



<http://researchspace.auckland.ac.nz>

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.

<http://researchspace.auckland.ac.nz/feedback>

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

Note : Masters Theses

The digital copy of a masters thesis is as submitted for examination and contains no corrections. The print copy, usually available in the University Library, may contain corrections made by hand, which have been requested by the supervisor.

The role of working memory in second language reading comprehension

Mohammadtaghi Shahnazari Dorcheh

A thesis submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Language Teaching and Learning,
Department of Applied Language Studies and Linguistics, The
University of Auckland, 2011

Abstract

This study examined the role of working memory capacity in the development of second language reading at beginning, intermediate and advanced levels. A total of 140 L1 Persian EFL learners at the beginning, intermediate and advanced levels participated in the study. They were all studying English as a foreign language in a private language school. All the participants completed an L1 reading span test based on Daneman & Carpenter's (1980), a math span test based on Robert and Gibson (2002) and Salthouse & Babcock (1991), and an English non-word recognition task adapted from Gathercole, Pickering, Hall, & Peaker (2001). The participants also completed measures of reading proficiency, including a cloze test, a short- answer comprehension test, and a reading recall (the texts were adjusted for different proficiency levels). Multiple regression analysis was applied to determine whether there was any significant relationship between working memory and reading at different levels of proficiency. The results of this study indicated a significant relationship between working memory capacity and L2 reading ability only among learners of lower proficiency, not for the intermediate and advanced levels. This suggests that working memory plays a diminishing role in discriminating performance on second language reading measures at higher levels of proficiency.

Dedication

To Zahra and Reza, my wife and son, who supported me incredibly during my Ph.D
programme

Acknowledgment

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr Rebecca Adams, for her patience, guidance, and expertise on this project. Rebecca has always been supportive and provided me with her insightful comments and constructive suggestions throughout this project. I would like to express my special thanks to my co-supervisor, Dr Jenefer Philp, for her time and insightful suggestions. Jenefer has always been so kind, encouraging and her support throughout this project has been invaluable.

I would like to thank my participants who kindly consented to take part in this study and put a lot of time into completing several reading and memory tasks. I would also like to thank Mrs Karshenas, the head of the Nasle-Farda language school, who kindly consented to this study being conducted in her language school. I wish to thank Joanna Niederer for proofreading this thesis.

Finally, I would like to express my appreciation of my examiners, Professor Michael Harrington and Dr. Paula Winke, for their careful reading of the thesis and insightful comments.

Table of Contents

Abstract	II
Dedication	III
Acknowledgment	IV
Table of Contents	V
List of Figures	IX
List of Tables	X
List of Acronyms	XII
Chapter One	1
Introduction	1
1.1 Overview	1
1.2 Introduction	1
1.3 Working Memory and Language Learning	2
1.4 Working Memory and Reading	3
1.5 Working Memory and L1 Reading Ability	3
1.6 Working Memory and L2 Reading Comprehension	4
1.7 Purpose of the Study	5
1.8 Significance of the Study	6
1.9 Organization of the Thesis	6
Review of Literature	8
2.1 Introduction	8
2.2 Reading Comprehension	8
2.3 Models of Reading Comprehension	10
2.3.1 Introduction	10
2.3.2 Bottom-up Model	12
2.3.3 Top-down Model	15
2.3.4 Rumelhart's (1977) Interactive Model of Reading	15
2.3.5 Stanovich's (1980) Interactive-Compensatory Model	16
2.3.6 Construction-Integration Model of Reading	17
2.3.7 Verbal Efficiency Theory of reading	18
2.3.8 The Compensatory-Encoding Model	20

2.4 Reading in a Second Language.....	22
2.5 Assessment of Reading Comprehension.....	25
2.6 Memory.....	30
2.7 Working Memory and the Assessments of its Components	31
2.8 Alternative Models of Working Memory	34
2.9 Working Memory and Language Processing.....	36
2.10 Working Memory and Second Language Learning.....	37
2.11 Phonological Short-Term Memory and Second Language Learning	38
2.12 The Role of Working Memory in Reading Comprehension.....	42
2.13 Research on Working Memory and L1 Reading	43
2.14 Working Memory in Child and Adult L1 Reading.....	47
2.15 Working Memory and Studies in the L1 and L2: An Overview.....	49
2.16 Working Memory and L2 Reading.....	49
2.17 Rationale of the Study.....	54
2.18 Research Questions.....	55
Chapter Three.....	56
Methodology	56
Section One.....	56
3.1.1 Introduction	56
3.1.2 Subjects	56
Section Two	59
Materials	59
3.2.1 Memory Measures.....	59
3.2.1. 1 Reading Span Test.....	59
3.2.1.2 Math Span Test.....	63
3.2.1.3 Non-Word Recognition Test	67
3.2.2 Reading measures.....	69
3.2.2.1 Introduction	69
3.2.2.2 Cloze Test.....	72
3.2.2.3 Short-Answer Reading Test.....	75
3.2.2.4 L1 Recall Reading Test	79
Section Three	82
3.3 Design of the Study.....	82

3.3.1 Data Analysis	85
3.3.2 Ethics Concern	87
Chapter Four	88
Results.....	88
4.1 Descriptive Statistics for Participants' WM Performance for Each Proficiency Level.....	88
4.2 Descriptive Statistics for Participants' Reading Performance for each Proficiency Level.....	90
4.3 Describing Memory Measures.....	93
4.3.1 Descriptive Statistics for the Combined Groups.....	93
4.3.2 Correlations between the Memory Measures for Combined Groups.....	94
4.3.3 ANOVA Results.....	95
4.3.4 Correlations between the Memory Measures for each Proficiency Level	96
4.4 Describing Reading Measures	100
4.4.1 Correlations for Reading Measures.....	100
4.5 Research Question One: Is there a Relationship between WMC and L2 Reading Ability? Does this Relationship Differ According to Proficiency Level.....	102
4.5.1 Inferential Analysis at the Beginning Level.....	102
4.5.1.1 Correlation Results	103
4.5.1.2 Regression Results.....	104
4.5.2 Inferential Analysis at the Intermediate Level	106
4.5.2.1 Correlation Results	106
4.5.3 Inferential Analysis at the Advanced Level	107
4.5.3.1 Correlation Results	108
4.5.4 Summary of the Inferential Analysis for each Proficiency Level.....	109
4.6 Research Question Two: Is there a Relationship between PSTM and L2 Reading Ability? Does this Relationship Differ According to Proficiency Level?	111
4.6.1 Inferential Analysis for each Proficiency Level.....	111
4.6.1.1 Correlation Results	111
4.7 Summary	112
Chapter Five.....	114
Discussion.....	114
5.1 An Overview	114
5.1.1 Research Questions	114
5.1.2 Participants, Design and Measures.....	114
5.2 Research Question One.....	115

5.2.1 Is there a Relationship between Working Memory and L2 Reading Ability? Does this Relationship Differ according to Proficiency Level?.....	115
5.2.2 Researching WM and Reading.....	120
5.3 Research Question Two	124
5.3.1 Is there a Relationship between Phonological Short-term Memory and L2 Reading Ability? Does this Relationship Differ according to Proficiency Level?.....	124
5.4 Discussion of the Dependent Variables	128
5.5 Discussion of the Independent Variables.....	130
5.6 Summary	132
Chapter Six.....	133
Conclusion	133
6.1 An Overview	133
6.2 Theoretical Implications	133
6.3 Limitations of the Study and Directions for Further Research	136
6.4 Summary	139
Appendix A	140
Participant Consent Form	140
Appendix B	142
Head of Language School Consent Form	142
Appendix C	144
Reading Measures	144
Appendix D.....	183
The List of Reading Span Test Items.....	183
Appendix E	193
Math Span Test Items	193
Practice Session	193
Set Nine.....	194
Appendix F.....	195
English Non-words	195
References.....	198

List of Figures

Figure 1: Baddeley's (2000) revised model of working memory	32
---	----

List of Tables

Table 3.1 Descriptive Statistics for Language Proficiency Scores for Proficiency Levels	58
Table 3.2 Readability and Vocabulary Information of the Reading Passages for each Proficiency Level	71
Table 3.3 Range of Idea Units for Each Passage Type for each Proficiency Level	72
Table 3.4 Distribution of Idea Units per Text and Proficiency Level	80
Table 3.5 Timetable Used for the Reading and Memory Measures in this Study	85
Table 4.1 Descriptive Statistics for the Memory Measures for each Proficiency Level	88
Table 4.2 Descriptive Statistics for the Working Memory Z-scores for each Proficiency Level	89
Table 4.3 Descriptive Statistics for the Composite Memory Z-Scores for each Proficiency Level	90
Table 4.4 Descriptive Statistics for Level Specific Reading Measures	91
Table 4.5 Descriptive Statistics for the Reading Measure Z-Scores	92
Table 4.6 Descriptive Statistics for the Memory Measures.....	93
Table 4.7 Descriptive Statistics for the Composite Memory Measure Z-scores.....	94
Table 4.8 Correlations between the Memory Measures.....	95
Table 4.9 One Way ANOVA between the Proficiency Groups for Memory Measures.....	96
Table 4.10 Correlations between the Memory Measures at the Beginning Level.....	97
Table 4.11 Correlations between the Memory Measures at the Intermediate Level.....	98
Table 4.12 Correlations between the Memory Measures at the Advanced Level.....	99
Table 4.13 Correlation Results for the Reading Measures at the Beginning Level.....	110
Table 4.14 Correlation Results for the Reading Measures at the Intermediate Level.....	101
Table 4.15 Correlation Results for the Reading Measures at the Advanced Level	102

Table 4.16 Correlation Results between the Reading and Memory Measures at the Beginning Level	93
Table 4.17 Regression Results for the WM Measures and Reading Composite at the Beginning.....	94
Table 4.18 Correlation Results between the Reading and Memory Measures at the Intermediate Level	95
Table 4.19 Correlations between the Reading and Memory Measures at the Advanced Level	96
Table 4.20 A Summary of the Correlations at the Beginning Level.....	97
Table 4.21 A Summary of the Correlations at the Intermediate Level	98
Table 4.22 A Summary of the Correlations at the Advanced Level.....	99
Table 4.23 Correlations between PSTM and the Reading Measures for each Proficiency Level	100

List of Acronyms

WM: Working Memory

WMC: Working Memory Capacity

PSTM: Phonological Short-Term Memory

L1: First Language

L2: Second Language

SLA: Second Language Acquisition

EFL: English as a Foreign Language

Chapter One

Introduction

1.1 Overview

This chapter provides an overview of the main focus of the research presented in this thesis. First, a brief theoretical overview is given of the role of working memory as a component of language learning aptitude and a source of individual differences in L2 reading ability. The next section provides a definition of working memory and explores the role it plays in L1 or L2 learning. Then the role of working memory in L1 and L2 reading is outlined, followed by the purpose of the study. The latter includes the rationale behind the study. Finally, the chapter presents the significance of the study and a brief overview of the organization of the thesis respectively.

1.2 Introduction

One primary purpose of reading comprehension is to learn from the text (Grabe & Stoller, 2002, p., 13). However, the extent to which each individual learns from a text differs from one to another (e.g., Cain, 2006; Cain & Oakhill, 2004; Grabe & Stoller, 2002; Koda, 2005; Macaro, 2003; McGuinness, 2005). The reasons underlying individual differences in reading comprehension have been a major research focus in cognitive psychology. The findings to date attribute some of these differences to language learning aptitude (Carroll, 1965; Skehan, 1989). Language learning aptitude is one of the sources of individual differences in second language learning (e.g., Carroll, 1965; Skehan, 1991). A four-factor model of aptitude was proposed by Carroll (1965), in which he attributed individual differences to phonemic coding ability, associative memory, grammatical sensitivity, and inductive language learning ability. Based on this view, individuals do not have an undifferentiated talent for learning languages, but rather a multi-component talent from which each component may vary relatively independently from the others (e.g., Carroll, 1965; Skehan, 1991). Carroll's early research led to the development of a number of tests to measure aptitude for language learning, including the MLAT and DLAB (Modern language Aptitude Test and Defense Language Aptitude Battery). Carroll's view of aptitude was associated with the audio-lingual method and therefore was not adopted so strongly after the 1960s due to the rise of the communicative

approach. As changes in our understanding of language learning have influenced our understanding of language learning aptitude, working memory has been identified as a likely central component of language learning aptitude (Carroll, 1981; Miyake & Friedman, 1998; Skehan, 1989, 1998). Following the introduction of a three-component model of working memory by Baddeley and Hitch (1974), research shifted towards this construct as an important variable in second language learning (Miyake & Friedman, 1998; Sawyer & Ranta, 2001; Skehan, 2002). The purpose of the current study is to enhance our understanding of the role of working memory in reading comprehension.

1.3 Working Memory and Language Learning

Working memory (WM) can be defined as a limited capacity temporary storage system with a limited capacity pool of attentional resources. Baddeley and his colleagues view WM as that which simultaneously maintains and processes the input it receives through different channels of communications (e.g., touch, long-term memory, sight, and hearing) (Baddeley, 1986, 1999, 2003, 2007; Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Gathercole & Baddeley, 1993).

In language processing, WM acts like a mental workspace where form and meaning could be connected (VanPatten, 2004) and as a result, learning takes place (Schmidt, 1990; VanPatten, 2004). Thus, WM plays an important role in L1 and L2 learning. However, this role differs across individuals because they do not possess the same pool of attentional resources which are required for noticing the coming input, which is, in turn, a pre-requisite condition for learning (e.g., Schmidt, 1990 - 2001). Thus, since attention is limited by working memory capacity (WMC), there may be a close relationship between the amount of learning and the size of WM (Sawyer & Ranta, 2001). This means that individuals with higher WMC may be advantaged in L1 or L2 learning.

This idea is supported by a body of research in L1 and L2 learning in general and in L1 and L2 reading in particular. For example, L1 research suggests that there are relationships between individuals' WMC and constructs such as fluency of speech (Daneman, 1991), the ability to learn new words (Daneman & Green, 1986), and L1 reading comprehension (e.g., Daneman & Carpenter, 1980; Waters & Caplan, 1996). There are also some L2 studies arguing that WM has an effect on word naming and vocabulary learning (Atkins & Baddeley, 1998), online parsing performance (Juffs, 2004), use of interactional feedback (e.g., Ando, Fukunaga, Kurahachi, Stuto, Nakano, & Kage, 1992; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Mackey, Adams, Stafford, & Winke, 2010; Shahnazari & Adams, in press), and L2 reading comprehension (Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004).

1.4 Working Memory and Reading

Both L1 and L2 studies suggest that WM plays an important role in reading ability by providing a workspace where the representations of recently read material are maintained for further processing such as integrating these representations with prior background knowledge in order to develop a plausible interpretation of the writer's message (e.g., Cain & Oakhill, 2006; Koda, 2005). These studies suggest that readers with higher WM processing and storage capacities outperform those with lower capacities. One general explanation for these studies may be that individuals with higher WMC possess better storage capacity skills, so they employ less of their limited cognitive resources for maintaining the encoded items, leaving more cognitive resources for processing abilities, which in turn leads to advantages in overall comprehension.

1.5 Working Memory and L1 Reading Ability

A large body of research has indicated that WM plays an important role for children learning to read in their L1 (e.g., Cain, 2006; Cain & Oakhill, 2004; Cain, Oakhill, & Bryant, 2000a). In particular, a component of WM, phonological short term memory (PSTM), seems to have an effect on L1 reading skill. Phonological short term memory (PSTM) is a more specific component of WM for temporary storage and processing of phonological features of language. PSTM plays an important role in language acquisition and early reading skills such as phonemic awareness (e.g., Baddeley & Logie, 1999; Roth, Speece & Cooper, 2002; Spear-

Swerling, 2007). As Baddeley (2003) suggested, PSTM plays this role by maintaining the representation of new phoneme sequences, while at the same time, the articulatory rehearsal system (inner speech) facilitates encoding of these sequences.

Research has also indicated that WM accounts for adults' L1 reading. For example, Daneman and Carpenter (1980) found that the individuals with higher WMC might maintain more items in their WM than those with lower WMC. This might help these individuals to develop more coherent representations of the encoded items in the form of chunks which are less demanding and better recalled.

Similarly, King and Just (1991) suggested that individuals with higher WMC were faster and more accurate in their comprehension of complex sentences such as *The reporter that the senator attacked admitted the error*. Miyake, Just and Carpenter (1994) found that individuals with higher WMC are better and faster at comprehending ambiguous words than those with lower WMC. They argued that those with higher WMC might resolve lexical ambiguity because they were able to retain both interpretations of ambiguous words in their WM for a longer period of time.

1.6 Working Memory and L2 Reading Comprehension

Similar to the findings for L1 research, some L2 studies have linked WMC to variability in L2 reading ability (e.g., Alptekin & Erçetin, 2009; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004). These studies suggested that WM, as a source of individual differences, plays an important role in L2 reading performance (see chapter two for further details). They found that individuals with higher WMC had a higher L2 reading ability than those with lower WMC. However, these studies differed in terms of how WMC was operationalized, in the methodology used in administering and scoring RSTs, in the proficiency level of their participants, and in L2 reading measures. These differences in research design may explain the divergent research findings on the role of WM in L2 reading. These studies measured WMC either through an L1 RST (Chun & Payne, 2004; Lesser, 2007), an L2 RST (Alptekin & Erçetin, 2009), or both L1 and L2 RSTs (Harrington & Sawyer, 1992; Walter, 2004). Two

studies used other memory measures such as a digit span test (Harrington & Sawyer, 1992) and a non-word recognition task (Chun & Payne, 2004). Thus, as these studies used different memory measures with different methodologies in administering and scoring, this might have in turn yielded such divergent findings. Furthermore, some studies operationalized WMC based on both processing and storage capacity scores (e.g., Lesser, 2007; Walter, 2004), whereas other studies established WMC based on the storage scores alone (e.g., Harrington & Sawyer, 1992; Osaka & Osaka, 1992). One other important difference between these studies is the proficiency of the participants, which varies from the advanced level (Alptekin & Erçetin, 2009; Harrington & Sawyer, 1992) to the intermediate (Chun & Payne, 2004), upper and lower intermediate (Walter, 2004) and high beginning levels (Lesser, 2007). Divergent findings on the role of WM in L2 reading may be related to a shifting role of WM as L2 reading ability develops.

Similar to the findings for WM, there are divergent research findings on the role of PSTM in L2 reading. While some research has linked PSTM to L2 reading ability (e.g., Masuora & Gathercole, 2005; Service, 1992), some others found no relationship between PSTM and L2 reading ability (e.g., Hummel, 2009; Harrington & Sawyer, 1992; Kormos & Sáfár, 2008). These studies were also different in terms of the methodology and the proficiency level of the participants used. Differences in the research findings point to the need to examine the relationship between PSTM and L2 reading for each proficiency level.

1.7 Purpose of the Study

Although current research provides evidence for a relationship between WMC and L2 reading ability, the findings on such a relationship are not consistent. In addition, it is as yet uncertain as to whether this relationship is mediated by L2 proficiency. For this reason, this study has been designed to examine the relationship between WMC and L2 reading ability for three proficiency levels. Thus, the main purpose of this study is to examine (1) the relationship between WMC and L2 reading comprehension, and (2) whether this relationship differs according to proficiency level.

Furthermore, while some prior studies suggested a relationship between phonological short-term memory and L2 reading ability at early stages of L2 development (e.g., Service, 1992; Service & Kohonen, 1995), some others did not establish such a relationship (e.g., Harrington & Sawyer, 1992; Hummel, 2009; Kormos & Sáfár, 2008). Thus, a further purpose of this study is to examine (3) the relationship between phonological short-term memory, and L2 reading ability, and (4) whether this relationship differs according to proficiency level.

1.8 Significance of the Study

This study has the potential to make a number of important contributions to the research area of WM and L2 reading ability. It may provide some theoretical implications for those researchers who investigate the role of cognitive variables, like WM, in second language acquisition in general and L2 reading comprehension in particular. It may improve our understanding of the role that WM and phonological short-term memory play in second language reading comprehension. This research will extend the existing research on WM and reading comprehension by examining how the role of WM may differ according to proficiency level. The findings of this study may also help to establish an appropriate procedure for the measurement, administration, and scoring of WM measures.

1.9 Organization of the Thesis

This dissertation comprises six chapters. As an introduction, Chapter One sets the scene and provides an overview of the rationale behind the study as well as the basic concepts of WM and reading ability. Chapter Two reviews the major theoretical issues that frame the study, and then provides an overview of prior studies in the field. The research questions of this study will also be presented at the end of this chapter. Chapter Three describes the materials, instruments and methodological issues in this study. It outlines the research design and specifies the procedures for data collection. It also includes the description of all pilot studies carried out in preparation for the main study on memory and reading comprehension. The results of the study are presented and discussed in Chapter Four and Chapter Five. Chapter Four provides answers to two research questions. It also includes the inferential analyses for dependent and independent variables. Chapter Five discusses the findings of the study for the two research questions. Chapter Six, the concluding chapter, summarizes the results, explains

where this study fits in the field, and discusses the theoretical implications of the study. Recommendations for further research are then offered and the limitations of the study are addressed.

Chapter Two

Review of Literature

2.1 Introduction

This section begins with an explanation of the theories and models of reading comprehension, the theoretical issues in reading assessment, the methods of reading comprehension, and the current shortcomings in this domain. Second, an overview of earlier studies on memory, short-term memory and WM is presented. These studies deal with the role of memory, particularly WM, in language learning. This includes the role that WM plays in L1 and L2 reading comprehension processes, in particular the correlations between the central executive and L1 and L2 reading comprehension. Then, the findings of the studies to date on WM with respect to L1 and L2 reading comprehension are elaborated. Finally, the research questions of the current study will be proposed.

2.2 Reading Comprehension

Reading is a way to draw information from a text to form an interpretation of that information (Grabe & Stoller, 2002, p., 4). Reading is a cognitive activity involving skills, strategies, attentional resources, knowledge resources and their integration (Grabe & Stoller, 2011). The reader's role is to decode the written symbols to allow for the recovery of information from long-term memory to construct a plausible interpretation of the writer's message (e.g., Hudson, 2007; Mitchell, 1982). Researchers argue that the reader goes through certain stages to comprehend a text (e.g., Koda, 2005; Cain, 2006; Beach, 1997; Mitchell, 1982; Grabe & Stoller, 2002; Cain & Oakhill, 2006).

The first stage is word recognition, through which the reader must decode the individual words on the page for their meanings. The literature provides positive evidence that the overall shape of the word, pronunciation and orthographic rules are associated with word recognition and can be used to access its lexical entry in the case of familiar words and pronunciation strategy to convert the printed form into a phonemic form in the case of unfamiliar words (e.g., Beech, 1997; Cain, 2006; Mitchell, 1982). The second stage is

syntactic parsing which is “the ability to recognize phrasal groupings, word ordering information, and subordinate and superordinate relations among clauses” (Grabe & Stoller, 2011, p., 16). In the process of syntactic parsing, the structure of sentences is analyzed to assemble the most logical clause-level meanings. So, it is at this stage that ambiguity is removed from the words that have multiple meanings out of context. The third stage is making propositions through which the grammatical structure and sense of each sentence is interpreted. Based on the evidence derived from memory experiments, researchers argue that when an individual reads a passage, they construct an internal representation which matches with the meaning of the text. This internal representation can be explained in terms of abstract statements (propositions) which are connected to one another and organized according to their importance. These propositions are constructed based on some superficial properties of the words which are used as clues to determine the underlying structure of the sentence (Bever, 1970; Foder, Bever & Carrett, 1974).

The fourth stage is integrating propositions and making inferences. To grasp the overall meaning of the passage, the reader needs to do more than derive the meaning of individual words and sentences. He or she should be able to combine these propositions to establish local coherence and to generate inferences to make sense of details that are only implicitly mentioned (Grasser, Singer, & Trabasso, 1994; Halivand & Clark, 1974; Kintsch, 1988, 1998; Mitchell, 1982; Long & Chong, 2001). In doing so, he or she makes use of cohesive ties (Halliday & Hasan, 1976) either in the form of “anaphoric processes” or “relatedness of reference”. The former refers to usual syntactic devices such as “he” or “she”, which are used to maintain cohesion between phrases, propositions, and sentences in a text. The latter is made through a procedure known as Given-New strategy (Halivand & Clark, 1974) to determine the relations between propositions.

According to this strategy, the reader first constructs the propositions in the current sentence and then assigns each of these propositions to one of the two categories depending on whether it consists of presupposed information (Given information) or information which has not already been mentioned in the text (New information). Then the reader searches his or her internal representation of the preceding text for an antecedent that exactly matches the Given information. If such an entity can be located in memory, it clearly provides a way of

connecting the current sentence to the material that has gone before it. This link is established by storing the new information in the current sentence with the Given information that has just been located (e.g., Beach, 1997; Cain, 2006; Cain & Oakhill, 2006; Grabe & Stoller, 2002; Koda, 2005; Mitchell, 1982).

Comprehension monitoring is the final stage. The reader's ongoing comprehension is monitored in order to take remedial steps such as re-reading in cases where he or she has not fully understood a point or when he or she finds a mismatch between a current proposition and his or her perception (e.g., Cain, 2006; Cain & Oakhill, 2006).

2.3 Models of Reading Comprehension

2.3.1 Introduction

Foreign or second language reading has been the focus of researchers' attention over the past twenty years (Macaro, 2003, p., 118), based on which some models of reading comprehension have been proposed. During the 1960s and early 1970s, a number of researchers proposed more or less formal models of reading comprehension. For example, Carroll (1964) suggested a definition of reading along with a simple one-way flow diagram from visual stimulus to an oral language recoding to meaning responses. Since his aim was to be illustrative, rather than definitive, many imprecisely specified stages were left in his model. Also, Levin and Kaplan (1970), Hockberg (1970), and Mackworth (1972) all argued about what a model explaining the processes of skilled reading must account for. This work heralded a change in conceptions of the reading constructs among researchers and practitioners.

In contrast, Goodman (1965, 1976) described reading as a psycholinguistic guessing game. Three distinctive characteristics distinguished Goodman's model from other models. First, he believed that the reader relies on existing syntactic and semantic knowledge rather than graphic information in the process of reading. Second, he used the term "decoding" differently from others. While others used this term to describe what happens when a reader translates a graphemic input into a phonemic input, Goodman used it to illustrate how either a

graphemic input or phonemic input gets translated into a meaning code. He also used the term ‘recoding’ to describe the process of translating graphemes into phonemes. Goodman’s and his colleagues’ efforts were mostly focused on indicating the strong procedural preference that readers of all ages have for depending on the meaning cues (rather than graphic and graphophonemic cues) available in the printed message. Third, his model has arguably had the greatest influence on conceptions about reading pedagogy, to the extent that ‘the psycholinguistic approach to reading’ or ‘the whole-language approach to reading’ have become commonly used terms in the language teaching field (Samuels & Kamil, 1988).

In summary, Goodman (1996) argued that when an individual reads a text, he or she makes a set of hypotheses about the upcoming text, samples minimally from the text, confirms hypotheses, and then produces new predictions. However, other researchers (e.g., Grabe, 2000, 2009; Koda, 2005; Pressley, 2006) impose some criticisms on this argument. They argue that there is no persuasive evidence in a fluent reading that good readers (a) sample from texts and make hypotheses about what words are coming next (b) controls their eye movements (direct the eye where to go during reading to sample from a text). They further argue that good readers do not usually guess upcoming words in a text, and make less use of context for word identification than poor readers. Grabe (2009) argues that Goodman’s Psycholinguistic Guessing Game model provides a possible explanation for an early stage of reading development. He further argues that Goodman’s Psycholinguistic Guessing Game Model cannot be a valid alternative to any other models of reading, which will be described in the following section (Grabe, 2009 p., 3).

In the following section, seven models of reading will be discussed in turn: bottom-up, top-down, Rumelhart’s Interactive Model (1977), Stanovich’s (1980) Interactive-Compensatory Model, Construction-Integration Model, Verbal Efficiency Theory of Reading, and Compensatory-Encoding Model. Discussion of these models follows by a critical overview with the focus on the most relevant models to this study. It should be noted that these models are limited to the cognitive aspects of reading. Other models characterize reading as a more complex process where motivational and emotional aspects play an important role; however, these aspects are beyond the scope of this study.

2.3.2 Bottom-up Model

A bottom-up reading model is a model that focuses on a single-direction, part-to-whole processing of a text. More specifically, in bottom-up models, the reader is assumed to be involved in a mechanical process where he or she decodes the ongoing text letter by letter, word by word, and sentence by sentence (Grabe, 2009). In these models, the reader decodes the text which has been previously encoded by the writer. Decoding of the text includes a visual focus on the identification of the letters, noticing the combination of the letters, recognition of the words, establishing sentences via their syntactic structures and finally integrating sentences into coherent discourse until the meaning of the text is eventually determined. The reader's world knowledge, contextual information, and other higher-order processing strategies play a minor role, particularly at beginning stages, in processing information in this model (e.g., Alderson, 2000; Beach, 1997; Dechant, 1991; Grabe & Stoller, 2002; Koda, 2005). The proponents of bottom-up models (e.g., Flesch, 1955; Gough, 1972; LaBerge & Samuels, 1974) argue that these models work on the premise that the written text is hierarchically organized, and the reader's job is to process the smallest linguistic unit first (grapho-phonetic) and then combine the smaller units to discover and comprehend the higher units (e.g., sentence syntax) (Alderson, 2000; Dechant, 1991; Field, 2003; Grabe & Stoller, 2002; Koda, 2005; Macaro, 2003; Mitchell, 1982).

Word recognition plays an essential role in reading comprehension. Koda (2005) defines it as "the processes of extracting lexical information from graphic displays of words" (p., 29). Studies on eye movement indicate that nearly every content word obtains direct visual fixation (Balota, Pollasek, & Rayner, 1985; Just & Carpenter, 1980, 1987), and the lack of even a single letter can be disruptive, largely decreasing reading efficiency (e.g., McConkie & Zola, 1981; Rayner & Bertera, 1979). Furthermore, based on developmental studies, researchers argued that poor readers could not extract visual information from print, and deficient word recognition is associated with poor comprehension (e.g., Perfetti, 1985; Stanovich, 1988). If inefficient word recognition continues, it may have adverse effects, directly or indirectly, on the acquisition of reading competence (e.g., Juel, 1988; Juel, Griffith, & Gough, 1986). Thus, word recognition efficiency can result in successful comprehension. Some studies also indicate that automaticity can be rather easily achieved in word recognition (e.g., Adams, 1994; LaBerge & Samuels, 1974; Perfetti & Lesgold, 1977, 1979). This may reduce the processing load in WM, leaving more capacity for the storage component, and

eventually facilitating conceptual manipulations of the extracted information (e.g., Daneman & Carpenter, 1980; Waters & Caplan, 1996).

Word recognition involves orthographic, phonological and semantic operations. While a word's meaning is obtained in semantic operation, the word's sound features are achieved in phonological operation. Both of these operations are activated through orthographic operation and achieved via an analysis of graphic symbols (Koda, 2005; Samuels & Kamil, 1988).

Orthographic knowledge plays an important role in word recognition. Research suggested that skilled readers were able to not only analyze and manipulate word-internal elements such as letters and letter clusters (e.g., Ehri, 1998; Shankweiler & Liberman, 1972), but also to pronounce both individual letters and nonsense letter strings (e.g., Siegel & Ryan, 1988; Wanger, Torgesen, & Rashotte, 1994). This is because orthographic knowledge is a powerful mnemonic device that connects the written forms of specific words to their pronunciation in memory (Ehri, 1998).

Phonological decoding may be the most essential competence for reading acquisition in all languages (Koda, 2005). It is defined as the processes involved in accessing, storing, and manipulating phonological information (Torgesen & Burgess, 1998). Researchers argued that deficits in phonological decoding could lead to poor comprehension in both alphabetic (Abu Rabbia, 1995) and nonalphabetic languages such as Japanese and Chinese (Kuhara-kojima, Hatano, Saito, & Haebara, 1996; Zhang & Perfetti, 1993).

The empirical evidence supports the idea that all of a word's known meanings are activated by its orthographic input, even when strong constraints are imposed by the context (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982). Then contextual facilitation helps to determine the appropriate meaning of the word in the immediate context at the sentence or discourse level. Researchers also argue that less skilled readers are more likely to be dependent on the context to retrieve word meanings than skilled readers (e.g., Biemiller, 1979; Becker, 1985; Perfetti, 1985; Stanovich, 1988). This supports the idea that poor readers use

contextual clues to compensate for their underdeveloped visual information sampling skills in order to decipher a word's meaning (e.g., Pring & Snowling, 1986; Stanovich, 1986).

There is an assumed relationship between contextual effects on word-meaning retrievals and language proficiency. It is suggested that as L2 proficiency improves, reliance on contextual effects to retrieve word meaning diminishes (Becker, 1985; Grabe, 2009; Pring & Snowling, 1986; Stanovich, 1986). A large body of studies also indicates that efficiency in extracting visual information differs among high and low-proficiency readers, suggesting that low-proficiency readers are slower and less accurate in a variety of word recognition tasks (e.g., Favreau & Segalowitz, 1982; Haynes & Carr, 1990; Macnamara, 1970).

Some other studies suggest that low-proficiency readers are more largely involved in word-level than discourse-level processing (e.g., Cziko, 1980; Horiba, 1990). Since low-proficiency readers rely on a word's visual information rather than its semantic information (Chamot & El-Dinary, 1999; Clarke, 1980), they are less likely to engage in conceptual manipulations (such as hypothesizing and predicting) than high-proficiency readers (e.g., Chamot & El-Dinary, 1999; Anderson, 1991).

In bottom-up models, the reader takes a serial order to process the text, and the processing of each component takes place independently of the others (e.g., Alderson, 2000; Grabe, 2009; Koda, 2005; Mitchell, 1982). For example, the perception of phonemes is not influenced by the words in which they appear (Carroll, 2008). Since there is a single and restricted meaning in the text driven and constructed by the writer, the reader needs to extract this meaning and cannot go beyond it (Alderson, 2000; Beach, 1997; Grabe & Stoller, 2002; Koda, 2005). Therefore, it is not possible to make use of higher-order reading skills such as making inferences, and consequently, background knowledge plays virtually no role in deriving and interpreting the meaning of the text in this model.

2.3.3 Top-down Model

A top-down reading model is a model that focuses on what the reader brings to the text to arrive at the meaning. In top-down models, it is assumed that the comprehension process is not mechanical, but actively controlled by the reader (Grabe, 2009). The proponents of these models (e.g., Schank, 1978; Smith, 1971) suggested that processing of a text begins in the mind of the reader with meaning-driven processes, or an assumption about the meaning of a text. From this viewpoint, readers identify letters and words only to confirm their assumptions about the meaning of the text (Dechant, 1991). In these models, the primary purpose of reading is deriving meaning from the text rather than mastery of letters, letter-sound correspondence, and words (e.g., Alderson, 2000; Grabe & Stoller, 2002; Macaro, 2003; Smith, 1971). Readers are supposed to use meaning and grammatical cues to identify unfamiliar words, and they are able to comprehend a passage even if they do not recognize each word. In this view, the meaning of a text, which is considered an important goal to achieve, is accessed by the reader's activation of prior knowledge of semantic, pragmatic, syntactic and discourse elements. Then he or she will be able to predict and infer the meaning underlying propositions and words (e.g., Alderson, 2000; Beach, 1997; Dechant, 1991; Grabe & Stoller, 2002; Koda, 2005). However, this view does not identify what mechanisms the reader draws on to generate inferences or how the mental composition of comprehension works (Grabe, 2009).

2.3.4 Rumelhart's (1977) Interactive Model of Reading

Since the information in top-down and bottom-up models is passed along in one direction only and the information contained in higher stages does not influence the information in lower stages, these models could not account for a number of well-known occurrences, such as making inferences, which take place while reading. Thus, to remove this deficiency, Rumelhart (1977) proposed an interactive model of reading. This model, which is a combination of both top-down and bottom-up strategies, is now widely considered a comprehensive explanation of how we derive the meaning of a written text. Rumelhart (1977) developed this model based on the fact that meaning does not reside in the text alone, but is a co-construction of the writer's text and the reader's interpretation. So, reading requires an interaction between the reader's mind and the writer's text. This allows the information contained in higher stages to interact with and influence the information in lower stages.

In this model, the process starts with the information picked up by the eyes in the form of visual features, registered in a visual information store, and then sent to the central component of the model, the pattern synthesizer, at the first stage. Then a wide variety of sources of information about letter shapes and orthography (including what is semantically and syntactically acceptable in the language, the contextual situation, and information in the mental lexicon) is drawn up from long-term memory into WM. Finally, the pattern synthesizer uses this information to work out the more probable interpretation of the text. During this process, the already-made hypothesis is confirmed, strengthening connections and built-up layers of interpretation by pausing over individual words and syntactic patterns and their relationship with other words and phrases (Macaro, 2003, p., 121). Therefore, the reader is involved in deriving the meaning of the text and making inferences through a constant interaction between the surface structure of the text and his own knowledge of the topic.

Since WM is the workspace for the temporary storage and processing of ongoing information (e.g., Baddeley & Hitch, 1974, 2000, 2007), it may play a significant role in the processes involved in deriving meaning from text (e.g., Alderson, 2000; Beach, 1997; Cain & Oakhill, 2006; Grabe & Stoller, 2002; Koda, 2005). These processes consist of maintaining the text information, activating the reader's world knowledge and retrieving it from long-term memory, integrating the information received from these two sources into coherent discourse, and finally deriving the meaning of the text. A substantial body of L1 (e.g., Daneman & Carpenter, 1980; Divesta & Dicintio, 1997; Waters & Caplan, 1996) and L2 research (e.g., Harrington & Sawyer, 1990; Lesser, 2007) supports the idea that good readers have higher WMC than poor readers.

2.3.5 Stanovich's (1980) Interactive-Compensatory Model

Stanovich's (1980) interactive-compensatory model was a refinement of Rumelhart's (1977) interactive model in explaining skilled and unskilled reading. It is based on the principle that a process at any level can compensate for deficiencies at any other level. In his words, "... a deficit in any knowledge results in a heavier reliance on other knowledge sources regardless of their level in the processing hierarchy" (p. 63). So, top-down processing, for a reader weak at word recognition, but good at the knowledge of the text topic, may compensate for this

deficit. On the other hand, a reader good at word recognition, but lacking knowledge of the topic may rely on bottom-up processes for this compensation (Samuels & Kamil, 1988). The research also supports the idea that prior knowledge of the topic can be used by the learner as a strategy to reduce the cognitive load when syntactic complexity makes access to meaning difficult (Barry & Lazarte, 1998). From a theoretical perspective, Stanovich (1988) made a unique contribution to reading models by providing an explanation of compensation strategies, which account for why poor readers show greater sensitivity to contextual constraints under some circumstances than good readers (e.g., Alderson, 2000; Beach, 1997; Grabe & Stoller, 2002; Samuels & Kamil, 1988).

Unlike the models described above, the following two models (Construction-Integration and Verbal-Efficiency models) are experimental/ behaviour models of reading where the researchers draw on a range of experimental evidence to develop and support their assumptions (Grabe, 2009). Moreover, they envision an important role for WM and automatic bottom-up processing in reading process as it will be described in the following section.

2.3.6 Construction-Integration Model of Reading

Construction-Integration Model was proposed by Kintsch and his colleague (Kintsch, 1988, 1998; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). Based on this model, automatic lower-level reading processes are combined with higher level reading processes to create a coherent discourse representation of a text, and these processes are supported by a limited capacity pool of attentional resources. There are two phases in this model; a construction phase and an integration phase. In the construction phase, a reader develops propositions from the incoming text information in order to generate a mental model of the text. This model is provisional and incoherent since it includes both relevant and irrelevant information which have been activated (when an individual reads a word, all the meanings of the word as well as the semantic associates of that word are automatically activated in his or her long-term memory, Graesser, Millis & Zwaan, 1997). In the integration phase, the reader evaluates the propositions he or she has developed within a global context with the goal of making a stable activation pattern or a coherent mental network. In doing so, the propositions which are compatible within the context are connected to form the network, and those which

are incompatible are disregarded. At this phase, the integration of text information with the reader's background knowledge yields a coherent mental model which captures the global and local relations and consequently results in comprehension. All these processes in construction and integration phases are manipulated by WM. More specifically, WM is involved in the processes of making propositions, suppressing irrelevant information, and developing a coherent mental network which result in reading comprehension. This suggests that WM plays a strong role in reading comprehension. However, it is not yet clear to what extent individual differences in WM may explain differences in reading performance, and also whether this role may be shifted as proficiency develops, one of the hypotheses that the present study has been designed to investigate.

The assumptions of cognitive capacity limitations in comprehension processes following by the integration processes (summarizing processes due to being overlapping associations among propositions) distinguish the Construction-Integration Model from the models reviewed before.

The automatic lower-level processes and limited pool of attentional resources in WM are also considered as important assumptions in the Verbal-Efficiency Model, proposed by Perfetti (1985-2007). However, the emphasis is on automatic word-recognition skills which result in reserving more attentional resources for higher level processes, and consequently better reading performance.

2.3.7 Verbal Efficiency Theory of reading

This model was proposed by Perfetti (1985, 1988, 1991, 1999, 2007). It is an example of an interactive model which is very constrained by the bottom-up view of reading (Hudson, 2007). Efficient word-recognition skills play a very important role in good reading performance in this model. It is argued that problems with higher-level comprehension skills originate from inefficient word-recognition skills which, in turn, stem from low-quality lexical representations (Perfetti, 2007). Perfetti and his colleague (Perfetti, 2007; Perfetti & Hart, 2001, 2002) argue that there are three constituent information sources for word

recognition including phonological, orthographic and semantic information. These constituents work together and share information until a word is recognized.

Based on Verbal Efficiency Theory of reading, skilled readers have automatic lower-level processes (e.g., efficient word recognition skills), and this allows them to draw on their limited attentional resources in WM for higher level comprehension skills. More specifically, there are two sets of processes in this model, local text processes and text-modelling processes, which have interactions in reading process. The central principle of this model is that the comprehension of a text is partially constrained by the efficient operation of the local processes. The local processes involve the processes that the reader uses to encode contextually appropriate meanings and propositions. When a text is read, first, the possible meanings associated with each word in the text are activated in working memory. Second, the most appropriate semantic meaning for the proposition in the context is selected. Third, initial propositions are created from the propositional encoding of each word and maintained in working memory. Finally, new propositions are integrated with previous propositions held in working memory to give a representation of the text. This representation remains active in working memory to be further processed by text-modelling processes (higher-level processes) (Perfetti, 1985 p., 111). Text-modelling processes are used to combine the representation of the text with a reader's background knowledge to fill the gaps in the propositional base and make him or her create inferences. It is at this stage that comprehension (text-modelling) takes place and causes the propositions to make sense as a whole. A continual updating process occurs during the reading by reconciling incoming text processing with background knowledge. For this process to be efficient, the processes for word-recognition components (phonological, orthographic, and semantic) must be automatic. As there are limited attentional resources in working memory, automatic processes reduces the amount of attentional resources for processing letter and word identification, and consequently leaves further attentional resources for processing higher level comprehension skills. This suggests that efficient WM processes play an important role in reading comprehension (Hudson, 2007), particularly for low-proficiency readers who have not obtained automaticity in their local processes. If this is the case, WM is expected to explain individual differences in reading comprehension. However, it is not clear yet whether or not the role of WM may change as a result of language proficiency development.

Overall, this model is compatible with the Construction-Integration Model where WM with a limited capacity pool of resources is central to manipulating reading processes. It appears that Verbal-Efficiency Model is more prominent in explaining efficient word-recognition skills which result in automaticity in reading process which leads to leaving more attentional resources in WM for higher-level processes such as making inferences.

The last model reviewed here is the Compensatory Encoding Model which also conceives of WM as having a key role in reading process. However, a compensatory process is assumed in this model which distinguishes it from the Verbal Efficiency Model. Moreover, unlike the Verbal Efficiency model, it is a descriptive model where a synthesis of the most important evidence is used to explain how a cognitive process like reading works (Grabe, 2009).

