE

35. Some files were deleted from my computer.
   CORRECT/INCORRECT/NOT SURE

36. The tires on the car were changed yesterday.
   CORRECT/INCORRECT/NOT SURE

37. You doesn’t love me anymore.
   CORRECT/INCORRECT/NOT SURE

38. The letter sent to London in the morning.
   CORRECT/INCORRECT/NOT SURE
39. Chess has been played since the middle ages.
   CORRECT/INCORRECT/NOT SURE

40. Drugs tested in a laboratory.
   CORRECT/INCORRECT/NOT SURE

41. The thief was catch by the police.
   CORRECT/INCORRECT/NOT SURE

42. The house is cleaned twice a week.
   CORRECT/INCORRECT/NOT SURE

43. Eurovision song contest have been broadcasting twice this week.
   CORRECT/INCORRECT/NOT SURE

44. Several applications received for the job yesterday.
   CORRECT/INCORRECT/NOT SURE

45. She is doing the dishes right now.
   CORRECT/INCORRECT/NOT SURE
Appendix D
Past Progressive OPT

Slide 1

The famous pop singer Christina Grey was killed last Saturday between 11 a.m. and 3 p.m.

Slide 2

Her rival, Canadian pop star, Greta Hilton has always hated Christina. Witnesses say that they saw a woman who looked like Greta Hilton near Christina’s house last Saturday. So the police have charged Greta with murder, and now she is in prison.
You are a “photographer” who always follows Greta and takes photos wherever she goes. You took several pictures of Greta and her boyfriend last Saturday, so you are a witness and can help prove her innocence.

These are some of the photos that show what Greta was doing last Saturday.
Now imagine that I am a police officer and tell me about your pictures and use the information at the top of the slides.

At 11:26 she was shopping.
At noon she arrived at Moonbean cafe.

While she was entering she picked up a newspaper.
She was reading the newspaper.

While she was working she was listening.
While she was drinking her friend came.

While she was talking on the phone he got bored.
Slide 13

When did they leave?

Slide 14

At 1:07 they were standing.
At 1.25 they bought ice-cream.

At 1.37 they were sitting.
While she was swinging they were eating ice-cream.
At 3:04 they left.

While she was walking she dropped her ice-cream.
They were driving.

Thank you!
Her mother mailed the packaged to Anna in Canada.

Anna moved to Canada from Mexico. The rest of her family still lives in Mexico. Her mother is sending Anna a package of her favourite things. This story is about what happened to this package.
Slide 3

The package was taken to post office.

Slide 4

The package was weighed.

Slide 5

The package was stamped.
The package was put on the truck.

The package was sent to Toronto, USA/the woman.

The package was returned to the post office.
The package was sent to Canada.

The package was delivered to Anna’s place in Toronto.
She opened the package.

Anna thanked her mother.

The package was recycled.
Appendix F
Language aptitude test
Appendix Appendix G
Digit Span Test

<table>
<thead>
<tr>
<th>5</th>
<th>8</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>4</th>
<th>3</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>2</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4</th>
<th>2</th>
<th>7</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

From the computer

<table>
<thead>
<tr>
<th>6</th>
<th>1</th>
<th>9</th>
<th>4</th>
<th>7</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>1</td>
<td>8</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