2.3.8 The Compensatory-Encoding Model

This model of reading was proposed based on verbal efficiency model and adopted its basic assumptions including automatized lower-level processing, well-developed lexical representations, and efficient working memory processes (Walczyk, 1995, 2000; Walczyk, Marsiglia, Bryan, Naquin, 2001). This model assumes an additional process that is a compensatory process. The compensatory process in this model differs from that of Stanovich's interactive compensatory model in that it is used continually to counter inefficiencies and weaknesses in reading skills. In Stanovich's model, higher-level skills and strategies are used only when needed. Based on the compensatory encoding model, the compensatory processes play an influential and a predictive role in reading performance when there is no time constraint on a reading task. This could work well particularly for readers with lower working memory capacity as they may employ these strategies to compensate for their inefficient working memory processes. These compensatory processes, similar to those in Stanovich's model, include higher-level skills and metacognitive strategies (e.g., goal checking, comprehension monitoring). When the process of reading proceeds under time pressure, the compensatory processes do not play a role and instead lower-level processes become influential and play a predictive role in reading performance (Grabe, 2009).

Overall, each model can contribute to our understanding of the reading process. With each new model building on previous work, a developing understanding of the reading process has emerged from this rich research history. The increasing specification of the role of cognitive processing in reading is of particular relevance to this study, and makes it possible to more clearly understand the role of WM in the reading process. Except for bottom-up and top-down models, there are some commonalities among the models reviewed above. They all conceive of the reading process as involving both lower-level (e.g., word-recognition skills, syntactic parsing) and higher-level (e.g., making inferences) processes. The proponents of these models suggest that word-recognition skills play a very important role in reading comprehension. As Perfetti (2007) suggests one explanation could be that word recognition involves the interaction of orthographic, phonological, semantic and syntactic processes which are cognitively demanding. Thus, those readers who are good at word recognition (due to possessing well-represented lexical information) leave much of their attentional resources for higher-level reading processes which in turn result in better comprehension. Moreover, they argue that basic grammatical information can be extracted to support clause-level meaning and proposition formation. However, these models differ in explaining the nature and role of these processes.

Reading is not a mechanical process, as assumed by bottom-up models, nor is it carried out in a serial order as it is envisioned by the bottom-up and top-down models, but it is an interactive cognitive process involving simultaneous lower-level and higher level processes which are manipulated in working memory. Of the reading models described above, Verbal-Efficiency (Perfetti, 1999, 2007) and Construction-Integration models (Kintsch, 1998) are those most relevant to this study. They specify the role of cognitive processes in reading comprehension more precisely than other models of reading. They provide a reasonably complete explanation of reading abilities in terms of cognitive processes with empirical evidence indicating how reading performance may vary under different conditions (Grabe, 2009). They specify how reading performance may vary due to individual differences in reading abilities. These individual differences could stem from either attentional resources or reading skills (e.g., word-recognition skills). For example, in Verbal-Efficiency model, skilled readers are distinguished from poor readers in terms of possessing automatized lower-level processes (e.g, more efficient word-recognition skills). Thus, the central assumption of Verbal Efficiency model could be used to explain the automaticity aspect of the reading

process in the present study. Based on this assumption, automatized reading processes are not very cognitively demanding of attentional resources, so more attentional resources are left in WM for higher-level reading processes, which, in turn, enhance the reading performance. This assumption could be used to explain what role WM plays in reading process and whether this role changes as proficiency increases. For example, lower proficiency learners may struggle with word-recognition processes, and attentional resources are directed towards lower-level processing. However, as familiarity with second language increases, reading becomes more automatized and greater attention can be given to higher-level skills such as making inferences. In Construction-integration model, comprehension processes are carried out within the attentional capacity limitations of WM. Of relevance to the current study, this model may explain how limited-attentional resources in WM could be drawn on to develop a local representation of a text (a set of main idea and supporting details), and then to integrate this representation with background knowledge to make an interpretation of the text in a global context.

There are still some limitations among all the models of reading described above. None of these models explain how executive control processes in WM work in fluent reading and how reading strategies are used when reading more difficult texts or learning from texts. In both phases of Construction-Integration model, the abilities of monitoring comprehension, using strategies, and reassessing and re-establishing goals are used to repair comprehension problems. However, it is not completely clear how the operation of monitoring, as an attentional demanding process and an aspect of executive control processing in WM, is manipulated cognitively. Moreover, these models do not explain how WM handles the cognitive processes in comprehending longer and more complex texts.

2.4 Reading in a Second Language

Like first language (L1) reading, reading in a second language is a complex process involving multiple operations. However, there are some important differences between L1 and L2 reading. Unlike most L1 readers, who begin to read with considerable tacit grammar knowledge, the cognitive and linguistic resources accessible to L2 readers vary considerably more than those available to L1 readers (Koda, 2005). There are a number of factors in the

linguistic and cognitive processing domain which indicate the differences between L1 and L2 reading.

This section will focus on eight important differences between the L1 and L2 reading process. The major difference between L1 and L2 reading lies in the fact that there are two languages involved in the L2 comprehension processes. Thus, all variables which play a role in L2 reading are influenced by the inevitable interplay between the L1 and L2. These variables include word recognition, the reading rate, the organization of the lexicon, the speed of the syntactic processing, strategies for comprehension, experiences in task performance, expectations of success and failure, motivations for reading, and a number of possible points of interaction (Segalowitz, Poulsen & Komoda, 1991).

The second difference is between L1 and L2 readers' knowledge of vocabulary, grammar and discourse at initial stages. In the L1, before the learners start to read formally around the age of five, they have developed their oral language to a good extent. They know most basic grammatical structures as well as an appropriate range of vocabulary, 5000 to 7000 words. However, this is not the case in the L2. Most L2 readers develop both reading and oral language at the same time. In other words, they do not begin to read in the L2 with as many words and grammatical structures as they do in their L1. Additionally, sometimes L2 readers' lack of discourse knowledge, text organization and genres impedes a more effective reading comprehension, although they may know most of the vocabulary and understand the main concepts of a text (Alderson, 2000; Cain & Oakhill, 2006; Grabe & Stoller, 2002; Koda, 2005; Leslie & Caldwell, 2009).

The third difference compares L1 and L2 readers in terms of metalinguistic and metacognitive awareness. In fact, adult L2 readers' metalinguistic and metacognitive awareness is greater than that of L1 readers (Grabe & Stoller, 2002). Since L2 readers' knowledge mostly results from direct instruction in the classroom, they develop greater metalinguistic knowledge, which is the knowledge of letters, sounds, sentences, texts and genres and how they are organized. L2 readers also bring the experience of L1 literacy skills and content knowledge into an L2 setting. Because of their L1 literacy, they have developed

greater metacognitive knowledge, which includes knowledge of planning, goal setting, processing of tasks, monitoring of progress, recognition of problems, repair of problems, and explicit and conscious use of reading strategies. This is why L2 readers can more easily use their metalinguistic knowledge at a conscious level to provide strategic support or understand comprehension failure (e.g., Alderson, 2000; Grabe & Stoller, 2002; Koda, 2005).

The fourth difference discusses L1 and L2 readers' exposure to print. While the extent of L1 readers' exposure to L1 print pre-literacy is more or less consistent, L2 readers' exposure to L2 print varies depending on the L1 orthography. Because L1 readers have been exposed to L1 print for years before reading, they have been able to develop sufficient fluency and automaticity in word and syntactic processing, while many L2 readers are not exposed to enough L2 print through reading to build fluent processing (Koda, 1996).

The fifth difference between L1 and L2 readers lies in the area of linguistic differences. Linguistic differences across any two languages may influence L2 reading comprehension considerably. Learners from different L1 backgrounds may vary from one another significantly in reading rates and fluency in word processing. For example, words in Arabic and Hebrew are read more slowly than words in English by their speakers because they include greater morphological complexity with embedded grammatical information (Share & Levin, 1999; Shimron & Sivan, 1994).

The sixth difference concerns the orthographic differences across L1s. For example, as Turkish and Finnish both have phonemically transparent orthography, their speakers are able to sound out the word and activate it in their WM even if the graphic form is previously unknown. However, this is not the case with languages such as Japanese and Chinese (Elley, 1992; Harris & Hatano, 1999b; Oney, Peter, & Katz, 1997). These languages use logographic writing systems, which do not encode phonological information. Word processing differs across different types of orthography. In general, L2 readers, particularly at the beginning level, tend to rely on L1 processing strategies when trying to decode the L2.

The next difference involves the role of cognates in the L2 reading process. Cognates may play an important role in L2 reading comprehension. Languages may share anywhere between a few to thousands of cognates. For example, an L1 French reader of English is able to make greater use of cognates to support his reading comprehension than an L1 Chinese reader as French shares far more cognates with English than Chinese does (Grabe & Stoller, 2002).

The last difference concerns the role of transfer of knowledge from one language into the other. Transfer of L1 knowledge distinguishes L2 from L1 reading. Transfer is defined as the ability to learn new skills by drawing on previously acquired resources (Genesee, Geva, Dressler, & Kamil, 2006). It may either support L2 comprehension or interfere with it. Transfer as interference usually plays a role at the beginning and intermediate levels of L2 reading, particularly when learners at these stages are required to read difficult material. This is due to the lack of sufficient L2 knowledge, which leads them to rely on any sources available to make sense of the text.

To summarize, reading is a complex cognitive ability which involves lower-level and higher-level processes. There are important differences between L1 and L2 reading. The main difference is that reading in the L2 involves two languages. Other differences are related to knowledge of vocabulary and grammar, metalinguistic and metacognitive awareness, the extent of exposure to print, linguistic aspects, orthographic features, the role of cognates in the reading process, and the role of transfer of knowledge from one language into the other. In the next section, a theoretical explanation will be given on the assessment of reading comprehension, and then different methods of reading assessment will be presented.

2.5 Assessment of Reading Comprehension

A large amount of research suggests that the development of reading assessment needs to be based on a model of reading comprehension as its foundation (Alderson, 2000; Cain & Oakhill, 2006; Koda, 2005; Leslie & Caldwell, 2009). However, since models of reading comprehension have rarely been clearly defined, they suffer from a lack of construct validity

(Alderson, 2000; Cain & Oakhill, 2006; Koda, 2005; Leslie & Caldwell, 2009). Construct validity is a unifying concept consisting of test score and consequences (Messick, 1989), and a reliable assessment of a reading comprehension model is to a large extent dependent on construct validity (Alderson, 2000; Cain & Oakhill, 2006; Koda, 2005; Leslie & Caldwell, 2009).

Thus, having a reliable reading comprehension model with good construct validity involves both a comprehensive description of the processes used to understand the text, and a set of systematic measures to assess these processes (Leslie & Caldwell, 2009, p., 404). Methods of assessing reading comprehension can be described along a continuum based on two criteria: the precision of the reading comprehension model underlying the assessment as well as the degree to which the measures are specified. At one end of the continuum, no explicit theoretical explanation of the processes involved in reading, how the reader constructs meaning, as well as how to assess these processes is suggested. However, at the other end of the continuum, some assessment models explicitly describe the theory underlying them and design measures of many sub-processes of comprehension (Leslie & Caldwell, 2009). In the formal assessment of reading comprehension, the components of reading are assessed differently. For example, the items assessing memory for word meanings and drawing inferences from the content (Davis, 1994; Thorndike, 1917) are suggested by standardized tests of reading comprehension (cf., Leslie & Caldwell, 2009).

All in all, testing reading is a complicated and problematic issue, since it addresses the reader's background knowledge and his or her cognitive processes on the one hand and the interrelationship of the text and task on the other (Macaro, 2003, p. 149). Moreover, research in testing reading suggests that individual differences mediate how a reader reacts to a test item differently from others (Alderson, 2000). For example, field-dependent individuals perform differently on cloze tests and multiple choice texts than field-independent individuals (e.g., Alderson, 2000, Macaro, 2003). Several studies (e.g., Alderson, 2000, Cain & Oakhill, 2006; Koda, 2005; Leslie & Caldwell, 2009) suggest no single comprehensive assessment of reading comprehension exists due to inadequacies in the construct validity of the current existing formal assessments of reading comprehension. For this reason, researchers and practitioners have turned their attention to indirect tests of reading comprehension. One

explanation could be that as reading ability is measured for divergent purposes, and each method of reading assessment may tap into a specific aspect of reading construct, then different indirect tests could be good options and used to meet the assessment purposes. They are tests such as cloze tests, true/false sentence recognition, sentence verification technique tasks, multiple-choice tasks, open-ended questions (Cain & Oakhill, 2006; Leslie & Caldwell, 2009), and short-answer reading tests (Alderson, 2000). For example, short-answer test may be a good measure to look at the cognitive aspect of reading, particularly when it involves organizational and inferential questions because this test could well tap into lower-and higher level processes. As cloze test is constrained by syntactic and semantic knowledge, it may be a good measure of vocabulary and grammar aspect of language.

In true/false sentence recognition, the examinees read a text and then they must determine whether a set of sentences are true or false based on what they have read. However, sentence recognition tasks are not able to measure high level processes of reading comprehension such as inference making. Therefore, this type of task is not able to distinguish poor readers from good readers (e.g., Alderson, 2000; Cain & Oakhill, 2006; Koda, 2005).

In a sentence verification technique task, the participants are required to read texts and judge which set of test sentences has the same meaning as the sentences they have just read. In this test, there are usually 12 to 16 test sentences consisting of original sentences, paraphrases of originals, meaning-changed items and distracters. Acceptance of the original and paraphrased sentences and rejection of others is considered to be good performance in this task (Royer, Greene & Sinatra, 1987). However, as this task requires a large number of test items to be accurate and does not assess reading skills beyond interpreting literal meaning, it is generally not used in standardized assessments of reading ability.

The free recall protocol task is an assessment technique as well as the most straightforward procedure for measuring the outcome of the reader-text interaction (Koda, 2005, p., 236). In this task, the participants are required to read a text and recall as many details as possible. One advantage of this task is that it is easy to construct and administer. However, there are some drawbacks to this method, such as scoring, which is quite time-consuming and

subjective. Moreover, no explanation is given for what is not recalled. In other words, this type of task does not clarify whether the lack of some text components is due to inadequate comprehension, retention difficulty or other factors. In addition, task conditions influence recall performance. For example, recall protocols differ both qualitatively and quantitatively depending on whether they are produced in the L1 or L2 (Lee, 1986, cf., Koda, 2005). Finally, since the free recall protocol task relies largely on memory, it makes it difficult to distinguish the recalled items extracted from the text from those retrieved from long-term memory stored knowledge (Koda, 2005, p., 237).

In open-ended reading tests, the participants read a passage and then are required to provide a completely subjective response to a series of questions following it. They need to provide verbal responses to these questions based on what they have understood. Thus, it is more likely that this type of reading test underestimates the comprehension of individuals with expressive language deficits (Cain & Oakhill, 2003; Cain, et. al. 2001; Norbury & Bishop, 2002). However, this kind of assessment is less a test of rote learning, instead measuring the participants' understanding of what they have covered. Moreover, it builds understanding of the factors affecting language comprehension by analyzing the incorrect responses to find out the source of errors

A cloze test is another technique for measuring reading comprehension. It consists of sentences from which words, usually every 7th word, are removed at regular intervals, leaving blanks of standard length. The participants must then read the passage and try to either fill in the blanks with appropriate words, or select a replacement word from a choice of usually 3-5 options (Cain, & Oakhill, 2006). The sentences can also be presented on their own (Hagley, 1987) rather than in a coherent text. The advantage of a cloze test is that it can be administered in groups where the reading comprehension of a large number of individuals needs to be assessed (Alderson, 1983; Brown 1992; Fotos, 1991). The cloze test is also used to judge the difficulty level of the texts, rate bilinguals, study textual constraints and evaluate teaching effectiveness. The cloze test may accurately assess local sentence-processing skills, but it may fail to measure global text-processing skills (Cain & Oakhill, 2006).

In short-answer reading tests or limited production responses (Bachman & Palmer, 1996), the participants are required to read a text and provide the given questions following it with only a brief response (Alderson, 2000). The rationale behind this test is that it is possible to interpret the participants' responses to see if they have really understood the content of the text or not; whereas this is not the case for multiple-choice items, and the participants may have chosen the correct answer by simply crossing out others (Alderson, 2000; Gronlund, 1973; Harris, 1969; Weir, 1990). Short-answer questions need to be carefully constructed to avoid ambiguous questions and wordings too similar to those found in the text. If short-answer questions are worded similarly to the text, learners may simply search for similar phrases and fill in answers from the surrounding text, without comprehending much of the text.

To summarize, there is no single, comprehensive model of reading assessment. This is because of the inadequacies in the construct validity of the current existing models of reading comprehension. In order to have a model of reading comprehension with good construct validity, a more explicit theoretical explanation of the processes involved in reading and the methods to assess these processes must be proposed. Since the current models of reading have described the extent of the underlying processes involved in reading comprehension differently, various measures have been accordingly proposed to assess these processes, all of which have limitations in terms of construct validity. For this reason, researchers have switched their attention to informal assessment of reading comprehension such as through cloze tests, multiple-choice tasks and short-answer reading tests. These are informal assessments since they measure students' performance under normal classroom conditions rather than through the use of standardized tests or other controlled methods of appraisal (Richards & Schmidt, 2002). Furthermore, the more general consensus today is to employ more than one test method to measure a construct like reading comprehension (Alderson & Bannerjee, 2002).

2.6 Memory

The study of memory has a long history in psychology. The first scientific study about memory was conducted by Ebbinghaus (1885). He suggested that the individual's interest, level of attention, relevant knowledge, and experience were the factors influencing the recall of events in life. He developed a task including nonsense syllables to investigate the memory under controlled conditions in a laboratory. Based on his findings, he proposed the "laws" of memory. These included the idea that an individual is able to hold only seven items in his mind long enough to recall them, and that an individual's memory decays exponentially with time. However, these laws were criticized by other researchers such as Miller (1956), who argued individuals could recall different amounts of items depending on the type of material presented to them. For example, participants could recall more letters when they were presented to them in word formats rather than in the form of individual letters. Furthermore, McGuinness (2005) explains that the participants were able to maintain the items in their memory longer when they were surprised or had new experiences.

A single unitary faculty of memory was not posited until the late 1950s, when a short-term memory system was suggested by Brown (1958), and Peterson and Peterson (1959) after they found that hindering the individual from rehearsing the encoded items in memory makes him rapidly forget small numbers of them (Baddeley, 2006). Atkinson and Shiffrin (1968) proposed a model based on dissociation between short-term and long-term memory. It consists of sensory, short-term, and long-term memory, and the information is processed through the basic mechanisms of encoding, maintenance, and retrieval. According to this model, the information taken from the environment is processed by a series of temporary sensory systems. Then it is sent into a limited capacity short-term memory. This memory is assumed to act as WM, where the encoded information is held temporarily and a wide variety of tasks such as transfer into and retrieval from long-term memory are conducted. Finally, rehearsal is operated on the encoded information to either maintain or transfer it to a more durable long-term memory (Shah & Miyake, 1999).

However, the model suffered from two shortcomings. First, the learning assumption in this model was based on little evidence; the assumption was that simply maintaining an item in the short-term store would facilitate learning. However, there was sufficient evidence to suggest that the way in which information was processed by the participants would determine the degree of learning. For example, elaborate semantic encoding resulted in much more learning than simply focusing on the sound of the word presented (Baddeley, 2006). Second, this model could not account for why patients with serious disruptions in their short-term store did not have any difficulties in feeding information into and out of their long-term memory.

To account for these problems, a three-component WM model was proposed by Baddeley and Hitch (1974). This model consists of a central executive and two “slave” components, the phonological loop and the visuospatial sketchpad (described below). This model was in use until 2000, when Baddeley added a new component to it, the episodic buffer, to account for the studies on densely amnesiac patients with long-term memory deficits.

2.7 Working Memory and the Assessments of its Components

WM is defined as a limited capacity temporary storage and processing system comprised of four components: the phonological loop, the visuospatial sketchpad, the central executive, and the episodic buffer (Baddeley, 2000). This model, as shown in Figure 1, differs from Atkinson and Shiffrin’s (1968) model in that it specifies a functional role of memory as well as an economical and coherent account of information on each memory component.

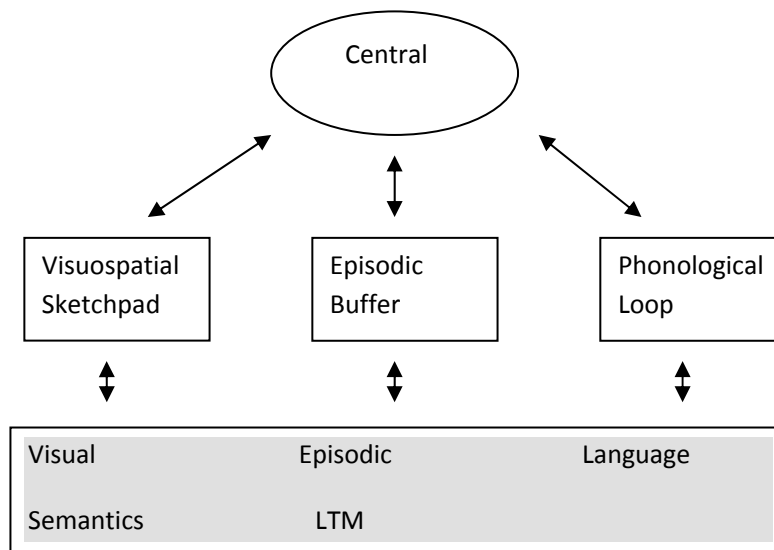


Figure 1

Baddeley's (2000) model of WM, revised to incorporate links with long-term memory (LTM) by way of both the subsystems and the newly proposed episodic buffer.

The most important component in this model is the central executive or supervisory attentional system, which is a limited capacity pool of general resources. According to N. Ellis, (2001), "It regulates information flow within WM, activates or inhibits the whole sequences of activities, and resolves potential conflicts between ongoing schema-controlled activities" (p., 33). WM is usually measured through a reading or listening span test in which the participants read a set of unrelated sentences and judge whether they make sense or are nonsense (processing assessment), and then try to recall the final word of each sentence at the end of the set (storage assessment). An index of WMC is calculated with the composite score of these two assessments (e.g., Friedman & Miyake, 2004; Waters & Caplan, 1996).

The phonological loop is in charge of the temporary storage and processing of verbal information. It plays a role as a phonological store by holding phonological representations of auditory information for a brief period of time, and as an articulatory rehearsal system by enabling the reader to use inner speech to refresh the decaying representations in the phonological store (Baddeley, 2007, 2000; N. Ellis, 2001). Phonological memory is often measured by presenting spoken lists of words (word span), digits (digit span) or non-words (non-word span), and asking participants to recall the lists of words and/or digits in the order in which they are presented. The maximum number of items that the individual can correctly recall is considered to be their phonological memory score.

The visuospatial sketchpad is an interface between visual and spatial information received either through the senses or from long-term memory (Baddeley & Hitch, 1974, p., 854). It is also involved in generating visual images, temporarily maintaining them, and manipulating information with visual or spatial dimensions. Furthermore, it can be activated by spoken words by using long-term knowledge to convert the auditory presented words into visuospatial code (Baddeley, 2007; Ellis, 2001). Della Sala, Gray, Baddeley, Allamano and Wilson's (1999) pattern span test is usually used to measure visual memory. In this test, the individual is presented with 2 x 2 matrixes, with two of the cells filled. Then after 3 seconds, the individual is asked to indicate which cells were filled in the stimulus matrix, using an empty 2 x 2 matrix. The size of the matrix is increased by two cells every three trials, with half of the cells of each matrix being randomly filled. The individual's pattern span is determined by the maximum number of the cells that the participant is able to recall correctly.

The Corsi Block task is typically used to measure spatial memory (Milner, 1971). In this test, the subject is presented with an array of nine cubes arranged at random locations on a board placed between the tester and the participant. The test starts with the tester initially tapping two of the blocks one after the other and then asking the subject to imitate the sequence. The sequence of taps gradually increases to a point at which performance breaks down.

The episodic buffer (Baddeley, 2000a), a component of WM, is a limited capacity temporary storage system. According to Baddeley (2007), “It combines information from the loop, the sketchpad, long-term memory, or indeed from perceptual input into a coherent episode” (p., 148). Moreover, it plays a role in interfacing between WM and long-term memory through the central executive, interacting phonological loop and sketchpad. It is also proposed that retrieval from the episodic buffer is through conscious awareness. However, no method of measurement has been proposed yet to assess the episodic buffer (Baddeley, 2007).

2.8 Alternative Models of Working Memory

Unlike Baddeley and his colleagues, who assumed a multi-component WM model, Engle and some other researchers (Engle, Kane & Tuholski, 1999; Conway & Engle, 1996) proposed a unitary, domain-free model of WM. This unitary model includes domain-specific codes and maintenance mechanisms such as a phonological loop. The ability of controlled attention is assumed to be the core of WM in this model. It is suggested that individual differences in this area explain learning outcomes more than a phonological loop or memory span capacity (Engle, Kane & Tuholski, 1999; Conway & Engle, 1996).

Waters & Caplan (1996) suggested that different levels of linguistic processing, particularly syntactic processing could not be adequately explained by Baddeley’s model and the measurement tools based on it. To clarify the point, they conducted research on both normal participants and participants with short-term memory deficits. Both groups were required to listen to and process sentences with relative clauses and garden path sentences. The results indicated no significant differences in the way of processing the sentences between the two groups. This is also supported by Robert and Gibson (2002), who argued that multiple resource pools are tapped in language comprehension, but not by span and loop tests developed and based on Baddeley’s model.

Ericson and his colleagues (Ericsson, 1996; Ericsson & Charness, 1994; Ericsson & Delaney, 1999) view WM as a portion of long-term memory, and for this reason they call it “long-term WM” rather than WM by itself. They established this model based on the expert performance

in different mental activities, such as playing chess. They suggest that the efficiency of long-term WM relies on knowledge and skills already stored in long-term memory. In their view, encoding, consolidation and retrieval of already learned items are considered to be very important cognitive functions. Based on this view, Ericsson and Kintsch (1995) conclude that the same basic processes might be involved in language learning and use.

Two relatively similar views of WM were proposed in the connectionist-oriented models by Schneider (1999) and O'Reilly, Braver and Cohen (1999). In Schneider's view, there are some modular processors, consisting of neuron-like units, in WM. These modular processors are activated by some short-patterns that are controlled by an executive. Unlike Daneman and Carpenter's (1980) and Just and Carpenter's (1992) view, which suggested that WM is limited by span capacity, Schneider believes that WM is limited by interference effects and the limits of the executive function. However, these are not long-lasting limits in WM. They diminish as the learner's expertise develops. This allows for slower, serial processing to be replaced by faster, parallel processing.

In O'Reilly et al.'s (1999) view, WM is an activated part of long-term memory. Thus, it is not an isolated construct located in a specific part of the brain. It is a system that is distributed over several parts of the brain, particularly in the prefrontal cortex and the hippocampus. Although encoding and retrieval of learned material are significant in this view, it does not play a significant role in learning outcomes.

To sum up, the connectionist-oriented models of WM proposed by Schneider (1999) and O'Reilly et al. (1999) are too different to be related to Baddeley's model of WM. This is because these models see no clear-cut boundary between WM and long-term memory. However, Engle and colleagues' as well as Waters and Caplan's models are, to some extent, compatible with Baddeley's model. For example, both Engle and colleagues' model and Baddeley's model benefit from a central executive and a phonological loop, although the phonological loop plays a minor role in determining learning outcomes in the former. Like Baddeley's model, Water and Caplan's model can be accommodated into the array of psychometric tools if more detailed measures of verbal WM are included in it.

2.9 Working Memory and Language Processing

Language processing involves active use of cognitive processes in WM. For example, in language processing for word recognition, links between orthographic and phonological information are activated. This is followed by the activation of appropriate semantic and syntactic resources. The orthographic and phonological information of a word activates all the words in the lexicon that have many of these visual and sound features. Word candidates begin to generate their meanings. These word candidates contribute orthographic, phonological, and semantic information to the process until the form which is the best is then accessed. Full information in that lexical entry then becomes available in WM. It is at this stage that word-recognition takes place (e.g., Grabe & Stoller, 2011). This word recognition process involves both registration of the linguistic item (e.g., Lee, 1998; Leow, 1997, 2001; Shook, 1994) and “making a connection between form and meaning” (VanPatten, 2004, p., 6). This means that a language learner notices a form and simultaneously determines its meaning. As described above, this registration or connection takes place in WM (Schmidt, 1990; VanPatten, 2004). WM is a limited capacity processing and storage system (e.g., Baddeley, 1986; Baddeley & Hitch, 1974). It is a workspace where the verbal input is processed and specific verbal material is maintained simultaneously. During processing, not only does the input need to be filtered and selectively attended to, but the activation and rehearsal of the information also needs to be controlled (e.g., Baddeley, 1986, 1997, 2000; Baddeley & Hitch, 1974). These are conducted by the central executive, which is a limited-capacity pool of attentional resources in charge of processing information (Baddeley, 1986; 1997). Attentional resources are allocated to the loops and episodic buffer by the central executive. This is to provide them with a specified amount of attention. Attention, according to Schmidt (1990, 1994, 1995, 2001), is required for noticing which is, in turn, a prerequisite condition for learning. Since attention, at any moment, is limited by WMC, there must be a close relationship between the amount of learning and the size of the WM (Sawyer & Ranta, 2001). This is also supported by the idea that some learners process input more effectively possibly because they have greater WMC (Skehan, 1998, p., 50). So far, there have been some studies arguing the role of WMC in processing and learning of both the L1 (e.g., Daneman, 1991; Daneman & Green, 1986) and the L2 (e.g., Ellis, 1996, 2005; Erlam, 2005; Harrington & Sawyer, 1992; Miyake & Friedman, 1998). A large body of research also suggests that PSTM could be a source of individual differences in processing information and learning the L1 (e.g., Adams & Gathercole, 1996; Gathercole & Baddeley, 1989; Gathercole & Baddeley, 1993) and L2 (e.g., Atkins & Baddeley, 1998; Brown & Hulme, 1992; Hummel, 2009,

Papagno, Valentine & Baddeley, 1991; Service & Craik, 1993). In the following two sections, studies on the role of WM and PSTM in second language learning will be reviewed respectively.

2.10 Working Memory and Second Language Learning

A considerable amount of research has linked WM to second language learning in areas including interactional feedback (e.g., Ando, Fukunaga, Kurahachi, Stuto, Nakano, & Kage, 1992; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Mackey, Adams, Stafford, & Winke, 2010; Shahnazari & Adams, in press) and reading comprehension (e.g., Alptekin & Erçetin 2009; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004).

Research on the role of WM in learning from L2 corrective feedback suggests that WM may shape, explain, and predict the way that learners respond to corrective feedback (e.g., Mackey et al., 2010; Trofimovich, 2007). Research conducted by Payne and Whitney (2002) suggests that learners with high WMC benefit more from feedback in face-to-face interaction, whereas those with low WM capacity benefit more from feedback delivered via computer-mediated communication..

Recent studies implicate WM in the production of modified output following interactional feedback. They suggest that individuals with higher WMC benefit more from corrective feedback and produce more modified output (e.g., Mackey & Sachs, 2011; Mackey et al., 2010), which in turn promotes L2 learning (e.g., McDonough, 2005; Swain, 2005). For example, Mackey et al. (2010) examined the relationship between learners, production of modified output and their WMC. In their study, a total number of 42 college-level, L1 English learners of Spanish participated in a dyadic task-based interaction. They were given opportunities to modify their erroneous utterances on a wide range of forms followed by a range of corrective feedback types such as recasts and clarification requests. They also completed a listening span test, as a measure of WMC, as well as an exit questionnaire, as an index for the level of the learners' involvement in the tasks. Mackey et al. suggested that learners with higher WMC produced more modified output following corrective feedback.

WM also predicts both linguistic accuracy on written post-tests and the amount of target-like modified output (Sagarra, 2007).

Shahnazari and Adams (in press) also investigated the relationship between modified output and WMC. In their study, a total of 56 L1 Persian EFL learners completed three WM measures, a reading span test, a math span test, and a non-word recognition task, as well as a grammatical judgment test. They also participated in a teacher-learner interaction task where they were given opportunities to modify their problematic utterances on simple present and past tense forms following corrective feedback in the form of simple clarification requests, as a type of elicitation. The results of their study, consistent with those of Mackey et al. (2010), indicated that learners with higher WMC, as measured by the reading span test and math span test, produced more modified output following oral corrective feedback.

In a recent study, Mackey and Sachs (2011) examined the relationship between WMC and interaction-driven learning with older adult learners. They recruited nine L1 Spanish adult ESL learners, four men and five women, ranging from 65-89 years of age, as their participants. The participants completed a listening span test as a measure of WMC, and a non-word recall test as a measure of PSTM. They also participated in some communicative tasks with native speakers of English, who provided corrective feedback in the form of recasts in response to their erroneous utterances on English question forms. They all completed a pretest, and three post-tests on the target structures, which also consisted of communicative tasks that elicited question forms. Mackey and Sachs (2011) found a significant correlation between WMC and L2 development. More specifically, they found that two thirds (6/9) of the participants indicated development in question formation on at least one post-test.

2.11 Phonological Short-Term Memory and Second Language Learning

A considerable body of evidence suggests that PSTM may be an essential cognitive mechanism underlying successful L2 learning (e.g., Atkins & Baddeley, 1998; Dufva & Voeten, 1999; French, 2006; Hummel, 2009; Hummel & French, 2010; French & O'Brien,

2008; Masoura & Gathercole, 1999, 2005; Papagno, Valentine & Baddeley, 1991; Papagno & Valler, 1992, 1995; Service 1992, Service & Kohonen, 1995). Many of these studies found an important role for PSTM in adult L2 vocabulary learning (e.g., Atkins & Baddeley, 1998; Papagno, Valentine & Baddeley, 1991; Papagno & Valler, 1992, 1995). For example, Atkins and Baddeley (1998) suggested that PSTM, as measured by digit-span or letter-span tasks, could explain individual differences in L2 vocabulary learning, as measured by 56 English-Finnish word pairs.

Studies with adults also indicated an important role for PSTM in L2 oral fluency production (e.g., O'Brien, Segalowitz, Collentine, & Freed, 2006; O'Brien, Segalowitz, Freed & Collentine, 2007). O'Brien et al. (2007) examined the relationship between PSTM and L2 fluency gains in native English-speaking adults learning Spanish. Their results suggested that PSTM, as measured by a serial non-word recognition test, could predict 4.5–9.7 % variability in L2 oral fluency gains, as measured by the speech rate and total number of hesitations or pauses in extracts from an oral proficiency interview.

Hummel (2009) found a significant relationship between PSTM and aspects of L2 proficiency other than L2 reading ability. She recruited 77 L1 French advanced ESL learners to complete a validated short form of the MLAT aptitude test, a non-word repetition task, and an L2 proficiency test including reading, vocabulary, and grammar sections. Her results indicated a significant correlation between PSTM and L2 proficiency in general ($r=.35^{**}$, $P < .05$), and PSTM and the vocabulary ($r=.36^{**}$, $p < .01$) and grammar ($r=.33^{**}$, $p < .01$) sections of the L2 proficiency test in particular, but no significant relationship between PSTM and L2 reading ability. Interestingly, the relationship between PSTM and L2 proficiency was strong for lower proficiency participants, but disappeared when proficiency increased.

However, Kormos and Sáfár (2008) found no significant correlation between PSTM and L2 proficiency. They investigated whether there is a relationship between PSTM and WMC and performance in L2 language skills, with an L2 proficiency test. They asked 121 secondary school students to complete a non-word repetition test, a Cambridge First Certificate Exam, and a backward digit span test after an intensive language training program. Their results

indicated that there was no significant correlation between PSTM and L2 language skills, but there was a significant correlation between WMC, as measured by a backward digit span test and L2 language skills (reading, listening, and speaking), with the exception of writing. Kormos and Sáfár (2008) suggested that PSTM and WM are distinct constructs, and play a different role in instructed second language acquisition.

In contrast to studies conducted with adult L2 learners, studies on the L2 learning of children have consistently shown a role for PSTM (e.g., Dufva & Voeten, 1999; French, 2006; Masoura & Gathercole, 2005; Service, 1992; Service & Kohonen, 1995). For example, in a longitudinal study that lasted for four years, Service (1992) examined the role of PSTM in English as a foreign language learning of 44 L1 Finnish primary school students. PSTM was measured through a pseudoword repetition task conducted each year of the study. In each task, the participants were required to listen to two lists of pseudowords, an English-sounding list and a Finnish-sounding one, and repeat aloud the pseudowords they heard as quickly as possible. Service (1992) found a strong relationship between PSTM, as measured by the English-sounding pseudoword lists, at the start of the English instruction and the performance on tests of listening, reading comprehension and writing 2.5 years later. She also suggested that PSTM underlies the acquisition of new vocabulary items in a foreign language.

In a follow-up longitudinal study, Service and Kohonen (1995) investigated whether the relationship between PSTM and foreign language learning is accounted for by vocabulary acquisition. They recorded 42 (9-10 year-old) Finnish participants' performance on pseudoword repetition, as a measure of PSTM, over four consecutive years. They also recorded the participants' performance on different individual L2 English tasks during the fourth year of the longitudinal study. These tasks measured the participants' L2 reading, listening, writing, vocabulary and knowledge of grammatical structures. Their regression analyses on pseudoword repetition and L2 tasks revealed significant correlations between pseudoword repetition and foreign language learning, even after a measure of general academic achievement had been partialled out. By varying second-step factors in their regression analysis, they were able to show that L2 vocabulary performance and pseudoword repetition accounted for the same variance in performance for foreign language measures. Service and Kohonen (1995) interpret these findings as an indication that PSTM influences

vocabulary learning, which in turn influences success in other areas of L2 performance. This data provides evidence of a specific relationship between PSTM (as measured by pseudoword repetition) and vocabulary learning.

Dufva and Voeten (1999) examined the effects of PSTM and native language literacy acquisition on learning English as a foreign language in a longitudinal study. A total of 160 Finnish school children were asked to complete measures of native language word recognition and listening comprehension in the first grade; word recognition, reading comprehension and PSTM in the second grade, and English skills in the third grade. Service's (1989) repetition task was used to measure the participants' PSTM.

Using the structural equation modelling, Dufva and Voeten (1999) found that both PSTM and native language literacy (word recognition and comprehension skills) could have positive effects on learning English as a foreign language. These skills accounted for 58% of the variance in the beginning stage of English proficiency. Dufva and Voeten (1999) suggested that diagnosing at-risk children and providing them with training in word recognition in their L1 may help to develop their L2 proficiency.

Furthermore, Masoura and Gathercole (2005) found an important role for PSTM in the L2 English vocabulary learning for Greek children. They investigated the contributions of PSTM and existing foreign vocabulary knowledge to the learning of new English words. Their L1 Greek children completed a paired-associate learning task as a measure of L2 vocabulary learning, two non-word repetition tasks as measures of PSTM, and a nonverbal ability task. Masoura and Gathercole (2005) found that PSTM made a large contribution to L2 vocabulary learning at earlier stages of L2 learning, but as the familiarity with L2 knowledge increased, the existing L2 knowledge played a mediating role in L2 vocabulary learning.

To summarize, the research conducted in various L2 contexts has linked PSTM to adults' L2 vocabulary learning (e.g., Atkins & Baddeley, 1998) and aspects of L2 proficiency such as grammar (e.g., Hummel, 2009). L2 studies with children have also reported a significant

relationship between PSTM and L2 vocabulary learning (e.g., Service, 1992; Masoura & Gathercole, 2005) as well as overall L2 achievement including reading and listening comprehension (e.g., Dufva & Voeten, 1999; Service & Kohonen, 1995). However, some other studies with adolescents and adults (e.g., Harrington & Sawyer, 1992; Kormos & Sáfár, 2008) found no significant relationship between PSTM and aspects of L2 learning.

2.12 The Role of Working Memory in Reading Comprehension

The research on the development of reading comprehension skills and sources of individual differences in comprehension indicates a strong relationship between L1 reading skills and cognitive variables such as WM (Just & Carpenter, 2002) and inhibitory control (Gernsbacher, Varner & Faust, 1990). Since WM is considered a mental workspace where the processes of retrieving, integrating, updating and revising of information are performed, it plays an important role in understanding a text. First, to identify the words, the reader needs to recode written symbols into phonological codes which are held in the phonological short-term memory to allow for computations to recognize linguistic structure. Then, he or she develops a coherent and integrated representation of the concepts by making links between successive sentences. This requires the reader to maintain the recently read material in the WM to make inferences (Schmalhofer, McDaniel, & Keefe 2002), while simultaneously processing the same or other information either recoded from the text or retrieved from the long-term memory. Finally, WM plays a role as a buffer. It is a limited capacity workspace for maintaining the recently read propositions in a text and the information retrieved from the long-term memory temporarily. This is to establish a local coherence between sentences and also to facilitate its integration into the activated background knowledge (Beech, 1997; Graesser et al., 1994; Ericsson & Kintsch, 1995).

However, comprehension may be impaired for two major reasons. First, since WMC is limited and can be overloaded by insufficient inhibitory processes, individuals may find it difficult to gain a full understanding of the text (Johnson & Barenès, 2008, p., 125; Baddeley, 2006, 2007). In other words, since inhibitory processes, controlled by the central executive, reflect a competition between what the memory follows to process and store and competing memory traces (Baddeley, 2006, p., 23), any deficiencies in their performance may diminish

comprehension. For example, some studies suggest that some individuals' weak reading comprehension is due to a deficiency in their inhibitory processes to control the irrelevant meanings of ambiguous words (Barnes, Faulkner, Wilkinson, & Dennis, 2004; Gernsbacher & Faust, 1991; de Beni & Palladino, 2000; de Beni, Palladino, Pazzaglia, & Cornoldi, 1998; cf., Cain, 2006). Second, poor comprehenders are believed to suffer from deficits in WM which may prevent them from retrieving or (re)activating information from long-term memory or within WM itself (Barnes, Huber, Johnston, & Dennis, 2007 cf., Johnson & Barnes, 2008, p., 128). Such deficits go beyond the short-term memory, phonological skills and vocabulary (Cain, 2006, Cain, Oakhill, & Bryant, 2004; Yuill, Oakhill, & Parkin, 1989).

2.13 Research on Working Memory and L1 Reading

A body of research suggests that there is a correlation between WMC and L1 reading comprehension (Carretti, Cornoldi, Beni & Romano, 2005; Daneman & Carpenter, 1980; Divesta & Dicintio, 1997; Waters & Caplan, 1996). Individual differences in reading comprehension and WM were first investigated among 20 university undergraduate native speakers of English who were enrolled in an introductory course in psychology by Daneman and Carpenter (1980). In this experiment, the participants took four tests: (A) a reading span test (RST) to measure the capacity of their WM, (B) a reading comprehension test that measured the subjects' ability to answer the questions about the facts and pronominal references, and (c) traditional simple span tests (a word and a digit span test).

Reading span and word span tests, as opposed to digit and letter span tests, tax both the processing and storage functions of WM. Daneman & Carpenter suggested that a potential source of individual differences in reading comprehension could be the amount of trade-off between processing and storage in WM. They also assumed that better readers benefit more from efficient processes because they can omit the intermediate steps (e.g., decoding, lexical accessing, parsing, and inferencing), thus placing fewer computational demands on WM. This allows them to have more WMC for storing, rehearsing and consolidating the essential outcomes of the reading process.

To measure WMC, Daneman & Carpenter's subjects read increasingly longer sets of sentences aloud at their own pace and recalled the last word of each sentence at the end of the set. The subjects' reading span score was based on two criteria: a) the level at which the subject was correct on two out of the three sets, and b) a word span test where he or she was required to recall increasingly longer sets of individual one-syllable common nouns. These common nouns were semantically and phonetically unrelated. The subjects' reading comprehension was examined through the verbal SAT and a series of passages asking two types of questions, asking about facts and pronominal reference (P., 454). A high correlation ($r = .72, .90, .59$) was found between the reading span test and all three reading comprehension measures. Daneman and Carpenter interpreted this as evidence that WM played an important role in reading comprehension.

Waters and Caplan (1996) presented two criticisms of Daneman and Carpenter's (1980) reading span test (RST). First, they claimed that it was not able to measure the amount of semantic and syntactic processing because it did not include semantic and syntactic acceptability judgments. Second, WMC was measured solely through recall, ignoring the possible trade-off between the components of the task. In other words, Daneman and Carpenter's (1980) subjects may have focused more on remembering words rather than processing sentences. Thus, Waters and Caplan (1996) argued that it would be difficult to interpret a WM score based on just the number of words recalled if there were trade-offs.

To remove these short-comings, Waters and Caplan (1996) proposed a computerized version of the RST. In this test, the participants' WM scores were based on three measures: (a) the mean reaction times for correct responses on the acceptability of the sentences, (b) the number of errors on the sentence judgments, and (c) the number of trials in which sentence-final words were incorrectly recalled. After applying correlation analyses to these three components, they found that a number of participants were trading speed for accuracy and speed for storage. This was based on some significant negative correlations found between the reaction time on the one hand and the recall, as well as sentence judgment errors on the other hand. Then based on these results, they questioned the reliability of WM measures that only consider scores from the recall component of the reading span test. Thus, they used a composite WM score by converting the scores for each measure, including the reaction time

for sentence processing, into a z-score, and then averaging the z-scores into one composite. Using test-retest reliability, they gained the reliability of 0.75 (for cleft subject sentences) and 0.83 (for subject-object sentences), compared with 0.41 for the Daneman and Carpenter test. Finally, Waters & Caplan (1996) concluded that considering sentence processing components, consisting of the reaction and sentence judgment, in the composite score decreased “trade-offs across tasks which contribute to the variable performance on the recall measure” (p., 75).

Divesta and Dicintio (1997) investigated the interactive effects of WM span and content on reading comprehension and retrieval. They argued that assigned perspectives (contexts) facilitate comprehension and retrieval by providing structures for: (1) activating relevant and related information, (2) specifying the focus of attention (3), specifying what information is to be selected, (4) relating new information to prior knowledge, and (5) facilitating the identification of word or passage meanings, conceptual relations and propositions. Their findings suggested that the assignment of a reading perspective enhances the recall of both relevant and irrelevant encoded information to a significantly greater extent than when this is not the case. Furthermore, high WM span readers benefit more from external support in the form of assigned perspectives than low WM span readers. That is because the additional processing demands of the switch in perspectives significantly hinder the recall of low WM span readers, although they depend on assigned perspectives at retrieval.

The ability of good and poor comprehenders to update information in WM was investigated by Carretti et al. (2005). Updating in WM is defined as the process through which the content of WM is modified to accommodate the new input (Morris & Jones, 1990). It is also an executive function through which WM is involved in psychological processing to meet the on-line request (Carretti et al., 2005). The reader is continuously involved in the process of maintaining important information and omitting irrelevant information (Gernsbacher et al., 1990). Carretti et al. (2005) hypothesized that selecting and updating information distinguishes individuals with efficient WM processes from the inefficient ones. A total of 109 poor readers and 109 good readers listened to a list of words, including nine concrete nouns. These concrete nouns were all displayed in the column of pictures. At the same time, the participants were required to look at a column of pictures and recall the highest or lowest pictures in the column named in the word list. They were required to recall 3 items and to

exclude 3 items immediately (immediate intrusion) and 3 items later (delayed intrusion). Poor comprehenders made more delayed intrusions than did good comprehenders, implying poor comprehenders' weaknesses in suppressing irrelevant information. Moreover, Swanson and Ashbakar (2000) suggested that individuals with some weaknesses in reading comprehension might suffer from deficits in WM. In addition, poor comprehenders are not able to perform on a task that requires the selection of relevant information and inhibiting irrelevant information efficiently. In other words, an insufficient inhibitory mechanism is attributed to poor comprehenders since they are not able to control activated information (Carretti et al., 2004; De Beni et al., 1998).

A significant correlation has been reported between WM resources and higher level skills in reading comprehension such as anaphor resolution and comprehension monitoring (Cain, 2006). For example, Yuill and Oakhill (1988) found that readers with low WM failed to work out all types of anaphor used in a narrative story. This may suggest that readers with low WM do not have sufficient attentional resources to maintain the representations of different kinds of anaphor long enough to be associated with their antecedents. This study is in line with Daneman and Carpenter's (1980) finding that adults with low WMC fail to establish a link between the anaphor and its antecedent, particularly when they appear in non-adjacent sentences.

The research with 7-8 year olds also indicates a link between WM and reading comprehension through the ability to infer the meaning of new words (Cain, Oakhill, & Lemmon, 2004). Moreover, WM is argued to be significantly associated with children's ability to monitor their comprehension by detecting deliberate inconsistencies in texts such as contradictory or scrambled sentences. For example, poor comprehenders' performance in monitoring processes and identifying some deliberate inconsistencies in texts is impaired when the WM demands of the task are high (Cain, 2006, p., 80-82).

However, WM is not the only determinant of children's reading comprehension ability (Cain, 2006, p., 85). Training programs can successfully influence children's comprehension without targeting WM resources. For example, Yuill and Joscelyne (1988) found that

teaching poor comprehenders to look for clues, while reading, enables them to make a larger number of inferences. Furthermore, familiarizing the children with the nature of the errors in the text as well as the use of mental imagery when reading enhances their ability to monitor comprehension (Paris et al., 1991).

2.14 Working Memory in Child and Adult L1 Reading

The study of children's WM and L1 reading is different from that of adults. This is because children's WM is in a developing state during early and middle childhood (Gathercole, 1998). Furthermore, a very different relationship between WM and reading comprehension in children is suggested. Several studies indicate that the relationship between WM and reading comprehension is direct for adults, whereas it is mediated by either phonological short-term memory or basic reading and language skills, or even both in children (e.g., Hatcher & Hulme, 1999; Stanovich, Nathan, & Vala-Rossi, 1986). For example, it is argued that any relationship between reading comprehension and the verbal WM will be mediated by word reading ability (Perfetti, 1985). This will be predicted by the verbal efficiency hypothesis proposed by Perfetti (1985, 1994). This hypothesis accounts for reading comprehension problems by using the relationship between word reading ability and reading comprehension. Assuming WM is a limited capacity system, the proponents of this hypothesis argue that the processes of reading words compete with the processes for comprehending text for the same amount of processing resources (Perfetti, 2007; Perfetti & Hart, 2001, 2002). When a child has an inefficient word reading ability, a greater deal of processing capacity may be allocated to reading words on the page, and consequently there would be an insufficient amount of processing capacity to figure out both the lower level of comprehension processes such as the relationship between words, phrases, propositions and sentences, and higher level processes such as making inferences and coherent representation of the meaning of the text (Cain, 2006; Perfetti, 1985, 1999, 2007; Perfetti & Hart, 2001, 2002). Because items are maintained in WM temporarily, and they fade away after a short while, stored information will have been lost by the time children with slow word reading ability resort to processing and integrating it completely, leading to deficiencies in their reading comprehension (Cain, 2006; Perfetti, 2007).

The other mediator in the relationship between WM and children's reading comprehension is phonological processing skills such as non-word repetition and phonemic awareness. The literature indicates that they are the best predictors of word reading ability, which is in turn highly correlated with reading comprehension (Hatcher & Hulme, 1999; Stanovich et al., 1986). The children with deficient phonological processing skills are not able to maintain a sound-based representation of the text they are reading in their WM. This leads to them experiencing difficulties when processing language because heavy demands are placed on their verbal WM resources (Cain, 2006; Gathercole, 1998).

Semantic knowledge is also an influential factor in the WM relationship with reading comprehension. The number of word meanings known by a child is highly correlated with his reading comprehension level (Carroll, 1993). This is because individuals with superior semantic knowledge are better able to activate representations of words from long-term memory and employ them to process verbal information in WM (Nation, Adams, Bowyer-Crane, & Snowling, 1999). Thus, knowing a large number of word meanings supports verbal WM and consequently leads to better comprehension.

Moreover, higher level skills such as making inferences, integration, anaphoric processing, use of context and comprehension monitoring may affect the relationship between WM and reading comprehension processes, and deficiencies in any of these skills can impact on this relationship (Cain, 2006; Cain & Oakhill, 2004). This is because these skills are involved in the construction of the representation of a text's meaning. To do this, newly read ideas are employed to update the mental representation of a text, which requires both new and old information to be active, thus implying an important role for WM.

There are children with good word reading, sight vocabulary and phonological processing skills but reading comprehension deficits (Cain, Oakhill, & Bryant, 2000a). Yuill et. al., (1989) suggested that these children suffer from WM deficits. It is argued that any WM impairments associated with this population's reading comprehension problems have not originated from lower level-word reading and vocabulary deficits (Oakhill, Cain, & Yuill, 1998). This implies such deficits could arise from higher-level skill deficits.

Weak inhibitory processes may be the source of WM deficits. This occurs when regulation of the activation of the content of memory is not conducted efficiently. In other words, maintaining the activation of the just-read important information and preventing the entry of irrelevant information into the WM are not efficiently adjusted. This account indicates a direct link between WM and reading comprehension (de Beni & Palladino, 2000; de Beni et al., 1998). Since constructing the meaning of the text requires efficient maintenance, storage, and updating of information, any inhibitory deficits in WM may lead to poor reading comprehension (de Beni & Palladino, 2000; de Beni et al., 1998).

2.15 Working Memory and Studies in the L1 and L2: An Overview

Individual differences in WM influence language learning by constraining noticing (Schmidt, 1990; VanPatten, 2004), which is a precondition for learning. Noticing is also controlled by attentional resources in WM. During the last decade, several studies have investigated the role of WM, which is often operationalized as Daneman and Carpenter's (1980) reading span test, in a number of L1 and L2 processes. They include individual differences in WM on one hand and L1 reading comprehension (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Waters & Caplan, 1996), L2 word naming and vocabulary learning (Kroll et al., 2002; Atkins & Baddeley, 1998), L2 online parsing performance (Juffs, 2004), and L2 grammatical rule learning (Miyake & Friedman, 1998; Robinson, 1995, 2002; Williams & Lovett, 2003) on the other hand.

A few studies have investigated the role of WM in L2 reading processes. They have assessed the relationship between WMC and L2 reading comprehension (Alptekin & Ercetin 2009; Chun & Payne, 2004; Gholamain & Geva, 1999; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004). Additional studies suggest WM is independent of language and can be measured in both the L1 and L2 (Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993).

2.16 Working Memory and L2 Reading

Seven prior studies on WMC and L2 reading comprehension have been carried out. The first study was conducted by Harrington and Sawyer (1992). They selected 34 Japanese native

speakers learning English as a foreign language at intermediate and advanced level as their subjects, and asked them to take three memory tests in their L1 and L2 as well as measures of L2 reading comprehension. The memory assessment consisted of digit span, word span, and reading span tests (RSTs). The L2 reading comprehension measures consisted of the grammar and reading sections of the TOEFL and a cloze passage. They found a significant, strong correlation between WMC (L2 reading span) and both the TOEFL reading test ($r=.54$) and TOEFL grammar test ($r=.57$). Furthermore, there was a weak correlation between the L2 reading span and cloze passage, ($r=.33$). However, no significant correlations were found between the digit span and word span measures on the one hand and the L2 English reading comprehension on the other.

Gholamain and Geva (1999) examined the role of WM, speed of letter naming, and global L2 oral language proficiency in understanding individual differences in the concurrent emergence of basic reading skills. A group of 70 children in Grades 1-5 learning to read concurrently in English (L1) and Persian (L2) was selected as the participants in the study. Then four measures including WM, letter naming speed, word recognition and word attack skills were administered, each in both languages. The results of the study indicated that there were significant correlations between parallel basic reading skills (such as word recognition and word attack skills) and cognitive skills (such as WM and letter naming speed) in English and Persian. In particular, WM and letter naming speed proved to be significant predictors of word recognition and word attack skills in both languages. It was also revealed that WM and speed of letter naming in the L2 were more stable predictors of both isolated word recognition and word attack skills in the L1 than were the parallel measures in English.

However, Walter (2004) examined the question of whether the transfer of the reading comprehension skill from the L1 to the L2 is linked to the development of verbal WM in the L2, which turned out to take place at a much lower level of L2 proficiency than that found by Harrington and Sawyer (1992). Two groups of L1 French ESL learners participated in her study. The first group consisted of 19 lower-intermediate ESL learners, while the second group consisted of 22 upper-intermediate ESL learners. Three measures were administered by Walter (2004), each in both languages (French and English): 1) a baseline comprehension assessment where the participants were required to complete a gapped summary of the story

they had just read, 2) a pro-form resolution test where the participants were told to read a story and stop when they encountered an expression printed in red, then read the word aloud, give the meaning of the word, and identify when it was first mentioned, 3) a verbal WM measure where the participants were asked to read an increasingly longer sets of sentences and judge if they were logical or illogical and then recall the sentence-final words across the sets.

The results indicated a significant correlation between the WM scores and the L2 summary completion scores. However, the correlation was higher for the lower-intermediate group (.79, $P < .0001$) than for the upper-intermediate group (.46, $P < .01$). This implies that the lower-intermediate group's success in the summary completion tasks relied significantly on their WMC. This supports the idea that there is a link between the development of verbal WM in the L2 and success in L2 reading comprehension. This study also revealed that success of the upper-intermediate group in L2 reading comprehension relied more on reading skills (the ability to build well-structured mental representations of texts) than on WM.

In the fourth study, Chun and Payne (2004) examined the role of individual differences in the L2 German reading comprehension and vocabulary acquisition of 13 L1 English students in a second year German language course. A computer-delivered version of Daneman and Carpenter's (1980) L1 reading span test as well as a non-word repetition task were used to measure WM. A German short story, including four sets of comprehension exercises followed by a recall protocol, was used as a measure of reading comprehension. The results indicated a strong relationship between phonological WMC as measured by word recognition based on non-word repetition and look-up behavior, measured as the number of annotations which had been looked up and recorded while reading an L2 text. Learners with low phonological short-term memory capacity looked for an average of three times more words than participants with high phonological short-term memory capacity. However, they did not report any significant findings for WM on any of the comprehension or vocabulary acquisition measures.

In the fifth, Lesser (2007) conducted research on the effect of topic familiarity and WMC on beginning Spanish learners' reading comprehension and their processing of future tense morphology. 94 high beginner L2 Spanish learners enrolled in an accelerated, elementary Spanish course at a large public university were chosen as the subjects. The subjects completed a computerized version of an L1 reading span test as a measure of WMC, a recall protocol task to measure passage comprehension, and form recognition and tense identification tests to determine the subjects' processing of future tense morphology. The results of the study suggested that topic familiarity was an important factor in L2 reading comprehension as it played a significant role in beginning L2 readers' recognition of target forms and their ability to make form-meaning connections. WM also played a significant role in learners' comprehension and processing of grammatical form, depending on the extent to which it interacted with learners' prior knowledge. In other words, a more significant role of WM in reading comprehension was observed as the participants' prior knowledge of the text topic increased.

In the sixth, Payne, Kalibatseva and Jungers (2009) examined whether domain experience compensate for WMC in second language comprehension. They asked 73 L1 English college students learning Spanish as a second language to report on their experience on L2 learning (e.g., number of Spanish courses and number of years actively learning L2), complete a measure of WM as well as measures of L1 and L2 reading. They used an operation span test to measure WMC. In this test, the students were required to count dark blue circles surrounded by some distracters (dark blue squares and light blue circles), repeat the total number of dark blue circles aloud, and remember this number for later recall at the end of a set. They operationalized WMC based on the total number of correct sets with complete correct recall. To measure L2 reading comprehension, they used six short passages in L2, each with a different difficulty level from the others. In this test, the students were required to read each passage and answer to 4-5 multiple-choice questions which assess their inferences on the text. To measure L1 reading comprehension, these researchers used *The Air Force Officer Qualifying Test*. This test was similar to L2 reading test they used and measured the students' inferences and reasoning. They found a significant correlation between L2 reading comprehension and WMC ($r = .24, P < .05$), L1 reading comprehension ($r = .54, P < .01$), and L2 experience ($r = .52$, & $r = .58, P < .01$ for number of years and courses respectively). They also found that WMC made no further contribution to the variance in L2 reading when L1

reading was already in regression equation. Based on these results, the authors suggested that WM has an indirect relationship with L2 reading (through a mediating role of L1 reading). L2 reading is influenced independently by cognitive ability and experience with the language. Moreover, as expertise in a domain develops, at some point experience could compensate for WMC differences. However, this study did not specify precisely at which point and how this could take place, implying the need for further research to be done.

In the most recent study of WM and L2 reading, Alptekin and Ercetin (2009) investigated the relationship between L2 reading ability and WMC. In their study, 30 L1 Turkish undergraduate students in English language teaching courses were required to complete two WM measures and a reading comprehension test. WM measures included two RSTs, one with recall-based and the other with recognition-based items designed to measure storage. In the recall-based RST, the participants were required to remember the sentence-final words at the end of the sets and enter them into a textbox on a computer screen. In the recognition-based RST, the participants were required to choose the sentence-final words from a list of three given options. The reading comprehension test was a narrative text followed by 20 multiple-choice questions. Of 20 multiple-choice questions that were presented in a mixed order, half of them were textually explicit (about information stated in the text) and the other half of the questions were implicit and required processing for either local or global coherence. The explicit and implicit questions were designed to measure the participants' literal and inferential understanding of the text respectively. The results of their study indicated a moderately significant correlation ($r=.40$, $P<.05$) between the WMC scores, as measured by the recall-based RST, and L2 reading ability scores, as measured by just the inferential comprehension questions. They did not find any significant correlations between WM, as measured by the recognition-based RST and L2 reading ability.

These studies provide preliminary evidence that, as with L1 reading comprehension, WM plays a role in L2 reading comprehension. However, there are two major limitations associated with these studies. First, the administration and scoring of the reading span tests were different across these studies. For example, the maximum number of words recalled for at least three sets was used as a WM score by Osaka and Osaka (1992) and Osaka et al., (1993), while the total number of words recalled was utilized as a WM measure by

Harrington and Sawyer (1992). Second, all the studies, except for Lesser (2007), have ignored the possibility that individuals trade off the components of a reading span test (Waters & Caplan, 1996; Friedman & Miyake, 2004) in calculating the WM score.

L1 reading research has indicated that the effect of WM on reading is most apparent in early stages of literacy development (e.g., Daneman & Carpenter 1980; Just & Carpenter, 1992; Waters & Caplan, 1996). It is possible that this same pattern exists for second language reading as well. However, unlike L1 readers who are able to communicate orally at the time of reading and possess sufficient L1 knowledge, L2 readers may differ in terms of their level of L2 knowledge, and this may impact the involvement of WM in L2 reading. Due to divergent research findings reported above, it is not yet clear to what extent individual differences in WM make a role in explaining variability in L2 reading, and whether this role is mediated by L2 proficiency or not. Thus, more studies which track the effect of WM on reading comprehension at different proficiency levels are needed. Overall, more research is needed to understand how WM influences the development of L2 reading comprehension.

2.17 Rationale of the Study

Because the role of WM is emerging as an area of concern in second language learning (e.g., Miyake & Friedman, 1998; Robinson, 1995, 2002, 2005), research is needed to clarify the role of WM in second language acquisition, and in particular in L2 reading comprehension. As a result of the few WM studies which have been conducted, evidence is emerging of the important role played by WM in L2 reading comprehension. However, these studies show divergent research findings on the role of WM in L2 reading comprehension. Furthermore, there are limitations associated with the testing and scoring of WM in several studies. There has been little attention paid to the role of WM in reading comprehension across language proficiency levels. Similarly, prior L2 studies indicate divergent research findings on the role of PSTM in L2 reading comprehension. Some of these studies suggest that there is a significant relationship between PSTM and L2 reading ability (e.g., Masuora & Gathercole, 2005; Service, 1992, Service & Kohonen, 1995), whereas some others indicate no significant relationship between PSTM and L2 reading ability (e.g., Hummel, 2009; Harrington & Sawyer, 1992; Kormos & Sáfár, 2008).

These limitations and differences in research findings point to the need to examine the relationship between WM and PSTM and L2 reading ability across proficiency levels. Thus, the current study was designed and proposed the following research questions to investigate whether WM and PSTM influence L2 reading ability at different levels of proficiency.

2.18 Research Questions

This study has been designed to address the following questions:

1. Is there a relationship between WMC and L2 reading ability?
 1. a If so, does this relationship differ according to proficiency level?
2. Is there a relationship between PSTM and L2 reading ability?
 2. a If so, does this relationship differ according to proficiency level?

Chapter Three

Methodology

Section One

3.1.1 Introduction

This chapter will explain the methodology used to conduct this study in three sections. In section one, the procedure used to select the participants and place them in three proficiency groups will be described. In section two, the stages of the material development for the study including memory and reading measures will be discussed. It also involves the description of pilot studies for each measure. Finally, the design of the study as well as the procedure used for data analysis will be described in section three.

3.1.2 Subjects

A total number of 140 L1 Persian EFL learners, at beginning (56), intermediate (43) and advanced level (41) participated in the study. They included both males and females, 16-35 years old, studying English as a foreign language in a private language school in Iran.

In the Iranian curriculum, most of the L2 learners who go through private language school courses are generally more proficient than their peers in public schools in the four L2 (English) skills, particularly in reading comprehension. This is indicated by their scores in the reading section of their final exams taken at the end of each semester. This is because they are continuously encouraged to read and are exposed to a lot of supplementary reading material by their teachers. The language school where this study was conducted is one of the most successful private English language schools in Iran. In this school, L2 learners start learning English as a foreign language from a very basic level and have the opportunity to go through to the intermediate and advanced levels based on their final exam results. At the time of conducting this study, the total population of the L2 learners enrolled in this school was around 2500. They included groups of children, teenagers, and adults with different academic and L2 background knowledge. Fifty English teachers were the school's academic staff, who had been selected based on their academic merits and job experience. The school was

equipped with a language lab, a library, and supplementary electronic resources such as L2 instructional CDs and movies available to all language learners. All L2 learners had the opportunity to communicate in English in the language school's environment as well as in their classes with their peers. However, English was generally not spoken in their social life.

Identification of the proficiency level of the participants was already done upon their enrolment. It was based on the New Interchange / Passages Placement test (Lesley, Hansen, & Zukowski/ Faust, 2003), a test used in the language school where the participants were selected. This test consists of three sections, including a placement objective, a placement conversation, and a placement essay. The placement objective measures the participants' listening and reading comprehension as well as their grammatical knowledge of English. The placement conversation and the placement essay measure the participants' speaking and writing skills respectively. To obtain the participants' proficiency level for the placement in an appropriate group, the average performance of the participants' on these three sections is obtained and then rated on a scale from 1 to 12. In this scale, every four ratings indicate a proficiency group. Thus, the first, second and third groups of four ratings represent the beginning, intermediate and advanced proficiency groups respectively. Each rating also indicates the recommended course level each participant should start with. Each New Interchange / Passage level corresponds to a different rating. For example, when a participant is given an overall rating of 9, this means that he / she could be placed in the advanced group, and it is recommended that he/she starts his / her course with the first half of Passage level 1.

Furthermore, a modified version of the Objective Placement test was administered to participants of this study to ensure that each participant was placed at the right proficiency level. The objective Placement Test is a 70-item multiple-choice test that lasts 50 minutes. However, to obtain an accurate proficiency score and not to allow the variable of time constraint to influence the results, an additional 40 minutes were given to the participants. Overall, 90 minutes were given for this test, although most of the participants completed it within the range of 60 to 80 minutes.

This test measures L2 learners' listening, reading, and grammar recognition of English. With the exception of two participants, the results were consistent with the original proficiency placement. To score the test, one mark was allocated to the participants' correct answer (total of 70). The participants were placed at an appropriate proficiency level depending on their scores in this test. Scores of 30, 49, and 70 were determined as the cut-off scores for the beginning, intermediate and advanced levels respectively. A one-way ANOVA was also run to see if there were any significant differences between the assigned groups. The results of this analysis indicated that there were significant differences between the three proficiency groups ($F(2, 137) = 512.952, p = .000$). More specifically, the one-way ANOVA analysis indicated that each proficiency group differed from the other two groups significantly and with a large effect size ($\eta^2 = .88$). Table 3.1 indicates the descriptive statistics of this analysis for each proficiency level and how the participants were distributed into three levels of proficiency based on the Objective Placement Test scores.

Table 3.1

Descriptive Statistics for the Language Proficiency Scores for each Proficiency Level

	Beginning	Intermediate	Advanced
N	56	43	40
Gender	M= 21, F= 35	M= 14, F= 29	M= 5, F= 35
Age Range	16-27	16-31	18-35
Years of Study	.5-1	1.5-2	3-3.5
Min	6	31	50
Max	29	49	65
Mean	20.08	39.06	55.68
S.D	5.70	5.60	4.88

Note. M= Male; F= Female

It should be noted that although the same proficiency test and memory measures were administered to the participants in this research, the reading measures were not the same for each proficiency level. Although consistent in type, different reading measures were created for each proficiency level because it was not possible to narrowly discriminate between levels of reading ability within each proficiency level using a general test appropriate for all. For this reason, this study should be considered as three independent studies at the beginning,

intermediate and advanced levels respectively. However, to avoid unnecessary repetition and for the sake of clarity, the results and implications of these studies are usually presented, described, explained and discussed together. In this way, we can compare the findings for each proficiency level and better consider the relationship between WM and L2 reading ability at the three proficiency levels. For the purpose of clarity and simplicity, the terms “beginning”, “intermediate”, and “advanced” are used throughout the thesis for the representation of these three independent studies.

Section Two

Materials

A battery of tests of memory and a reading comprehension of 4-5 hours in duration were used. Memory measures included a reading span test, a math span test, and an English non-word recognition task. The reading measures included a cloze test, two short-answer reading comprehension passages and two L1 recall reading tests.

3.2.1 Memory Measures

3.2.1. 1 Reading Span Test

Reading span tests were first introduced by Daneman and Carpenter (1980). They were used to measure WMC and give an index of processing and storage, the components of WM. In a reading span test (RST), participants are asked to read sets of sentences, report on the semantic acceptability of each sentence (processing assessment), and then recall the final word of each sentence when prompted (storage assessment). This type of test has been used in several studies as a measure of WMC (e.g., Babcock, 1991; Chun & Payne, 2004; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Lesser, 2007; Light & Anderson, 1985; Osaka & Osaka, 1992; Salthouse & Frisk & Milner, 1990; Swanson, 1993).

In this study, a Persian reading span test was used to measure WMC. Prior research indicates that WM is language independent (Miyake & Friedman, 1998; Osaka, Osaka & Groner, 1993; Osaka & Osaka, 1992). Therefore, WM can be measured in the L1, and measuring WM in the L1 helps to avoid conflating WM and L2 proficiency. Thus, the language independent nature of WM as measured by tests like the RST more precisely pertains to Verbal WM (Juffs & Harrington, 2011). Therefore, this study used a Persian reading span test. This test was developed by the researcher, an L1 Persian speaker, and piloted on 74 Persian EFL learners at three proficiency levels.

The test was designed with 64 items. For each item, the participants were required to judge whether the sentence made sense or not and also to remember the final word. After sets of 3, 4, 5, or 6 sentences, the participants were asked to orally recall the final words.

To identify the potential problems with the RST, three pilot studies were administered to three different groups of L1 Persian EFL learners. In the first pilot, a group of 12 participants completed the RST, followed by a retrospective report. In their retrospective report, they all reported that the transition time, 6 seconds, for each slide was too short to read through the sentence. They also wrote that a few sentences were too vague for them to determine whether they made sense or not. The results of an item analysis indicated that there were some poor test items in the test. They were identified as being too difficult. These results indicated that the participants had performed poorly on both the processing and recall components. The sentences which the students had identified as too vague were located among the ones which had been identified as too difficult by the item analysis. In consultation with a Persian language expert, these sentences were either revised or replaced with new sentences. Then the transition time for each slide was increased to 8 seconds as well.

In the second pilot study, similar to the procedure in the first pilot study, a group of 18 L1 Persian EFL learners completed the revised RST followed by a retrospective report. In their retrospective report, they wrote that they had had sufficient time to read through the sentence on each slide and even rehearse the sentence final words (target). They also reported a case where two sentence final words were semantically related, and they had been able to make an

association between them for better recall. The results of this study supported the participants' claims. Their performance on the RST was better than the prior group's. Most of them were also able to obtain the scores for the two semantically related targets. Since the participants' rehearsing could have inflated the recall scores, the transition time for each slide was decreased to 7 seconds. Furthermore, one of the sentences including a semantically related word was replaced with a new sentence including a different target word. The new sentence was developed and proposed by the same Persian language expert.

In the third pilot study, the revised reading span test was administered to a group of 44 participants. They reported that the transition time for each slide was just enough to read the sentence through and decide whether it made sense or not. No one reported any opportunity for rehearsing the targets. Moreover, they believed that all sentences throughout the test had been neither too easy nor too difficult for them. The results of the item analysis also indicated that each item made a good contribution to the test. The internal reliability for this test, as indicated by Cronbach's Alpha, was .834 & .737 for the RST recall and processing respectively. This was consistent with the results of the main study where the internal reliability for this measure, as indicated by Cronbach's Alpha, was .844 and .790 for the RST processing and recall respectively.

The final test included 64 sentences, 10 practice session sentences and 54 test sentences, all of which were in an active and affirmative form within a range of 13-16 words. Half of the sentences were constructed as 'nonsense' sentences. This was done by rearranging a few content words in such a way that sentences were syntactically possible, but semantically anomalous (Chun & Payne, 2004; Harrington & Sawyer; Lesser, 2007; Turner & Engle, 1989; Waters & Caplan, 1996). This was to make sure that the participants processed sentences for meaning without focusing only on the retention of recall items. This test was administered individually using a computer-based format. Because Persian sentences follow SOV syntax (the sentences initiate with a subject followed by an object and a verb respectively), each sentence ends in a verb, similar to the reading span tests in Japanese (Osaka & Osaka, 1992) and German (Osaka et al., 1993; Roehr & Ganem-Gutierrez, 2008). Each verb appeared only once in the test. Therefore, the final words in this test were 64 different three syllable verbs. The verbs in each set were not semantically related. The sentences in the test were arranged

in three sets of 3, 4, 5, and 6 sentences. Half of the sentences in each set were nonsense. Each sentence appeared on screen for 7 seconds, when the computer transitioned to the next slide. After each set, a slide with 3 hash keys and a two-second auditory prompt appeared. This was to signal to the participants to recall the final word of each sentence in the set.

The test was in PowerPoint format and was taken individually. It assessed two WM components, processing and storage (e.g., Chun & Payne, 2004; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Lesser, 2007; Waters & Caplan, 1996). The participants had to read each sentence aloud, judge whether or not it made sense and say their judgment aloud while their answer was recorded. This was the measure of WM processing, and was intended to ensure that they actually engaged in processing the text and did not simply rehearse the targets. They also had to remember the last word of each sentence up to the end of the set until a visual prompt (three hash keys) along with a two-second auditory prompt appeared on the computer screen. The pilot study results suggested that these two simultaneous prompts could well put a clear boundary between the sets and help the participants not to miss the recall time. At this time, the participants had to recall the sentence-final words and say them out loud while their answers were recorded by the researcher. This was the measure of the WM storage component. To control the recency effect, the participants were required to recall the words in the order in which they appeared (Baddeley & Hitch, 1993; Waters & Caplan, 1996).

A test instruction guide followed by an oral explanation which included an example set of three sentences was given to the participants prior to the test. Then they were given a practice session consisting of 10 sentences in two sets of three and one set of four sentences. Then the test began with a set of 3 sentences, and as the test progressed, the number of sentences presented on each trial increased successively from three to six, with three trials being presented at each series length. The prompt slide transitions increased accordingly from 12 to 18 seconds based on the length of each set.

To score the test, one mark was allocated to the participants' correct judgment and one mark to their correct recall of the test session items, with the total of 54 each. Thus, since there were 54 sentences across all the trial sets, the range of the participants' processing and recall scores was between 0 and 54 for each participant. No marks were given to the practice session items. This was consistent with the scoring method in recent studies (e.g., Alptekin & Erçetin, 2009, 2010). A composite WM score was used as an indicator of the participants' WMC (e.g., Lesser, 2007). The composite WM was obtained by adding the processing and recall z-scores. This is a more reliable scoring method of WMC compared to the traditional span scores that quantify the highest set size completed or the number of words in correct sets (Friedman & Miyake, 2005).

3.2.1.2 Math Span Test

In this study, in order to measure WM independently of reading, a math span test was developed as the second memory measure. This type of test was first developed by Turner and Engle (1989) to measure WMC. They called it an operation span test. In this test, a set of simple arithmetic equations such as $(6/2) + 5 = 8$, $(3 \times 4) - 5 = 7$, and $(4/2) + 2 = 6$ are presented to the participants. For each equation, the participants' task is to verify whether the stated solution is correct or incorrect (processing assessment), and at the end of the set, they have to recall the stated solutions from each equation in the set (here, 8, 7, and 6) (storage assessment). The number of arithmetic problems on each set is successively increased from one to seven, with three sets being presented at each series length. The total number of stated solutions recalled from the perfectly recalled set is regarded as the participant's math span. This test has been used as a measure of WMC in several prior studies (e.g., Daneman & Merikle, 1996; Mizera, 2006; Hambrick & Engle, 2002; Robert & Gibson, 2002; Salthouse & Babcock, 1991; Turner & Engle, 1989). Further support for the use of operation span test, as a reliable measure of WMC is provided by a recent study in cognitive psychology (Sanchez, Wiley, Miura, Golflesh, Rioks, Jensen & Conway, 2010). This study suggests that an operation span test can be used to effectively assess WMC and could be a predictor of a fluid intelligence test like RAPM (Raven's Advanced Progressive Material).

The math span test for the current study was based on Salthouse and Babcock's (1991) and Robert and Gibson's (2002) math span tests. This test was comprised of some simple arithmetic problems in the form of $X + Y = ?$ or $X - Y = ?$ type. X and Y can be single digit numbers between 1 and 9, and none of the answers to the problems were negative. There were no identical (repetitive) arithmetic problems across the test or any repeated target digits for two consecutive problems. However, whereas Salthouse and Babcock (1991) provided three possible answers and asked their subjects to check off one, the participants here had to take the test individually and provide the answers orally, like in Roberts and Gibson's version of this task (2002), and their production was recorded by the researcher. This format of the math span test was used to make sure that the participants' correct answers would not be subjected to guessing as well as to control for the recency effect. Furthermore, it would ensure that the participants had recalled the target digits at the end of the set and not earlier during the processing time.

Thus, in this test, the participants viewed simple addition and subtraction problems (i.e., $4 + 2 = ?$ or $9 - 6 = ?$) on a computer screen. Each problem appeared on the screen for 2.5 seconds. The participants were required to state the answer to the problem aloud immediately (processing) and remember the second digit in each problem for later recall (storage). Unlike the sentences in RST, math problems were not read aloud because they could be read in different ways in Persian. So, to control the speed of processing, and consequently possible rehearsal of the targets, each participant was required to view each math problem, and does it in his or her mind within a very constrained given time.

This test was developed by the researcher and piloted with different groups of participants (overall 48 participants) at three levels of proficiency over five pilots. On each occasion, a different combination of participants at different proficiency levels completed the test, followed by a retrospective report. Based on their reports and results on each occasion, the shortcomings of the test, which were mostly related to the slide transition times, were removed until no further shortcomings were reported by the participants.

During the first pilot, the test was administered to a group of 10 participants. The slide transition for each math problem was set on 5 seconds. However, the participants reported that they had some extra time to rehearse the targets (second digit at each math problem). Furthermore, they claimed that there had been two consecutive math problems within one set with the same target digits. To remove this problem, the positions of the digits were reversed in one of the math problems. Furthermore, the slide transition was decreased to 4 seconds. Then the revised test was piloted with another group of participants during the second pilot to see whether it worked well or not.

During the second pilot, the test was administered to a group of 9 participants. They reported that they had no problems with the test. However, they said that the slide transition for each math problem had been too long, so they had time to rehearse the targets. The results of study also indicated that the participants' scores were very high. Then it was concluded that it might be the extra time that had led to inflated scores here. This problem was removed by decreasing the slide transition for each math problem to 3 seconds. Then the revised test was piloted with a new group of participants to see whether there was still extra time for the participants to rehearse the targets or not.

For the third pilot, the test was administered to a group of 11 participants. The results of the study indicated two ranges of scores, with some participants scoring quite highly and others quite low. Participants with low scores reported that the updated slide transition times had been just enough for them, while those with high scores said that they had had a little extra time for rehearsing the targets. The range of scores was wider than before.

During the fourth pilot, the slide transition was decreased to 2 seconds and the revised test was administered to a group of 8 participants. However, the participants all reported that the slide transition had been too fast for them to do the math problem. Thus, they had had to skip some math problems and focus just on the targets for better recall. The results of the study also indicated that the participants had not obtained consistent scores for the processing and recall components, like those in the prior pilot studies.

During the fifth pilot, the slide transition was set to 2.5 seconds and the test was administered to a group of 10 participants. The participants' scores demonstrated the widest spread of the pilot tests (31-60 & 16-58 for processing and recall capacities respectively). They also reported that they had had sufficient time to process each math problem but had had no more time to rehearse the second digits. Thus, the duration of 2.5 seconds was established as an appropriate slide transition for the final test. The results of this pilot were consistent with the findings of the main study, which showed a wide spread of participant scores (33-60 & 13-59 for processing and recall capacities respectively). A satisfactory internal reliability, as indicated by Cronbach's Alpha, was found for this measure in the main study. The reliability was .850 and .863 for the MST processing and recall respectively.

The final version of the test was comprised of 60 simple addition and subtraction problems, 30 each, distributed equally in 3 sets of 2, 3, 4, 5, and 6 math problems. There was also a practice test including 10 math problems at the beginning of the test session. This was to familiarize the participants with the test procedure. The participants were told that they would receive no points for the practice test items. Following this, they went through increasingly longer sets of math problems. At the end of each set, a prompt (three hash keys) appeared on the computer screen. This was to signal to the participants to recall the target digits aloud while their production was recorded. To control for the recency effect, the participants were instructed not to say the last target digit first.

To score the participants' math span test, each participant's score for the processing and storage components of WM was calculated. The processing score was the total number of correct answers given to the math problems. The storage score included the total number of target digits recalled correctly across the test (Friedman & Miyake, 2005). Thus, since there were 60 math problems in this test, and one mark was allocated to each correct answer, the range of each participant's processing and recall score was between 0 and 60. A composite WM score was obtained (Turner & Engle, 1989; Waters & Caplan, 1996). The composite WM score was calculated by adding up the z-scores of the WM components. This was an index for each participant's WMC.

3.2.1.3 Non-Word Recognition Test

A non-word recognition task was used to measure phonological short term memory (e.g., Gathercole, Pickering, Hall & Peaker, 2001; Gathercole, Frankish, Pickering, Peaker, 1999; Trofimovich, Ammar, & Gatbonton, 2007). Phonological-short term memory controls the temporary storage and processing of verbal information (e.g., Baddeley, 2000, 2007; Baddeley & Hitch, 1974; Ellis, 2001). It is a component of WM model proposed by Baddeley & Hitch (1974). Research suggests that learning the sound structures of new words in L2 is mediated by this component (e.g., Gathercole & Baddeley, 1990a; Gathercole, Thorn, & Bristol, 1998; Gathercole, Service, Hitch, Adams & Martin, 1999; Masoura & Gathercole, 1999; Miyake & Friedman, 1998; Valler & Papango, 2002; Skehan, 1989). In the non-word recognition task, the participants hear two consecutive sequences of pronounceable non-words and judge whether they are in the same or different order (e.g., Gathercole et al., 2001; Trofimovich et al., 2007). Non-words are used since they minimize the influence of vocabulary knowledge on phonological short-term memory and yield a relatively accurate estimate of it (Gathercole et al., 2001).

The non-words employed in this test were adapted from Gathercole et al., (2001) by the researcher. The test consisted of 22 pairs of sequences of English non-words. The length of each sequence was gradually increased across the pairs within the range of 4 to 7 non-word syllable length. There were 4, 5, 6, and 7 sets of 4, 5, 6, and 7 non-word sequences respectively in this test. This test was conducted in a classroom environment. The participants were required to listen to each pair of sequences to determine whether the order of non-words in both sequences was the same or different by checking the boxes next to each choice in their answer sheet. To score the test, the total number of correct answers was calculated. This was an index of phonological short-term memory. The participants' phonological short-term memory scores ranged between 0 and 22 in this task.

This test was piloted on a total of 114 participants over four pilots, each time on a different group of participants. The first pilot included 52 participants at three proficiency levels. During the test session, the researcher noticed that some participants were taking notes while they were listening to the sets of English non-words. These participants later reported that the test instruction, which was given orally without any practice tests, had not been very clear to

them. Since this could be a variable influencing the participants' performance, the second pilot study was conducted on another group of participants with a different design.

During the second pilot, the test was administered to a group of 44 participants in two separate sessions. Prior to the exam, the test instructions were given orally to the participants in a clear way. The instructions included two sets of four-English non-word length examples which were given to the participants to make sure that they had fully understood the test procedure. This followed by a practice test including two four-English non-word length sets. The practice test was given orally with the same interval as the one in the main test by the researcher. Finally, the participants' questions were answered before the test started. After the test session, a few students reported that they had been confused by the oral explanation of the test instructions. They suggested that it would have been much better if they had been given the test instruction guide followed by an example in writing some time prior to the test.

In response to these concerns, the test was administered to a group of 10 participants during the third pilot. These participants were at three proficiency levels and selected from the same language school as the main study. Here the participants were given a written test instruction guide 5 minutes before the test session. Then their questions were answered and a practice test including two sets of four English non-words was conducted prior to the test session. Unlike the test session in which English non-words were played on a tape recorder, in the practice test, they were read aloud by the researcher. Similar to the procedure during the test session, the participants were required to determine whether the order of the English non-words was the same or different. Then the participants completed the test session. After the test session, some participants reported that they had been confused by the test instruction guide as it had been too detailed. They suggested that it might have been clearer if it was shorter and followed by a brief oral explanation.

To see whether this worked better or not, the test was administered to another group of participants during the fourth pilot. They included 8 participants, both males and females, at three proficiency levels. They were given a brief test instruction guide 5 minutes prior to the test session. Then an oral explanation followed by an example was given at the beginning of

the test session. The participants were given an opportunity to ask questions, as well. Unlike the prior times, there were no questions here. After the participants had completed the test, similar to the times before, they were given a retrospective report to complete. They reported that they had had no problems with the test instruction guide or the test itself, except that the test had been memory demanding. This was consistent with the participants' reports during the prior three pilots.

The results of this study also indicated that the participants' scores were in a wide spread range. The internal reliability for this measure, as indicated by Cronbach's Alpha was .706. This indicated a satisfactory internal consistency for this test. Furthermore, an item analysis was conducted on the test here to examine the contribution each item was making to the test. The results indicated that the discrimination index for the test items was in a desirable range of .43 to .97. Thus, this test, as described in the last pilot study, was established as a reliable and good measure of PSTM for the main study.

As a result, the test was administered to 140 participants at three proficiency levels in the main study. The results of the main study were consistent with the ones in the pilot studies. The participants all reported that they had had no problems with the test, except that it had been demanding for them. The participants' scores showed a wide spread as well. Furthermore, the discrimination index for the test items was in a desirable range of .45-.87. This suggested that these test items could discriminate well between weak and strong participants. The internal reliability for this test ($r = .683$), as indicated by Cronbach's Alpha, was a bit lower than that of the last pilot study ($r = .706$). This might have been due to test fatigue, likely as a result of using English non-words, as the participants had found this measure challenging.

3.2.2 Reading measures

3.2.2.1 Introduction

Recent studies suggest that there is no optimal method to assess reading comprehension because all methods of reading assessment have some inadequacies in terms of construct validity (e.g., Hudson, 2007; Leslie & Caldwell, 2009; Cain & Oakhill, 2006; Koda, 2005).