---

184
Appendix Appendix H
Motivation questionnaire

Following are a number of statements with which some people agree and others disagree. We would like you to indicate your opinion after each statement by putting an “X” in the box that describes the extent to which you agree or disagree with the statement. Thank you very much for your help.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Not at all true</th>
<th>Not really true</th>
<th>Partly true</th>
<th>Partly untrue</th>
<th>Mostly true</th>
<th>Absolutely true</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I would like to learn as many languages as possible.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>After finishing English I'd like to start learning another language.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>For me language learning is a hobby.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sometimes language learning is a burden for me.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Language learning is an exciting activity.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I don't particularly like the process of language learning and I do it only because I need the language.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Language learning often gives me a feeling of success.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Language learning often makes me happy.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Studying English is important to me because it provides an interesting intellectual activity.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>English proficiency is a part of the general culture.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I am learning English to become more educated.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>English proficiency is important to me because it allows me to learn about the current intellectual trends of the world, thus to broaden my view.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Everybody in Turkey should learn English to at least an intermediate level.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>English proficiency is indispensable for a Turkish person to be able to live a valuable and colorful life.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>The more I learn about the British/Americans, the more I like them.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Most of my favorite artists (e.g. actors, musicians) are either British or American.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Britain and America are among the most exciting countries of the world.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>British/American culture is of vital importance in the world nowadays.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>If I spoke English I could do a more interesting job.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>If I spoke English I could travel more for official purposes.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>English proficiency would have financial benefits for me.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>I don't think it is very important to speak elaborately in a foreign language; the point is only to be able to express my thoughts.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>Pronunciation in a foreign language is important only insofar as one can make himself/herself understood.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>It doesn't matter if I make mistakes in a foreign language; the point is only to be fluent.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>I think I have a good sense for languages.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>I think language learning is more difficult for me than for the average learner.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>I had some bad experiences with learning languages.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td>I think I belong to the class of learners who can completely lose their interest in learning if they have had a bad teacher.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>My colleagues usually speak a foreign language on at least intermediate level.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>30.</td>
<td>English proficiency is important to me because it is indispensable for establishing an international reputation.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>31.</td>
<td>There would be a serious gap in my life if I couldn't learn English.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td>I believe I'll be able to learn English to an extent that satisfies me.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>33.</td>
<td>At present learning English is one of the most important things to me.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>34.</td>
<td>English proficiency is important to me because it will allow me to get to know various cultures and people.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>Studying English is one of the most important to me because it offers a new challenge in my life, which has otherwise become a bit monotonous.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>I am learning English because I would like to spend a longer period abroad.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Appendix I
Learner Perception Questionnaire

Please fill in the information requested or check the appropriate option(s).

Gender: M □ F □
Age: ...........
Level: ........
I have been studying English for ........... years.
Native Language: Turkish .......... other ..........
Have you ever lived in an English-speaking country? Yes □ No □
If so, for how long? ...........................................

Section 2: Perceptions about English Language Structure
The following is a list of English language structures that you have learnt in this course. Please indicate whether you think each structure is more or less difficult to learn. For example, if you think a structure is more difficult to learn, circle 3 or 4 (3=difficult; 4=very difficult). If you think a structure is less difficult to learn circle number 1 or 2 (1=simple; 2=very simple). If you don’t recognize the structure or have never learned it, check “Not Sure”.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Easy</th>
<th>Difficult</th>
<th>Not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present simple (I do)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Past simple (I did)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Past progressive (I was doing)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Present perfect (I have done)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Passive voice (it was borrowed)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Future tenses (I will call)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Section 3:

1. How do you learn difficult structures? For example, do you use memorization techniques or other strategies?

2. Do you use the same techniques for easy structures? If not, how do you learn easy structures?

3. Did you notice any breakthroughs in your learning of that difficult structure during this course?

Two structures you learned in this course are:

a) Passive voice

b) Past Progressive Tense

3. Did you get any help from Turkish when learning these structures? Did you look for corresponding structures in Turkish? Did it help at all?

4. What do you think is the easiest language structure you learned in this course?
b) Why do you think that structure is easy?

c) Can you verbalize the rule for that easy structure?

5. a) What do you think is the most difficult language structure you learned in this course?

b) Why do you think that structure is difficult?

c) Can you verbalize the rule for that difficult structure?
Aşağıdaki her cümleyi dikkatle okuyunuz. Cümlelerden bazıları dilbilgisi hataları içermektedir. Eğer cümlelerde bir yanlış olduğunu düşünüyorsanız, sağ taraftaki yanlış şıklını yuvarlak içine alıp, yanlışı düzeltiniz. Eğer cümlede yanlış olmadığını düşünüyorsanız, doğru seçeneğini yuvarlak içine alınız. Bazı cümleler hiç yanlış içermez; yanlış içeren cümlelerde sadece bir tane yanlış vardır. Yanlış içeren cümlelerde hatayı düzelttikten sonra, yanlışa neden olan bozulduğu aşağıda verilen boşluğa açıklamanız gerekmemektedir. **Örnek:** Hatalarımı her zaman kabul eder, özür dilemesini bilirim.

**DOĞRU/YANLIŞ**

-Hatalarımı olması gerekir; çünkü iki sessiz arasındaki sahiplik eki -im'dir ve -i düşmez.