All methods of measurement have some advantages and disadvantages. So, researchers advise the use of more than one test method to measure a construct like reading comprehension (e.g., Cain & Oakhill, 2006; Alderson & Bannerjee, 2002; Alderson, 2000; Koda, 2005; Leslie & Caldwell, 2009). In this study, a rational deletion format cloze test (Koda, 2005), two short-answer reading comprehension passages (Alderson, 2000), and two L1 recall reading comprehension tests were used to measure the participants' reading comprehension at each proficiency level. The reason for choosing one cloze test was to prevent test fatigue as it included 40, 45 and 50 test items at the beginning, intermediate and advanced levels respectively. The reason for choosing two passages for the short-answer and L1 recall tests was to control the effect of the participants' background knowledge as well as text genres on comprehension (Alderson, 2000; Beach, 1997; Koda, 2005; Mitchell, 1982; Nation, 2009). All passages included social and science topics, instructions were given in the participants' L1, and consistent scoring procedures were used. These methods of reading measures were chosen because they matched well with the aspect of reading (role of WM with a limited pool of cognitive resources in reading process) this study is attempting to measure.

A general test like the TOEFL or IELTS reading was not used in this study for two reasons. First, these tests, as measures of L2 general proficiency, should be administered to higher proficiency learners. Therefore, they are not appropriate for measuring L2 general knowledge at lower proficiency levels. Second, reading measures in these tests are all in a multiple-choice format, and do not include variety types of reading test methods for use at higher proficiency levels. For these reasons, three types of reading tests were developed for each proficiency group. This could help us to control the readability of the passages from which the tests were developed for each proficiency level.

At each proficiency level, one cloze passage, two short-answer, and two L1 recall readings were selected from two ESL textbook series available in the market. They included the series of Active Skills for Reading, level 1 to 4 (Anderson, 2007, 2008) as well as the New Cutting Edge, elementary and intermediate levels (Cunningham & Moor, 2005). The length of the passages was in the range of 254 – 284; 318 – 409, and 370 – 616 words for the beginning, intermediate, and advanced levels respectively. To control the difficulty level of the reading

passages at each proficiency level, Cobb's (2002) VocabProfiler, which classifies the words by word family, type, and token, was used. This software indicates how many words each passage contains from the following four frequency levels proposed by Laufer and Nation's (1995) lexical frequency profiler: (a) the list of the most frequent 1000 word families (K1), (b) the second 1000 (K2), (c) the academic word list, and (d) words that do not appear on the other lists (off-list).

Table 3.2

Readability and Vocabulary Information of the Reading Passages for each Proficiency Level

Vocab Information	Beginning level	Intermediate level	Advanced level
K1	83.10- 93.06%	77.80- 86.20%	72.62- 77.44%
K2	2.61- 8.18%	1.04- 7.32%	4.38- 9.82%
K1+K2	90.34- 97.23%	85.01- 89.35%	80.07- 82.53%
AWL	0.39- 1.49%	3.66- 9.31%	3.92- 6.81%
Off-List	1.74- 8.97%	3.82- 11.22%	13.04- 13.82%
Readability	68- 80	58.4- 62.8	48.7- 57.6

As indicated in Table 3.2, while the proportion of the academic word list (AWL) and off-list is higher for the advanced level, the proportion of the first and second thousand level words (K1+K2) is higher for the beginning level with the intermediate level in between. This suggests that the difficulty level of the passages increases for each proficiency level as the proportion of most frequent words decreases. Second, the density of the propositions in the passages was controlled. It is a variable which affects understanding and recall (Alderson, 2000; Nation, 2009; Beach, 1997; Koda, 2005; Cain & Oakhill, 2006). This was done by keeping the same range of propositions adjusted for length in each reading passage. To do this, first, several reading passages, which included similar numbers of sentences, were selected from a considerable number of reading passages which were the same length and difficulty level at each proficiency level. Then the total number of major and minor idea units in each sentence was determined. These were added up to return the total number of major and minor idea units per passage. Finally, those passages which included more or less the

same number of idea units were selected for each proficiency level. Table 3.3 indicates the range of idea units for each passage type for each proficiency group.

Table 3.3

Range of Idea Units for Each Passage Type at each Proficiency Level

	Cloze	Short Answer 1	Short Answer 2	L1 Recall 1	L1 Recall 2
Beginning	35	42	40	35	36
Intermediate	43	57	50	53	55
Advanced	49	76	70	70	79

A pilot study was conducted among the participants at each proficiency level to gain the participants' feedback on the appropriateness of a sample of the selected passages. Since the participants may have discussed similar content in class time, they were asked to read through each passage and fill out a retrospective report. In this report, the participants were asked to determine the extent to which they had already been familiar with the content of the passages, on a Likert scale of 4 items (very familiar, rather familiar, less familiar, and unfamiliar). Their reports indicated that 100% of the participants had been unfamiliar with the content of the passages for the three proficiency groups.

3.2.2.2 Cloze Test

A cloze test is defined as a passage with a few sentences of introduction followed by text with deleted words (gaps) with a consistent number of words (from five to eleven) between them. The test taker's responsibility is to predict the deleted words based on the words given in the text (e.g., Alderson, 2000; Brown, 1989, 1992; Cain & Oakhill, 2006; Koda, 2005; Lee, 1985; Nation, 2009; Oller, 1975, 1979). It was devised as a technique to measure reading comprehension. The rationale behind a cloze is that the reader must be sensitive to both semantic and syntactic constraints in each context to fill in the blanks. Such sensitivity can be taken as a reliable indicator of reading ability since text information processing is supported by these constraints (Koda, 2005, p., 239). A multiple-choice format, rather than a fill in the blank one, was used in this study. This was to improve scoring reliability. Moreover, it was more practical to administer (e.g., Bachman, 1990; Heaton, 1988; Madsen, 1983) and a more familiar format for the participants. It consisted of 40 to 50 empty spaces where the readers'

task was to read the passages and choose a replacement word from a choice of 4 options (Alderson, 2000; Cain & Oakhill, 2006; Koda, 2005).

Cloze tests can be constructed in two ways, by standard fixed ratio deletion or rational deletion format (Koda, 2005). In standard fixed ratio deletion format, every 5th or 7th word is deleted from the text. The missing word can be from any category of function or content words. However, the standard fixed ratio format has some drawbacks: such clozes are more sensitive to surface linguistic forms than the underlying text meaning, they say little about global-text information processing competencies, and may reflect memory performance rather than reading ability (e.g., Alderson, 2000; Beach, 1997; Cain & Oakhill, 2006; Koda, 2005). For these reasons, a rational deletion format was preferred. In this test, there is no fixed ratio to delete words from the text; rather, a prespecified category of words is deleted to test a specific facet of reading ability. Therefore, to measure text-meaning understanding, a proportion of content words are chosen and deleted. Contextual support availability rather than the number of deleted content words will determine the rational cloze test's difficulty level (Koda, 2005, pp., 240-1).

At each proficiency level, an appropriate passage (as described above) was chosen, and 40-50 content words were deleted. The number of deleted words differed for each proficiency level depending on the length of each text. There were 40, 45, and 50 deleted words in the selected passages for the beginning, intermediate and advanced levels respectively. The difficulty level was also controlled by leaving sufficient and equal contextual support for the learners to restore the deleted content words (Abraham & Chapelle, 1992). This was done by averaging the total number of words in the passage over the total number of gaps. The results indicated that there was an average of 6.35, 7.06, and 7.4 words as contextual support for each gap at the beginning, intermediate and advanced levels respectively. The learners were required to read the passage and choose the best answer fitting with the gaps from the choices given for each blank. One mark was allocated to each correct answer, with a total of 40-50 marks for this task.

The test at each proficiency level was piloted three times, each time with a different group of L1 Persian EFL learners chosen from the language school, but not from the participant pool for the main study. During each pilot, they completed the test followed by a retrospective report where they were asked about their overall viewpoints on the test, test items, and content of the test. Then their tests were scored and an item analysis was conducted to examine the contribution that each test item made to the test. Finally, based on results of the item analysis and the participants' reports on the test, the poor items were identified and revised until no further ones were identified.

During the first pilot, the test at the beginning, intermediate and advanced levels was administered to 9, 7 and 10 participants respectively. As they had prior experience with cloze reading tests, they raised no questions about the test during the exam session. They reported that they had not been familiar with the content of the test and that some of the test items had been too easy and some too difficult. They said that the main problem with the overly difficult items was that more than one answer could have been correct. And the main problem with the too easy ones was that they had been able to dismiss the distracters very easily. The results of an analysis indicated a range of 15, 21 and 17 poor items which had been either too easy or too difficult at the beginning, intermediate, and advanced levels respectively. To determine the poor items, the facility value (e.g., Alderson, Clapham & Wall, 1995) for each item was computed. This was used to measure the difficulty of an item, and was obtained by dividing the participants who scored correctly by the total number of participants. The cut-off levels used on the item analysis to determine these poor items were .22-.77; .14-.71; and .2-.8 for the beginning, intermediate, and advanced levels respectively. These items were revised by replacing the distracters with more appropriate options. More specifically, the distracters for the overly easy items were replaced with more challenging distracters and those for the overly difficult items were replaced with less ambiguous distracters.

To determine how the revised tests would work, they were administered to a group of 9 participants, 3 in each proficiency group, during the second pilot. The participants at each proficiency level reported that they had had no problems with the test. The results of an item analysis indicated that of the revised test items, those at the beginning level had worked well. The facility value index for this test was in a desirable range of .33 - .66. However, a few of

the revised test items were still not performing well for the tests at the intermediate and advanced levels. They were either too easy or too difficult. These items were revised once more. The distracters for these items were replaced with more appropriate options.

Finally, the revised tests were piloted with a group of 4 intermediate and 3 advanced participants during the third pilot. The participants reported there were no overly difficult or overly easy test items at this stage. The results of the item analysis also indicated that the facility value index for these tests at the intermediate and advanced levels were in a desirable range of .25 - .75 and .33 - .66 respectively. An estimate of the duration of the exam session was also obtained for the tests. This was the average time for each test to be completed by the participants at each proficiency level. It included 40, 45 and 50 minutes for the cloze test at the beginning, intermediate and advanced levels respectively. These tests were then finalized for use in main study.

In the main study, each cloze test was administered to the participants at the respective proficiency level. The participants completed the test within the proposed time. They reported they had had no problems with the test. The results of an item analysis also indicated a satisfactory range for the internal reliability of the tests. They, as indicated by Cronbach's Alpha, included .730, .814, and .872 for the beginning, intermediate and advanced levels respectively.

3.2.2.3 Short-Answer Reading Test

A short-answer test or a limited production response (Backman & Palmer, 1996) is a semi-objective alternative to the multiple-choice question form (Alderson, 2000). In this test, the participants are simply asked a question which requires a brief response. The rationale behind this test is that it is possible to interpret participants' responses to see whether they have really understood the content of the text or not, whereas this is not the case on multiple-choice items, where the participants may have chosen the correct answer by crossing out others (Alderson, 2000; Gronlund, 1973; Harris, 1969; Weir, 1990). However, this is less likely to apply to a multiple-choice cloze test. This is because the sentences in a cloze test are

interrelated and together make a unified coherent text. Thus, one must take into account the whole context to be able to choose the answer for each blank space.

Two short-answer reading passages were given to the participants at each proficiency level. To make sure that the participants' performance on this task would not be affected by the possible inadequacies in their English proficiency, the instructions and questions were given in the participants' L1. Moreover, they were given sufficient time to read each passage and answer the respective questions in L1. However, to avoid conflating the L1 and L2 reading processes, the participants were required to read each passage first, and then answer the questions. They were allowed to look back at the text.

To develop the questions for each passage, Day's and Park's (2005) taxonomies of types of comprehension and forms of questions were used. This taxonomy was influenced by the work of Pearson and Johnson (1972) and Nuttall (1996). In this taxonomy, they proposed six types of comprehension (literal, reorganization, inference, prediction, evaluation and personal response) and five forms of questions (Yes / No, Alternative, Multiple-choice and True or False questions as well as Wh-questions). However, in this study, only one form (Wh-questions) and three types of comprehension (literal, reorganization, and inference) questions were used. The Wh-question form was used because they could elicit what the participants had really understood, whereas the participants' answers to other forms of questions (Yes / No, True / False, Alternative, and Multiple-choice questions) were subject to guessing (e.g., Alderson, 2000; Cain & Oakhill, 2006; Day & Park, 2005; Koda, 2005; Leslie & Caldwell, 2009). Literal, reorganization and inference types of comprehension questions were used since the answer to these questions was more controlled (a single word or phrase), and accordingly the scoring would also be more objective than other types of comprehension questions. Questions were developed for 2 passages at each proficiency level.

Each passage was followed by 9 questions including literal, reorganization and inference questions, 3 of each. Overall, there were 18 questions for the two passages at each proficiency level. Then they were piloted twice with 3 competent speakers of English. They were also piloted three times with a group of L1 Persian participants, each time with a

different group from the language school, but not from the participant pool for the main study. More specifically, the competent speakers of English included 3 experienced EFL teachers in Iran, and they were given the test to complete to see whether the questions were answerable and whether they had any criticism. The participants included a different number of participants during each pilot, and they completed the test followed by a retrospective report. An item analysis was run for each pilot to identify poor items. Then the poor items were revised and the test was replicated again until no further poor items were identified. During all the pilots, the participants at each proficiency level reported that they had been unfamiliar with the content of the tests.

First, the test was piloted with the competent speakers of English. They were given just the test items without the texts to see whether they could answer them based on prior knowledge rather than knowledge of the texts. The results of the study indicated that they were able to answer 5, 3, and 4 questions correctly at the beginning, intermediate and advanced levels respectively. They reported that they had answered these questions based on their prior knowledge. Then these poor items were replaced with more challenging test items.

During the second pilot, the same competent speakers as in the first pilot were given both the text and the questions. They completed the test; however, they felt that a few test items at the intermediate and advanced levels were poorly written and not clear. Then these poor test items were revised by simplifying their grammatical structures. Then the revised items were given back to the competent speakers to see whether they were clear enough or not. They confirmed that they looked much better; however, they needed to be piloted with language learners to see how they would work.

Then the test was administered to a group of 5, 3 and 4 participants at the beginning, intermediate and advanced levels respectively during the third pilot. These participants reported there were a few items which had been either overly difficult or overly easy for them. They reported that as the overly easy questions had been worded similarly to the text, they could easily recognize the answer. They also said that the overly difficult test items had been vague and they could find more than one answer for these questions. These poor items were

identified and replaced with new items. The overly easy test items were replaced with new items, and the overly difficult questions were revised into clearer questions. To do this, they were sometimes broken into two statements.

To see how the revised tests would work, they were piloted with a different group of participants. They included 9, 7 and 10 participants at the beginning, intermediate and advanced levels respectively. The participants at the advanced level reported that they were confused by two very similar test items. The participants at the beginning and intermediate levels also reported that some of the test items had been difficult. This was consistent with the results of the item analysis. These results indicated that there were 11, 13, and 10 poor items, which were identified as either too easy or too difficult at the beginning, intermediate and advanced levels respectively. The cut-off levels, as indicated by the facility values used on the item analysis to determine these poor items were .22 - .72; .35 - .85; and .16 - .77 for the beginning, intermediate, and advanced levels respectively. These poor items were either revised or replaced with the new items of the same type, and administered to a group of 9 participants in each proficiency group during the third pilot. The participants reported that they had had no problems with the test items, the text and the test instructions. The results of the item analysis also indicated that the test items include a desirable range of facility value and could be used in the main study. The range of facility value was .27 - .66, .33 - .77 and .38 - .83 for the beginning, intermediate and advanced levels respectively.

In the main study, the tests were administered to the study participants at each proficiency level. The participants completed the tests within the proposed time. They reported they had had no problems with the tests and that the tests had been quite new and exciting. One mark was allocated for each correct answer and .5 of a mark to partially correct ones with a total of 18 possible marks for this task (9 per passage) at the beginning and advanced levels. The same method of scoring was also applied to the tests at the intermediate level. However, since the answer for 4 questions included the relevant examples, which could also be correct, these questions were broken into two parts, (a) and (b). Then the overall score for the tests here was 22 (11 each). The results of an item analysis also indicated a satisfactory range for the internal reliability of the tests. This, as indicated by Cronbach's Alpha, included .826, .731, and .711 for the beginning, intermediate and advanced level tests respectively.

3.2.2.4 L1 Recall Reading Test

The L1 recall reading test is a modified version of the immediate comprehension recall protocol. However, unlike the immediate comprehension recall protocol, the variable of memory is controlled in L1 recall test. In immediate comprehension recall protocol, to determine what the participants have understood from a text, they are required to read the text as often as they like and then to write down in either their L1 or L2 as much detail as they can remember from it after the text has been taken away (e.g., Bernhardt, 1983). In this test, therefore, reading ability and memory are measured together. However, in the L1 recall reading test, to control the memory variable, the participants are allowed to look back at the text and check their comprehension (e.g., Beach, 1997; Grabe & Stoller, 2002; Koda, 2005; Nation, 2009).

In the L1 recall test, the participants were required to write down what they had understood from an L2 text in their L1. This was consistent with the prior research (e.g., Bernhardt, 1991b; Shohamy, 1984; Swaffar, Arens, & Byner, 1991; Wolf, 1993), where it was suggested that measuring L2 comprehension in recall protocol through the L1 could be a valid measure of comprehension. This is because it could allow us to measure reading comprehension without allowing the participants' performance to be influenced by insufficiencies in their English writing ability, particularly at the beginning and intermediate levels, and obtain fuller accounts of their comprehension. The L1 was also used in the comprehension recall protocol by prior L2 studies (e.g., Chun & Payne, 2004; Lesser, 2007).

In this study, the L1 recall tests included two L2 reading texts given to the participants at each proficiency level. They included a different range of idea units (propositions) depending on the length of the passages at each proficiency level. As indicated in Table 3.4, the total number of idea units in the passages was 71, 108, and 149 for the beginning, intermediate and advanced levels respectively. They included both the major and minor idea units. An idea unit is defined as a message segment consisting of a topic and comment that is separated from contiguous units syntactically and / or intonationally (Ellis & Barkhuizen, 2005). Major idea units were those that were required to convey the essential content of the message. Minor idea units related to the details which embellished the message, but were not essential. Both the major and minor idea units were established by analyzing the propositions in the text.

Table 3.4

Distribution of Idea Units per Text and Proficiency Level

	Beginning Level Texts		Intermediate Level Texts		Advanced Level Texts	
Idea Units	One	Two	One	Two	One	Two
Major	13	11	17	21	30	27
Minor	22	25	36	34	40	52
T. p. Text	35	36	53	55	70	79
T. p. Level	71		108		149	

Note. T. p.= Total per

The tests at each proficiency level was piloted with a group of 3 competent speakers of English first to examine their perceptions of the test as well as to obtain a baseline for scoring. The competent speakers read through each text and gave a summary of it rather than a detailed description of their comprehension. Two of the competent speakers suggested that written passages with a simpler structure be used at the beginning level. To obtain a more detailed production from the participants, they further suggested that the test be administered in a written form. The reading passages at the beginning level were replaced with simpler passages. After they were checked with the competent speakers, they, along with the passages at the intermediate and advanced levels, were piloted three times, each time with a different group of participants from the language school, but not from the participant pool for the main study.

During the first pilot, the tests were piloted with a group of 4, 2, and 3 participants at the beginning, intermediate and advanced levels respectively. Each participant read through each test and said what he / she had understood. However, their production, similar to the competent speakers, included just a summary of the reading passages. Some of the participants at the beginning and intermediate levels reported that while they were reading the reading passage, they were worrying about the next stage where they were required to speak. Then they could not concentrate well on the passage. These participants reported that they would have felt relaxed if the production had been in a written form, and if they had not been required to say what they had understood.

During the second pilot, the tests at each proficiency level were administered to a group of 5, 4 and 3 participants at the beginning, intermediate and advanced levels respectively. Contrary to the prior time, the participants were required to read each reading passage and write down what they had understood from the passage in detail. The participants reported that they had felt relaxed during the exam, and had had enough time to complete the test. They produced more language than in the previous pilot study. However, it was more like a sentence by sentence translation. To allow the participants to produce just their understanding of the text, another pilot study with a new design was conducted.

During the third pilot, a clear set of written test instructions was developed for each proficiency group. In this instruction, the participants were encouraged to show their comprehension within a proposed range of idea units which differed for each group of proficiency. The participants were also allowed to express their understanding of each sentence in more than a statement or two if needed. For this reason, the proposed range of idea units for each proficiency group was higher than the total number of idea units. For example, a range of 40-50 idea units was proposed to the beginning proficiency group. An estimate of the time duration for administering the tests at each proficiency level was measured. To make sure that all the participants understood the test direction, in addition to clear, written test instructions, an oral explanation followed by an example sentence was given to the participants at the beginning of this pilot study.

The participants reported they had been quite relaxed and had had no problems with the test. They also claimed that this kind of test could measure their comprehension in a detailed way. The results of the study indicated that they had produced much more than the participants during the prior tests. Furthermore, an analysis of their production indicated that no direct translations had been done as there were many instances where sentences had been expressed through several statements including those which had been inferred from the reading passages. To obtain a baseline for scoring the tests, they were given to two competent speakers of English. Their production for each reading passage was higher than the prior group of competent speakers, whose production was in an oral form. Thus, this test with such a design was accepted for use in the main study.

In the main study, the tests at each proficiency level were given to the participants. An oral explanation followed by an example sentence was given to them as well to make sure they had all understood the test instructions. They were also given an opportunity to raise their questions if there were any. Then the test was administered in a class, similar to the one in the prior pilot studies. The participants were given sufficient time, 45 to 90 minutes depending on their proficiency level, to read and write down what they had understood from the text on their answer sheets. However, they were advised that they would be given additional time if they needed any. Except for a few cases where they were given 10 to 30 minutes' additional time, all the participants completed this measure within the proposed time constraint. The participants were told not to translate the text, but rather to write down their comprehension of the text in as many sentences as they wished, including the ones they could infer from the text. Then the data was analyzed into idea units and scored by the researcher. 25 % of the test papers were scored by both the researcher and an L1 Persian EFL teacher. Inter-rater reliability to see how consistent the raters were on identifying idea units reached 95.3%. Then the rest of the data was scored by the researcher. Scoring was based on the number of idea units (propositions) uttered by each participant. After the idea units were identified, Zaki and Ellis's (1999) propositional analysis was used. In this analysis, the total number of major and minor ideas was counted and then one mark was allocated to each correct proposition and .5 of a mark to partially correct propositions. As indicated in Table 3.4, the range of the possible total scores was 35-36, 53-55, and 70-79 for the texts at the beginning, intermediate and advanced levels respectively. The results of the study indicated a range of 21.51, 44.55 and 64.47 mean scores for the participants' performance on this measure at the beginning, intermediate and advanced levels respectively.

Section Three

3.3 Design of the Study

After the reading and memory measures were developed and piloted by the researcher, and their potential problems were identified and removed, the main study was conducted. As indicated in Table 3.5, this study was carried out in 3 sessions over a month. In the first session, all the participants were initially required to complete a roughly 15-minute PSTMT test followed by a 90-minute general proficiency placement test. These tests were conducted

in the language school where the participants had been chosen for the study. In the PSTM test, the participants were required to listen to an audio file to see whether the sequences of English non-words were the same or different, and then to check the correct answer in their answer sheets. In the language proficiency placement test, they completed a listening section first, and then reading and grammar sections. Since the test was in a multiple-choice format, they were just required to choose the correct answer and mark it in their answer sheets. Then the participants were distributed into three proficiency groups based on their scores in this test. As mentioned before, those participants who obtained scores within a range of 1-30, 31-49, and 50-70 were placed at the beginning, intermediate and advanced levels respectively.

In the second session, 10 days after the first session, two WM measures, including a RST and a MST, were administered consecutively with a 15-minute break in between. These measures were developed in a computer software program format and conducted individually within two weeks. The participants were just required to say the answers to the researcher, who was keeping a record of their answers. This was to make sure that they read each sentence or math problem for the purpose of processing and recalled the sentence final words / numbers at the end of the sets. Moreover, it could help to avoid distracting their attention or skipping any test items.

All the participants followed the same consistent procedure in conducting these measures. First, they were required to complete a roughly 15-minute Persian RST. Then they were allowed to take a 15-minute break in between, and have some food and a drink. This was to allay any potential test fatigue for the participants and refresh them for the following test. Second, they were required to complete the 10-minute MST, which had a relatively different processing and storage procedure from the RST. This could help to avoid any training test effects. Unlike in the RST, the participants were involved in processing non-linguistic information here, and calculating simple arithmetical problems. Also, the targets included the second digits, not the last digit located at the end of the problem. Finally, the participants explained in a retrospective report that this had been the first time any of them had completed WM measures, and they had found them to be quite interesting. Furthermore, they reported that the RST had been more memory-demanding than the MST, although both of the measures needed their full concentration. Finally, they believed that they had been quite

relaxed once they started completing the second memory measure, the MST, and their prior experience in completing the RST was quite different to the procedure followed in completing the MST.

In the third session, the participants were required to complete a set of reading measures, including a cloze test, two short answer tests and two L1 recall tests. The participants were required to complete the appropriate test versions for their proficiency level. They all completed these tests in the same order. They completed the cloze test, the short-answer and the L1 recall tests in turn. Since these measures were different for each proficiency level, they were conducted in three consecutive rounds. In the first round, the participants at the beginning level completed the measures in a 100-minute session. In the second round, the participants at the intermediate level did the same in a 130-minute session. In the third round, the participants at the advanced level were allocated 160 minutes to complete their reading measures. Differences in allotted time at each proficiency level was due to the reading measures which were different in terms of length (number of words and propositions), and readability. During the test sessions, the participants were given some food and drink to refresh them and prevent them getting tired. The participants were given additional time when they needed some. In this study, there were just a few participants who needed an extra 10 to 15 minutes.

Table 3.5

Timetable Used for the Reading and Memory Measures in this Study

Session	Tests	Duration
One	PSTM	15 minutes
	Language Proficiency Placement Test	90 minutes
Two	WM Measures (RST & MST)	40 minutes including 15 minutes break time in between
Three	Reading Measures (Cloze, L1 Recall & Short-Answer)	100, 130 & 160 minutes at beginning, intermediate and advanced levels respectively

Once each participant completed their reading measures, they gave their answer sheets to the researcher along with the retrospective report on these measures. They all believed that short-answer reading measure had been quite new and very interesting. They suggested that a short-answer test was a better measure of their L2 reading ability. They also reported that the L1 recall measure had been quite new and interesting. Furthermore, since this test required the participants to recall whatever they had understood and write a detailed answer in their L1, it was more time-consuming. The participants suggested that it would be better if the test was a bit shorter. Overall, they were happy with both of these measures, although it was the first time they had completed such reading measures.

3.3.1 Data Analysis

After the memory and reading measures were completed, each participant was assigned an ID code. Then each measure was scored by the researcher as described above. Two different methods were used to score the RST and MST tests. This was to see how the results would vary for the three proficiency levels. First, a lenient method was applied to score both the RST and MST. In this method, one point was allocated to each correctly recalled item irrespective of the sequence of the recalled items. In the strict method, no point was given to

the last target if it was recalled first. Based on the results obtained from each method of scoring, two sets of correlations within and between groups were taken for the three proficiency levels. They were to a large extent different from each other. In this study, the results obtained from the strict method were used. This was because they are more reliable than the ones obtained from the lenient method (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). They were also consistent with the theoretical assumptions of this study and those for the prior research (e.g., Lesser, 2007; Walter, 2004).

After all measures were scored, an item analysis was conducted on reading and memory measures to remove any poor items. In this analysis, all the test items were given values of either 1 if they were correct, or 0 if they were incorrect. Internal consistency reliability coefficients for both memory and reading measures were calculated. The poor items in the reading measures were removed. They were in the cloze tests at the beginning (14 items) and intermediate (20 items) levels. The total poor responses removed from the cloze tests were 23.6 % for the beginning level and 43.5 % for the intermediate level. The cut-off levels used to determine these poor items were .30 - .71 and .39 - .62 for the beginning and intermediate levels respectively. The range of the Cronbach's Alpha for WM and reading measures were quite high, .790 - .863 and .711 - .872 respectively. This suggests that these measures are reliable enough to be employed in future studies.

The z-scores of all the measures were calculated using SPSS software. This was to weight all the tests equally. Correlations between and within the independent and dependent variables were obtained. There were significant correlations between the WM measures and the RST and MST for the three proficiency levels. To provide more stable measures of the participants' WMC, composites were created from the unit-weighted z-scores of the storage and processing measures (e.g., Lesser, 2007; Turner & Engle, 1989; Waters & Caplan, 1996). Some further memory composites, other than the RST and MST, were also created from the WM z-scores at each proficiency level. They included the RSTMST, processing and recall composites.

There were also significant correlations between the reading measures, cloze, short-answer test and L1 recall test at each proficiency level. To have a more stable index for the participants' L2 reading ability, a composite reading score was also created from the reading measure z-scores at each proficiency level. Finally, an inferential analysis was conducted at each proficiency level. This included the correlations and regressions for the WM and reading measures. A different range of significant correlations between the WM and reading measures was found for each proficiency level. Finally, a regression was run for each significant correlation to determine how much variability in the reading measure could be explained by the memory measure.

3.3.2 Ethics Concern

Ethics approval for this study was obtained from the University of Auckland Ethics Approval Committee in August, 2009 (Ref. 2009/332).

Chapter Four

Results

In order to answer the research questions, first the descriptive statistics of the participants' WM and reading performance are described in sections one and two. Then inferential analyses including the correlation and regression results for each proficiency group are discussed in sections three and four. Finally, the correlation results for the WM and reading measures are described in sections five and six.

4.1 Descriptive Statistics for Participants' WM Performance for Each Proficiency Level

140 EFL participants, 40 males and 100 females, from 16-35 years of age, comprised the participants for three proficiency levels in this research. They completed two WM measures, a RST and a MST tasks, as independent variables. The participants' performance on these measures was analyzed distinctly for each proficiency group. The descriptive statistics of the participants' performance on basic WM measures, the RST and MST processing as well as the RST and MST recalls at each proficiency level are indicated in Table 4.1.

Table 4.1

Descriptive Statistics for the Basic WM Measures for each Proficiency Level

			N	Min	Max	Mean	S.D.
RST	Pro.	Beg.	56	22	52	41.8	7.21
		Int.	43	29	52	41.86	6.27
		Adv.	41	37	54	47.68	3.89
	Recall	Beg.	56	1	30	14.19	6.66
		Int.	43	3	30	15.76	7.07
		Adv.	41	2	28	15.63	5.99
MST	Pro.	Beg.	56	33	60	54.76	5.50
		Int.	43	35	60	55.11	5.41
		Adv.	41	46	60	56.65	3.46
	Recall	Beg.	56	13	59	34.91	9.60
		Int.	43	16	55	36.72	9.32
		Adv.	41	16	51	37.78	8.00
PSTM		Beg.	56	4	19	13.00	2.84
		Int.	43	7	20	12.23	2.89
		Adv.	41	6	19	13.36	2.82

Note. RST= Reading Span Test; MST= Math Span Test; PSTM= Phonological Short Term Memory; Pro.= Processing; Beg.= Beginning; Int.= Intermediate; Adv.= Advanced

As shown in Table 4.1, the mean scores for all the measures, except for the RST processing, are close to each other. The mean scores for the RST processing at the beginning level are very close to that of the intermediate level (See section 4.3.3 for more information). However, the advanced participants had slightly higher means scores on the RST (six points higher than the intermediate participants). There is also a steady, smooth increase in the mean scores of the MST processing and MST recall for each proficiency level. Furthermore, the standard deviations and the means of PSTM are very close to each other in each proficiency group. This suggests that the participants' performance on this measure was nearly the same for each proficiency level.

The raw scores of the basic WM measures were converted into z-scores for each proficiency group, and then converted to composite scores. To do this, all the means were set to zero and all the standard deviations to one. Then the z-scores were added up to obtain a composite score. The results of the descriptive statistics for the basic WM z-scores for each proficiency group are presented in Table 4.2.

Table 4.2

Descriptive Statistics for the Working Memory Z-scores for each Proficiency Level

			N	Min	Max
RST	Pro.	Beg.	56	-2.75	1.40
		Int.	43	-2.04	1.61
		Adv.	41	-2.74	1.62
	Recall	Beg.	56	-1.97	2.37
		Int.	43	-1.80	2.01
		Adv.	41	-2.27	2.06
MST	Pro.	Beg.	56	-3.94	.967
		Int.	43	-3.71	.902
		Adv.	41	-3.07	.963
	Recall	Beg.	56	-2.28	2.50
		Int.	43	-2.22	1.96
		Adv.	41	-2.71	1.65
PSTM		Beg.	56	-3.16	2.10
		Int.	43	-1.80	2.68
		Adv.	41	-2.60	1.99

Note. Mean= 0; Standard Deviation= 1 (for all Measures); RST= Reading Span Test; MST= Math Span Test; PSTM= Phonological Short Term Memory; Pro.= Processing; Beg.= Beginning; Int.= Intermediate; Adv.= Advanced

As indicated in Table 4.2, the minimum and maximum z-scores for the RST and MST processing are lower than those for the recalls in these measures.

To obtain an index for the participants' memory capacity, composite WM z-scores (RST & MST) were created for each proficiency level. Furthermore, composite RSTMST z-scores were created to provide more stable scores for participants' memory capacity. The results of this analysis are displayed in Table 4.3.

Table 4.3

Descriptive Statistics for the Composite Memory Z-Scores for Each Proficiency Level

		N	Min	Max	Mean	S.D.
RST	Beg.	56	-3.45	3.77	0	1.47
	Int.	43	-2.58	3.19	0	1.47
	Adv.	41	-4.25	2.76	0	1.44
MST	Beg.	56	-4.35	3.11	0	1.60
	Int.	43	-4.65	2.68	0	1.64
	Adv.	41	-3.92	2.61	0	1.64
RSTMST	Beg.	56	-6.01	5.58	0	2.43
	Int.	43	-5.74	5.32	0	2.62
	Adv.	41	-5.88	4.75	0	2.64

Note. RST= Reading Span Test; MST= Math Span Test; com.= Composite; Beg.= Beginning; Int.= Intermediate; Adv.= Advanced

As indicated in Table 4.3, the standard deviations and maximum z-scores for the RSTMST composites are higher than the other composites at each proficiency level.

4.2 Descriptive Statistics for Participants' Reading Performance for each Proficiency Level

The participants in each proficiency group completed a different set of reading measures as dependent variables. Their performance was analyzed in each proficiency group to determine whether it could be predicted by their WMC. Descriptive statistics for reading performance were obtained for each proficiency level. The results of this analysis are provided in Table 4.4.

Table 4.4

Descriptive Statistics for the Level Specific Reading Measures

		N	Score Total	Min	Max	Mean	S.D.
Cloze	Beg.	56	26	4	25	10.25	4.30
	Int.	43	25	4	21	12.04	4.24
	Adv.	41	50	8	44	30.00	8.39
Short Answer	Beg.	36	18	0	16	7.75	4.70
	Int.	43	18	2	13.5	8.59	2.81
	Adv.	40	23	1.5	16	10.50	2.90
L1 Recall	Beg.	53	51	0	43	21.51	10.33
	Int.	43	80	13	71	44.55	13.91
	Adv.	40	121	30	90	64.47	16.66

Note. Beg.= Beginning; Int.= Intermediate; Adv.= Advanced

As indicated in Table 4.4, the reading measures include the scores of one cloze test, the sum of the scores for the two short-answer tests as well as those for the two the L1 recall tests at each proficiency level. There are some consistencies and inconsistencies among the descriptive statistics for the participants here. While the number of participants remain consistent for the reading measures at the intermediate level, this fluctuates for each reading measure at the beginning and advanced levels. This is because some participants, particularly at the beginning level, did not complete the entire battery of reading measures, and left one of the reading measures unanswered. In SPSS generated analyses, any unfilled answers to questions are counted as missing values for the entire test, and are thus not recorded.¹ This was the case particularly for the short-answer tests at the beginning level. Furthermore, the test performance within each group was similar across the test battery, with the exception of the advanced learners' performance on the cloze test, in which they scored higher than they did on the other tests. However, there is an inconsistent pattern in standard deviations for the participants' reading performance for proficiency level. The advanced participants have a wider standard deviation for the cloze tests than the beginning and intermediate participants

¹ SPSS can estimate missing values and it works well statistically if there is a low number of missing values. In this study, as there was a large number of missing values (20), SPSS could not yield a reliable estimated value for the missing scores.

do. Furthermore, while the intermediate and advanced participants have closer standard deviations for the short-answer tests, the beginning participants have a wider standard deviation for this measure, as indicated in Table 4.4. This suggests that the short-answer test was better able to discriminate among the participants at the beginning level and indicate individual differences by yielding a wider range of scores than for the intermediate and advanced participants.

A composite score was created for the participants' reading performance. This was to obtain a more reliable index for the participants' L2 reading ability. To create the composite score, the raw scores for each reading measure were first converted into z-scores. This allowed us to create the composite score across tests with different number of items. Then the z-scores for three kinds of reading measures were added up at each proficiency level. Descriptive statistics for the basic and composite reading z-scores were obtained for each proficiency level. The results of this analysis are displayed in Table 4.5.

Table 4.5

Descriptive Statistics for the Reading Measure Z-Scores

		N	Min	Max	Mean	S.D.
Cloze	Beg.	56	-1.45	3.42	0	1
	Int.	43	-1.89	2.10	0	1
	Adv.	41	-2.61	1.66	0	1
Short Answer	Beg.	36	-1.64	1.75	0	1
	Int.	43	-2.34	1.74	0	1
	Adv.	40	-3.09	1.89	0	1
L1 Recall	Beg.	53	-2.08	2.07	0	1
	Int.	43	-2.26	1.89	0	1
	Adv.	40	-2.06	1.53	0	1
Readingcom.	Beg.	34	-4.66	5.36	.1	2.62
	Int.	43	-5.04	5.10	0	2.31
	Adv.	39	-7.43	4.19	.01	2.31

Note. Beg.= Beginning; Int.= Intermediate; Adv.= Advanced

As indicated in Table 4.5, the reading composite has a larger range of scores than the basic reading measures. There is also a consistent pattern for the range of minimum and maximum z-scores in the cloze and reading composite for proficiency level. The range of these scores is higher at the lower proficiency levels than the higher proficiency ones.

4.3 Describing Memory Measures

In this part, a presentation of the descriptive and inferential statistical analyses of WM and reading measures is given for each proficiency group in order to describe the relationship between these measures. More specifically, first, the correlations between the WM memory measures for combined groups and for each proficiency level are described. Then the correlations between the reading measures are described for each proficiency level.

4.3.1 Descriptive Statistics for the Combined Groups

As all the participants in each proficiency group took the same measures of memory capacity, their scores were initially combined to analyze the effectiveness of the WM test battery. The memory measures comprised two complex span tasks, an RST and an MST, for the measurement of WMC and a simple span task, non-word recognition for the measurement of PSTM. The complex span tasks involved both processing and recall, and these components of WM were scored separately. In the simple span task, the participants' recall of English non-words was measured. Descriptive statistics for the five basic memory measures are provided in Table 4.6.

Table 4.6

Descriptive Statistics for the Memory Measures

	N	Min	Max	Mean	s.d.
RSTpro.	140	22	54	43.57	6.62
RSTre.	140	1	30	15.1	6.60
MSTpro.	140	33	60	55.39	4.99
MSTre	140	13	59	36.3	9.09
PSTM	140	4	20	12.87	2.87

Note. RST= Reading Span Test; MST= Math Span Test; pro.= Processing; re.= Recall; PSTM= Phonological Short Term memory

The basic memory measures comprise the RST processing and recall, the MST processing and recall, and PSTM. The mean scores for recall capacity in both WM measures are lower than the mean scores for processing capacity. Furthermore, while none of the participants were able to correctly recall every item in either the RST or the MST measures, there were some participants who managed to correctly process every item in one or other of the measures.

In addition to the RST and MST composites, to reduce the task-specific factors and have a more reliable measure of WMC (e.g., Conway et al., 2005), one further WM composite was also created. It included the RSTMST. Descriptive statistics for the composite scores were obtained. The results of this analysis are displayed in Table 4.7.

Table 4.7

Descriptive Statistics for the Composite Memory Measure Z-scores

	N	Min	Max	Mean	S.D.
RSTMSTcom.	140	-7.13	5.32	0	2.57
RSTcom.	140	-4.03	3.53	0	1.48
MSTcom.	140	-5.06	3.02	0	1.63

Note. RST= Reading Span Test; MST= Math Span Test; com.= Composite; PSTM= Phonological Short Term Memory

As indicated in Table 4.7, of the WM composites, RSTMST includes the highest standard deviation. This suggests that the RSTMST could indicate the widest range of the participants' performance on memory measures. Other composites differ minimally from each other in indicating the range of the participants' performance on memory measures.

4.3.2 Correlations between the Memory Measures for Combined Groups

Correlations between the WM measures were obtained for the combined groups. This was to determine the strength of relationship between these measures, which in turn, allowed us to see which WM measures shared the same kind of cognitive processes. The results of this analysis are presented in Table 4.8. They include the correlations for both the basic and composite WM scores.

Table 4.8

Correlations between the Memory Measures

	RSTMST com.	RST com.	MST com.	RSTpro.	RSTre.	MSTpro.	MSTre.	PSTM
RSTMST com.	1							
RST com.	.808**	1						
MST com.	.845**	.367**	1					
RSTpro.	.583**	.740**	.248**	1				
RSTre.	.613**	.740**	.295**	.096	1			
MSTpro.	.647**	.225**	.816**	.281**	.052	1		
MSTre.	.732**	.374**	.816**	.123	.430**	.331**	1	
PSTM	.227**	.204*	.173*	.070	.232**	.066	.216*	1

Note. n=140; RST= Reading Span Test; MST= Math Span Test; com.= Composite; PSTM= Phonological Short-Term Memory; **p <.01 (2-tailed); *p <.05 (2-tailed)

As indicated in Table 4.8, there are significant correlations between the scores of the RSTMST, RST, and MST composites and those of the basic memory measures. Of the basic memory measures, RST processing correlates significantly just with MST processing ($r = .281, p < .01$). Furthermore, the RST recall correlates with the MST recall as well as with PSTM. More interestingly, there are no significant correlations between PSTM and the processing measures, but there are significant correlations between PSTM and the recall measures. This suggests that PSTM may be more like the recall measures, and less like the processing measures in the RST or the MST. Moreover, while there is a significant correlation between the MST processing and recall ($r = .331, p < .01$), there is no significant correlation between the RST processing and recall.

4.3.3 ANOVA Results

A one-way ANOVA was run for both the basic and composite WM scores to see whether there are any significant differences between the participants' WM performance for three proficiency levels. The results of this analysis are displayed in Table 4.9.

Table 4.9

One Way ANOVA Results between the Proficiency Groups for Memory Measures

	F	Sig.	η^2
RSTMSTcom.	6.42	.002	.08
RSTcom.	7.44	.001	.09
MSTcom.	2.29	.105	.03
RSTpro.	13.1	.000	.16
RSTre.	.876	.419	.01
MSTpro.	1.97	.142	.02
MSTre.	1.24	.290	.01
PSTM	1.74	.178	.02

Note. n= 140; df= 2; RST = Reading Span Test; MST= Math Span Test; PSTM= Phonological Short Term Memory; com.= Composite

As indicated in Table 4.9, at the first level of ANOVA results, there were significant differences for the participants' performance on the RST processing ($F(2, 137) = 13.10$, $P = .000$), the RST composite ($F(2, 137) = 7.44$, $P = .001$), and the RSTMST composite ($F(2, 137) = 6.42$, $P = .002$) between proficiency groups. Post hoc Tukey analysis indicated that the advanced group had significantly higher RST processing than the beginning ($t = -4.676$; $p = .000$) and intermediate groups ($t = -5.081$; $p = .002$). It also indicated that the advanced group had significantly higher RST composite ($t = -2.988$; $p = .039$) scores than the intermediate group. However, the difference for the participants' performance on the RSTMST composite between the proficiency groups did not reach a significant level. Furthermore, there were no significant differences between the beginning and intermediate groups on the memory measures.

4.3.4 Correlations between the Memory Measures for each Proficiency Level

As differences were found between the proficiency groups, correlations were then run separately for each group. This was to determine the strength of relationship between both the basic and WM composites in each proficiency group. It was also consistent with the design of the study where correlations between the memory and reading measures could be obtained for each proficiency group. Furthermore, the correlation results for each proficiency group could be compared with those of the other groups. The results of this analysis are presented in Table 4.10, 4.11, and 4.12 for the beginning, intermediate and advanced levels respectively.

Table 4.10

Correlations between the Memory Measures at the Beginning Level

	RSTMST com.	RST com.	MST com.	RSTpro.	RSTre.	MSTpro.	MSTre.	PSTM
RSTMST com.	1							
RST com.	.768**	1						
MST com.	.807**	.241	1					
RSTpro.	.503**	.740**	.080	1				
RSTre.	.633**	.740**	.276*	.095	1			
MSTpro.	.584**	.088	.803**	.103	.028	1		
MSTre.	.711**	.298*	.803**	.026	.415**	.288*	1	
PSTM	.143	.192	.039	.045	.239	.058	.005	1

Note. n= 56; RST= Reading Span Test; MST= Math Span Test; com.= Composite; Pro.= Processing; re.= Recall; PSTM= Phonological Short-Term Memory; ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

As seen in Table 4.10, there are two medium correlations between the basic memory measures between the RST and MST recall ($r = .415$, $p < .01$), and between the MST processing and the MST recall ($r = .288$, $p < .05$). There are also significant correlations between the RSTMST composite and all the WM measures except for PSTM. Further correlations are between the RST and MST composites and the MST and RST recalls respectively. This suggests that the RST and MST composites might tap the same capacities. Furthermore, while there is a significant correlation between the MST processing and recall ($r = .288$, $p < .05$), there is no significant correlation between the RST processing and recall.

The results of the correlation analysis to determine the strength of the relationship between the WM measures at the intermediate level are presented in Table 4.11. The results include the correlations for both the basic and composite WM scores.

Table 4.11

Correlations between the Memory Measures at the Intermediate Level

	RSTMST com.	RST com.	MST com.	RSTpro.	RSTre.	MSTpro.	MSTre.	PSTM
RSTMST com.	1							
RST com.	.822**	1						
MST com.	.860**	.417**	1					
RSTpro.	.626**	.737**	.339*	1				
RSTre.	.586**	.737**	.275	.085	1			
MSTpro.	.657**	.254	.821**	.397**	-.022	1		
MSTre.	.755**	.430**	.821**	.160	.474**	.348*	1	
PSTM	.383*	.366*	.283	-.015	.555**	-.032	.497**	1

Note. n= 43; RST= Reading Span Test; MST= Math Span Test; com.= Composite; Pro.= Processing; re.= Recall; PSTM= Phonological Short-Term Memory; ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

Similar to the results at the beginning level, the RSTMST composite scores correlated with all the other WM scores. Furthermore, the MST processing scores correlated with those of the MST recall ($r=.348$, $p < .05$), but the RST processing scores did not correlate with those of the RST recall. There are also medium to high correlations for the basic WM scores here. However, unlike the results at the beginning level, PSTM correlated with all the basic memory measures with the exception of the processing measures. Of these memory measures, the RST recall had a high correlation with PSTM ($r=.555$, $p < .01$). There was also a high correlation between the recall composite and PSTM ($r=.613$, $p < .01$). This suggests that PSTM may be more like the recall measures and less like the processing measures at the intermediate level.

The correlations between the WM measures for the advanced level are presented in Table 4.12. The results include the correlations for both the basic and composite WM scores.

Table 4.12

Correlations between the Memory Measures at the Advanced Level

	RSTMST com.	RST com.	MST com.	RSTpro.	RSTre.	MSTpro.	MSTre.	PSTM
RSTMST com.	1							
RST com.	.832**	1						
MST com.	.873**	.456**	1					
RSTpro.	.598**	.725**	.322*	1				
RSTre.	.608**	.725**	.338*	.051	1			
MSTpro.	.728**	.393*	.822**	.377*	.193	1		
MSTre.	.707**	.356*	.822**	.153	.363*	.352*	1	
PSTM	.127	-.059	.255	.063	-.148	.192	.228	1

Note. n= 41; RST = Reading Span Test; MST= Math Span Test; com.= Composite; Pro.= Processing; re.= Recall; PSTM= Phonological Short-Term Memory; ** p < 0.01 (2-tailed); * p < 0.05 level (2-tailed)

There are three medium correlations between the basic WM measures at the advanced level. More specifically, these correlations are between the RST and MST processing ($r = .377$, $p < .05$), the RST and MST recalls ($r = .363$, $p < .05$) and the MST processing and the MST recall ($r = .352$, $p < .05$). There are also significant correlations between the scores of the RSTMST, RST, and MST composites and those of basic and composite WM except for PSTM. This suggests that both the WM measures (RST & MST) tap into the same construct. Furthermore, similar to the results at the beginning and intermediate levels, while there is a significant correlation between the MST processing and recall ($r = .352$, $p < .05$), there is no significant correlation between the RST processing and recall. Unlike the results at the beginning and intermediate levels, there are also significant correlations between the scores of the MST composite and those of the RST processing ($r = .322$, $p < .05$) and recall ($r = .338$, $p < .05$).

To summarize, the results for the WM composites were nearly the same for each proficiency level except for the correlation between the RST and MST composites. Unlike the results at the beginning level, there was a significant correlation between the RST and MST composites at the intermediate and advanced levels. Furthermore, unlike the results at the beginning and advanced levels, there were significant correlations between PSTM and the basic and composite WM scores, except for the MST composite and processing ones at the intermediate level. Finally, there was a significant correlation between the RST processing and the MST processing at the intermediate and advanced levels, but there was no correlation between these measures at the beginning level.

4.4 Describing Reading Measures

4.4.1 Correlations for Reading Measures

To determine the strength of the relationship between the reading measures at the beginning level, correlation results were obtained between all the measures including the basic and composite reading scores. The basic reading scores were obtained from the cloze test, the two short-answer tests and the two L1 recall tests. The results of this analysis are presented in Table 4.13.

Table 4.13

Correlation Results for the Reading Measures at the Beginning Level

	Reading com.	Cloze	Short-Answer	L1 Recall
Reading com.	1			
Cloze	.880**	1		
Short-Answer	.872**	.644**	1	
L1 Recall	.800**	.428**	.546**	1

Note. Reading com.= Reading Composite; n= 34 (Reading com.); n= 56 (Cloze); n= 53 (L1 Recall); n= 36 (Short Answer); ** $p < 0.01$ (2-tailed)

As indicated in Table 4.13, there are significant correlations between all these measures. This suggests that they all measure similar constructs. More specifically, there were medium to high correlations between the basic reading measures here. The high correlations were between the cloze and short-answer ($r = .644, p < .01$) as well as the short-answer and the L1

recall tests ($r = .546, p < .01$). There was a medium correlation between the cloze and the L1 recall tests ($r = .428, p < .01$). This suggests that the kind of cognitive and linguistic processes in the short-answer test tend to be more similar to that of the cloze and the L1 recall tests.

Correlations between the reading measures were obtained for the intermediate level. Similar to the results at the beginning level, there were significant correlations between these measures. This suggests that they all measure the similar construct. However, unlike the results at the beginning level, there were just medium correlations between the basic reading measures, specifically, between the cloze and the short answer tests ($r = .445, p < .01$), the cloze and the L1 recall tests ($r = .303, p < .05$), and the L1 recall and the short-answer tests ($r = .434, p < .01$). The results of this analysis are displayed in Table 4.14.

Table 4.14

Correlation Results for the Reading Measures at the Intermediate Level

	Reading com.	Cloze	Short-Answer	L1 Recall
Reading com.	1			
Cloze	.755**	1		
Short-Answer	.812**	.445**	1	
L1 Recall	.750**	.303*	.434**	1

Note. $n=43$; Reading com.= Reading Composite; ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

Correlations between the reading measures were also obtained for the advanced level. The results of this analysis are presented in Table 4.15.

Table 4.15

Correlation Results for the Reading Measures at the Advanced Level

	Reading com.	Cloze	Short-Answer	L1 Recall
Reading com.	1			
Cloze	.750**	1		
Short-Answer	.803**	.431**	1	
L1 Recall	.730**	.269	.399*	1

Note. Reading com.= Reading Composite; n= 39 (Reading com.); n= 41 (Cloze); n= 40 (Short-Answer); n= 40 (L1 Recall); ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

At the advanced level, all the measures correlate with one another, with the exception of the L1 recall and cloze test. It is in this respect that the advanced level performance differs from the beginning and intermediate levels. This suggests that the kind of cognitive and linguistic processes required by the L1 recall test are different from those for the cloze test at this level. Similar to the results at the intermediate level, there are just medium correlations between the basic reading measures here. More specifically, these correlations are between the cloze and short-answer tests ($r = .431, p < 0.01$) as well as the short answer and L1 recall tests ($r = .399, p < .05$). However, unlike the results for the beginning level, there is not a high correlation between the basic reading measures at the intermediate and advanced levels.

4.5 Research Question One: Is there a Relationship between WMC and L2 Reading Ability? Does this Relationship Differ According to Proficiency Level?

This question was explored through three inferential analyses; the results of each analysis will be discussed in turn in sections three, four and five.

4.5.1 Inferential Analysis at the Beginning Level

To view how the results could be generalized to the population that the sample was randomly drawn from, an inferential analysis was conducted. To obtain the results for the inferential analysis at the beginning level, first correlations between dependent (L2 reading ability) and independent (WMC) variables were achieved. Then regressions were run for any significant

correlations between these variables to determine the contribution of the independent variables to explaining the variability in the dependent variables.

4.5.1.1 Correlation Results

To determine the strength of the relationship between L2 reading ability, as a dependent variable, and WMC, as an independent variable, correlations between the measures of these two variables were obtained for each proficiency level. The results of this analysis at the beginning level are displayed in Table 4.16. The results include the correlations between the basic and composite reading scores on the one hand and the basic and composite memory scores on the other. The basic memory scores include the scores of the RST processing and recall, as well as the MST processing and recall measures. The composite memory scores include those of the RSTMST, RST, and MST composites. The basic reading scores include the scores of the cloze test, the two short-answer tests and two L1 recall tests. The composite reading score was created from the basic reading z-scores.

Table 4.16

Correlation Results between the Reading and Memory Measures at the Beginning Level

	Reading com. (n= 34)	Cloze (n= 56)	Short Answer (n= 36)	L1 Recall (n= 53)
RSTMST com.	.324	.248	.464**	.234
RST com.	.485**	.395**	.501**	.308*
MST com.	.007	.011	.242	.068
RSTpro.	.541**	.375**	.489**	.378**
RSTre.	.181	.209	.249	.079
MSTpro.	.171	.115	.206	.108
MSTre.	-.122	-.097	.193	-.001

Note. RST= Reading Span Test; MST= Math Span Test; com.= Composite; pro.= Processing; re.= Recall; ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

As indicated in Table 4.16, there are some significant correlations between the reading and memory scores here. More specifically, the RST processing and RST composite scores significantly correlate with each of the basic and composite reading scores. There is also a

significant correlation between the RSTMST composite and the scores on the short- answer test. All the correlations are in the medium to high range for the beginning level.

As the results indicate in Table 4.16, there are no significant correlations between the scores of the basic WM measures (the MST processing and recall, RST recall) or those of the MST composite, and the scores of the reading measures. This suggests that WM, as measured by the MST, may have required a different kind of cognitive process from L2 reading ability measures at the beginning level.

4.5.1.2 Regression Results

Regression analyses were run to determine to what extent the various WMC measures make a unique contribution to predicting performance on the different reading measures. They were run where there were significant correlations between the WM and L2 reading measures.

Before the analyses were run, the regression assumptions were checked by looking at the data for multicollinearity, normality including outliers, linearity, and homoscedasticity and independence of residuals. This was done by looking at the Normal- Probability Plot (P-P) of the regression standardized residual and the scatterplot. In the Normal P-P Plot, the points lay in a reasonably straight diagonal line from bottom left to top right with no distinct curvature in them. In the Scatterplot of the standardized residuals, there was no systematic pattern (e.g., curvilinear or higher on one side than the other) to the residuals, and they were rectangularly distributed, with most of the scores concentrated in the center. This suggests that there were no major deviations from normality, linearity, and homoscedasticity. The outliers were checked by looking at the standardized residuals in the residual statistics output, where no points should be above 3.0 or below -3.0 (Larsen-Hall, 2010). The standardized residuals for all models were within this range (-3.0 to 3.0). Independence of residuals was checked by the Durbin-Watson value, where a value around 2 is considered satisfactory. In all the models at the beginning level, this value was in a satisfactory range of 1.623 – 2.70. This suggests that there were no correlations between the errors in these models. Finally, multicollinearity (perfect correlations between predictors) was examined by looking at the strength of the correlations between the predictors. This assumption was also met as the strength of the

correlations was satisfactory (.768) for this level. Overall, the models could be reliable and suggest that the RST could be the best predictor for the L2 reading ability at the beginning level.

Overall, 5 regressions were run at the beginning level. The results of this analysis are displayed in Table 4.17.²

Table 4.17

Regression Results for the WM Measures and Reading Composite at the Beginning Level

		R	R ²	B	F	t	p
Reading Composite	Model 1						
	RST com.	.49	.24	.82	9.85	3.13	.004
Cloze	Model 2						
	RST com.	.40	.16	.26	9.96	3.15	.003
Short Answer	Model 3						
	RSTMST com.	.46	.22	.19	9.35	3.05	.004
L1 Recall	Model 4						
	RST com.	.50	.25	.32	11.39	3.37	.002
L1 Recall	Model 5						
	RST com.	.31	.10	.20	5.35	2.31	.025

Note. n= 34 (Reading Composite); n= 56 (Cloze); n= 36 (Short Answer); n= 53 (L1 Recall); RST= Reading Span Test; com.= Composite; pro.= Processing

As indicated in Table 4.17, the L2 reading measures as well as the reading composite were put as dependent variables, the WM composites (RSTMST and RST) were considered as independent variables (predictors).

To determine how much effect can be attributed to the influence of WM measures on the reading measures, the effect size for each model was obtained from the R^2 values in the regression models. As displayed in Table 4.17, R^2 values indicated RST composite accounted for 16% of the variance for the cloze test, 25% of the variance for the short-answer test, and

² The figures for R and R^2 were rounded up to two decimal places for all regressions in this study.

10% of the variance for the L1 recall test. It accounted for 24% of the variance for the reading composite. The reading composite model had a β of .82, indicating that each increase of 1 point in the RST composite scores predicted a less than 1 point (.82) increase in the reading composite scores. The R^2 value for the RSTMST composite accounted for 22% of the variance for the short-answer test. The short-answer test model had a β value of .19, indicating that each increase of 1 point in the RSTMST composite scores predicted around a one fifth point increase in the short- answer test scores.

4.5.2 Inferential Analysis at the Intermediate Level

To obtain the results for the inferential analysis at the intermediate level, correlations between the dependent and independent variables were achieved. However, as there were no significant correlations between WM composites and reading measures, no regressions were run at the intermediate level.

4.5.2.1 Correlation Results

Correlations between WM and reading measures were obtained at the intermediate level. In comparison to the beginning level, there were no significant correlations between WM composite and reading scores. The results of this analysis are displayed in Table 4.18 where the rows present r-values with asterisks denoting significance at the $p < 0.05$ (2- tailed).

Table 4.18

Correlation Results between the Reading and Memory Measures at the Intermediate Level

	Reading com.	Cloze	Short-Answer	L1 Recall
RSTMST com.	.148	-.088	.238	.194
RST com.	.149	-.046	.292	.098
MST com.	.104	-.100	.118	.222
RSTpro.	.121	-.088	.284	.085
RSTre.	.098	.021	.146	.060
MSTpro.	.252	.000	.230	.354*
MSTre.	-.082	-.164	-.036	.010

Note. n=43; RST= Reading Span Test; MST= Math Span Test; com.= Composite; pro.= Processing; re.= Recall; * $p < 0.05$ level (2- tailed)

As indicated in Table 4.18, there is only one significant correlation between the MST processing and the L1 recall test. This is unlike the results at the beginning level, where medium to large correlations were found between the WM and the L2 reading scores. Moreover, except for the correlation between the RST composite and the short-answer test, which was reasonably significant ($r = .292$, $p = .057$), there are no significant correlations between the RST composite and the reading measures here. All these suggest that the kind of cognitive processes demanded in processing the WM and reading measures here differ from those at the beginning level.

4.5.3 Inferential Analysis at the Advanced Level

As for the procedures at the beginning and intermediate levels, inferential analysis was conducted at the advanced level. To obtain the results, correlations between the dependent and independent variables were obtained. Unlike the results at the beginning level, there were no significant correlations between WM composite and reading scores.

4.5.3.1 Correlation Results

Similar to the results at the intermediate level, the results for the correlations between WM and reading measures at the advanced level indicated fewer numbers of significant correlations than those at the beginning level. The MST processing correlates with one of the reading measures. This is consistent with the results at the intermediate level. The results of the correlation analysis are displayed in Table 4.19, where the first and second rows of figures indicate the r-values and p-values respectively.

Table 4.19

Correlations between the Reading and Memory Measures at the Advanced Level

	Reading com. (n= 39)	Cloze (n=41)	Short-Answer (n=40)	L1 Recall (n= 40)
RSTMST com.	.296	.100	.285	.280
RST com.	.239	.048	.278	.182
MST com.	.276	.118	.213	.294
RSTpro.	.272	.215	.234	.140
RSTre.	.052	-.145	.173	.108
MSTpro.	.361*	.158	.235	.419**
MSTre.	.095	.036	.116	.065

Note. RST = Reading Span Test; MST= Math Span Test; com.= Composite; pro.= Processing; re.= Recall; ** p < 0.01 (2-tailed); * p < 0.05 level (2-tailed)

As indicated in Table 4.19, there are two medium correlations between the scores of the L1 recall and reading composite on the one hand and those of the MST processing ($r = .361$, $p < .05$ & $r = .419$, $p < .01$) on the other. These results are consistent with those at the intermediate level, where the variability in reading measures could be accounted for by the math-related WM measure. However, MST processing is just one aspect of WM and as a result MST processing scores cannot be interpreted as an index for WMC.

4.5.4 Summary of the Inferential Analysis for each Proficiency Level

The results of this study found a relationship between WMC and L2 reading comprehension, and that this relationship differs according to the proficiency level. The results for each proficiency level will be summarized in turn.

At the beginning level, medium to high correlations were found between WMC, as measured by the RST and L2 reading ability. More specifically, these correlations were between the WMC, as measured by the RST and RSTMST composites, and L2 reading comprehension, as measured by the basic and composite reading measures. Medium to large effect sizes ($R^2 = .10 - .25$) were found for the regression models run for the correlations here. However, there was no significant correlation between WMC, as measured by the MST, and L2 reading ability. Table 4.20 presents a summary of the correlations found between the measures of WMC and those for L2 reading ability at the beginning level.

Table 4.20

A Summary of the Correlations at the Beginning Level

Memory Measure	Reading Measures			
	Cloze	Short answer	L1 recall	Reading com.
RST processing	.375**	.489**	.378**	.541**
RST recall				
RST composite	.395**	.501**	.308*	.485**
MST processing				
MST recall				
MST composite				
RSTMST com.		.464**		

Note. com.= composite; ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

Furthermore, two medium correlations were found between: L2 proficiency, as measured by an objective placement test and two of the L2 reading measures: the cloze test ($r = .469$, $p < .01$); and the reading composite ($r = .421$, $p < .05$). Medium to large effect sizes ($R^2 = .18 - .22$) were found for the regression models run for these correlations at this level. The results indicated that the contribution of WMC to explaining variability in L2 reading comprehension was higher than that of L2 proficiency at this level. The WMC explained, as indicated by the effect size, a wider range of variability (up to 25%) in L2 reading

comprehension than L2 proficiency (up to 22%). It also explained the variability in L2 reading comprehension as measured by a wider range of reading task types (cloze, short-answer, L1 recall and reading composite) than L2 proficiency (cloze and reading composite).

However, at the intermediate level, there were no significant correlations between WMC and L2 reading comprehension. There was just a medium correlation between the MST processing and L2 reading, as measured by the L1 recall test. Table 4.21 provides a summary of the correlations found between the measures of WMC and those for L2 reading ability at this level.

Table 4.21

A Summary of the Correlations at the Intermediate Level

Memory Measure	Reading Measures			
	Cloze	Short-answer	L1 recall	Reading com.
RST processing				
RST recall				
RST composite				
MST processing			.354*	
MST recall				
MST composite				
RSTMST com.				

Note. com.= composite; * $p < 0.05$ level (2-tailed)

The results at the intermediate level indicated that WMC made no contribution to explaining the variability in L2 reading comprehension. However, there were two medium correlations between L2 proficiency and two of the L2 reading measures: the cloze ($r = .437, p < .01$); and the reading composite ($r = .408, p < .01$). L2 proficiency accounted for 17 - 19% of variability in L2 reading comprehension.

The results for the advanced level were similar to the intermediate level. There was no significant correlation between the WMC and L2 reading ability. There were just two significant correlations between MST processing and reading scores. Table 4.22 displays a summary of the correlations found between the measures of WMC and those for L2 reading ability at this level.

Table 4.22

A Summary of the Correlations at the Advanced Level

Memory Measure	Reading Measures			
	Cloze	Short answer	L1 recall	Reading com.
RST processing				
RST recall				
RST composite				
MST processing			.419**	.361*
MST recall				
MST composite				
RSTMST com.				

Note. com.= composite; ** $p < 0.01$ (2-tailed); * $p < 0.05$ level (2-tailed)

Similar to the intermediate level, the results at the advanced level indicated that WMC made no contribution to accounting for variability in L2 reading comprehension. There were significant correlations between L2 proficiency and L2 reading measures: the reading composite ($r = .512, p < .01$); the cloze ($r = .462, p < .01$); the short-answer ($r = .375, p < .05$); and the L1 recall ($r = .310, p < .05$). Medium to high effect sizes were found for L2 proficiency here. It explained a high proportion of variability (10 – 26%) in L2 reading comprehension.

4.6 Research Question Two: Is there a Relationship between PSTM and L2 Reading Ability? Does this Relationship Differ According to Proficiency Level?

This question was addressed through correlation analysis.

4.6.1 Inferential Analysis for each Proficiency Level

4.6.1.1 Correlation Results

To determine the strength of the relationship between L2 reading ability as a dependent variable and PSTM as an independent variable, correlations were obtained between the measures of these variables for each proficiency level. L2 reading ability was measured with a cloze test, two short-answer and two L1 recall tests at each proficiency level. A composite reading score was also created from the scores of the basic reading measures. PSTM, as discussed in chapter 3, was measured through a non-word recognition task. The results for this analysis at each proficiency level are indicated in Table 4.23, where the first and second rows of figures indicate the r-values and p-values respectively.

Table 4.23

Correlations between the PSTM and the Reading Measures for each Proficiency Level

Measure		Reading Composite	Cloze	Short Answer	L1 Recall
PSTM	Beginning	.063	.089	.001	.086
	Intermediate	-.018	-.063	.100	-.077
	Advanced	.081	.005	.213	-.035

Note. PSTM= Phonological Short Term Memory

As indicated in Table 4.23, there is no significant correlation between L2 reading ability and PSTM for each proficiency group. This suggests that there is no relationship between PSTM and L2 reading ability for the three proficiency levels.

4.7 Summary

This study examined the relationship between WMC and L2 reading comprehension, and whether this relationship differs according to proficiency level. The results of this study indicated that there is a relationship between WMC and L2 reading comprehension, and that this relationship differs according to each proficiency level.

At the beginning level, medium to high correlations were found between WMC, as measured by the RST, and L2 reading comprehension. Overall, 5 regressions were run at the beginning level with medium to large effect sizes, which indicated that WMC accounted for 10 – 25% variability in L2 reading ability. However, at the intermediate and advanced levels, no significant correlations were found between WMC and L2 reading comprehension. These results indicated that WMC made a large contribution to explaining individual differences in L2 reading comprehension just at the beginning level.

The results of this study also indicated there were medium to high correlations between reading measures for the three proficiency levels. This suggested that they might have measured similar constructs. Except for the RST processing and the recall scores, there were medium to large correlations between the basic and composite WM scores for the three proficiency levels. This suggested that WM measures tapped the same construct.

Chapter Five

Discussion

5.1 An Overview

This chapter will start with a brief overview of the research questions, participants, design, and measures. The findings for research question one will then be discussed. This is followed by the sub-section entitled “Researching WM and Reading”, where the advantages of the methodology used in the current study to measure WMC, administer and score span tasks are discussed. The findings for research question two will be elaborated in the third section. The main findings for the dependent and independent variables will be discussed in the fourth and fifth sections respectively. Finally, a summary of the chapter will be presented.

5.1.1 Research Questions:

- 1) Is there a relationship between working memory and L2 reading ability? Does this relationship differ according to proficiency level?
- 2) Is there a relationship between phonological short-term memory and L2 reading ability? Does this relationship differ according to proficiency level?

5.1.2 Participants, Design and Measures

A total of 140 L1 Persian EFL learners, at the beginning, intermediate and advanced levels, participated in the study. All the participants were studying English as a foreign language in a private language school in Iran, and carried out the same tasks for WMC and PSTM. These consisted of a Persian reading span test (based on Daneman & Carpenter, 1980), a math span test (Salthouse & Babcock, 1991; Robert & Gibson, 2002), and an English non-word recognition task (adapted from Gathercole, Pickering, Hall, & Peaker, 2001). The participants carried out similar tasks as measures of their L2 reading comprehension, with the texts adjusted for proficiency level. The participants completed these measures in three sessions over a month in the language school.

5.2 Research Question One

5.2.1 Is there a Relationship between Working Memory and L2 Reading Ability? Does this Relationship Differ according to Proficiency Level?

The results of the present study found a relationship between WMC and L2 reading comprehension. These results suggest an important role for WMC in explaining individual differences in L2 reading comprehension. However, as indicated in chapter four, this role differs according to proficiency level. At the lower proficiency levels, WMC played a large role in explaining individual differences in L2 reading comprehension, whereas it played no significant role at the higher levels of proficiency. This may be due to a higher proportion of attentional resources being used in L2 reading comprehension at lower levels of proficiency. This study follows the common experimental models of reading (Grabe, 2009), i.e., Verbal-Efficiency Model and Construction-Integration Model, in explaining the processes involved in reading comprehension. These models suggest that reading is a complex cognitive task, requiring cognitive resources for processing both the low (e.g., feature analysis of the print, word identification, parsing of phrases and sentences) and high level (e.g., integrating background knowledge with pragmatic, semantic and syntactic information from the text to make inferences) aspects of reading. A good body of research (e.g., Alderson, 2000; Bernhardt, 1991; Cain & Oakhill, 2006; Grabe & Stoller, 2002; Hudson, 2007; Leslie & Caldwell, 2009) supports this view and suggests that processing of both the low and high level aspects of reading is required for successful reading. To achieve this goal, all readers need to employ their background knowledge and cognitive resources to process text information. Processing text information differs for low and high proficiency readers. For those at lower levels, much of the language processing may still be controlled, non-automatic and effortful (capacity-demanding) (e.g., N. Ellis, 2001; Schmidt, 2001; Skehan 1989). It is challenging for these readers to attend to both the lower level aspects of reading at the same time as higher level aspects. For example, research shows that word identification can be very challenging and effortful for lower proficiency readers (e.g., Perfetti, 1985; Segalowitz, Poulsen & Komoda, 1991; Stanovich, 1988). This is because word identification involves processing orthographic, phonological, and semantic information. Since these readers possess limited L2 knowledge for processing such information, while also processing the information of other aspects of reading, they may need to rely on their cognitive resources and employ much of their cognitive capacity here. In other words, the reading process may be more cognitively demanding for lower proficiency readers as it requires them to process the semantic (e.g., word meanings, assembly of idea units) and syntactic (e.g., complex

grammatical relations) information, along with that of higher level aspects of reading (e.g., making inferences). Thus, successful readers may depend more strongly on greater cognitive capacities like WM as a source of cognitive resources (based on Baddeley & Hitch's 1974 model) and individual differences (Miyake & Friedman, 1998; Sawyer & Ranta, 2001; Skehan, 2002). These results provide further support for this notion because it was those with greater WMC who did better on reading tests at the lower proficiency levels. One explanation may be that readers with higher WMC may have more cognitive resources available for processing both lower and higher aspects of reading and storing a coherent representation of the concepts (e.g., Cain & Oakhill, 2006; Daneman & Carpenter, 1980; Waters & Caplan, 1996).

In the present study, the participants with higher WMC outperformed those with lower WMC in L2 reading ability, and such outperformance was indicated by the RST and RSTMST composite scores. This suggests that reading processes at the lower proficiency levels were as capacity-demanding as the processes in the RST, and so better matched and correlated with them. However, the MST on its own did not predict individual differences in L2 reading ability at lower proficiency levels. This may have been because the MST was not as capacity-demanding as the RST, and therefore was not as good at discriminating between high and low WM participants. In other words, reading processes used at lower proficiency levels, as explained before, were capacity-demanding and effortful, whereas the processes employed for the MST, based on the participants' reports on the WM measures, may have been less cognitively demanding. The participants reported that the RST was more demanding than the MST. This was also borne out by the significant differences between the mean scores, as displayed by the t-values, for the RST and MST processing and recalls. Thus, if the MST is not demanding enough, it will not be able to distinguish participants with high WMC from those with low WMC. As a result, it was not mathematically possible to show a relationship between reading and WM using this measure.

While WM played a clear role in reading ability for the beginning level participants, at the higher proficiency levels, WMC made no contribution to explaining individual differences in L2 reading ability. As mentioned before, this might be due to the different kind of processes involved in the L2 reading ability. One possible explanation could be that, due to higher L2

knowledge, much of the language processing at these levels may be less controlled, less effortful (less capacity-demanding) and more automatic (e.g., N. Ellis, 2001; Schmidt, 2001; Skehan 1989). More specifically, since the majority of the lower level aspects of reading such as processing orthographic, phonological, semantic and syntactical information is operated automatically (i.e., without adding attention to process meaning), and automatic processes are much less cognitively demanding (e.g., Segalowitz, Segalowitz, & Wood, 1998), these participants do not rely very much on cognitive resources in processing text information. For example, research suggests that most of the high proficiency participants are able to analyze and manipulate word-internal elements such as letters and letter clusters much more easily than those at lower proficiency levels (e.g., Ehri, 1998; Segalowitz, Poulsen & Komoda, 1991; Shankweiler & Liberman, 1972). They may have also automatized the cognitive components underlying word recognition (e.g., Kintsch, 1998; Perfetti, 1985, 1999, 2007; Segalowitz, Poulsen & Komoda, 1991; Segalowitz & Segalowitz, 1993). Thus, they are able to devote their cognitive resources for higher level skills such as drawing upon prior knowledge and integrating the information in the text with such knowledge (e.g., Perfetti, 1985; Segalowitz, Poulsen & Komoda, 1991; Stanovich, 1988), and inference-making or comprehension-monitoring (e.g., Cain & Oakhill, 2006; Kintsch, 1988, 1998). As a result, successful readers at the higher proficiency levels are less likely to be dependent on their cognitive resources like WM to complete the reading tasks. This is because their greater L2 knowledge allows them to process text information much more easily. In other words, they are able to do a greater part of the reading processes automatically, which means they do not need to depend on WM resources to the same extent as the lower proficiency participants do. For this reason, WMC is not a discriminating factor between learners with different levels of reading ability at higher proficiency levels to the extent found at the lower proficiency levels. It should also be noted that the difference in the results may be due to the different reading materials used by the three groups or may be due to differences in the reading materials used by each group, that is, by differences in learner-material interaction.

These results provide further evidence to support the prior studies (Alptekin & Erçetin 2010; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004) that found a significant relationship between WMC and L2 reading ability. However, as described in chapter 2, none of these studies examined this relationship for three proficiency levels. Furthermore, none of them employed both multiple memory measures and multiple reading measures as the present

study did. For example, Lesser (2007) found a significant role for L1 WMC, as measured by an RST, in accounting for variability in post-beginner L2 Spanish learners' reading comprehension, but only within conditions where the participants were familiar with the text topics. Walter's (2004) found a high correlation ($r = 0.79$, $P < .0001$) between WMC, as measured by an L2 RST, and L2 reading comprehension for lower-intermediate participants. She argued that the structure-building ability might have played an important role in transferring reading comprehension skill from the L1 to the L2, and has linked the structure-building ability to the development of L2 WMC. She also suggested that the success of the upper-intermediate group in L2 reading comprehension relied more on the L2 reading skills (the ability to build well-structured mental representations of texts). Alptekin and Erçetin (2010) found a moderate correlation ($r = .40$, $P < .05$) between WMC, as measured by a recall-based RST, and L2 reading comprehension at the higher proficiency levels, but not between WMC, as measured by the recognition-based RST and L2 reading ability. Similar to the Alptekin and Erçetin (2010), Chun and Payne (2004) examined the relationship between WMC, as measured by a recognition-based RST, and L2 reading comprehension. However, contrary to the findings of the current study and the prior studies in L2 (e.g., Harrington & Sawyer, 2004; Lesser, 2007; Walter, 2004), they found no significant relationship between WMC and L2 reading ability. One explanation may be that similar to the results of the current study where there was no evidence of relationship between WM and reading at the intermediate level, Chun and Payne's participants were at an intermediate level, and they may have been too advanced for WM to play a significant role in reading. Furthermore, there were only 13 participants in their study; and correlations are less likely in studies with small sample sizes.

Harrington and Sawyer (1992) found a significant correlation ($r = .54$, $p < .001$) between WMC, as measured by an L2 RST, and L2 reading ability, as measured by a TOEFL reading measure at the higher proficiency level. One explanation for the inconsistency in their results might have been due to the methodological procedure they took in scoring WM measures. They operationalized WMC as the ability to temporarily maintain the sentence-final words and recall them once cued, overlooking the processing function of WM and the possible trade-offs between these two functions (Waters & Caplan, 1996). However, as explained before, the present study considered both functions (recall versus processing) in the total

score for WMC, and there was no trade-off effect for time, as this was consistent for both span tests.

It should be noted that unlike prior studies showing significant correlations between WM and reading comprehension at advanced (e.g., Alptekin & Erçetin, 2009; Harrington & Sawyer, 1992) and upper-intermediate (e.g., Walter, 2004) levels, this study found no significant correlations between WM measures and reading measures at these levels. One explanation may lie in the different types of reading measures which were used in these studies. In Harrington and Sawyer (1992), a global (TOEFL) reading measure was used; in Alptekin & Erçetin (2009), a narrative reading measure following by 20 multiple-choice questions was used; and in Walter (2004), a baseline comprehension test, where the participants were required to complete a gapped summary of the story they had just read, was used. The reading measures in the present research (cloze, short-answer, and L1 recall) were not global, but developed and adjusted for each proficiency level by the researcher to obtain a varied picture of reading comprehension. Of these measures, cloze requires implicit knowledge of vocabulary and grammar for processing text information (e.g., Alderson, 2000; Koda, 2005), while L1 recall is predominantly direct (e.g., Bernhardt, 1983; Koda, 2005), and the short-answer lies in between. All these suggest that a global and a more direct test may tap into the underlying reading construct better than a measure like cloze, at least at higher proficiency levels. One explanation may be that implicit-oriented reading measures may be more sensitive to syntactic constraints than other types of reading measures, and consequently they are more of a test of grammar than a test of reading. Future research might explore this through a parallel study that also makes use of global reading measures.

Taken together, the results of the current study concur with those of the prior studies. However, beyond the findings in those studies, the present study proposes a unique implication for further understanding the relationship between WMC and L2 reading ability. The findings of this study imply that WM, as a source of individual differences, plays an important role in second language reading ability, particularly at lower proficiency levels. This suggests that participants with higher WMC are advantaged in the early stages of becoming L2 readers.

5.2.2 Researching WM and Reading

The methodology used in the measurement of WMC in the current study differentiates it from the prior studies (e.g., Alptekin & Erçetin, 2010; Alptekin & Erçetin, 2009; Chun & Payne, 2004; Harrington & Sawyer, 1992; Lesser, 2007; Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993; Walter, 2004) in that multiple WM span tasks, a language-related (RST) task and a math related (MST) task were employed, to reduce the task-specific factors and give a more reliable index for WMC (Conway et al., 2005). As the results of this study suggested, these span tests managed to account for a wider range of variability in L2 reading ability at lower proficiency levels. Furthermore, they helped us to distinguish the reading processes at the lower and higher proficiency levels, and to determine how demanding they would be. In other words, if there had only been one span test in this study, it would have been difficult to explain the kind of reading processes for each proficiency level. This lends further credence to Conway et al.'s (2005) claim that using multiple, reliable measures of WM could be an ideal method to obtain an index for WMC (p., 780).

The findings of this study suggest that the RST and MST, as two complex span tasks, could be reliable measures of WMC. This is because, as indicated in the methodology chapter, these two span tests had a high internal consistency (as indicated by Cronbach's Alpha) and correlated significantly, implying that they tap the same construct. Moreover, the RST could be a good predictor for L2 reading ability as it accounted for the variability in it at the beginning level. It should be noted that the present study used the MST as a measure of WMC, and found it not to be a good predictor of L2 reading ability. Like the results in prior studies, this study found the RST may better explain variability in L2 reading ability, but for lower proficiency levels.

Furthermore, unlike some prior studies (e.g., Harrington & Sawyer, 1992; Osaka & Osaka, 1992; Osaka et al., 1993) which established WMC based on storage capacity scores alone, this study operationalized WMC based on the storage and processing capacity scores. It also created composite z-scores from either span tasks (e.g., RSTMST). Correlation among the span tasks ranged from .281**, $p < .001$ to .430**, $p < .001$, suggesting that they tap the same ability, but they are not identical (Conway et al., 2005). Some prior L2 studies (e.g., Alptekin & Erçetin, 2010; Alptekin & Erçetin, 2009; Lesser, 2007; Walter, 2004) also used

WM composite scores as an index for WMC. In contrast, the WM composites used in these studies were created from just a single type of span task, an L1 or an L2 RST or both.

Overall, as the results of this study suggested, the WM composites, created from two different types of span tasks, may explain the wide range of individual differences in L2 reading ability. For example, in the current research, the RSTMST composite correlated significantly with the short-answer test at the beginning level. However, it did not correlate with the other reading measures. One explanation may be due to the processing component in this composite score, which might have employed the same kind of processes as the short-answer test, given that the RST and MST processing scores correlated for all the measures. More specifically, the reading processes in each reading test might have been different from those in the others. For example, the cloze test involves implicit knowledge of vocabulary and grammar, and it requires that the participants make use of their language analytic ability to arrive at the right choice. The L1 recall test involves explicit knowledge of lexical and grammatical items. The short-answer test lies in between. It involves processing a proportion of explicit and implicit information of the given text, as one-third of the questions in the short-answer test were inferential, and needed the execution of higher level processes. So, it seems that only the processes in the short-answer test matched better with those in the RSTMST composite.

The methodology employed in administering the span tasks in the current study was also different from those used in prior studies (e.g., Alptekin & Erçetin, 2010; Alptekin & Erçetin, 2009; Chun & Payne, 2004; Lesser, 2007; Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993; Walter, 2004). Unlike prior studies where participants were required to write down their judgements and immediately recalled targets on an answer sheet, or type them into the computerized version program, participants in this study were required to just focus on the task and orally produce their judgments or immediately recalled items for the researcher for recording purposes. This method of administration may have an advantage over the prior methods, in terms of timing. Reflecting on oral judgements and oral immediate recalls requires less time than written responses or typing responses into the computer program. Accordingly, as the content of WM is subject to decay, then the immediately recalled items in the written form and computerized version could be further exposed to decay. As a result, it

may influence the participants' actual performance on the span tasks. This is regardless of the fact that some participants may be faster at typing the immediately recalled items than the others in the computerized versions of RST, and this simply requires lower reaction times which in turn leads to higher WM scores for these participants. In contrast, the participants in this study all had the same amount of the processing and recall measurements, and they mostly reported that they were able to recall the targets they had stored temporarily, and that yielding oral responses had not interrupted the information flow. This is consistent with the methodology used in a recent study (Mackey & Sachs, 2011) in which the participants' immediately recalled items in a listening span test were audio-recorded. To elucidate further how these methods of administration could influence the participants' memory performance, further research should be conducted.

The scoring procedure used in the current research also distinguishes it from the prior studies (e.g., Harrington & Sawyer, 1992; Osaka & Osaka, 1992; Osaka et al., 1993) in three respects. First, the processing and storage capacities of WM were scored separately, but not interpreted or discussed individually because WM is recall *while* processing or the ability to *simultaneously* store and process information (e.g., Baddeley, 2007). Consistent with Conway et al.'s (2005) view, the processing scores were high enough to suggest participants were processing items while recalling: WM was being tapped into. The range of mean scores for RST processing and MST processing were 77.5 – 88.2% and 91 – 94% respectively. However, unlike Conway et al. (2005), processing scores were included in overall score of WMC. Conway et al. (2005) argue that processing scores must be maintained at 85% or above to make sure that WM is being measured. If this requirement is met, they suggest that recall scores could be obtained and counted as an index for WMC. However, as the current research indicated, there are individual differences in both processing and recall abilities of WM. Conway et al. (2005) included speed of processing as a variable; however, in this study time was set as a constant. Thus, using Conway et al.'s method of scoring may result in overlooking individual differences in processing ability of WM. Conway et al.'s method may work well in scoring computerized version of span tests where there is a variable of time for processing ability as it may differ across individuals.

Second, following Friedman and Miyake (2005), this study employed the total number of words recalled as it was a more reliable method for scoring the storage capacity of WM. In this method, the sum of the correctly recalled elements from all sets, regardless of whether the elements in each set are all recalled or not, is counted for the storage capacity score. In Conway et al.'s (2005) term, this is "partial-credit scoring" which is used to obtain recall scores for individuals whose processing scores meet the requirement (85% or above). In the current research, one point was allocated to each perfectly recalled item and half a point to each partially recalled one for the RST. This method of scoring is supported by the most recent research (Juffs & Harrington, 2011) where it is argued to provide "a finer discrimination between individuals and be more reliable" (p., 144). To control any recency effect (Baddeley & Hitch, 1993), no points were given to the targets in sentences or math problems appearing in final positions in sets if they were recalled first. In contrast, a less reliable method of scoring, the maximum number of words recalled for at least three sets (Friedman & Miyake, 2005), was used to obtain the span score in Osaka & Osaka (1992), and Osaka et al. (1993).

The same method was also used in the scoring of processing capacity in this study. The total number of correct judgements on the sentences, regardless of whether the target in each of them had been recalled correctly or not, was regarded as the processing capacity score. The advantage of this scoring procedure for processing and storage capacities, other than being more reliable (Friedman & Miyake, 2005), is that it may involve a wider range of scores, better discrimination between high and low capacity participants, as it counts even partially-correct responses, as well as all correct judgments in the total scores of storage and processing respectively.

Overall, the scoring method used in the current research is supported as being the best current practice by the most recent research (Juffs & Harrington, 2011) in the literature. Juffs and Harrington (2011) argue that as processing and recall scores can both contribute to the measurement of WM, inclusion of these scores in L2 WMC studies could result in widespread recognition of individual differences in WM and consequently should be paid further attention (p., 144). This method of scoring had a large impact on the results in the current research. Processing and recall correlated differently with the reading measures,

suggesting the benefits of making this distinction. Furthermore, it helped to see the predictive power of each of these abilities in the overall WMC scores. However, it is hard to see this predictive power in Conway et al.'s (2005) method of scoring, as explained above, as well as in Waters and Caplan's computerized version of RST. In the latter, this may be due to the reaction reading time scores which are included in creating WM composite scores besides processing and recall scores. These reaction time scores could vary and result in trade-offs between processing and recall abilities. They may also vary as a function of variables other than the ability of processing sentences (e.g., anxiety; imagination of an event during processing; parallel flow of thought). In both conditions, it may result in not having true scores of individual's WM composite scores which in turn impact on the reliability of the measure. However, in the current research, processing time was consistent and this time was just enough to read each sentence aloud or do each math problem without having extra time to rehearse the targets. This suggests that variability in processing and recall scores could have been due to individual differences in each of these abilities as there had not been any trade-off effect here.