1. Bu kadar üşiyecek ne var anlamadım.  **DOĞRU/YANLIŞ**

2. Ankaralılar ulaşım zamlarına isyan ettiler.  **DOĞRU/YANLIŞ**

3. Ali tüm gün sokakda oynadığı için eve yorgun geldi.  **DOĞRU/YANLIŞ**
4. Kapının kiliti bozulduğu için çilingir çağırdık. DOĞRU/ YANLIŞ

5. Anne ve babasını özleyince tatilden erken dönmüş. DOĞRU/ YANLIŞ

6. Bursa’lılar şampiyonluğa çok sevindiler. DOĞRU/ YANLIŞ

7. Burcu’lar bu gece bize katılamayacakmış. DOĞRU/ YANLIŞ

8. Ahmetler tekne turu ile Marmaris’e gitmişler. DOĞRU/ YANLIŞ

9. Başarı için çok çalışman gerektiğini söylememiştim? DOĞRU/ YANLIŞ
10. Bu gece evdeysen seni arayabilir miyim? DOĞRU/ YANLIŞ

11. Her gün spor yap ki sağlıklı kalasın. DOĞRU/ YANLIŞ

12. Eşyalarını hep yer de bırakıyorsun. DOĞRU/ YANLIŞ

13. Senem’in Londra’da evine hırsız girmiș. DOĞRU/ YANLIŞ

14. Sende gelmek istersen çabuk olmalısın. DOĞRU/ YANLIŞ

Teşekkürler!

192
## Appendix K
### Scoring Scheme for Passive GJT

<table>
<thead>
<tr>
<th>Grammaticality</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Score</strong></td>
<td><strong>Correct</strong></td>
</tr>
<tr>
<td>Grammatical</td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., The tires on the car were changed yesterday.</td>
</tr>
<tr>
<td><strong>Incorrect</strong></td>
<td>change to a wrong agreement</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., The tires on the car <em>was</em> changed yesterday</td>
</tr>
<tr>
<td></td>
<td>change to passive with different tense</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., The tires on the car <em>are</em> changed yesterday</td>
</tr>
<tr>
<td></td>
<td>change to passive with a wrong PP (minor spelling)</td>
</tr>
</tbody>
</table>

|                | 2 points                                                               |
|                | E.g., The tires on the car were *changed*                              |
|                | No attempt to provide past participle                                  |
|                | 1 point                                                                |
|                | E.g., the tires on the car were *change*                               |

<table>
<thead>
<tr>
<th><strong>Ungrammatical</strong></th>
<th><strong>Correct</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Change to form other than passive. item discarded</td>
</tr>
<tr>
<td></td>
<td>0 points</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Missing AUX</strong></th>
<th><strong>Correct</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g., Aspirin taken for a headache</td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., Aspirin is taken for a headache.</td>
</tr>
<tr>
<td></td>
<td>change to target passive</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., Aspirin are taken for a headache.</td>
</tr>
<tr>
<td></td>
<td>change to passive with wrong agreement</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., Aspirin were taken for a headache.</td>
</tr>
<tr>
<td></td>
<td>change to passive with wrong tense</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., Aspirin were taken for a headache.</td>
</tr>
<tr>
<td></td>
<td>change with a wrong past participle</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., Aspirin <em>was</em> taken for a headache.</td>
</tr>
<tr>
<td></td>
<td>change with PP and agreement error</td>
</tr>
<tr>
<td></td>
<td>1 point</td>
</tr>
<tr>
<td></td>
<td>E.g., Aspirin were taken for a headache.</td>
</tr>
<tr>
<td></td>
<td>change to form other than passive.</td>
</tr>
<tr>
<td></td>
<td>ID</td>
</tr>
</tbody>
</table>

|                   | 1 point                                                                  |

<table>
<thead>
<tr>
<th><strong>Past Participle</strong></th>
<th><strong>Correct</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g., The report wasn’t sign by the teacher</td>
<td>1 point</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., The report wasn’t signed by the teacher.</td>
</tr>
<tr>
<td></td>
<td>change to passive with wrong agreement</td>
</tr>
<tr>
<td></td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., The report weren’t signed by the teacher.</td>
</tr>
</tbody>
</table>

193
change to passive with wrong tense 2 points
E.g., The report isn’t signed by the teacher.
change with a wrong past participle 2 points
E.g., The report was signet by the teacher.
Change to form other than passive. ID

Ungrammatical → Correct 0 point
Missing PP
E.g., Facebook is using by many young people.