5.3 Research Question Two

5.3.1 Is there a Relationship between Phonological Short-term Memory and L2 Reading Ability? Does this Relationship Differ according to Proficiency Level?

The results of this study indicated that there is no relationship between PSTM and L2 reading comprehension. One possible explanation could be that PSTM which involves maintaining phonological information via rehearsal mechanisms (e.g., Baddeley, 1986, 2007; Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Gathercole et al., 1992) taps just the storage capacity of WM. This is supported by the results of this study: there was a significant correlation between just the storage component of WM, as measured by the two span tasks (RST & MST), and PSTM, as measured by a non-word recognition task (.232**, $P < .01$; .216*, $p < .05$). This suggests that the storage capacity of span tests and PSTM may tap the same construct. In contrast, reading comprehension, as a complex cognitive task, involves activating world knowledge, processing text information and maintaining their representations in WM for further analysis, such as integrating idea units and making inferences (e.g., Bernhardt, 1991; Cain & Oakhill, 2006; Hudson, 2007; Kintsch, 1998; Koda,

2005). Thus, reading comprehension, as a complex cognitive task, taps both components of WM, processing and storage capacities. For this reason, PSTM and L2 reading measures may not employ the same kind of cognitive resources. Together with the above results, these findings suggest that storage alone does not play a significant role in reading comprehension, providing divergent validity for the RST as a measure of WM which taps both processing and storage.

As indicated by the results of this study, there was a significant correlation between just the recall scores and PSTM scores. This provides further support for the idea that WM is a separate construct from PSTM and also what aspect of WM differs from or resembles to PSTM.

The results of this study are consistent with prior results in L1 (e.g., Daneman & Carpenter) and L2 studies (e.g., Harrington & Sawyer, 1992; Hummel, 2009; Kormos & Sáfár, 2008) which specifically investigated whether there is a relationship between PSTM (as well as WMC) and reading comprehension. For example, Daneman and Carpenter (1980) investigated the role of individual difference in WMC and PSTM in L1 reading comprehension. The results in their study indicated that the variability in L1 reading ability could be explained by the WMC, but not by the PSTM, as measured by a word or digit span test. Similarly, Harrington and Sawyer (1992) examined the relationship between PSTM and WMC on the one hand and L2 reading comprehension on the other. The results of their study indicated that WMC, rather than PSTM, could explain individual differences in L2 reading comprehension. Further support comes from Hummel's (2009) study, which investigated whether there is a relationship between PSTM and aptitude on the one hand and second language proficiency on the other. Her results indicated a significant correlation between PSTM and L2 proficiency in general ($r = .35$, $P < .05$), and PSTM and the vocabulary ($r = .36$, $p < .01$) and grammar ($r = .33$, $p < .01$) sections of the L2 proficiency test in particular, but not with the reading section. She also found that the relationship between PSTM and L2 proficiency remained significant in lower proficiency participants, but disappeared in participants with higher proficiency levels.

Unlike the results in Hummel (2009), Kormos and Sáfár (2008) found no significant correlation between PSTM and L2 language skills, including reading. They investigated whether there is a relationship between PSTM and WMC and performance in L2 language skills, as measured by an L2 proficiency test. Their results indicated that there was no significant correlation between PSTM and L2 language skills, but there was a significant correlation between WMC, as measured by a backward digit span test, and several L2 language skills (reading, listening, and speaking), with the exception of writing. Kormos and Sáfár (2008) suggested that PSTM and WM are distinct constructs, and play a different role in L2 language skills.

In contrast to the results of this study and prior studies discussed above, Chun and Payne (2004) found evidence of a relationship between PSTM and L2 reading comprehension in their study of the role of individual differences in PSTM and WMC in L2 reading comprehension and vocabulary learning. The results of their study indicated that there was a high negative correlation ($-.853^{**}$, $p = .000$) between PSTM and look-up behavior while reading an L2 text. They argued that participants with low PSTM looked up L2 words more than participants with high PSTM, implying the participants with higher PSTM have a better command of L2 vocabularies, and consequently a better L2 reading ability. Chun and Payne suggested that the participants with low PSTM capacity used other factors (e.g., features of the software application) to compensate for memory limitations while reading an L2 text. This supports the idea that with multiple abilities in WM, individuals make up for weaknesses in some areas by using other strengths (e.g., Mackey et al., 2002). While this study did show a connection between PSTM and a specific reading strategy, it should be noted that Chun and Payne did not find a direct link between PSTM and reading comprehension.

Overall, the current study fits with prior L2 studies which found no significant relationship between PSTM and L2 reading. However, it should be noted that vocabulary research suggests that there is a relationship between PSTM and L1 reading (e.g., de Jong & de Jong, 1996; Engle, Carullo & Collins, 1991; Gathercole, Willis, Emslie & Baddeley, 1991, 1992) and L2 reading (e.g., Masuora & Gathercole, 2005; Papagno, Valentine & Baddeley, 1991; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995), particularly at lower

proficiency levels, mediated by L1 and L2 vocabulary development respectively, which in turn impacts on L2 reading ability. For example, in a longitudinal study, Service (1992) found a strong relationship between PSTM and L2 reading and listening comprehension as well as writing ability over two years later for Finnish-speaking elementary school children. In a follow-up study, Service and Kohonen (1995) explored the relationship between PSTM and L2 reading comprehension as evidence of a relationship between PSTM and L2 vocabulary learning. This relationship may be established through the processes where readers convert letters into sounds while reading a text, and store them temporarily in PSTM (or verbal short-term memory) until the last letter is translated. Then they blend the full sequence of sounds into a word. This explanation is consistent with the Baddeley's (2006) study in which he asserts there is a relationship between PSTM and reading ability "possibly in a number of ways, ranging from learning letter-sound correspondences, through sound blending, possibly up to the level of text comprehension (p., 13)." Phonological awareness or sensitivity is also argued to play an important role in reading ability, even more than that of PSTM (e.g., de Jong & Olson, 2004; Pennington, Van Orden, Kirson, & Haith, 1991). Phonological sensitivity is strongly related to PSTM, and defined as "the ability to detect and manipulate the sound units of one's oral language" (de Jong, 2006, p., 37).

However, at the higher proficiency levels, the role of PSTM diminishes, and as Masoura and Gathercole (2005) argued, may be replaced with "other factors which impose increasingly significant constraints on the ease of vocabulary expansion (p., 423)." One of these factors, they added, could be existing long-term phonological knowledge which mediates L2 vocabulary learning once participants' familiarity with the language increases, reducing the role previously played by PSTM. This argument is consistent with Cheung's (1996) study, in which he found a relationship between PSTM and L2 vocabulary learning with lower proficiency EFL learners, but not with those at higher proficiency levels. Further support comes from Hummel's (2009) findings that PSTM could explain individual differences in L2 proficiency among ESL learners at lower levels of proficiency, but not among those at higher levels.

Thus, if this is the case, one anticipates a relationship between PSTM and L2 reading comprehension. However, as the results of the current study and those of prior studies indicated, there should be at least two conditions available to explore this relationship: (1) lower proficiency participants, and (2) a measure or measures of L2 vocabulary because the relationship between PSTM and reading seems to be mediated by vocabulary learning. As the current study, similar to Kormos and Sáfár (2008), included lower proficiency participants, but no L2 vocabulary measures, it is not then surprising that no direct connection to reading comprehension was found. Hummel (2009) included an L2 vocabulary measure, but his participants may have been too advanced for a relationship between PSTM and vocabulary learning to be significant. In contrast, Chun and Payne (2004) found evidence for the relationship between PSTM and L2 reading comprehension, although they did not include any lower proficiency participants in their study. However, their results should be interpreted cautiously because they were analyzed based on the amount of look-up behavior for new words, and not on a direct measure of L2 vocabulary knowledge. Furthermore, their study included a low number of participants, and L2 proficiency measures were not used. Overall, consistent with studies conducted with children learning to read in their L1, we can conclude that the role of PSTM in L2 reading is likely to be restricted to vocabulary learning at early stages. Future research could investigate the interplay between PSTM, L2 vocabulary learning and L2 reading comprehension at earlier stages of language learning, including children and adults as well as different types of reading, vocabulary, and simple span measures.

5.4 Discussion of the Dependent Variables

This section of the chapter considers the contribution of the reading measures used in the present study to our understanding of the processing involved in L2 reading ability in general and different L2 reading measures in particular. L2 reading ability was measured through three types of basic reading measures including a cloze test, two short-answer and two L1 recall tests for each proficiency level. A composite reading score was also created from the basic reading z-scores for each proficiency level. The results of the study indicated that there were significant correlations between these measures. As indicated in chapter four, there were medium to high correlations between the reading measures at the beginning level, and medium correlations at the intermediate and advanced levels, with the exception of a low

correlation between the cloze and L1 recall tests at the advanced level ($r = .269, p > .05$). Overall, these findings suggest that as there were significant correlations between these measures, they likely involved similar types of cognitive processing.

Of the correlations between the basic reading measures at each proficiency level, the one between the cloze and short-answer tests was the highest. One explanation could be that the cloze and short-answer tests both involved giving restricted answers, while the L1 recall required learners to write an expanded response. This gives more room for participants to produce different types of mistakes. As a result, the strength of the correlation between the L1 recall test and the cloze as well as the short-answer test tends to be weaker. This is consistent with the participants' reports on the L1 recall test as they believed that they had approached the L1 recall test differentially from the cloze or short-answer tests. They said that the L1 recall test had been more time-consuming, but they had not found it stressful to complete as they had been able to answer freely and had not been limited to certain specific questions. Moreover, they had been allowed to write down whatever details they had comprehended from the text including those they had inferred beyond it. This suggests that the L1 recall test could indicate a broader range of variability in L2 reading ability.

Furthermore, the participants added that the cloze and the short-answer tests had been more memory demanding because they had to simultaneously maintain the text information and the given options for each test item to arrive at the most appropriate answer. This was also indicated by the stronger strength of correlations between the short-answer and memory scores at the beginning level. This could be explained by the memory operations involved in conducting these two tasks. As the prior research on reading assessment (e.g., Cain & Oakhill, 2006; Koda, 2005; Alderson, 2000) points out, a cloze test of this type involves the selection of an answer from a choice of usually 3-5 options. The short-answer test also requires a kind of selection, not from a limited number of given options, but from a range of possible options within the text. Since the selection of the answer requires the inhibition of irrelevant information followed by updating memory for the next test item, additional memory resources could be employed here (Carretti et al., 2005; Gernsbacher et al., 1990; Morris & Jones, 1990) than those by the L1 recall test. For this reason, these two tests might have employed the same kind of cognitive processes, and consequently possessed a higher

correlation. However, in the L1 recall test, the participants reported that they could easily distinguish the sequence between the ideas in the text and show that they had understood in a coherent way. This suggests that the type of reading tasks should be viewed as an important factor in the participants' performance as they may demand different cognitive resources even if they are created from the texts within the same range of readability.

5.5 Discussion of the Independent Variables

This section of the chapter considers the contribution of the memory measures used to our understanding of the processes involved in L2 reading ability. WMC was measured through two cognitively complex span tests, as a language-related test (RST) and as a math-related test (MST). Since each of these span tests measured the processing and storage components of WM, a composite WM (RST and MST composites) score was created from the scores of the processing and storage components to have an index for WMC. To reduce the task-specific factors from each of the span tests and have a more reliable measure of WMC (e.g., Conway et al., 2005), a further WM composite, RSTMST, was created from the composite WM scores. The results of the study indicated that there were high correlations between the basic (RST processing, RST recall, MST processing, and MST recall) and composite WM scores. There were also medium correlations between the basic WM measures. This suggests that these measures use the same kind of processes and tap the same construct.

As the results of the one-way ANOVA indicated, there was a significant difference between the WM mean scores, as measured by the RST processing, between the participants at the advanced level on the one hand, and those at the beginning and intermediate levels on the other. This may be due to more efficient processing at the advanced level. It is more likely that there were unexpectedly more participants with higher WMC in the advanced group than in the other groups, and therefore, they outperformed the participants at the beginning and intermediate groups in the RST processing. In other words, if this was the case, the advanced level participants had more attentional resources, and they most probably used such additional resources for processing and comprehension of the RST sentences because their recall mean scores were in the same range of 14.19, 15.76, 15.63 for the beginning, intermediate and advanced levels respectively.

Overall, these findings imply that complex span tasks (RST and MST) are two reliable cognitive tasks tapping the same construct and measuring both the processing and storage components of WMC. This adds further support to prior studies where WMC was operationalized as the performance on the complex span tests such as the RST, operation span test or counting span test both in the L1 (e.g., Daneman & Carpenter, 1980; Turner & Engle, 1989; Waters & Caplan, 1996) and the L2 (e.g., Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004). Furthermore, as the results of the present study indicated, complex span tasks may be more challenging and memory-demanding (e.g., RST) or less challenging and memory-demanding (e.g., MST). A composite WM score created from different types of span test scores could be a more reliable index for WMC and increase the construct validity of the WMC once measured (e.g., Conway et al., 2005; Turner & Engle, 1989). Finally, the results of this study support the idea that there are differences between processing and storage capacities, and that differentiating between these two could have some benefits. One of these differences could be that processing may be faster and less demanding than storage. Differentiating between processing and storage could help detect the predictive power of these abilities (e.g., Juffs & Harrington, 2011), and also the extent to which a particular task depends on each of these capacities.

The second independent variable in this study was PSTM. It is a more specific sub-component of the multi-component WM model (e.g., Baddeley & Hitch, 1974; Baddeley & Logie, 1999) for temporary storage of acoustic and speech-based information. It was measured through a non-word recognition task in this study. As the results in this study indicated (see Chapter Four, Table 4.8), there was a significant correlation between PSTM and WMC. More specifically, there were significant correlations between the PSTM scores and all the basic and composite WM scores, except those of processing (RST and MST processing as well as processing composite). This lends further credence to the multi-component model of WM proposed by Baddeley and his colleagues (Baddeley, 1986; 2007; Baddeley & Hitch, 1974; Baddeley & Logie, 1999), in which they differentiate between the functions of the phonological loop (operationalized as PSTM) and the central executive (operationalized as WMC). As they describe, the former is involved in maintaining just phonological information, whereas the latter is involved in processing information from different sub-systems of WM and long-term memory. This suggests, based on the results of this study, that PSTM involves little processing, or even if it involves processing in the form

of rehearsal (e.g., Baddeley, 1986; Turner & Engle, 1989), it is not as demanding as that of the central executive.

5.6 Summary

This chapter addressed the two research questions and provided a separate discussion for each. It was established that there is a significant relationship between WMC and L2 reading comprehension just at the lower proficiency levels. Furthermore, it was argued that there is less evidence of a relationship between PSTM and L2 reading ability: this relationship is significant just at the earlier stages of first or second language learning. This may be explained by the omission of a direct vocabulary learning measure in the current study.

It was also argued that the three tests of L2 reading ability used in this study, as a dependent variable, measured a similar construct, although they may demand cognitive resources differentially. The cloze and short answer tests were more demanding than the L1 recall test as the procedure to complete the former was different from that of the L1 recall test.

Finally, the measures of WMC (RST & MST), as an independent variable, were found to be reliable as complex cognitive tasks that tapped both the processing and storage capacities of WM. The second independent variable, PSTM, tapped just the storage capacity. These findings provided further support for the multi-component model of WM proposed by Baddeley and his colleagues (Baddeley, 1986; 2007; Baddeley & Hitch, 1974; Baddeley & Logie, 1999).

Chapter Six

Conclusion

6.1 An Overview

In this chapter, two central theoretical implications of the study will be presented in the first section. This section also includes some further contributions which may be helpful for future research into WM and reading. Then the limitations of the study are explained and directions for further research are proposed. The conclusion ends with the summary of the issues presented in the chapter.

6.2 Theoretical Implications

This study examined whether the relationship between WMC and L2 reading ability is mediated by L2 proficiency. Similar to prior studies (e.g., Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004), the present study indicated that there is a relationship between WMC and L2 reading ability. However, this study is distinguished from the prior studies in that it adds two unique theoretical implications to the research area of WM and L2 reading ability.

The first implication of this study is that the relationship between WMC and L2 reading ability differs according to proficiency level. There is a strong relationship between WMC and L2 reading at lower proficiency levels, and no significant relationship at higher levels of proficiency. In other words, WMC makes more contribution to explaining individual differences in L2 reading ability at the lower proficiency levels than those at higher proficiency levels. This suggests that low proficiency readers rely on cognitive resources such as WM more than higher proficiency readers, and therefore, at lower proficiency levels, readers with higher WMC are advantaged in their L2 reading ability than those with lower WMC.

A further implication of this study relates to explanations of reading variability, and possible distinctions between reading processes (capacity-demanding versus automatic) employed by learners of differing proficiency levels. The results of the current study imply that the L2 reading process differs according to proficiency level. At lower levels of proficiency, due to limited L2 knowledge, the reading process may still be controlled, non-automatic, and capacity demanding; whereas much of the reading process at higher proficiency levels is less controlled, less capacity demanding and more automatic because readers rely on their greater L2 knowledge. Further research on cognitive capacities and reading at different proficiency levels may clarify this possibility. It is also possible that while WMC was not found to differentiate between more and less competent readers at higher levels of proficiency, other factors such as more reading experience could play a greater role at these levels. For example, the reading process at higher proficiency levels tends to require a greater level of inference and recognition of more complex aspects of genre-based characteristics; it may be that the readers with higher order reading skills and strategies are more likely to outperform those who have not yet developed such skills and strategies sufficiently. Future research could consider the role of L2 reading experience in differentiating learners at higher levels.

Prior studies (Alptekin & Erçetin 2009; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004) found a relationship between WMC and L2 reading ability, as did the current study. However, unlike the present study, none of them examined this relationship for three proficiency levels. Two of these studies illustrated this relationship at the advanced level (Alptekin & Erçetin 2009; Harrington & Sawyer, 1992), one at the lower-intermediate level (Walter, 2004) and one at the high beginning level (Lesser, 2007). One study found no relationship at the intermediate level (Chun & Payne, 2004). Moreover, as discussed in chapter 5, the operationalization of WM as well as the methodology used in administering and scoring of WM measure differed in these studies. Because of these differences, no clear and certain explanations could be given for the kind of relationship between WM and L2 reading for each proficiency level as well as the kind of reading and memory processes at each proficiency level based on the findings of the prior studies. Finding a clear distinction between the role of WM in L2 reading by lower and higher level proficiency learners is the major new finding of the research reported here.

The current study also provides some important implications for research processes for WM and L2 reading. The findings of this study suggest that using multiple WM span tasks (in the current study), a language related (RST) task and a math related (MST) task is better than using just a single span task. One of the benefits of using two types of span tasks is that it helps to diminish task-specific factors and gives a more reliable index for WMC.

A further benefit of using multiple WM measures is that it helps to capture the complex nature of WMC. For example, this study suggests that the RST is more demanding than the MST. Therefore, the RST could be used as a reliable predictor for L2 reading ability at lower proficiency levels. This study suggests that there are differences between the processing and storage capacities of WM, and measuring these separately can help both to see the predictive power of WM measure and to detect the extent to which a particular task depends on each of these capacities.

Furthermore, the results of this study support prior research which has suggested that processing and storage capacities should both be considered for the overall memory score, and they must not be interpreted alone as an index for WMC because WM is the ability to recall while processing (e.g., Lesser, 2007; Walter, 2004).

As there are few Persian span tests, one further contribution of this study is the development and validation of an RST which can be used for L1 Persian participants. This span test was developed in the participants' L1 as prior research indicates that WM is language independent (Osaka, Osaka & Groner, 1993; Osaka & Osaka, 1992). Moreover, measuring WM in the L1 avoids conflating WM and L2 proficiency. To have an independent WM measure of reading, an MST was also developed and validated for the participants in this study, in order to ensure appropriateness for L1 Persian participants, particularly in terms of timing. As this span test successfully accounted for the kind of reading process (e.g., more automatic) at higher proficiency levels, it can also be used as a measure of WMC in future studies where native speakers of English are included.

The methodology used for administering span tasks in this study can be used for future research. As discussed in chapter five, this methodology has an advantage over those used by prior studies in terms of timing which in turn impacts on the participants' overall score and the reliability of the test. In prior studies, the participants' judgments or immediately recalled items were either written or typed which is more time-consuming than the oral responses used in the current study. Since the content of memory is subject to decay, then the immediately recalled items in either written or typed forms are likely to be exposed to greater decay than those in an oral form. Less exposure to decay may in turn allow for a more accurate measurement of WMC.

The present study also investigated whether the relationship between PSTM and L2 reading ability is mediated by L2 proficiency. This study provides further support for Harrington and Sawyer's (1992) and Hummel's (2009) studies, which suggest that PSTM does not play a direct role in L2 reading ability. This is likely to be because simple processing in PSTM (articulatory rehearsal) may not be a good predictor for multi-level processing in L2 reading ability. Moreover, it provides further support for prior studies, which suggest that just those WM tasks which involve both storage and processing functions, irrespective of their task modalities, correlate more strongly with reading comprehension measures, and not those tasks which involve just the storage capacity like digit or word span tests (e.g., Carretti et al., 2009; Daneman & Merikle, 1996). While not implied by the results of the current study, it should be noted that these findings do not preclude the possibility that PSTM may play an indirect role in L2 reading ability at least at lower proficiency levels by mediating L2 vocabulary development, which in turn impacts on L2 reading ability, as found in prior research (e.g., Masuora & Gathercole, 2005; Papagno, Valentine & Baddeley, 1991; Service, 1992).

6.3 Limitations of the Study and Directions for Further Research

Although every effort was made to conduct all the stages of this study as well as possible, this study had some limitations. The first limitation was related to the lack of an L2 vocabulary test for each proficiency level to see whether the relationship between PSTM and L2 reading ability could be mediated by L2 vocabulary learning ability as proposed by some studies

(Masuora & Gathercole, 2005; Papagno, Valentine & Baddeley, 1991; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995). Due to the large number of reading tasks (one cloze, two short-answer and two L1 recall tests) and memory measures (a RST, a MST, and a non-word recognition task), it was not logistically possible to include an L2 vocabulary test in the current study. Furthermore, the volunteer participants had limited time, and the threat of test fatigue restricted the choice of tasks used in this study. Given these constraints, a further separate study is needed to specifically investigate this relationship. The findings of future research may clarify the extent to which the relationship between PSTM and L2 reading ability is mediated by L2 vocabulary development, and any variance due to proficiency level.

Furthermore, the beginning participants in this study were not true beginners; it may be that their proficiency level was already too high to examine a relationship between PSTM and L2 reading ability. To accurately examine such a relationship, further studies are needed. These studies should focus participants with a wider range of proficiencies, particularly at the beginning level to determine how the relationship between PSTM, vocabulary learning and L2 reading ability might vary for each proficiency level.

While two measures of WM were used in the study, only one measure of PSTM was included due to the time constraints mentioned above. As different aspects of the relationship between reading and WM for each proficiency level emerged on the different WM measures, it is possible that additional PSTM measures might be helpful to illuminate any relationship between PSTM and L2 reading. These studies might use the non-word repetition test or non-word recognition test as measures of PSTM. Research suggests that non-words may yield a more reliable index of PSTM (e.g., Gathercole, Frankish, Pickering, Peaker, 1999; Trofimovich, Ammar, & Gatbonton, 2007) because they minimize the influence of background knowledge (e.g., L2 vocabulary knowledge) on PSTM as opposed to a word or digit span test. L2 reading research has not yet included studies of PSTM including multiple measures.

Another limitation with this study was related to the logistical constraints which prevented spreading WM testing over several days to avoid test fatigue. The RST was conducted first followed by the MST with a 15-minute break in between for each proficiency level. Although the participants relaxed during the break time, had some juice and food to refresh themselves, and did not report any test fatigue prior to the next test, it might be a good idea for future studies to conduct these measures with at least a day-long interval in between. This would avoid test fatigue, if there is any, as well as minimize any training effect of the first test on the second one, if there is any. Of course, the memory measures in this study were quite different in test content as one of them was verbal and the other one was mathematical. Moreover, based on the participants reporting, the RST and the MST were perceived as being different in their test procedure. In other words, in the RST, the participants were required to read each sentence and judge on their semantic acceptability and then recall the final word of each sentence. In the MST, they were required to see whether each mathematical calculation was correct or not, and then recall the second digit of the math problem, and not the last one.

Finally, in the present studies, a multiple-choice cloze test was used as one of the measures of L2 reading ability. However, as cloze tests are strongly associated with grammar knowledge, the inclusion of a more global measure may have been more appropriate. The extent to which the lack of correlation found between reading and WM at higher proficiency levels was a function of the reading measures should be further explored by future research with the inclusion of more global measures as used in previous studies. This would also allow clearer comparisons with earlier research.

Overall, this study showed a relationship between WMC and L2 reading ability and how this relationship differs according to proficiency level. In the current study, the need to develop tests at three different levels of proficiency made it necessary to limit the number of measures used, so this study did not examine the relationship between WMC and any specific aspects of the L2 reading process. To further explore the relationship between WMC and L2 reading ability, future studies should specifically examine the relationship between WMC and these specific aspects of the L2 reading process. Such aspects may include lower-level processes such as word identification and integration of idea units as well as higher-level processes such as inference making and comprehension monitoring. These studies might be able to use

a wide variety of L2 reading measures in examining the role of WMC in explaining individual differences in lower and higher aspects of reading processes.

6.4 Summary

To summarize, this study elucidated the role of WM in L2 reading ability for three proficiency levels and contributes in this way to the literature (Alptekin & Ercetin, 2009; Chun & Payne, 2004; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004). It found that WMC makes a larger contribution to explaining individual differences in L2 reading ability at lower proficiency levels because of the students' limited L2 knowledge here.

This study made some contributions to research in the area of WM and reading. It suggested that using multiple WM measures accounted for a wider range of individual differences in L2 reading ability for different proficiency levels. The methodology used in administering and scoring span tests in the current study yielded a more reliable measure of WMC. This study found no relationship between PSTM and L2 reading ability, and suggested there might be an indirect relationship between these two abilities. Using very beginner-level learners and a lack of L2 vocabulary measures were the limitations of this study to see whether there is a relationship between PSTM and L2 reading. Directions for further studies were proposed. These studies should examine the relationship between WMC and the specific abilities in the lower and higher aspects of L2 reading.

Overall, it can be concluded that the relationship between WM and L2 reading ability is stronger for lower proficiency learners because their lower-level processes may still be controlled and cognitively demanding. At higher proficiency levels, due to greater L2 knowledge, much of language processing may be more automatic and thus there is argued to be less reliance on WM for lower-level processes in comparison to lower proficiency learners. This suggests that WM may play a diminishing role in second language reading ability as language proficiency increases: the less proficient the learners are in their L2, the more dependent they will be on WMC when reading in their L2.

Appendix A

Participant Consent Form



THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Department of applied language studies and Linguistics
Fisher Building, 18 Waterloo Quadrant
+64 9 3737599 (Ext., 86252)

The University of Auckland
Private Bag 92019
Auckland, New Zealand

CONSENT FORM

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project Title: The role of short-term working memory in second language reading comprehension

Principal Researcher: Mohammadtaghi Shahnazari Dorcheh, Ph.D student, Department of Applied Language Studies and Linguistics, University of Auckland

Supervisor: Dr. Rebecca Adams, Department of Applied Language Studies and Linguistics, University of Auckland

I have read the Participant Information Sheet, have understood the nature of the research and why I have been selected. I have voluntarily participated in this research. I have had the opportunity to ask questions and have them answered to my satisfaction.

- I agree to take part in this research.
- I understand that I am free to withdraw participation at any time, and to withdraw any data traceable to me up to 30 April, 2010.
- I agree to participate in approximately 100 minutes research (two 50 minute-consecutive sessions) including no audio or videotaping.

- I understand that the information obtained from my participation in the project will be reported or published in a way that does not identify me as its source.
- I wish / do not wish to receive a feedback of my performance on research measures via a sealed envelope.
- I wish / do not wish to receive the summary of findings.
- I understand that data will be kept for 6 years, after which they will be destroyed.
- I understand that the head of school has given assurance that my participation or non-participation in this research will in no way influence my grades in the course nor my relationship with my teachers or the school.

Name:.....

Signature:.....

Date:.....

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS
ETHICS COMMITTEE ON AUGUST 15, 2009 FOR A PERIOD OF THREE YEARS,
REFERENCE 2009 / 332.

Appendix B

Head of Language School Consent Form



THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Department of applied language studies and Linguistics
Fisher Building, 18 Waterloo Quadrant
+64 9 3737599 (Ext., 86252)

The University of Auckland
Private Bag 92019
Auckland, New Zealand

CONSENT FORM

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project Title: The role of short-term working memory in second language reading comprehension

Principal Researcher: Mohammadtaghi Shahnazari Dorcheh, Ph.D student, Department of Applied Language Studies and Linguistics, University of Auckland

Supervisor: Dr. Rebecca Adams, Department of Applied Language Studies and Linguistics, University of Auckland

I have read the Participant Information Sheet, have understood the nature of the research and why this language school has been selected. I have had the opportunity to ask questions and have them answered to my satisfaction.

- I agree that you may conduct your research in this language school.
- I agree that you may approach the students who are interested in participating in your research.
- I understand that your research will take approximately 100 minutes in two 50 minute- consecutive sessions and will include no audio or videotaping.
- I provide / do not provide you with enough facilities such as classroom, language lab, etc to conduct your research.

- I wish / do not wish to receive the summary of findings in a way that does not identify its source.
- I understand that data will be kept for 6 years, after which they will be destroyed.
- I understand that participants are free to withdraw participation at any time and to withdraw any traceable data up to 30 April, 2010.
- I confirm that language learners' participation or non-participation will in no way influence their grades in their course nor their relationship with their teachers or the school.

Name:.....

Signature:.....

Date:.....

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS
ETHICS COMMITTEE ON AUGUST 15, 2009 FOR A PERIOD OF THREE YEARS,
REFERENCE 2009 / 332.

Appendix C

Reading Measures

Beginning Level

Cloze Test

متن درک مطلب زیر را بخوانید و بهترین پاسخ را برای هر جای خالی از بین گزینه های داده شده انتخاب کرده و در پاسخنامه علامت بزنید.

(Instruction: Please read the following reading passage and select the best answer from the choices given for each empty space)

Changing Ideas about Art

A lot of modern art is not realistic or beautiful. Artists want to show people a(1).... way of seeing things. They want to say(2)..... about the world they live in.(3).... paintings and sculptures disturb some(4)..... because they do not(5).... or understand(6)..... the artist is trying to(7)..... It often takes many(8)..... before an artist's work is accepted by the(9).....

In 1870, Vincent Van Gogh and Claude Monet were painting in France. They and some other painters(10).....many ideas about(11)..... They used strong,(12).... colors, and their(13).....were not always very(14)..... Their art was very(15).... from what people were(16)..... with, and few people liked it. The(17).... did not make a lot of(18)....with their paintings. Van Gogh(19).... only one painting while he was(20).....

Today, we(21)..... this group the Impressionists. Van Gogh's and Monet's paintings are in(22).... all over the world. The Impressionists are not only(23)....., they are also(24)..... People everywhere buy their(25).... in inexpensive posters and(26)..... When museums have special(27)..... of Impressionist paintings, people stand in line(28)..... for hours to(29)..... in.

Artists' ideas change(30)....., but the public's ideas(31)..... much more slowly.(32)...., artists who are(33).... popular in their(34).... are not well respected. Those who are(35).... can become(36)....., but they are(37)....

popular. Most of today's(38)..... artists will be(39).... in 100 years. Only a few of them will be(40).....

1) a) real	b) secret	c) clear	d) different
2) a) information	b) something	c) stories	d) jokes
3) a) Beautiful	b) Expensive	c) Modern	d) Amusing
4) a) people	b) experts	c) officials	d) parents
5) a) watch	b) like	c) indicate	d) follow
6) a) which	b) when	c) what	d) whom
7) a) observe	b) consider	c) improve	d) express
8) a) years	b) ages	c) experiences	d) records
9) a) public	b) government	c) college	d) museum
10) a) described	b) shared	c) planted	d) received
11) a) culture	b) cinema	c) art	d) photography
12) a) modern	b) fair	c) artificial	d) bright
13) a) papers	b) posters	c) pictures	d) positions
14) a) realistic	b) efficient	c) energetic	d) systematic
15) a) separated	b) different	c) far	d) away
16) a) associated	b) content	c) filled	d) familiar
17) a) group	b) swarm	c) crowd	d) bunch
18) a) medals	b) treasures	c) money	d) gold
19) a) gave	b) watched	c) started	d) sold
20) a) famous	b) alive	c) rich	d) young
21) a) provide	b) tell	c) guess	d) call
22) a) companies	b) universities	c) museums	d) exhibitions
23) a) respected	b) helped	c) offered	d) employed
24) a) social	b) popular	c) honest	d) emotional
25) a) plan	b) performance	c) statue	d) work
26) a) prints	b) lines	c) stamps	d) instructions
27) a) ceremonies	b) exhibits	c) semesters	d) hobbies
28) a) carefully	b) skillfully	c) patiently	d) seriously
29) a) take	b) get	c) fall	d) drop
30) a) heavily	b) perfectly	c) quickly	d) basically

31) a) change	b) make	c) influence	d) continue
32) a) Ideally	b) Immediately	c) Recently	d) Generally
33) a) truly	b) silently	c) likely	d) slightly
34) a) studies	b) schools	c) centuries	d) lifetimes
35) a)educated	b) admired	c) welcomed	d) examined
36) a) strong	b) excited	c) rich	d) tired
37) a) often	b) seldom	c) sometimes	d) usually
38) a) new	b) native	c) tall	d) thin
39) a) promoted	b) rejected	c) forgotten	d) supported
40) a) decisive	b)eager	c) surprised	d) immortal

Beginning Level

Short-answer Test (1)

متن درک مطلب زیر را بخوانید و به هر کدام از سوالات ان پاسخ کوتاه به زبان فارسی دهید. لطفا توجه داشته باشید که پاسخ شما باید در حد یک یا چند کلمه یا حداکثر یک جمله باشد.

(Instruction: Please read the following reading comprehension passages and answer to the questions following each passage in Persian. The length of your answer may vary from one question to the other and should not exceed a single word, phrase or sentence)

What's Your Sign?

Thousands of years ago, the ancient people of Babylon and Egypt studied the stars in the sky and created the zodiac. It was first used to keep track of time. Later, many used the stars to describe a person's personality and to say what would happen in the future.

A person's zodiac sign is connected to his or her birth date. Some believe this sign can tell us about a person's personality. For example, some think that a person born under the sign of Aries (between March 21 and April 20) is adventurous and isn't afraid to take risks. A person born under Cancer (between June 22 and July 23) is kind and happiest in the home.

In many countries in Asia, people believe the Chinese zodiac describes personality and can reveal the future. In the Chinese zodiac, there are twelve animals. A person's animal sign is connected to his or her birth year. Every animal stands for a different type of personality. People born in the year of the rat are friendly, but careful. Those born in the year of the monkey are smart and good at making money. Many believe that the rat and monkey are a good match.

In Asia, a person's blood type is also used to describe personality. People with the blood type A are calm and serious, but they can be selfish. Type Bs are independent but can be lazy. ABs are honest, and type Os are loving and talkative.

Not everybody believes that your birth sign or blood type describes your personality. In fact, some people disapprove of using the zodiac; they say it's just foolishness. But, if reading your horoscope amuses you, go ahead and read it!

1. از لحاظ سال تولد، کدام دسته از مردم بیشتر ممکن است پول در بیاورند؟

2. علامت های منطقة البروج چینی چه چیزی را توضیح می دهد؟

3. برای ساختن نخستین جدول منطقة البروج از چه چیزی استفاده شد؟

4. از نخستین جدول منطقة البروج چه استفاده ای می شد ؟ امروزه بیشتر چه استفاده ای دارد؟

5. مردمان آسیای شرقی از چه دو روشی برای توصیف شخصیت یک نفر استفاده می کنند؟

6. دو دوره ی حمل و سرطان در جدول منطقة البروج در چه تاریخی به پایان می رسند؟

7. از لحاظ تاریخ تولد، کدام افراد برای دست یافتن به هدفشان دست به کارهای خطر ناک می زنند؟

8. در جمله زیر، معنی عباراتی که زیرش خط کشیده شده چیست؟

(Chinese Zodiac can reveal the future.)

9. بر اساس متن، بانکها برای استخدام کارمند، بیشتر افراد با چه گروه خونی را ترجیح می دهند

English equivalent of Persian sentences for Short-answer Test 1

- 1- People born under what sign will probably make a lot of money?
- 2- What do Chinese zodiac signs explain?
- 3-What was used to create the first zodiac?
- 4- What was the zodiac first used for? What is it used for more often now?
- 5-What are the two ways that Asians may use to describe a person's personality?
- 6- When do the period of the sign of Aries and that of Cancer end?
- 7-A person born under what sign will probably do something dangerous to accomplish a goal?
- 8-What does the underlined phrase mean in the following sentence?
“Chinese Zodiac can reveal the future.”
- 9-Based on the passage, people of which blood type would banks most likely prefer to hire?

Beginning Level

Short-answer Test (2)

Three Centuries of Hoaxes

In 1769, long before computers were invented, a man from Hungary built a wonderful machine that could play chess very well indeed. It delighted people all over Europe and beat nearly everyone it played. Many people believed that it was truly a thinking machine. However, other people thought it was a hoax. The secret of the machine was finally revealed in 1837. An article explained that there was a man, who carefully avoided being seen, inside the machine playing chess.

On November 9, 1874, an article in an important New York newspaper claimed that all the animals in the New York Zoo had suddenly escaped and were running all over the city. It also said that there were 27 people dead and 200 harmed. It said the police were working to rescue people from the terrible situation. The entire city of New York was terrified, but there was really no reason to be frightened. The article was a cruel hoax from beginning to end.

On April 21, 1980, the first woman to cross the finish line of the Boston Marathon was 23-year-old Rosie Ruiz. However, as she climbed the stairs to receive her prize, people started to become suspicious because she didn't even seem to be out of breath. None of the other runners remembered seeing her, and her picture never appeared in photographs or TV broadcasts of the race. Later, several people revealed that they had seen her join the race near the very end. She had run only one half of a mile! Her prize was taken away, of course.

10. چرا مردم فکر می کردند که ماشین شطرنج باز یک ماشین متفکر است؟

11. چه چیزی باعث وحشت مردم در شهر نیویورک شد؟

12. چرا روزی Rosie در پایان مسابقه نفس نفس نمی زد؟

13. چقدر طول کشید تا مردم پی به راز ماشین شطرنج باز بردند؟

14. در این متن چند دلیل عنوان شده که به برنده شدن روزی در مسابقه شک کنیم . چند دلیل برای این شک ذکر شده است؟

15. کدام یک از این سه حقه بازی بیشترین استرس را به همراه داشته است؟

16. براساس متن، معنی واژه ی Hoax چیست؟

17. مردم نیویورک بالاخره به چه چیزی پی بردند؟

18. کدام یک از این حقه بازی ها سریعتر افشا شد؟

English equivalent of Persian questions for Short-answer Test 2 at the beginning level

10-Why did people believe that the chess-playing machine was a thinking machine?

11-What frightened the people in New York city?

12-Why wasn't Rosie out of breath after the race?

13- How long did it take the people to learn the secret of the chess-playing machine?

14-The article explains a few reasons to doubt that Rosie had truly won the race. How many reasons are listed?

15-Which of the three hoaxes probably caused the most stress?

16-Based on the passage, what is the meaning of the word "hoax"?

17-What did the people of New York finally find out?

18-Which of the three hoaxes was discovered most quickly?

Beginning level

L1 recall reading Test (1)

متن زیر شامل 18 تا 20 جمله میباشد. لطفاً آن را با دقت بخوانید و سپس میزان درک و فهم خود از متن که شامل مفاهیم اصلی و جزئیات موجود در متن میباشد را به یاد آورید و بین حداقل 35 تا 45 جمله به فارسی خوانا بنویسید. لطفاً توجه داشته باشید که نباید متن را ترجمه کنید، بلکه باید میزان درک و فهم خود از هر جمله متن را با کلماتی از خودتان بنویسید. همچنین میتوانید هر چند بار که لازم باشد به متن برگردید و آن را بخوانید.

(Instruction: There are 18 – 20 sentences in the following passages. Please read each passage carefully, and then based on your understanding, write down as much detail as you can recall from the text in a range of 35 – 45 sentences in Persian. The detail should include both main ideas and supporting ideas. Please note that you should never translate the text, but you need to give the information from the text in your own words. You are allowed to look back at the text to check your comprehension)

Looking into the Future

People like to think about what the future is going to be. But did you know that some people predict the future as a job? Futurists are scientists who study and predict the future. They can guess what is going to happen in the future from the information they know now. In this way, futurists can help people and businesses get ready for the future.

What do futurists say about the future? They think there are going to be big changes in people's everyday lives-and soon! For example, cars are going to drive us to places without a driver. We are going to tell the car where we want to go, and the car is going to get us there! Some futurists think we're going to talk to people on small video telephones that we can wear, like a watch. They say we're not going to use as much paper as we do today, and we're not going to need money at all- we're going to use special cards for everything we buy.

Some futurists predict that scientists are going to be able to make changes in the weather. And botanists are going to make new kinds of plants so there is more food for more people. And there are going to be new ways to cook food fast.

Of course, even scientists do not really know what the future is going to be. They can only guess. But we can be sure of one thing: the future is going to be very different, and futurists can help us get ready.

Beginning Level

L1 Recall test (2)

Student life

College is an exciting time to learn and to make friends that will last a lifetime. Many students do not like to worry about money, and they would rather not think about it. But, it doesn't matter whether a student's parents pay for everything, or whether the student works part-time to help pay for his or her education.

All students can get into money trouble if they're not careful.

The cost of a college education can be quite expensive. In English-speaking countries, the average cost per student per year is well over US\$10,000.

Students must also pay for books, paper, pens, etc. These can \$500 to \$1,000 per year. Students who live in university housing pay thousands dollars more per year for room and board. Add money for clothes, travel, and other personal expenses, and the average cost of one year at a university can be \$20,000 to \$ 30,000 or more.

Students need to spend their money carefully. At most universities, advisors can give students advice on how to budget their money. They suggest this: at the start of a school semester, write down your income; for example, money you will get from your family or a part- time job. Then, list all of your expenses. Put your expenses into two groups: those that change (food, phone, books, travel), and those that will stay the same (tuition, room and board). Add together all of your expenses. Are they more than your income? Do you have enough money, or do you need more?

Learning not to spend more money than you have is not always easy. But for many, it is easier than borrowing money from family or friends.

Intermediate Level

Cloze Test

متن درک مطلب زیر را بخوانید و بهترین پاسخ را برای هر جای خالی از بین گزینه های داده شده انتخاب کرده و در پاسخنامه علامت بزنید.

(Instruction: Please read the following reading passage and select the best answer from the choices given for each empty space)

Roles in Human Society

Human beings are creatures of society. They take part in a(1).... social system which expects them to(2)..... certain roles. Social scientists(3)..... that without roles, society could not(4).....

To be(5)....., members of society need to know how others(6)..... them to act so that they can act, or not act, in those(7)..... Let us take student life at a(8)..... as an example. When new students(9)....., they do not yet know what their(10).....roles are. That is, they do not know what their(11)....., teachers, or(12)..... want them to do. To help them(13)..... quickly and correctly,(14)..... make them attend a(n)(15).....program. They learn the expected(16)..... for college students. We can label this their(17)..... role.

In addition to the role, social(18)..... talk about the(19)..... role. For our(20)..... students, this means the(21)..... that each one has about what(22)..... behavior at a university is. In order to(23)....., he or she must know or(24)..... what others' roles are and then(25)..... his or her own(26)..... in relationship to them.

When members of a society(27)... perceive the rules of that society, and when their subjective roles are(28)..... to their prescribed roles, they(29)..... act in the ways that society(30)..... them to. That is, they do and say what is(31)..... correct. The actual(32)... of a role, with its(33)..... behavior, is called

the(34).....role. Our college students, if they(35)..... similar prescribed and subjective roles, will(36)..... obey university roles and(37)..... with their professors as(38)..... and with their roommates as(39)..... Their behavior will fall into the(40)..... of acceptable student behavior.

Social(41).... say that in order to(42)..... itself and make sure that its(43)..... perform their roles, society(44)..... those members to(45)..... others' behavior as acceptable or unacceptable.

1) a) comfortable	b) complex	c) correct	d) continuous
2) a) persuade	b) compress	c) support	d) perform
3) a) affirm	b) evaluate	c) generalize	d) collect
4) a) promote	b) collide	c) function	d) combine
5) a) successful	b) important	c) well-known	d) civilized
6) a) assure	b) assist	c) engage	d) expect
7) a) strategies	b) unions	c) ways	d) shares
8) a) home	b) university	c) barrack	d) hospital
9) a) graduate	b) survive	c) observe	d) arrive
10) a) appropriate	b) associative	c) collective	d) speculative
11) a) parents	b) officers	c) roommates	d) professors
12) a) researchers	b) advisors	c) players	d) volunteers
13) a) construct	b) travel	c) adapt	d) tolerate
14) a) authorities	b) governments	c) carpenters	d) adherents
15) a) celebration	b) orientation	c) formation	d) repetition
16) a) issue	b) introduction	c) schedule	d) behavior
17) a) memorized	b) reinforced	c) migrated	d) prescribed
18) a) researchers	b) scientists	c) officials	d) instructors
19) a) subjective	b) corrective	c) transitive	d) selective
20) a) native	b) sophomore	c) college	d) junior
21) a) inventions	b) statistics	c) documents	d) expectations
22) a) functional	b) cultural	c) suitable	d) relative
23) a) conduct	b) achieve	c) calculate	d) compound

24) a) tie down	b) find out	C) climb up	d) run into
25) a)make up	b) kneel down	c) take out	d) look at
26) a) perceptions	b) conditions	c) distinctions	d) traditions
27) a) wildly	b) clearly	c) occasionally	d) heavily
28) a) related	b) complicated	c) similar	d) significant
29) a) politely	b) recently	c) formerly	d) normally
30) a) relieves	b) lifts	c) expects	d) amuses
31) a) considered	b) designed	c) elaborated	d) included
32) a) opportunity	b) performance	c) influence	d) perspective
33) a) ultimate	b) specific	c) obvious	d) unusual
34) a) weakened	b) internalized	c) enacted	d) decisive
35) a) consult	b) review	c) surprise	d) possess
36) a) probably	b) largely	c) really	d) accurately
37) a) explain	b) interact	c) consume	d) collect
38) a) farmers	b) b) tutors	c) students	d) lecturers
39) a) kids	b) classmates	c) actors	d) friends
40) a) hole	b) range	c) circle	d) space
41) a) experts	b) adherents	c) creators	d) candidates
42) a) compute	b) persuade	c) protect	d) perceive
43) a) resources	b) opponents	c) advisors	d) members
44) a) identifies	b) encourages	c) characterizes	d) criticizes
45) a) certify	b) organize	c) observe	d) judge

Intermediate Level

Short-answer Test (1)

متن درک مطلب زیر را بخوانید و به هر کدام از سوالات ان پاسخ کوتاه به زبان فارسی دهید. لطفا توجه داشته باشید که پاسخ شما باید در حد یک یا چند کلمه یا حداکثر یک جمله باشد.

(Instruction: Please read the following reading comprehension passages and answer to the questions following each passage in Persian. The length of your answer may vary from one question to the other and should not exceed a single word, phrase or sentence)

Oh, No, Not another Test!

Have you ever felt that you were just studying to pass a test? Today, many schools and companies around the world use standardized tests such as the TOEFL test to measure students abilities. Thousands of people every year take these kinds of tests in order to enter or graduate from a school, or get a better job.

There are two types of tests that are commonly used in academic institutions. The first type of test is the achievement test. This is used at the end of a term, for example, to evaluate how much students have learned about a subject. The other commonly used test is the proficiency test. This kind of test measures students overall skill in a language, and is not directly related to a particular course or school. IELTS, TOEIC, TOEFL, and university entrance exams are all examples of language proficiency exams.

In many educational systems throughout the world, students take regular multiple- choice achievement tests. They have to pass these tests in order to move to a higher level or graduate from an educational institution. Many advocates of this type of traditional evaluation believe that students learn best, and increase their knowledge, by memorizing facts and information. These people also believe that teaching students to pass a test teaches them discipline, and prepares them for the working world.

Educational reformers, however, believe that standardized testing alone is limited and can only measure some of a student's ability. They realize that many students, though intelligent, are not always good at taking tests or at memorization. This makes it harder for them to

achieve passing grades, get into good schools, or get the right job. In addition, factors such as how a person feels on the day of a test can also affect his or her score. Using only a multiple-choice style of test to evaluate students' abilities cannot always tell us what they have learned, or how they may do in the future.

Even test experts agree that current tests are not perfect. Paul Barton, the director of Educational Testing Services' (ETS) Policy Information Center, has suggested that standardized testing needs to change. Students should take fewer standardized tests. The ones they do take, he says, should tell us what students really know, and how they can use that knowledge in real life.

To use standardized tests or not to- the debate continues. Meanwhile, many students' futures are still decided based on their test results.

1. در این متن تعدادی نمونه تست مهارت زبانی ذکر شده است. این نمونه ها چند تاست؟
2. تست های استاندارد معمولاً در کدام دو موقعیت به کار می روند؟
3. چرا از تست های استاندارد استفاده می شود؟
4. از نظر چه کسی یا کسانی تست های استاندارد نمی توانند اطلاعات درستی درباره ی امتحان دهندگان به ما بدهند؟
5. تفاوت بین تست پایان دوره و تست مهارت کدام است؟
6. از نظر طرفداران آموزش سنتی، دانش آموزان در چه صورت بهترین یادگیری را خواهند داشت؟
7. یک دانش آموز در طول تحصیل احتمالاً چه نوع امتحانی را بیشتر از بقیه می دهد؟
8. چگونه باید از امتحان های استاندارد استفاده کرد؟
9. طرفداران بازنگری در آموزش و پرورش چه نوع تستی را به عنوان بهترین معیار سنجش آموخته های دانش آموزان توصیه می کنند؟

English equivalent of the Persian sentences for Short-answer Test One at the intermediate level

1-The article lists some examples of language proficiency tests. How many examples are listed?

2- In what two places are the standardized tests usually used?

3- Why are the standardized tests used?

4-Who thinks that the standardized tests cannot give us a reliable score on the learners' general ability?

5-What is the difference between achievement and proficiency test?

6-How do supporters of traditional education believe that students learn in the most excellent manner?

7-What kind of test does a student probably take more during his education?

8-What is the rationale behind taking standardized tests?

9-What kind of test do the educational reformers recommend as the best measure of what students have learned?

Intermediate Level

Short-answer Test (2)

Is Your Diet Destroying the Environment?

Many people know that eating a vegetarian diet has important health benefits.

Vegetarians usually have lower levels of heart disease, and studies have also shown that they have a lower risk of diabetes than people who eat meat. Most people don't realize, however, that a vegetarian diet is also better for the health of our environment.

Recently, researchers from the Union of Concerned Scientists in the U.S.A gave a report on how consumer behavior influences the environment. Their study showed that meat consumption is one of the main ways that humans can damage the environment, second only to the use of motor vehicles.

So, how can a simple thing like eating meat have a negative effect on the environment? The most important effect of meat production is through the use of water and land. Two thousand five hundred gallons of water are needed to produce one pound of wheat. By producing crops instead of animals, we can make more efficient use of the land and water. One acre of farmland that is used for raising livestock can produce 250 pounds of beef. One acre of farmland used for crops can produce 40,000 pounds of potatoes, 30,000 pounds of carrots, or 50,000 pounds of tomatoes.

Furthermore, farm animals add to the problem of global warming. All livestock animals such as cows, pigs, and sheep release methane by expelling gas from their bodies. One cow can produce up to sixty liters of methane each day.

Methane gas is the second most common greenhouse gas after carbon dioxide. Many environmental experts now believe that methane is more responsible for global warming than carbon dioxide. It is guessed that twenty-five percent of all methane sent into the atmosphere comes from farm animals.

People are becoming aware of the benefits of coming to a vegetarian food, not just for health reasons, but also because it plays a very important role in protecting the environment. Some people go further, and eat a vegan diet, which excludes all products from animal sources, such as cheese, eggs, and milk.

However, some nutritionists believe that a vegan diet can be deficient in some of the vitamins and minerals that our bodies need daily. Today, many people are concerned about improving their health, and about protecting the environment. Switching to a vegetarian diet- or just eating less meat- is a good way to do both of these things at the same time.

10. بر اساس متن، دام برای تولید گوشت به کدام دو چیز نیاز دارد؟

11. اشکال گیاهخواری چیست؟

12. کدام گاز گلخانه ای بیشترین سهم را در گرمایش زمین دارد؟

13. در این متن دو نمونه رفتار بشر آمده که به محیط زیست صدمه می زند. آن دو کدامند؟

14. بر اساس متن حیوانات دامی ممکن است به دو طریق برای محیط زیان آور باشند. آن دو کدامند؟

15. برای تولید 750 پوند گوشت گاو به چند هکتار مرتع نیاز است؟

16. افرادی که از گوشت تغذیه می کنند احتمالاً به چه نوع بیماری مبتلا می شوند؟

17. دو نوع غذایی که نویسندگان متن توصیه می کند، کدامند؟

18. بر اساس متن، چند لیتر متان توسط 2 گاو در طول 4 روز تولید

می شود؟

English equivalent of the Persian sentences for Short-answer Test Two at the intermediate level

10- According to the article, what two things do livestock need to produce meat?

11-What is a problem with a vegan diet?

12- Which greenhouse gas is more responsible for global warming?

13- The article lists two ways that human behavior damages the environment. What are these?

14-The article explains that livestock animals may be disadvantageous for the environment in two ways. What are these?

a) Raising livestock animals need more water and land b) They produce methane

15- How many acres of farmland are needed to produce 750 pounds of beef?

16-What kind of diseases are people who eat meat more likely to suffer from?

17-What two diet options does the author of the passage recommend?

18-Up to how many liters of methane is probably released by two cows during four days?

Intermediate Level

L1 recall reading Test (1)

متن زیر شامل 27 تا 34 جمله میباشد. لطفاً آن را با دقت بخوانید و سپس میزان درک و فهم خود از متن که شامل مفاهیم اصلی و جزئیات موجود در متن میباشد را به یاد آورید و بین حداقل 53 تا 63 جمله به فارسی خوانا بنویسید. لطفاً توجه داشته باشید که نباید متن را ترجمه کنید، بلکه باید میزان درک و فهم خود از هر جمله متن را با کلماتی از خودتان بنویسید. همچنین میتوانید هر چند بار که لازم باشد به متن برگردید و آن را بخوانید

(Instruction: There are 27 – 34 sentences in the following passages. Please read each passage carefully, and then based on your understanding, write down as much detail as you can recall from the text in a range of 53 – 63 sentences in Persian. The detail should include both main ideas and supporting ideas. Please note that you should never translate the text, but you need to give the information from the text in your own words. You are allowed to look back at the text to check your comprehension)

Work Hard, Play Hard?

People today seem to have increasingly hectic lifestyles. Results of a 2001 Harris Poll on free time. Conducted in the United States, showed that the average work week for many Americans is fifty hours. In addition, many people spend up to two or three hours a day commuting to and from work. With the time spent eating, sleeping, taking care of household chores, and looking after the family, there is little time left for leisure activities for many Americans.

However, having free time to relax and pursue hobbies and interests is important, and good for a person's well-being. People need time away from the pressures of study or work in order to relax, and enjoy time with friends and family.

In different countries and cultures around the world free time is spent in different ways. The results of the Harris Poll showed that reading was the most popular spare-time activity in the United States. This was followed by watching TV, then spending time with family. In a U.K. survey on leisure-time activities, watching TV and videos was the most popular pastime; listening to the radio came second. In a similar survey conducted in Japan, the most popular free-time activity was eating out. The second most popular activity was driving. Karaoke,

which ranked fourth, was more popular than watching videos, which came fifth. Listening to the radio or music ranked sixth.

There were also differences in the most popular outdoor pursuits between the three countries. The most popular outdoor activity for Americans was gardening. In the U.K., it was going to the café, followed by visiting the cinema. In Japan, going to cafes was ranked eighth in popularity, and gardening was ranked ninth.

Although people around the world may enjoy doing similar things in their free time there is evidence to suggest that these interests are changing. In the U.S., for example, the popularity of computer activities is increasing.

Many more people in the States are spending their free time surfing the Web, e-mailing friends, or playing games online. In the 2004 Harris Poll, computer activities ranked sixth in popularity; in 1995, only 2 percent of people mentioned them. Currently, listening to music is ranked eleventh.

With more people downloading music from the Internet, it is possible that, in the future, music and computer activities will become the same pastime for many Americans.

Intermediate Level

L1 Recall Test (2)

How often do you eat chocolate? If you answered "every day," you may be addicted to chocolate, but is this addiction damaging your health?

Eating too much chocolate is often thought to be the cause of tooth decay, weight gain, headaches, and skin problems such as acne. On the other hand, chocolate is known to make people feel happier. Eating chocolate releases a different flavour that gives us a pleasant, positive feeling. What is it that causes this feeling?

Chocolate contains over three hundred known chemicals. Like a drug, these chemicals stimulate areas of the brain that enable us to feel pleasure. The most well-known chemical is caffeine, which is also found in coffee, tea, and some types of soda. Theobromine, a weak stimulant, is present in higher amounts than caffeine. It is believed that the combination of these two chemicals causes the short-lived "lift" we experience after eating chocolate.

However, does eating too much chocolate cause any real health risks? The popular opinion of chocolate is that it is a fattening food that gives you spots, and contains no nutritional qualities. The fact is, chocolate does contain saturated fat. This type of fat can help heart disease by increasing levels of bad cholesterol in the blood. On the other hand, scientists at the University of California have discovered that chocolate also contains high levels of chemicals called phenolics, also found in red grapes, coffee, and tea. Some phenolics if used in small amounts, are believed to lower the risk of heart disease.

Evidence that eating chocolate does not cause acne comes from two studies: one by the Pennsylvania School of Medicine, the other by the U.S. Naval Academy. Their research showed some interesting results.

They found that there was no difference in the skin condition between the study participants who did, or did not, eat chocolate. There is also no proof that chocolate is the cause of tooth decay. In fact, it is believed that the cocoa butter in chocolate forms a coating over teeth that may help to protect them. The sugar in chocolate can cause cavities, but no more than sugar in any other food or drink.

Consumption of cocoa around the world now amounts to almost three million tons a year. That's equivalent to 500 grams of cocoa for every person on earth! No one has ever died of eating too much chocolate, so this is one addiction that might be okay to have.

Advanced Level

Cloze Test

متن درک مطلب زیر را بخوانید و بهترین پاسخ را برای هر جای خالی از بین گزینه های داده شده انتخاب کرده و در پاسخنامه علامت بزنید.

(Instruction: Please read the following reading passage and select the best answer from the choices given for each empty space)

Human Adaptation to Space

It is important for the human race to spread out into space for the survival of the species, "said world- renowned astrophysicist Steven Hawking. He is far from being alone in his(1).... of humans learning to live in places(2)..... on Earth. A Space Odyssey(3)..... the possibility of(4)..... human life in(5)..... space, and presented a very realistic(6)..... of spaceflight. Since astronaut Yuri Gagarin(7)..... the first man to(8)..... in space in 1961,(9)..... have researched what(10)..... are like beyond Earth's atmosphere, and what(11)..... space travel has on the human(12).....

Although most astronauts do not(13)..... more than a few months in space, many experience(14)..... problems when they(15)..... Earth. Some of these(16)..... are short-lived: others may be(17).....

More than two-thirds of all astronauts suffer from(18)..... sickness while traveling in space. In the(19).....- free environment, the body cannot(20)..... up from down. The body's internal(21)..... system sends(22)..... signals to the brain, which can result in(23)..... lasting as long as a few days. A body that is(24)..... gravity also experiences changes in the(25)..... of bodily fluids. More fluid than normal(26)..... up in the face, neck, and chest, resulting in a(27)..... face, bulging neck(28)....., and a slightly(29)..... heart.

Throughout the(30)..... of a mission, astronauts' bodies(31)..... some potentially dangerous(32)..... One of the most common is(33)..... of muscle mass and bone(34)..... Another effect of the(35)..... environment is that astronauts(36).....

not to use the muscles they(37)..... on in a gravity environment, so the muscles(38)..... atrophy. This, combined with the(39)..... of fluid to the(40)..... body and the resulting loss of essential(41)..... such as calcium, causes bones to(42)..... Bone density can(43)..... at a rate of one to two percent a month and, as a result, many astronauts are unable to walk(44)..... for a few days upon their return to Earth. Exposure to(45)..... is another serious(46)..... that astronauts face. Without the Earth's(47)..... to protect them, astronauts can be exposed to(48)..... radiations from the sun and other(49)..... bodies, leaving them at risk of(50).....

1) a) vision	b) vote	c) vessel	d) venture
2) a) away from	b) similar to	c) next to	d) other than
3) a) established	b) stimulated	c) explored	d) exacerbated
4) a) creating	b) sustaining	b) exclaiming	d) exploiting
5) a) lighter	b) upper	c) larger	d) outer
6) a) theme	b) viewpoint	c) portrayal	d) content
7) a) promoted	b) become	c) submitted	d) negotiated
8) a) travel	b) mediate	c) relocate	d) unite
9) a) physicians	b) psychologists	c) scientists	d) archaeologists
10) a) surroundings	b) arguments	c) certificates	d) conditions
11) a) considerations	b) effects	c) circumstances	d) results
12) a) body	b) personality	c) brain	d) heart
13) a) concentrate	b) spend	c) sleep	d) accept
14) a) analytical	b) celestial	c) theoretical	d) psychological
15) a) go around	b) walk into	c) return to	d) reside in
16) a) ailments	b) advantages	c) adjustments	d) adventures
17) a) taking	b) changing	c) staying	d) long-lasting
18) a) rigid	d) motion	c) modern	d) emotional
19) a) pressure	b) oxygen	c) gravity	d) human
20) a) distinguish	b) encapsulate	c) ventilate	d) enhance
21) a) blood	b) bladder	c) balance	d) breathing
22) a) confusing	b) discriminating	c) circulating	d) coinciding
23) a) battle	b) nausea	c) diamond	d) funeral

24) a) supplied with	b) identified with	c) supported by	d) deprived of
25) a) deduction	b) destination	c) distribution	d) declination
26) a) ends	b) sets	c) backs	d) draws
27) a) choppy	c) puffy	c) catchy	d) cagey
28) a) eyes	b) cheeks	c) laps	d) veins
29) a) endured	b) enlarged	c) enriched	d) engaged
30) a) duration	b) extension	c) formation	d) computation
31) a) accommodate	b) behold	c) experience	d) illustrate
32) a) situations	b) circumstances	c) consequences	d) disorders
33) a) loss	b) injury	c) mixture	d) excess
34) a) severity	b) debility	c) security	d) density
35) a) defenseless	b) featureless	c) weightless	d) tactless
36) a) tend	b) like	c) allow	d) order
37) a) drag	b) rely	c) lay	d) touch
38) a) unlikely	b) consistently	c) surprisingly	d) gradually
39) a) shift	b) generation	c) risk	d) promotion
40) a) controversial	b) imaginary	c) upper	d) supplementary
41) a) elements	b) medicines	c) matters	d) minerals
42) a) reinforce	b) weaken	c) reconstruct	d) expand
43) a) decrease	b) confront	c) mediate	d) amplify
44) a) steadily	b) hesitantly	c) properly	d) unwillingly
45) a) radiations	b) chemicals	c) assaults	d) vacuums
46) a) opportunity	b) hazard	c) characteristic	d) mediocrity
47) a) gravity	b) wideness	c) atmosphere	d) balance
48) a) curious	b) inspired	c) furnished	d) intense
49) a) galactic	b) symbolic	c) telescopic	d) traumatic
50) a) extinction	b) cancer	c) isolation	d) catastrophe

Advanced Level

Short-answer Test (1)

متن درک مطلب زیر را بخوانید و به هر کدام از سوالات ان پاسخ کوتاه به زبان فارسی دهید. لطفا توجه داشته باشید که پاسخ شما باید در حد یک یا چند کلمه یا حداکثر یک جمله باشد.

(Instruction: Please read the following reading comprehension passages and answer to the questions following each passage in Persian. The length of your answer may vary from one question to the other and should not exceed a single word, phrase or sentence)

Like individual animals, animal species also eventually die out. It is estimated that, until the 18th century, one species disappeared from the Earth every four years.

By the 19th century, this had increased to one species per year. By 1975, it was 1,000 species per year, and today animals are disappearing at the appalling rate of more than 40,000 species per year. Most species are threatened by pollution, habitat destruction, and unreasonable exploitation caused by humans. The International Union of Conservation of Nature has created a number of categories that describe the danger level of animal species:

- A species that has died out completely is called extinct. Examples are the dinosaurs and the dodo bird.
- Species that only live in zoos or on farms, etc. fall into the category extinct in the wild, for example, various horse species.
- A species is labeled critically endangered when it is in immediate danger of dying out completely. Its numbers are dangerously low, and it needs protection in order to survive. The Siberian tiger and the snow leopard are two examples.
- Species that have a high, but not immediate, risk of dying out are simply labeled endangered. The giant panda is a famous example.
- A vulnerable species is in less trouble than an endangered one, but its numbers are still certainly declining. The cheetah and the African elephant are vulnerable species.
- Animal species that aren't particularly endangered and have high numbers of individuals are labeled less concern.

There are many factors that can cause an animal or plant species to become endangered.

The main cause of species endangerment is humanity's destruction of both water and land habitats. Deforestation and soil, air, and water pollution can all destroy habitat. This can then cause a large number of animals or plants to die. The critically endangered Sumatran orangutan has seen a 50% decline in its population over the last eight years as farmers clear more and more forests for agriculture.

Another cause of endangerment is the unreasonable exploitation of animals. Uncontrolled hunting of whales in the last century, for example, resulted in many whale species becoming critically endangered.

The very high demand for animal parts for use in certain foods or medicines is another example. The horn of the rhinoceros can be sold at a high price in some places where it is thought to be medicinal: a price so high that people will kill the animals even though it is against the law.

Introducing a non-native species to an environment can also cause species endangerment. A native species is one that develops naturally in a particular area, and has done so for a long time. A non-native species might be introduced into a new environment by humans, either intentionally or by accident. The brown tree snake, unknowingly introduced by cargo ships stopping at Guam, has managed to kill off ten of the eleven species of birds native to the island's forests. In Florida, large pet snakes such as the anaconda and the python have been released into the large Everglades swamp. The snakes have been quite successful, and now compete with and even threaten the swamp's alligators.

Organizations such as the World Wildlife Fund and the International Union for Conservation of Nature try to raise awareness of threatened animals and plants. These organizations work with government agencies to save threatened or endangered species, and to make new laws that will protect these species.

Many of these plans work, but some do not. Public awareness of this issue is important. To preserve the quality of our lives and the lives of future generations, we must also protect plant and animal species now and in the future.

1. دسته بندی آی یو سی ان (IUCN) از جانوران چگونه است؟
2. کدام گروه از گونه های جانوری نیازمند کمترین اقدام مراقبتی برای بقا هستند؟
3. چرا طی سال های گذشته تعداد اورانگوتان کاهش یافته است؟
4. چگونه از حیوانات سوء استفاده می شود؟ (دو مثال بیاورید)
5. سازمانها برای حمایت از گونه های جانوری چه اقداماتی انجام می دهند؟
6. در متن چند علت برای انقراض گونه های جانوری یا گیاهی ذکر شده است؟
7. کدام دسته از جانوران را باید به محیط زیست خودشان باز گرداند؟
8. بهترین روش محافظت از گونه های جانوری از خطر (انقراض) کدام است؟
9. بر اساس متن بعضی از گونه های جانوری در معرض خطر انقراض هستند. حیوانات زیر را بر اساس نیاز به فوری ترین اقدام مراقبتی (برای بقا) از یک تا چهار مرتب کنید.

African elephant, giant panda, snow leopard, horse species

English equivalent of Persian sentences for Short-answer test (1)

1- How has the IUCN categorized animals?

2- Which group of animal species needs the least urgent care for its survival?

3- Why has the population of orangutan declined over the past years?

4- How are animals exploited? (Give two examples)

5- What do the Organizations do to protect animal species?

6- How many reasons does the article list for the endangerment of an animal or plant species?

7- Which animals need to be moved to their own environment?

8- What is the best way of protecting animal species from endangerment?

9- The article explains that some animal species are in danger of dying out. How do you rate the following animals based on the most immediate care they need?

African elephant, giant panda, snow leopard and horse species

Advanced Level

Short-answer Test (2)

From Comic Books to Graphic Novels

Originally they were called comic books, or comics for short.

These were those picture or cartoon stories with dialogue that became very popular with young people in the United States in the 1940s and 1950s. Despite their name, they were not always funny, and they were more like magazines than books because they were shorter, cheaper, and appeared periodically (often once a month) in newsstands at local stores. Many were about the adventures of superheroes such as Superman, Batman, and Wonder Woman. These were individuals with special powers who fought against evil and injustice. The end of every story was never in doubt: the superhero, as the representative of the "good guys," always defeated the "bad guys," or the characters that represented the evil forces in the world".

Comic books probably started in America, but they became popular around the world, and different comic book traditions developed in various countries. In Japan, comic books are called manga. Manga developed out of a combination of ukiyo-e (a drawing style dating back to the end of the 18th century) and Western techniques. Comic book artists in several European countries, including France, Belgium, and Italy, also developed their own styles with original themes. For example, in the 1930s the Franco- Belgian artist Herge created the character Tintin, a reporter who travels to different countries to investigate stories. Many of the stories in the Tintin series were based on real historical events such as Hitler's conquest of Austria in 1938. As a result, it was much more political and reality- based than the typical superhero- type comics found in the United States.

You can still buy the magazine- like comic books, but by the 1970s a new form had appeared. This was a longer, more expensive, hardbound publication often sold in regular bookstores.

The content clearly aimed at an older, more mature reader. Adventure stories with superheroes remained popular, but the new characters were deeper and more true-to-life and the stories were more complex. To distinguish this new type of picture book from the earlier youth- oriented form, it was sometimes called a graphic novel.

The appearance of a number of successful cartoon stories with mature content in the mid-1980s helped to establish the popularity of the new form. The transition from old to new style is perhaps clearest in Frank Miller's Batman: *The Dark Knight Returns*. In his story, an older Batman lives in a world where both the "good guys" and the "bad guys" have to make difficult moral choices.

Similarly, *Watchmen* by Alan Moore and Dave Gibbons presents a series of complex, interesting stories in which the superheroes are more like regular human beings with all their faults and problems. *Watchmen* became very popular and was the only graphic novel on Time magazine's list of the best 100 books since 1923. Finally, in *Maus*, by Art Spiegelman, we have an original graphic novel that totally breaks with the superhero theme. *Maus* tells us the story of mice, inspired by Spiegelman's Jewish parents, who struggled against the Nazis during World War II.

People have become more familiar with the new graphic novels, but many still find the term confusing. To some, graphic novels are just comic books, only longer. And the fact that the expression has become an important advertising tool hasn't helped. For example, some publishers have put together some of their old superhero series and are now selling them as graphic novels.

As a result, some of the new creators have decided not to use the term to describe their original works. Instead, they prefer expressions such as picture novella, graphic album, or original graphic novel.

10. بر اساس متن کتابهای کمیک ابتدا در امریکا و سپس در کشورهای دیگر منتشر شدند. چگونه میتوان کتابهای منتشر شده در یک کشور را از انهایی که در کشورهای دیگر منتشر شده اند تشخیص داد؟

11. کتابهای کمیک در آغاز چه نکته مشترکی داشتند

12. رمان های گرافیک برای چه مخاطبانی تهیه می شود

13. تفاوت کتابهای کمیک و رمان های گرافیک چیست

14. تفاوت بین سری داستانهای Watchman و اولین سری کتابهای کمیک امریکایی چه بود؟

15. بر اساس متن، محتوای آثار زیر از لحاظ داشتن اتفاقات روزمره ی زندگی متفاوت می باشد. آن ها را از بیشترین به کم ترین از 1 تا 3 مرتب کنید.

Magna, Titin stories, Maus

16. چرا برخی از نویسندگان امروزی رمان گرافیک ترجیح می دهند نام های دیگری مانند رمان تصویری ، آلبوم

گرافیک و یا رمان گرافیک اورجینال روی آثار خود بگذارند؟

17. مردم امروزه چه نوع رمان های گرافیکی را می خوانند؟

18. بر اساس متن بعضی از نویسندگان ها یک تصور منفی از عبارت "رمان گرافیک" دارند. این تصور منفی چیست؟

English equivalent of the Persian questions in the Short-answer Test One at the advanced level

10-The article explains that comic books were first created in U.S and then they were developed in other countries. How can the comic books in one country be distinguished from those published in other countries?

11-What did all comic books have in common in the beginning?

12-Who was a graphic novel written for / intended for /aimed at?

13-How are comic books and graphic novels different?

14-What was the difference between Watchman story series and early comic books found in U.S?

15- Based on the daily real-life events, how do you rate the content of the following publications from one to three?

Magna, Titin stories, Maus

16-Why do some of the new creators of the graphic novel prefer to call their works “picture novella, graphic album or original graphic novel”?

17-What kind of graphic novels do people read nowadays?

18- What is the negative perception that some authors have of graphic novels?

Advanced Level

L1 recall reading Test (1)

متن زیر شامل 29 تا 31 جمله میباشد. لطفاً آن را با دقت بخوانید و سپس میزان درک و فهم خود از متن که شامل مفاهیم اصلی و جزئیات موجود در متن میباشد را به یاد آورید و بین حداقل 70 تا 85 جمله به فارسی خوانا بنویسید. لطفاً توجه داشته باشید که نباید متن را ترجمه کنید، بلکه باید میزان درک و فهم خود از هر جمله متن را با کلماتی از خودتان بنویسید. همچنین میتوانید هر چند بار که لازم باشد به متن برگردید و آن را بخوانید.

(Instruction: There are 29 – 31 sentences in the following passages. Please read each passage carefully, and then based on your understanding, write down as much detail as you can recall from the text in a range of 70 – 85 sentences in Persian. The detail should include both main ideas and supporting ideas. Please note that you should never translate the text, but you need to give the information from the text in your own words. You are allowed to look back at the text to check your comprehension)

Bring Back the Woolly Mammoth?

In the 1993 film *Jurassic Park*, several species of dinosaurs have been brought back to life using DNA millions of years old. The dinosaurs are placed in an animal theme park as a tourist attraction. However, when a group of scientists arrives for a visit, the dinosaurs escape and attack them. After many scary encounters, only a few of the visitors remain alive.

The story is of course fiction, but it reflects recent advances in genetic engineering which are getting ever closer to reality. At this point no one really suggests bringing back dinosaurs, but there are a number of serious proposals to revive extinct species. The animals on this possible comeback list include the woolly mammoth, an elephant-like creature that wandered the plains of Siberia; the moa, a giant flightless bird from New Zealand; the thylacine, a dog-like hunter also known as the Tasmanian tiger because of the dark stripes down its back; and the bucardo, a mountain goat from Spain.

These animals had very little in common and in most cases lived eras apart. The woolly mammoth, for example, died many thousands of years ago while the bucardo became extinct only around the year 2000.

But, all these species lived at the same time as humans, and humans have been largely responsible for their destruction. So it seems somehow fitting that we are now thinking of reviving them.

Scientists have proposed reviving an extinct species using one of two possible methods. In the first method, sex cells (sperm or eggs) are obtained from the extinct animal and are used to fertilize the sex cells of a closely related living relative in a laboratory. For example, sperm from a woolly mammoth could be used to fertilize an egg from a modern-day elephant. The fertilized egg would then be placed in the womb of a live female elephant where it would live and grow until it is ready to be born. The second method involves a type of cloning. In cloning, the DNA of one individual replaces the DNA of another. In the woolly mammoth example, scientists could inject DNA from a mammoth into an egg cell from an elephant. The cloned egg cell would then be placed into a living elephant and allowed to develop in the same way as a fertilized egg.

Many difficulties remain before it will be possible to revive an extinct species by either method. In fact, some scientists believe that because of all the problems, species revival will never happen. One of the major challenges is to obtain enough high-quality DNA from an extinct species to conduct an experiment. While it is theoretically possible to preserve genetic material for thousands of years under ideal conditions, these conditions are very hard to find in real life. For example, researchers have obtained a number of samples of mammoth DNA, but none have been usable. And the cloning procedure presents its own problems. Scientists have been able to clone only a few species of animals, and most cloned creatures are short-lived and frail.

And there is a final, ethical consideration. Even if we learn how to reproduce an example of an extinct species, that individual could never have a normal life. Its natural environment is most likely gone, and it would have no parents to show it how to behave as a member of its species. So it would remain a curiosity, and probably live out its life in a zoo. People question whether it would be ethical to revive one of nature's creatures for such a purpose.

Advanced level

L1 recall reading test (2)

The World's Most Popular Beverage

Coffee is indeed the world's most popular drink- over 400 billion cups of coffee are happily consumed each year. So, just who is drinking all this coffee? People from every country around the world, certainly, have coffee even though the Americans, the French, and the Germans consume over 65 percent of the world's yearly production by themselves. According to where they are from and their personal preference, coffee lovers the world over enjoy this aromatic beverage iced or steaming hot; black or with milk; with lemon peel or whipped cream; with spices such as cinnamon, ginger, or cardamom or with chocolate; with brandy or whiskey; and even with a pinch of salt or pepper. Coffee drinkers start their day with it, drink it as a mid-morning pick-me-up, sip it after lunch, and linger over it after dinner or dessert.

The first coordinated attempt to cultivate the coffee plant was made on the Arabian Peninsula around 1100 A.D. Arabs extracted the beans from the red coffee berries, roasted them, and then boiled them in water to make *qahwa*. By the 1500s, coffee had taken hold all over the Middle East, and the world's first coffee shop was established in Constantinople. Coffee came to Europe through the port of Venice around 1600 and only seven years later was introduced to America by Captain John Smith, the founder of Virginia. The Dutch became the first to see the potential in transporting and cultivating coffee as a business in 1690, but they had to contend with smugglers, who secretly took coffee plants and seeds to Brazil from Ceylon to sell. Many of the important coffee-producing countries owe their beginnings to those first smugglers.

For all practical purposes, coffee beans are of two types-Arabica and Robusta. As the name indicates, Arabica derives from the earliest cultivated species in the Middle East. Arabica beans cost more because they are highly prized for their rich flavor and aroma, and also because they require a great amount of care to cultivate properly. Robusta, on the other hand, requires less care, and so grows more successfully in West Africa and Southeast Asia. It tolerates different climates better than Arabica and contains twice the caffeine of Arabica as well.

Unfortunately, the flavour and aroma of Robusta can't compare, and so it accounts for about 25 percent of the coffee grown around the world, while Arabica accounts for 75 percent.

As people consume more and more coffee, they wonder about its effect on their health. Coffee contains caffeine, an alkaloid compound also found in tea and cola nuts, among other products. Coffee drinkers worry that coffee might be addictive and that caffeine might have harmful effects. Studies have shown that caffeine does stimulate the central nervous system, as well as the cardiovascular system. It increases blood pressure up to a point and also increases the secretion of gastric acid, thus aiding digestion. It is highly valued for its ability to make people more alert and less tired, and for that reason it is a favorite of students and office workers. It also helps fight migraine headaches, and for that reason may be a favorite of teachers and bosses! Measurement of caffeine varies according to the plant- a five-ounce cup of coffee made from Arabica beans will contain around 1.53 percent, while the same amount made from Robusta beans will contain around 6 percent. Other factors that affect caffeine levels include the amount of coffee used and the method used to brew it. Most experts agree, however, that caffeine does not pose a danger if consumed in moderate amounts. So, friend, if you are a coffee drinker who is worrying about caffeine, relax. Have a “cuppa joe”.

Appendix D

The List of Reading Span Test Items

This list includes both the sense and nonsense Persian sentences as follows:

Practice Session

Set One

گاه فردي تصميم به انجام کاري مي گيرد که فوق قدرت و توانايي او است.
احتمالا ناتوان ترين افراد بشر کسي است که نتواند کسي با ديگري دوست شود.
به نوشته هاي هنری که نمایش دقيقا از روي آن اجرا مي شود نمايشنامه مي گويند.

Set Two

بر عهده هر انساني در اين دنيا است که استعدادهاي را خدادادي باز شناسد.
خوب است که ما در يك ورزش خاص براي رشد استعدادهاي خود ماهر شويم.
در گرفتاري ها به تو اميد بسته ام و در نا امني ها تو پناه گاه را خويش يافته ام.

Set Three

اگر من تجربه ي اين چند سال را داشتم شايد ديگري را راه بر مي گزیدم.
من حاضرم براي بدست آوردن مدرک علمی بالاتر با همهي مشکلات روبرو بچنگم.
قرآن کتاب آسماني ما مي گوید ارزشمندترين مردم نزد خدا با تقواترين آنها است.
گياهان نور و دي اکسيدکربن را جذب مي کنند تا با آن براي غذا خود بسازند.

Test Session

Set One

اگر در آمد بهتري در ماه های آینده داشته باشم، شاید امسال يك کامپیوترکیفی به روز بخرم.

گل نرگس زیبای من بر اثر سهل انگاری خودم جلوي من چشمان پژمرد.

تازه وارد اتاق او شده بودم که پسردایی دوستم سراسیمه از جای خود برخاست.

Set Two

تیم ایران تلاش زیادی برای پیروزي کرد، اما حریف متأسفانه دوباره به باخت.

با شیلنگ آب زیادی روی زمین پاشیدم به این که امید گرد و خاک بخوابد.

من با آشنا شدن با فلسفه بسیاری از افکار قدیمی خود را به دور انداختم.

Set Three

من در کودکی از تاریکی زیاد می ترسیدم هنوز هم گاهگاهی تاریکی مرا می ترساند.

زنبور زیبایی در نقطه ای کور از پارک دیدم که با ظرافت هر چه تمام تر شهد گل ها را می نوشید.

پدر همیشه به ما می گفت: با بعضی امراض و بیماریها باید آخر عمرتا ساخت.

Set Four

ابو علي سینا حکیم معروف سرزمین علم و تمدن و هنر، قرنهایست که در آسمان علم جهان می درخشد.

بعضی از مارها می توانند زهر خود را تا دو متر طرف به دشمن بپاشند.

اگر مانند سال های گذشته امسال هم بارندگی کم باشد دچار کم آبی می شویم.

دکتر دست شکسته من را گچ گرفت و با يك پارچه به آن گردنم را آویخت.

Set Five

هوشنگ هنوز مانند قدیم ، هر روز شیر بیست گاو را با دست رأس می دوشد.

پدرم همیشه می گفت به دلیل مرگ مادرت اینگونه از پرو پا افتاده شده ام.

اگر باغبان وظیفه شناس مهربان ، مهربانانه اشاره می کرد، کسی این گل های زیبا را نمی چید.

ای کاش هر کودکی در زمان نیاز به محبت ، در آغوش پر مهر و وفای مادر خود آرام بگیرد.

Set Six

دوستان خاص و عام خود را بار دیگر ارزیابی دقیق می کنم و آنها را با معیارهای جدید می سنجم.

با اینکه روی لوله ای آب را پوشانیدم، باز هم زمستان امسال لوله ترکید.

در مهمانی ، به ظاهر آرام نشسته بودم اما دلم و سرکه مثل سیر می جوشید.

حسن پرچم را به دست گرفته بود و نشانه را به پیروزی آن می چرخاند.

Set Seven

اشتباه لفظی و غیر قابل پیش بینی فرهاد در کلاس باعث شد سال گذشته همه به او بخندند.

پسرم لوله بخاری را کشید و با این کار از لوله در جای خود رفت.

حمیده یکی از جواهرات خود را که پارسال گم کرده بود، پشت کمد لباس ها یافت.

مادر با لبخند گفت، بسیار خوب امیدوارم که راه همه کارها به رو باشد.

هنوز بسیاری از دختران خردسال روستایی با زحمت و مشقت زیاد هر سال يك فرش دستی می بافند.

Set Eight

در کتابخانه ای آن شهر کتاب های بسیاری بودند که دشمنان راهمه در آتش سوزانده اند.

پیرمرد کهنسال هر روز صبح از طلوع افتاب تا پاسی از شب ظرافت سنگهای قیمتی را می تراشد.

این نشانه ی رشد فکری است که یادآوری عیب هایمان را به عنوان هدیه ای گرانبها از دوست بپذیریم.

فرهاد با افتخار مي گفت كه از پارسال تاكنون دو هزار مترزمين كشاورزي خريده ام.

ديشب درهاي اتاق باز بود، به همين دليل پشه ها بدن تا صبح مرا گزيدند.

Set Nine

پرستار مهربان و خوشرو، هر روز صبح اول وقت قطره چشم پدر من را مي چكاند.

بسياري از كارمندان رسمي يا ازمايشي هستند كه براي امضاي يك پرونده ساده ماهها مردم را مي دوانند.

بعضي ها براي بدست آوردن ثروت بيشتر صرف نظر از حلال يا حرام ان دست به هر كاري مي زنند.

با اينكه راننده همه تلاش خود را كرد ، باز هم جلسه به ما نرسيد.

دكتر به محمد گفته بود بايد چند بار در ماه ، خود سر با سدر را بشويي

Set Ten

بدون اينكه قصدي داشته باشم مادر را با نسنجيده خود سخنان رنجاندم.

خاله زهرا مرتب اصرار مي كرد، كمی استراحت كن، تو از راه دور آمده اي.

تنها تفاوت فيلم نامه با نمايش نامه اين است كه تمام حرركات بازيگران را فيلم نامه در مي نويسند.

زمستان كه مي شد ، هر شب تا سحر پدر بزرگ براي ما شاهنامه را مي خواند.

از پنجره اتاقم او را در حياط خانه اش مي ديدم كه آب قدم مي زد و باغچه ها را مي داد.

بي شك آنان كه خوب مي نويسند، كساني مي توانند باشند كه خوب بر علوم بشر تسلط دارند.

Set Eleven

اختر فيزيك شاخه اي از علم فيزيك است كه به مطالعه اجرام آسماني دور، مي پردازد.

من براي جلوگيري از افزايش بي رويه وزن خود مجبور شده ام هر روز مقدار قابل توجهی راه بپيمايم.

به محض اينكه آب جوش را در ليوان ريختم، ليوان با صداي بلندي شكست.

متاسفانه دستم به ظرف غذا خورد و ظرف پايين غذا به ميز افتاد.

ديروز هر چه تلاش كردي نتوانستي خوب توپ بسكتبال را سبد درون بياندازي.

هرچه با مينا حرف مي زديم ، به ما نگاه مي كرد و هم دندانهايش را به مي سايبید.

Set Twelve

همیشه شوق دیدن کشوري بيگانه دلهره اي شگفت در رگ من و پي بدن مي آفريند.

گرچه كوچك خانگي ما ، گاهي با همان پنجه هاي مرا ظريف و شكنده اش صورت مي خراشد.

رنگين كمان پس از بارش سيل اسای باران در آسمان ظاهر مي گردد و بعد از اندكي ناپديد مي شود.

مردم محله ی ما با وجود بدن كوچك ، حاج مرتضي را پهلوان مرتضي مي ناميدند.

هنوز صدای سرود ملي بلند نشده بود كه پرچم را سربازان در ميدان حاضر برافراشتند.

من براساس عادت گذشته، هنگام غذا به خوردن روي نمك آن مي پاشم.

English equivalent of Persian sentences used in the Reading span test (It should be noted that the following sentences were not used in this study).

Practice Session

Set One

- An individual sometimes decides to do something which is beyond his abilities.
- The most miserable person is the one who is not able to make friends.
- The playwright is a piece of draft based on which a performance is played.

Set Two

- It is an individual's responsibility to realize his own gifted capabilities.
- It is a good idea that I try to become a professional in just one sport I am interested in.
- In my difficult and unsafe situations, I expect your assistance and support.