Incorrect → change to target passive 2 points
E.g., Facebook is used by many young people.
change to passive with wrong agreement 2 points
E.g., Facebook are used by many young people.
change to passive with wrong tense 2 points
E.g., Facebook has been used by many young people.
change with a wrong past participle 2 points
E.g., Facebook is uset by many young people.

Change to form other than passive. ID
## Appendix L
### Scoring Scheme for Past Progressive GJT (Type A Items)

<table>
<thead>
<tr>
<th>Grammaticality</th>
<th>Rating</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical</td>
<td>Correct</td>
<td>3 points</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>was singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>Change to bare / present progressive.</td>
<td>1 point / 2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>singing / is singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a non-progressive form</td>
<td>0 points</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>was singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No correction</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>Change to form other than past. prog.</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>was singing on</em> 7 yesterday.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ungrammatical</th>
<th>Correct</th>
<th>2 points</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>is/are-ing</em></td>
<td>E.g., She <em>is singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>Change to past progressive</td>
<td>3 points</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>is singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to bare progressive</td>
<td>1 point</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a non-progressive IL form</td>
<td>0 points</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>is singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to form other than pres. prog.</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>is singing on</em> 7 yesterday.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ungrammatical</th>
<th>Correct</th>
<th>0 point</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>bare -ing</em></td>
<td>E.g., She <em>singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>Change to present / past progressive</td>
<td>2 points /3 points</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>is singing / was singing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a non-progressive form</td>
<td>0 point</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>sing</em> at 7 yesterday.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to form other than bare prog.</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., She <em>singing on</em> 7 yesterday.</td>
<td></td>
</tr>
</tbody>
</table>
Scoring Scheme for Past Progressive Grammaticality (Type B items)

<table>
<thead>
<tr>
<th>Grammaticality</th>
<th>Rating</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical</td>
<td>Correct</td>
<td>3 points</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.g., While I was waiting at the bus stop it started raining.</td>
</tr>
<tr>
<td>Incorrect</td>
<td>Change to bare / present progressive.</td>
<td>1 point /2 points</td>
</tr>
<tr>
<td></td>
<td>E.g., While I waiting…/ While I am waiting…</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a non-progressive form</td>
<td>0 point</td>
</tr>
<tr>
<td></td>
<td>E.g., While I waited at the bus stop it started raining</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to form other than past. prog.</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., While I was waiting in the bus stop it started raining.</td>
<td></td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>Correct</td>
<td>1 point</td>
</tr>
<tr>
<td></td>
<td>E.g., While I studying in the library I met an old friend.</td>
<td></td>
</tr>
<tr>
<td>Bare -ing</td>
<td>Incorrect</td>
<td>2 points</td>
</tr>
<tr>
<td></td>
<td>Change to present progressive.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.g., While I am studying in the library I met an old friend.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a non-progressive IL form</td>
<td>0 point</td>
</tr>
<tr>
<td></td>
<td>E.g., While I studied in the library I met an old friend</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to form other than past. prog.</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., While I studying at the library…/I was meeting.</td>
<td></td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>Correct</td>
<td>2 points</td>
</tr>
<tr>
<td>Is/are- ing</td>
<td>E.g., While they are eating in a café the waiter dropped the dishes.</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>Change to bare progressive / past progressive</td>
<td>1/3 points</td>
</tr>
<tr>
<td></td>
<td>E.g., While they eating in a café the waiter dropped the dishes.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>While they were eating in a café the waiter dropped the dishes.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a non-progressive form</td>
<td>0 points</td>
</tr>
<tr>
<td></td>
<td>E.g., While they eat in a café, the waiter dropped the dishes.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to form other than past. prog.</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., While they are eating in a café waitress dropped the dishes</td>
<td></td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>Correct</td>
<td>2 points</td>
</tr>
<tr>
<td>Missing -ed</td>
<td>E.g. My friend was having dinner when I phone her last night.</td>
<td></td>
</tr>
<tr>
<td>Incorrect</td>
<td>Change to correct past tense</td>
<td>3 points</td>
</tr>
<tr>
<td></td>
<td>E.g., My friend was having dinner when I phoned her.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to a progressive form</td>
<td>0 point</td>
</tr>
<tr>
<td></td>
<td>E.g., My friend was having dinner when I was calling her last night.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change to form other than simple past</td>
<td>Item discarded</td>
</tr>
<tr>
<td></td>
<td>E.g., My friend was having when I called him last night.</td>
<td></td>
</tr>
</tbody>
</table>