Set Three

- If I had such a kind of experience, I might choose another way.
- I am happy to challenge with all potential problems to develop my knowledge.
- Our holy book says that the best people are those who promote the dignity of mankind.
- Plants absorb sunlight and carbon dioxide to make food for themselves.

Test Session

Set One

- If I have an income raise this year, I may buy a laptop.
- My beautiful flower got withered due to my own carelessness.
- I had just arrived in the room where my friend's cousin embarrassingly stood up.

Set Two

- The national team made a lot of attempts, but they lost the game again.
- I spread a lot of water on the floor to remove the dirt placed on it.
- By getting familiar with philosophy, I threw away all my illogical thoughts.

Set Three

- When I was a child, I was frightened of darkness, and I am still terrified by that.
- I saw a beautiful honey bee which was delicately sucking up the nectar of a flower.
- My father always kept telling us that we have to tolerate some diseases up to the end of life.

Set Four

- Avicenna, a famous Iranian scientist, has been shining in the world of science for centuries.
- Some snakes are able to throw away their poison to their enemies up to three meters.
- If there is the shortage of raining as before, we will face with the shortage of water again.
- The physician dressed my arm with plaster cast and to keep it safe.

Set Five

- Hooshang is able to milk twenty cows traditionally every day.
- My dad told me that he had become so depressed due to my mother's death.
- If the gardener was warning kindly, nobody could dare to pick up such beautiful flowers.
- I wish each kid could relax in his mom's affectionate arms once he needs motherly love.

Set Six

- I evaluate all my friends meticulously based on my new criteria once more.
- Although I had covered the pipe in winter, it would have been broken away.
- It seemed that I was in a relaxed way in the party, but in fact I was in an agitated mood.
- Hassan was holding the flat and got it around as the sign of victory.

Set Seven

- Ahmad's slip of tongue made everybody laugh at him.
- As my son pulled away the chimney, it removed from its original place.
- Hamideh found her jewellery that was lost last year behind the wardrobe.
- My mom smiled and told us, "All right, I hope everything is going on well with you".
- There are still some rural girls who are making rugs by their hands.

Set Eight

- There were a lot of books in the city's central library that were burnt by the enemy.
- The elderly man shaves that precious piece of stone every day from sunlight to the sunset.
- Accepting our friends' advice on our shortcomings is the sign of intellectuality.

- Farhad enthusiastically said that he had bought 2000 meters farmland since last year.
- Since the doors of the room were left open last night, the mosquitoes bit me a lot.

Set Nine

- The kind nurse looked after my mother by putting the drops in her eyes each morning.
- There are a lot of administration staffs who trump up an excuse for just a signature.
- Some people embark on gaining more money irrespective whether it is legal or illegal.
- Despite the fact that the driver made all his efforts, we did not get to the meeting on time.
- Doctor had told Mohammad to wash his hair with lotus several times a month.

Set Ten

- Without any particular intention, I disturbed my mother by some rash remarks.
- Due to coming from a long distance, Aunt Zahra insisted on relaxing for some time.
- The only difference between a playwright and a film script is that all actions are recorded in the film script.
- In winter, our grandfather narrated a long story from Shahnameh each night.
- I could view him through the window walking in the yard watering the flowers in the garden every day.
- With no doubts, those who are able to write well are those who know a lot.

Set Eleven

- Astronomy is a particular branch of physics which is concerned with the study of far planets and stars.
- To control my weight, I have to walk for some kilometres every day.

- As soon as I poured the boiling water in the glass, it cracked and broke away.
- Unfortunately I hit the dish on the table accidentally and it fell off the floor and broke down.
- You did not manage to put the ball in the basket despite that you made a lot of efforts.
- While we looked at Mina, she was just looking at us and rubbing her teeth together.

Set Twelve

- Visiting a foreign country has always been enthusiastic to me.
- Our domestic kitten sometimes scratches us with its delicate paws.
- Rainbow appears in the sky a little bit after raining and then disappears.
- The people in our local area called out Morteza as a hero although he was very short.
- The flag had been erected by the soldiers before the national anthem was broadcasted.
- As a habit, I usually pass salt on my dish before I serve it.

Appendix E

Math Span Test Items

Simple math problems used in the Math Span Test are as follow:

Practice Session

Set One

$$7 + 9 = ?$$

$$8 - 1 = ?$$

Set Two

$$9 + 4 = ?$$

$$5 - 2 = ?$$

Set Three

$$3 + 7 = ?$$

$$6 - 1 = ?$$

$$3 + 9 = ?$$

Set Four

$$4 - 2 = ?$$

$$7 + 8 = ?$$

$$9 - 3 = ?$$

Test Session

Set One

$$2 + 1 = ?$$

$$9 - 6 = ?$$

Set Two

$$8 - 5 = ?$$

$$3 + 2 = ?$$

Set Three

$$1 + 3 = ?$$

$$7 - 4 = ?$$

Set Four

$$4 + 3 = ?$$

$$9 - 5 = ?$$

$$6 + 2 = ?$$

Set Five

$$8 - 4 = ?$$

$$2 + 5 = ?$$

$$7 - 6 = ?$$

Set Six

$$4 + 8 = ?$$

$$5 + 3 = ?$$

$$6 - 5 = ?$$

Set Seven

$$9 - 4 = ?$$

$$5 + 1 = ?$$

$$6 - 2 = ?$$

$$5 + 7 = ?$$

Set Eight

$1 + 9 = ?$

$2 + 4 = ?$

$3 - 1 = ?$

$8 - 6 = ?$

Set Nine

$5 - 3 = ?$

$8 - 2 = ?$

$2 + 7 = ?$

$6 + 9 = ?$

Set Ten

$5 + 4 = ?$

$7 - 1 = ?$

$3 + 6 = ?$

$9 - 8 = ?$

$6 + 5 = ?$

Set Eleven

$8 - 7 = ?$

$1 + 4 = ?$

$9 - 3 = ?$

$2 + 8 = ?$

$5 - 1 = ?$

Set Twelve

$6 - 3 = ?$

$9 - 7 = ?$

$8 + 1 = ?$

$7 - 5 = ?$

$1 + 6 = ?$

Set Thirteen

$9 - 1 = ?$

$6 + 4 = ?$

$3 - 2 = ?$

$7 + 6 = ?$

$4 - 3 = ?$

$8 + 5 = ?$

Set Fourteen

$7 - 2 = ?$

$4 - 1 = ?$

$8 - 3 = ?$

$4 + 7 = ?$

$9 + 5 = ?$

$6 + 8 = ?$

Set Fifteen

$2 - 1 = ?$

$5 + 6 = ?$

$7 - 3 = ?$

$2 + 9 = ?$

$6 - 4 = ?$

$1 + 7 = ?$

Appendix F

English Non-words

1- peb kib bon deet

(Identical sequence)

2- peeb kol goob mab

-peeb goob kol mab

3-pib kom gook tam

(Identical sequence)

4-neeg gop doob jat

-neeg doob gop jat

5-pim goot neeb kig doog

-pim neeb goot kig doog

6-meb teeb dook cam jawn

-meb teeb cam dook jawn

7-teel nog gub pem chad

(Identical sequence)

8-jep cham tudge meech pag

(Identical sequence)

9-noog teed gadge pab chud

-noog teed pab gadge chud

10-mep teeg keb chim nup jit

-mep keb teeg chim nup jit

11-jick mip chool lod nug tep

(Identical sequence)

12-teeg chan mig padge dop nam

-teeg chan mig dop padge nam

13-geed mun peb cheem tep nuck

(Identical sequence)

14-bick meep tooch leck nam gab

-bick tooch meep leck nam gab

15-choom mit gab tidge pag nool

-choom gab mit tidge pag nool

16-jeck leem gan chut bock mon tud

(Identical sequence)

17-mitch tem jeeg lib cug bup neb

-mitch jeeg tem lib cug bup neb

18-pock mun tob juck lidge ged coom

(Identical sequence)

19- toock jeel peeb modge dack lig neeb

(Identical sequence)

20-lon cam deech mot jooch ked gock

-lon cam deech mot ked jooch gock

21-dook mip chon teep jal noog goot

-dook mip teep chon jal noog goot

22-kom chen meb lud tam dit loog

(Identical sequence)

References

- Abraham, R.G., & Chapelle, C.A. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal*, 76, 468-479.
- Abu Rabia, S. (2003). The influence of working memory on reading and creative writing processes in a second language. *Educational Psychology*, 23(2) 209-222.
- Adams, M.J. (1994). Modelling the connections between word recognition and reading. In R.B. Ruddell, M.R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed. pp. 830-863). Newark, DE: International Reading Association.
- Adams, A.M., & Gathercole, S.E. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology*, 41 (1), 216-233.
- Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Alderson, J.C. (1983). The cloze procedure and proficiency in English as a foreign language. In J.W. Oller (Eds.): *Issues in Language Testing Research* (pp. 205-217). Rowley, Massachusetts: Newbury House.

Alderson, J.C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.

Alderson, J. C. & Banerjee, J. (2002). Language testing and assessment (part 2).

Language Teaching, 35, 79-113.

Alderson, J.C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*.

Cambridge University Press.

Alptekin C., & Erçetin G. (2009). Assessing the relationship of working memory to L2 reading: Does the nature of the comprehension process and reading span task make a difference? *System*, 37, 627-639.

Alptekin C., & Erçetin G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, 33 (2), 206-219.

Anderson, N.J. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal*, 75, 460-472.

Anderson, N. (2007/2008/2009). *ACTIVE skills for reading: Intro, Books 1-4*.
Boston: Heinle.

- Ando, J., Fukunaga, N., Kurahashi, J., Stuto, T., Nakano, T., & Kage, M. (1992). A comparative study on two EFL teaching methods: The communicative and the grammatical approach. *Japanese Journal of Educational Psychology*, 40, 247-256.
- Atkins, W.B. & Baddeley, A.D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19, 537-552.
- Atkinson, R.C., & Shiffrin, R.M. (1968). Human Memory: A proposed system and its control processes. In K. W. Space (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp., 89-195). New York: Academic Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49 A, 5-28.

Baddeley, A.D. (2000a). The episodic buffer. A new component of working memory?

Trends in cognitive sciences, 4, 417-423.

Baddeley, A.D. (2003). Working memory and language: An overview. *Journal of*

Communication Disorders, 36, 189-208.

Baddeley, A.D. (2006). Working memory: An overview. In Pickering, S. J.(Ed.),

Working memory and education (pp. 3-26). Elsevier Inc.

Baddeley, A.D. (2007). *Working Memory, Thought and Action*. Oxford University

Press.

Baddeley, A.D., & Hitch, G. (1974). Working memory revised. *American Psychologist*,

November (2001), 851-864.

Baddeley, A.D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit

retrieval? *Memory and Cognition*, 21 (2), 146-155.

Baddeley, A.D., & Logie, R.H. (1999). Working memory: The multiple component model. In

A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active*

maintenance and executive central (pp. 28-61). Cambridge: Cambridge University

Press.

- Balota, D., Pollasek, A., & Rayner, K. (1985). The interaction of contextual constraints and Parafoveal visual information in reading. *Cognitive Psychology*, 17, 364-390.
- Barnes, M. A., Faulkner, H., Wilkinson, M., & Dennis, M. (2004). Meaning construction and integration in children with hydrocephalus. *Brain and Language*, 89, 47-56.
- Barnes, M.A., Huber, J., Johnston, A. M., & Dennis, M. (2007). A model of comprehension in spinabifida meningomyelocele: Meaning activation, integration, and revision. *Journal of the International Society*, 13, 854-864.
- Barry, S. and Lazarte, A.A. (1998). Evidence for mental models: how do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *Modern Language Journal*, 82, 2,176-193.
- Becker, C.A. (1985). What do we really know about semantic context during reading? In D. Besner, T. Walker & G. Mackinnon (Eds.), *Reading research: Advances in theory and practice*, 5 (pp. 125-166). NewYork: Academic Press.
- Beech, J.R. (1997). Assessment of memory and reading. In Beech, J.R. and Singleton, C. (Eds.), *The Psychological Assessment of Reading*. London: Routledge.

- Bernhardt, E.B. (1983). Testing foreign language reading comprehension: The immediate recall protocol, *Die Unterrichtspraxis*, 16, 17-33.
- Bernhardt, E.B. (1991). *Reading development in a second language*. Norwood, NJ: Albex.
- Bernhardt, E.B. (1991b). *Reading development in a second language: Theoretical research and classroom perspectives*. Norwood, NJ: Albex.
- Bever, T.G. (1970), The cognitive basis of linguistic structures. In J.R. Hayes (Eds.), *Cognition and development of language*, New York, John Wiley & Sons.
- Biemiller, A. (1979). Changes in the use of graphic and contextual information as functions of passage difficulty and reading achievement level. *Journal of Reading Behavior*, 11, 307-319.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12-21.
- Brown, J.D. (1989a). *Cloze item difficulty*. JALT Journal, 10, 15-42.
- Brown, J.D. (1992). What text characteristics predict human performance on cloze test items. *Proceedings of the 3rd Conference on Second Language Research in Japan*, (pp. 1-26) Niigata, Japan: International University of Japan.

- Brown, G.D.A., & Hulme, C. (1992). Cognitive psychology and second language processing: The role of short-term memory. In R.J. Harris (Ed.), *Cognitive approaches to bilingualism* (pp. 105-122). Amsterdam: North Holland/ Elsevier Science Publishers.
- Cain, K. (2006). Children's reading comprehension: The role of working memory in normal and impaired development. In Pickering, S. J. (Ed.), *Working memory and education* (pp. 61-91). Elsevier Inc.
- Cain, K., & Oakhill, J.V. (2004). Reading comprehension difficulties. In T. Nunes & P. Bryant (Eds.), *Handbook of Children's Literacy* (pp. 313-338). Dordrecht: Kluwer Academic Publishers.
- Cain, K. & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76, 697-708.
- Cain, K., Oakhill, J.V., Barnes, M.A., & Bryant, P.E. (2001). Comprehension skill, inference making ability and their relation to knowledge. *Memory and Cognition*, 29, 850-859.
- Cain, K., Oakhill, J.V. & Bryant, P.E. (2000a). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing*, 13, 31-56.

Cain, K., Oakhill, J., & Bryant, P.E. (2004). Children's reading comprehension ability: Concurrent prediction by working memory. Verbal ability and component skills. *Journal of Educational Psychology*, 96, 31-42.

Cain, K., Oakhill, J.V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, 30, 681-694.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, 96, 671-681.

Carroll, D.W. (2008). *Psychology of Language*. Thomson Wadsworth: Canada.

Carroll J.B. (1964). *Language and thought*. Englewood Cliffs, N.J. Prentice Hall.

Carroll, J.B. (1965). The prediction of success in foreign language teaching. In R. Glaser (Ed.), *Training, research, and education* (pp. 87-136). New York: Wiley.

Carroll, J.B., (1981). Twenty-five years of research on foreign language aptitude. In K.C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, M.A: Newbury House.

Carroll, J.B. (1993). *Human Cognitive Abilities: A survey of Factor Analytic Studies*.

New York : Cambridge University Press.

Carretti, B., Borella, E., Cornoldi, C., De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and individual differences*, 19, 246-251.

Carretti, B., Cornoldi, C., De Beni, R., & Palladino, P. (2004). What happens to information to be suppressed in working memory tasks? Short and Long term effects. *Quarterly Journal of Experimental Psychology*, 57 A, 1059-1084.

Carretti, B., Cornoldi, C., De Beni, R., Romano, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology*, 91, 45-66.

Chamot, A.U., & El-Dinary, P.B. (1999). Children's reading strategies in language immersion classrooms. *Modern Language Journal*, 83, 319-338.

Cheung, H. (1996). Non word span as a unique predictor of second language vocabulary learning. *Developmental Psychology*, 32, 867-873.

Chun, D.M., & Payne, J.S. (2004). What makes students click: Working memory and look-up behavior. *System*, 32, 481-503.

- Clarke, M.A. (1980). The short circuit hypothesis of ESL reading: Or when language competence interferes with reading performance. *Modern Language Journal*, 64, 203-209.
- Cobb, T. (2002). *Web VocabProfile*. Retrieved October, 2009 from [http:// www.Lextutor.ca/vp/](http://www.Lextutor.ca/vp/).
- Conway, A.R.A., Cowan, N., & Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8, 331-335.
- Conway, A.R.A. & Engle, R.W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, 4, 577-590.
- Conway, A.R.A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12 (5), 769-786.
- Cunningham, S., & Moor, P. (2005). *New Cutting Edge: Elementary and Intermediate levels*. Pearson Education.
- Cziko, G. (1980). Language competence and reading strategies: A comparison of first and second language oral reading errors. *Language Learning*, 30, 101-116.

- Daneman, M., (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, 20, 445-464.
- Daneman, M. & Carpenter, P.A. (1980). Individual Differences in Working Memory and Reading, *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1-18.
- Daneman, M., & Merikle, P. M. (1996). Working memory and Language comprehension: A Meta-analysis *Psycholinguistic Bulletin and Review*, 3, 422- 433.
- Davis, F.B. (1994). Fundamental factors of comprehension in reading. *Psychometrika*, 9, 185-197.
- Davis, F.B. (1994). Fundamental factors of comprehension in reading. *Psychometrika*, 9, 185-197.
- Day, R.R., & Park, J.S. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, 17 (1), 60-73.
- de Beni, R., & Palladino, P. (2000). Intrusion errors in working memory tasks: Are they related to reading comprehension ability? *Learning and Individual Differences*, 12, 131-143.

- de Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *Quarterly Journal of Experimental Psychology*, 51, 305-320.
- Dechant, E. (1991). *Understanding and teaching reading: An interactive model*. Hillsdale, NJ: Lawrence Erlbaum.
- de Jong, P. & de Jong, P.F. (1996). Working memory, intelligence, and reading ability in children. *Personality and Individual Differences*, 21, 1007- 1020.
- de Jong, P.F., & Olson, R.K. (2004). Early predictors of letter knowledge. *Journal of Experimental Child Psychology*, 88, 254-273.
- de Jong, P.F. (2006). Understanding normal and impaired reading development. A working memory perspective. In Pickering, S.J. (Ed.), *Working memory and Education* (pp. 33-60). Elsevier Inc.
- Della Sala, S., Gray, C., Baddeley, A. Allamano, N., Willson, L. (1999). Pattern span: A means of unwelding visuo-spatial memory. *Neuropsychologia*, 37, 1189, 1199.
- Divesta, F. J. & Dicintio, M. J. (1997). Interactive effects of working memory span and text context on reading comprehension and retrieval. *Learning and Individual Differences*, 9 (3), 215-231.

Dufva, M., & Voeten, M.J.M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, 20, 329-348.

Ebbinghaus, H. (1964/1885). *Memory: A contribution to experimental Psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover (original work published 1885).

Ehri, L.C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J.L. Metsala & L.C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3-40). Mahwah, NJ: Lawrence Erlbaum Associates.

Elley, W.B. (1992). *How in the world do students read?* Hamburg: International Association of Evaluation of the Educational Achievement (distributed through the International Reading Association).

Ellis, N.C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91-126.

Ellis, N.C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 33-68.

- Ellis, N.C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27 (2), 305-352.
- Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, 84, 253-262.
- Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory mechanisms of active maintenance and executive*.
- Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, 9 (2), 147-171.
- Ericson, K.A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K.A. Ericson (ed.), *The road to excellence: The acquisition of expert performance in the arts and science, sports and games*, 1-50. Mahwah, N.J.: Erlbaum.
- Ericson, K.A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49, 725-747.

- Ericson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Favreau, M., & Segalowitz, N. (1982). Second language reading in fluent bilinguals. *Applied Psycholinguistics*, 3, 329-341.
- Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in reading a second language. *Memory and Cognition*, 11, 565-574.
- Field, J. (2003). *Psycholinguistics: A resource book for students*. Routledge: London.
- Flesch, R. (1955). *Why Johnny can't read?* New York: Harper and Row.
- Fodor, J.A., Bever, T.G., and Garrett, M.F. (1974). *The psychology of language: An Introduction to Psycholinguistics and Generative Grammar*, New York, McGraw-Hill.
- Fotos, S. S. (1991). The Cloze test as an integrative measure of EFL proficiency: A substitute for the essay on the college -entrance examination.? *Language Learning*, 41 (3) 313-336.
- Friedman, P.N., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51, 136-158.

- Friedman, P.N., & Miyake, A. (2005). Comparison of four scoring methods for the reading span, *Behavioral Research Method*, 37 (4), 581-590.
- French, L.M. (2006). *Phonological working memory and L2 acquisition: A developmental study of Québec Francophone children learning English*. New York: Edwin Mellen Press.
- French, L.M., & O'Brien, I. (2008). Phonological memory and children's L2 grammar learning. *Applied Psycholinguistics*, 29, 463-487.
- Frisk, V., & Milner, B. (1990). The relationship of working memory to the immediate recall of stories following unilateral temporal or frontal lobectomy. *Neuropsychologia*, 28 (2), 121-135.
- Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, 39, 3- 27.
- Gathercole, S.E. & Baddeley, A.H. (1989). Evaluation of the role of phonological short-term memory in the development of vocabulary: A longitudinal study. *Journal of Memory and Language*, 28, 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29 (3), 336-360.

Gathercole, S.E. & Baddeley, A.D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition. *European Journal of Psychology of Education*, 8, 529-572.

Gathercole, S.E., Frankish, C.R., Pickering, S.J., & Peaker, S.M. (1999). Phonetic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 25, 84-95.

Gathercole, S.E., Pickering, S.J., Hall, M., & Peaker, S.M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology*, 54A, 1, 1-30.

Gathercole, S.E., Service, E., Hitch, G.J. Adams, A.M., & Martin, A.J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13, 65-77.

- Gathercole, S.E., & Thorn, A.S.C. (1998). Phonological short-term memory and foreign language learning. In Healy, A.F., & Bourne, L.E., (Eds.). *Foreign Language Learning: Psycholinguistics Studies on Training and Retention* (pp. 141-158). Mahwah, NJ: Erlbaum.
- Gathercole, S.E., Willis, C.S., Emslie, H., & Baddeley, A.D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887-898.
- Gathercole, S. E., Willis, C. S., Emslie, H. & Baddeley, A.D. (1991). The influence of the number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12, 349-367.
- Genesee, F., Geva, E., Dressler, C., & Kamil, M.L. (2006). Synthesis: Cross-linguistic relationships. In D. August & T. Shanahan (Eds.), *Developing literacy in second language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 153-183). Mahwah, NJ: Erlbaum.
- Gernsbacher, M.A., & Faust, M. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 245-262.

- Gernsbacher, M.M., Varner, K.R. , & Faust, M.E. (1990). Investigating individual differences in general comprehension skill. *Journal of Experimental Psychology: Learning Memory and Cognition*, 16, 430-445.
- Gholamain, M. & Geva, E. (1999). Orthographic and cognitive factors in the concurrent development of basic reading skills in English and Persian. *Language Learning*, 49 (2), 183-217.
- Goodman, K.S. (1965). *A linguistic study of cues and miscues in reading*. Elementary English 42, 639-643.
- Goodman, K.S. (1976), Reading: A Psycholinguistic guessing game. *Journal of the Reading Specialist*, 6, 126-135.
- Goodman, K. (1996). *On reading*. Portsmouth, NH: Heinemnn.
- Gough, P.B. (1972). One second of reading. In J.F. Kavanagh and I.G. Mattingly (Eds.), *In Language by ear and by eye* (pp. 331-358). Cambridge, Mass: MIT Press.
- Grabe, W. (2000). Reading research and its implications for reading assessment. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (Studies in Language Testing 9, pp. 226-62). Cambridge: Cambridge University Press.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Grabe, W. & Stoller, F. L. (2002). *Teaching and Researching Reading*. London: Longman.

Grabe, W. & Stoller, F.L. (2011). *Teaching and researching reading (2nd Ed.)*. London: Longman.

Graesser, A., Millis, K., & Zwaan, R. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.

Gronland, N.E. (1973). *Preparing Criterion-Referenced Tests for Classroom Instruction*. New York: MacMillan.

Hagley, J. F. (1987). *The Suffolk Reading Scale*. Windsor: NFER-Nelson.

Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. Longman English Language Series 9. London: Longman.

- Hambrick, D.Z., & Engle, R.W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is – power hypothesis. *Cognitive Psychology*, 44, 339-387.
- Hambrick, D.Z., Kane, M.J., & Engle, R.W. (2005). The role of working memory in higher level cognition: Domain specific versus domain general perspectives. In R. Sternberg & J.E. Pertz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 104-121). New York: Cambridge University Press.
- Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25-38.
- Harris, D.P. (1969). Testing English as a second language. New York: McGraw-Hill.
- Harris, M. and Hatano, G. (Eds.) (1999b). *Learning to read and write: A cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Hasher, L. & Zacks, R.T. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H. Bower (Ed.), *The psychology of learning and motivation*, 22, (pp. 193-225). Academic Press: San Diego.

Hatcher, P.J. & Hulme, C. (1999). Phonemes, rhymes, and intelligence as predictors of children's responsiveness to remedial reading instruction: Evidence from a longitudinal intervention study. *Journal of Experimental Child Psychology*, 72, 130-153.

Haviland, S.E., and Clark, H.H. (1974), What's New? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512-521.

Haynes, M., & Carr, T.H. (1990). Writing system background and second language reading: A component skills analysis of English reading by native speaking-readers of Chinese. In T.H. Carr & B.A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-421), San Diego: Academic Press.

Heaton, J.B. (1988). *Writing English Language Tests*. London: Longman.

Hockberg, J. (1970). Components of literacy speculations and exploratory research. In H. Levin & J.P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books.

Horiba, Y. (1990). Narrative comprehension processes: A study of native and non-native readers of Japanese. *Modern Language Journal*, 74, 188-202.

Hudson, T. (2007). *Teaching second language reading*. New York, NY: Oxford University Press.

Hummel, K.M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, 30, 225-249.

Hummel, K.M., & French, L. M. (2010). Phonological memory and implications for the second language. *The Canadian Modern Language Review*, 66 (3), 371-391.

Johnson, A. M. & Barnes, M.A. (2008). Reading comprehension: *Developmental Processes, Individual Differences, and Interactions*. *Canadian Psychology*, 49, 2, 125-132.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grade. *Journal of Educational Psychology*, 80, 437-447.

Juel, C., Griffith, P.L., & Gough, P.B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78, 243-255.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, 102, 199-225.

- Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixation to comprehension, *Psychological Review*, 87, 329-354.
- Just, M.A., & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Just, M.A., & Carpenter, P.A. (2002). A capacity theory of comprehension: Individual differences in working memory. In T.A. Polk & C. M. Seifert (Eds.), *Cognitive Modelling* (pp. 131-177). Cambridge, MA: MIT Press.
- King, J. & Just, M.A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580-602.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Kintsch, W. (1998). *Comprehension: A framework for cognition*. New York: Cambridge University Press.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production.

Psychological Review, 85, 363-394.

Koda, K. (1996). L2 word recognition research: a critical review. *The Modern Language*

Journal, 80, (5), 450-460.

Koda, K. (2005). *Insights into Second Language Reading: A cross-linguistic approach*.

Cambridge University Press.

Koda, K. (2007). *Reading and Language Learning*. Blackwell Publishing: U.S.A.

Kroll, J., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of

lexical fluency in a second language. *Second Language Research*, 18, 137-171.

Kormos, J. & Sáfár, A. (2008). Phonological short-term memory, working memory and

foreign language performance in intensive language learning. *Language Cognition*, 11

(2), 261-271.

Kuhara-Kojima, K., Hatano, G., Saito, H., & Haebara, T. (1996). Vocalization latencies

of skilled and less skilled comprehenders for words written in Hiragana and Kanji.

Reading Research Quarterly, 31, 158- 171.

- LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293- 323.
- Larsen-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge, Taylor & Francis.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16 (3), 307-322.
- Lee, F. (1986). On the of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 83-93.
- Lee, J.F. (1998). The relationship of verb morphology to second language reading comprehension and input processing. *The Modern Language Journal*, 8, 33-48.
- Lee, G. (2002). The influence of several factors on reliability for complex reading comprehension tests. *Journal of Educational Measurement*, 39, 149-164.
- Lesser, J.M. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory, *Language Learning*, 57, 2, 229-270.

- Leow, R. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition, *Applied language learning*, 8, 151-182.
- Leow, R. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, 84, 496-509.
- Lesley, T., Hansen, C., & Zukowski/ Faust, J. (2003). *Placement and Evaluation Package*. Cambridge: Cambridge University Press.
- Leslie, L., & Caldwell, J. (2009). Formal and Informal Measures of Reading Comprehension. In Israel, S.E., & Duffy G.G. (Eds.), *Handbook of Research on Reading Comprehension* (pp., 403-427). New York: Routledge.
- Leslie, L., & Caldwell, J. (2006). *Qualitative reading inventory-4*. Upper Saddle River, NJ: Pearson Education.
- Levin, H. & Kaplan, E.L. (1970). Grammatical structure and reading. In H. Levin & J.P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books.
- Light, L.L., & Anderson, P.A. (1985). Working memory capacity, age and memory for discourse. *Journal of Gerontology*, 40, 737-747.

- Long, D.L., & Chong, J.L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1424-1429.
- Macaro, E. (2003). *Teaching and Learning a Second Language: A guide to recent research and its applications*. Continuum: London.
- Mackey, A., Adams, R., Stafford, C. & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, 60 (3), 501-533.
- Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing interactional feedback and L2 development. In P. Robinson (Eds.), *Individual differences and instructed language learning* (pp.181-209). Philadelphia: Benjamins.
- Mackey, A. & Sachs, R. (2011). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, 1-37.
- Mackworth, J.F. (1972). Some models of the reading process: Learners and skill readers. *Reading research Quarterly*, 7, 701-733.
- Macnamara, J. (1970). Comparative studies of reading and problem solving in two languages. *TESOL Quarterly* 4, 107-116.

- Madsen, H.S. (1983). *Techniques in Testing*. New York: Oxford University Press.
- Masoura, E., & Gathercole, S.E. (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology*, 34, 383-388.
- Masoura, E. V., & Gathercole, S. E., (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, 13, 422-429.
- McConkie, G.W., & Zola, D. (1981). Language constraints and the functional stimulus in reading. In A.M. Lesgold & C.A. Perfetti (Eds.), *Interactive process in reading* (pp. 155-175). Hillsdale, NJ: Erlbaum.
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27 (1), 79-103.
- McGuinness, D. (2005). *Language development and learning to read: The scientific study of how language development affects reading skill*. MIT Press: Cambridge, Massachusetts. London, England.
- Messick, S. (1989). Validity. In R.L. Linn (Eds.), *Educational measurement* (pp. 13-104). Washington D.C: American council on Education.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, 27, 272-277.

Mitchell, D.C., (1982). *The process of reading: A cognitive analysis of fluent reading and Learning to Read*. John Wiley & Sons Ltd.

Miyake, A. & Friedman, N.P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In Healy, A.F. & Bourne L.E., *Foreign Language Learning: Psycholinguistics Studies on Training and Retention* (pp. 339-361).

Miyake, A., Just, M.A., & Carpenter, P.A. (1994). Working memory constraints on the resolution of lexical ambiguity: maintaining multiple interpretations in a neutral context. *Journal of Memory and Language*, 33, 175-202.

Mizera, G.J. (2006). *Working memory and L2 oral fluency*. (Ph.D Dissertation, University of Pittsburgh). Retrieved from <http://etd.library.pitt.edu/ETD/available/etd-04262006-124945>.

Morris, N., & Jones, D.M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, 81, 111-121.

Nation, I.S.P. (2009). *Teaching ESL/EFL Reading and Writing*. Routledge, Taylor & Francis.

Nation, K., Adams, J.W., Bowyer-Crane, C.A., Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139-158.

Norbury, C.F., & Bishop, D.V.M. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment, and high-functioning autism. *International Journal of Language and Communication Disorders*, 37, 227-251.

Nuttall, C. (1996). *Teaching reading skills*. Oxford: Heinemann English Language Teaching.

Oakhill, J.V. & Cain, K. (1997). Assessment of comprehension in reading. In Beech, J.R. and Singleton, C. (Eds.), *The psychological awareness in reading* (pp. 102-123). London : Routledge.

- Oakhill, J.V., Cain, K., & Yuill, N. (1998). Individual differences in young children's comprehension skill: Toward an integrated model. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 343-367). Mahwah, NJ: Lawrence Erlbaum Associates.
- O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language and production by adult learners. *Applied Psycholinguistics*, 27, 377-402.
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29, 557-582.
- Oller, J.W. (1975). *Pragmatics mappings*. *Lingua*, 35, 333-34.
- Oller, J.W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Oney, B., Peter, M. & Katz, L. (1997). Phonological processing in printed word recognition: Effects of age and written system. *Scientific Studies of Reading*, 1, 65-83.

O'Reilly, R. C., Braver, T. S., & Cohen, J.D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory mechanisms of active maintenance and control* (pp. 375-411). Cambridge: Cambridge University Press.

Osaka, M., Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of Psychonomic Society*, 30, 287-289.

Osaka, M., Osaka, N. & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychometric Society*, 31, 117-118.

Papagno, c., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effects of phonological similarity and item length. *Quarterly Journal of Experimental Psychology*, 44A, 47-67.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331-347.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, 48A, 98-107.

Paris, S.G., Wasik, B. A., & Turner, J.C. (1991). The development of strategic readers.

In R. Barr, M. L., Kamil, P. B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (pp. 609-640). New York: Longman.

Payne, T.W., Kalibatseava, Z., & Jungers, M.K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension? *Learning and Individual Differences*, 19-123.

Payne, J. S., Whitney, P.J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *Journal of CALICO*, 20 (1), pp. 7-32.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.

Pennington, B.F., Van Orden, G., Kirson, D., & Haith (1991). What is the casual relation between verbal STM problems and dyslexia? In S.A Brady & D.P. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 173-186). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C.A. (1985). *Reading Ability*. Oxford: Oxford University Press.

Perfetti, C.A. (1994). Psycholinguistics and reading ability. In M.A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 849-894). San Diego: Academic Press.

Perfetti, C. (1999). Comprehending written language: A blueprint for the reader. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 167-208). Oxford: Oxford University Press.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 8, 293-304.

Perfetti, C.A., & Lesgold, A.M. (1977). Discourse comprehension and sources of individual differences. In M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141-183). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Perfetti, C., & Hart, L. (2001). The lexical basis of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67-86). Washington, DC: American Psychological Association.

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamins.

Perfetti, C.A., & Lesgold, A.M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In L.B. Resnick & P. Weaver (Eds.), *Theory and Practice of early reading* (pp. 57-84). Hillsdale, NJ: Erlbaum.

Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.

Pressley, M. (2006). *Reading instruction that works* (3rd ed.). New York: Guilford Press.

Pring, L. & Snowling, M. (1986). Developmental changes in word recognition: An information-processing account. *Quarterly Journal of Experimental Psychology*, 38, 395-418.

Rayner, K., Bertera, J.H. (1979). Reading without a fovea. *Science*, 206, 468-469.

Richards, J.C., & Schmidt R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. Pearson Education, Inc.

Roberts, R., & Gibson E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research*, 31, 573-598.

Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning*, 45, 283-331.

- Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Philadelphia: Benjamins.
- Robinson, R. D. (2005). *Readings in reading instruction: Its history, theory, and development*. Pearson Education, Inc.
- Roher, K., & Ganem-Gutierrez, G. A. (2008). *Metalinguistic knowledge in instructed L2 learning: An individual difference variable?* University of Essex: Essex Research Reports in Linguistics.
- Roth, F.P., Speece, D.L. & Cooper, D.H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research*, 95(5), 259-272.
- Royer, J.M., Greene, B.A. and Sinatra, G.M. (1987). The sentence verification technique: the practical procedure teachers can use to develop their own reading and listening comprehension. *Journal of Reading*, 30, 414-423.
- Rumelhart, D.E. (1977). Toward an interactive model of reading. In S. Dornic (Eds.), *Attention and Performance* (pp. 573-603). New York: Academic Press.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 229-248). New York: Oxford University Press.

Salthouse, T.A., & Babcock, R.L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, 27, 763-776.

Samuels, S.J. & Kamil, M.L. (1988). Models of the reading process. In Carrell, P., Devine, J. & Eskey, D. (Eds.), *Interactive approaches to second language reading* (pp. 22-36). Cambridge: Cambridge University Press.

Sanchez, C.A., Wiley, J., Miura, T.K., Colflesh, G.J.H., Ricks, T.R., Jensen, M.S., & Conway, A.R.A. (2010). Assessing working memory capacity in anon-native language. *Learning and Individual Differences*, 20, 488-493.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319-353). Cambridge: Cambridge University Press.

Schank, R.C. (1978). Predictive understanding. In R.N. Campbell & P.T. Smith (Eds.), *Recent advances in the psychology of language—formal and experimental Approaches* (pp. 91-101). New York, NJ: Plenum Press.

- Schankweiler, D., & Liberman, I. Y. (1972). Misreading. A research for causes. In J.F. Kavanaugh & I. G. Mattingly (Eds.), *Language by eye and by ear* (pp. 293-317). Cambridge: MIT Press.
- Schmalhofer, F., McDaniel, M.A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, 33, 105-132.
- Schmidt, R. (1990). The role of consciousness in L2 learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, S.R. (1994). Effects of humor on sentence memory. *Journal of experimental Psychology: Learning Memory and Cognition*, 20 (4), 953-967.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu: University of Hawai'i.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.
- Schneider, P. (1999). Working memory in a multilevel hybrid connectionist control architecture (CAP2). In A. Miyake & P. Shah (Eds.), *Models of working memory mechanisms of active maintenance and control*, 340-374. Cambridge: Cambridge University Press.

Segalowitz, N. (1986). Skilled reading in a second language. In J. Vaid (Eds.), *Language processing in bilinguals, Psycholinguistic and neuropsychological perspectives* (pp., 3-19). Hillsdale, N. J: Erlbaum.

Segalowitz, N., Poulson, C. and Komoda, M. (1991). Lower level of components reading skill in higher level bilinguals: Implications for reading instruction. *Reading in Two Languages, ALLA Review*, 8, 15-30.

Segalowitz, N.S., & Segalowitz, S.J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369-385.

Segalowitz, S.J., Segalowitz, N.S., & Wood, A.G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, 19, 53-67.

Seidenberg, M.S., Tanenhaus, M. K., Leiman, J.M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14, 489-537.

Service, E. (1992). Phonology, Working memory, and foreign language learning. *Quarterly Journal of Experimental Psychology*, 45A, 21-50.

- Service, E., & Craik, F. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, 32, 608-623.
- Service, E., & Kohonen, V. (1995). Is the relationship between phonological memory and Foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, 16, 155-172.
- Shah, P., & Miyake, A. (1999). Models of working memory. In Miyake, A., & Shah, P. (Eds.), *Models of working memory: Mechanisms of Active maintenance and Executive Control*, pp. (1-27). New York: Cambridge University Press.
- Shahnazari, M. & Rebecca, A. (forthcoming). The relationship between modified output and working memory capacity.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-406.
- Stanovich, K.E. (1988). The language code: Issues in word recognition. In S.R. Yussen & M.C. Smith (Eds.), *Reading across the life span*, New York: Springer-Verlag.
- Share, D. and Levin, I. (1999). Learning to read and write in Hebrew. In M. Harris and G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 89-111). Cambridge: Cambridge University Press.

- Shimron, J. and Sivan, T. (1994). Reading proficiency and orthography: Evidence from Hebrew and English. *Language Learning*, 44, 5-27.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 77, 473-489.
- Shook, J. D. (1994). FL/ L2 reading, grammatical information, and the input to intake phenomenon. *Applied language learning*, 5, 57-93.
- Siegel, L.S., & Ryan, E.B. (1988). Development of grammatical sensitivity, phonological, and short-term memory in normally achieving and learning disabled children. *Developmental Psychology*, 24, 28-37.
- Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.
- Skehan, P. (1991). Individual differences in second language learning. *SSLA*, 13, 275-298.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

- Skehan, P. (2002). Theorizing and Updating Aptitude. In Robinson, P., *Individual differences and instructed language learning*. John Benjamins Publishing Company: Amsterdam/ Philadelphia.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). New York: Cambridge University Press.
- Smith, F. (1971). Understanding reading. New York, MY: Holt, Rinehart and Winston.
- Stahl, S.A. (2003). Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in Language Disorders*, 23, 241-274.
- Spear-Swerling, L. (2007). The research-practice divide in beginning reading. *Theory into practice*, 46 (4), 301-308.
- Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16,1, 32-65.
- Stanovich, K.E., Nathan, R.G., & Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly*, 21, 267-283.

Swaffar, J., Arens, K. & Byner, H. (1991). *Reading for meaning*. Englewood Cliffs, NJ: Prentice Hall.

Swanson, H.L. (1993). Working memory in learning disability sub-groups. *Journal of Experimental Child Psychology*, 56, 87-114.

Swanson, H.L. (1994). Short-term memory and working memory: do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities*, 27, 34-50.

Swanson, H.L. & Berninger, V. (1995). *The role of working memory in skilled and less-skilled readers' comprehension*. *Intelligence*, 21, 83-108.

Swanson, H.L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology*, 93, 720-734.

Swanson, H.L., & Ashbaker, M.H. (2000). Working memory, short-term memory speech rate, word recognition, and reading comprehension in learning disabled readers: Does the executive system have a role? *Intelligence*, 28, 1-30.

- Swanson, H.L., Howard, C.B., & Saez, L. (2007). Reading comprehension and working memory in children with disabilities in reading. In Cain, K. & Oakhill, J. (Eds.), *Children's comprehension problems in oral and written language* (pp. 157-189). A cognitive perspective New York: Guilford Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language* (pp. 471-483). Mahwah, N.J: Erlbaum.
- Thorndike, R.L. (1917). Reading as reasoning. A study of mistakes in paragraph reading. *Journal of Educational Psychology*, 8, 323-332.
- Torgesen, J.K., & Burgess, S.R. (1998). Consistency of reading related phonological processes throughout early childhood: Evidence from longitudinal, correlational and instructional studies. In J.L, Metsala & L.C. Ehri (Eds.), *Word recognition in beginning Reading* (p. 161-188). Hillsdale, NJ: Erlbaum.
- Trofimovich, P. Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 171-195). Oxford: Oxford University Press.
- Turner, M., & Engle, R.W. (1989). Is working memory task dependent? *Journal of Memory and Language*, 28, 127-154.

- Valler, G., & Papango, C. (2002). Neuropsychological impairments of verbal short-term memory. In A. D. Baddeley, M. D. Kopelman, B. A. Wilson (Eds.), *Handbook of memory disorders*, 2nd ed., (pp. 249-270). Chichester: Wiley.
- Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academia Press.
- VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5- 31). Mahwah, NJ: Lawrence Erlbaum.
- Wagner, R.K., Torgesen, J.K. and Rashott, C. (1994). Development on reading related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, 30, 73-87.
- Walczyk, J. (1995). Testing a compensatory-encoding model. *Reading Research Quarterly*, 30, 396-408.
- Walczyk, J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly*, 35, 554-566.
- Walczyk, J., Marsiglia, C., Bryan, K., & Naquin, P. (2001). Overcoming inefficient reading skills. *Journal of Educational Psychology*, 93, 750-757.

- Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, 25, 315-339.
- Waters, G.S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology*, 49A, 51-79.
- Weir, C.J. (1990). *Communicative language testing*. London: Prentice Hall International (UK) Ltd.
- Williams, J.N. & Lovatt P. (2003). Phonological Memory and Rule Learning. *Language Learning*, 53, 1, 67-121.
- Wolf, D. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal*, 77, 473-489.
- Yuill, N.M., Oakhill, J.V. and Parkin, A.J. (1989). Working memory, comprehension skill and the resolution of text anomaly. *British Journal of Psychology*, 80, 351-361.
- Yuill, N., & Joscelyne, T. (1988). Effect of organizational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, 80, 152-158.

Yuill, N., & Oakhill, J., (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology*, 79, 173-186.

Zaki, H. and Ellis, R. (1999). Learning vocabulary through interacting with a written text. In Ellis, R. (ed.), *Learning a second language through interaction*, (153-169). Amsterdam: John Benjamins.

Zhang, S., & Perfetti, C.A. (1993). The tongue-twister effect in reading Chinese. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 1-12.