EFFECTS OF WORKING MEMORY AND LANGUAGE PROFICIENCY ON L2 PREDICTIVE INFERENCE GENERATION: AN EYE-MOVEMENT STUDY

FATİH SİVRİDAĞ

BOĞAZİÇİ UNIVERSITY

2019

EFFECTS OF WORKING MEMORY AND LANGUAGE PROFICIENCY ON L2 PREDICTIVE INFERENCE GENERATION: AN EYE-MOVEMENT STUDY

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfilment of the requirements for the degree of

Master of Arts

In

Cognitive Science

by

Fatih Sivridağ

Boğaziçi University

2019

Effects of Working Memory and Language Proficiency on Second-Language Predictive Inference Generation: An Eye-Movement Study

The thesis of Fatih Sivridağ

has been approved by

Prof. Gülcan Erçetin
(Thesis Advisor) _____

Assist. Prof. Nazik Dinçtopal Deniz
(Thesis Co-Advisor) _____

Assoc. Prof. Albert Ali Salah _____

Assist. Prof. Pavel Logačev _____

Assist. Prof. Duygu Özge
(External Member) _____

June 2019

DECLARATION OF ORIGINALITY

I, Fatih Sivridağ, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature 

Date 25.06.2019

ABSTRACT

Effects of Working Memory and Language Proficiency on L2 Predictive Inference Generation: An Eye-Movement Study

In this study, the effects of language proficiency and working memory capacity on predictive inference generation during reading in a second language (L2) were investigated by analysing L2 readers' early and late eye-movements while they were reading predictive or neutral passages. The results suggested that while L2 readers can make predictive inferences on-line during reading regardless of their proficiency level or working memory (WM) capacity, these two factors and their interaction determine the time course of inference generation through different mechanisms. While WM capacity facilitates referent-antecedent resolution and readers with higher WM capacity can benefit para-foveal processing more, proficiency increases reading speed and makes lower level processes less resource consuming. As a result, high WM readers showed facilitation effects of prediction even before they encounter the to-be-predicted word, especially when the pre-target required referent-antecedent association. On the other hand, while high language proficiency readers can show the effect of prediction during early processing of the target word, low proficiency readers can show facilitation during late processing of pre-target word. During the late-processing of pre-target word, WMC and language proficiency will have an interaction effect due to differences in mechanisms through which they contribute to predictive inference generation. Although all groups of readers showed facilitation effects relatively early, the greatest facilitation emerged during late processing of the sentence final word, where sentence wrap-up processes take place. This is in line with L1 studies and its implications for L2 reading were discussed.

ÖZET

İşleyen Bellek Kapasitesi ve Dil Becerisinin Yabancı Dilde Tahminsel Çıkarımlar Üretme Üzerindeki Etkileri: Bir Göz-İzleme Çalışması

Bu çalışmada dil seviyesinin ve işleyen belleğin yabancı dilde okuma sırasında tahminsel çıkarımlar üretme üzerindeki etkisi katılımcıların yabancı dilde tahminsel çıkarımlara imkân veren ve vermeyen metinleri okurkenki göz hareketleri analiz edilerek incelenmiştir. İşleyen bellek kapasitesi gönderge-öncül çözümlemesini kolaylaştırmakta ve yüksek işleyen bellekli okurlar foveal olmayan işlemeden daha fazla yararlanmaktadır. Diğer taraftan, dil seviyesi okuma hızını artırmakta ve alt seviye işlemelerini daha az kaynakla yürütmeyi mümkün kılmaktadır. Sonuç olarak, özellikle tahmin edilecek hedef sözcükten bir önceki sözcük gönderge-öncül çözümlemesi içeriyorsa, yüksek işleyen bellekli okurlar tahminin kolaylaştırıcı etkisini hedef sözcükle karşılaşmadan önce göstermişlerdir. Diğer bir taraftan yüksek dil seviyeli okurlar tahminin kolaylaştırıcı etkisini hedef sözcüğünün erken işlemesi sırasında gösterebilirlerken, düşük dil seviyeli okurlar kolaylaştırıcı etkiyi hedef öncesi sözcüğün geç işlemesi sırasında göstermişlerdir. Hedef öncesi kelimenin geç işlenmesi sırasında işleyen bellek kapasitesi ve dil becerisi tahminsel çıkarım üretmeye farklı mekanizmalarla katkı sağladığı için iki değişkenin etkileşim etkisi ortaya çıkmıştır. Bütün gruplardaki okuyucular kolaylaştırıcı etkiyi göreceli olarak erken gösterse de, en büyük kolaylaştırma etkisi cümle toparlama süreçlerinin gerçekleştirildiği cümle sonundaki kelimenin geç işlenmesi sırasında ortaya çıkmıştır. Bu gözlem anadil çalışmalarının sonuçlarıyla aynı yöndedir ve bunun yabancı dilde okuma açısından anlamı tartışılmıştır.

ACKNOWLEDGEMENTS

This thesis is the result of a very long process which has taken several years. With the contribution of many people, I could not only manage to produce this thesis, but also gain invaluable knowledge and skills at each step. Therefore, I would like to present my sincere gratitude to those who did not abstain from helping me under any conditions.

First of all, I would like to thank my advisors Prof. Gülcan Erçetin and Assist. Prof. Nazik Dinçtopal Deniz for their excellent guidance and support throughout the whole process. Prof. Gülcan Erçetin provided me with the initial idea for the topic of my thesis and she was always ready to help even while she was commuting home after a tiring day. Assist. Prof. Nazik Dinçtopal Deniz gave me invaluable guidance about eye-tracking and she presented me the opportunity to work with the equipment. Without their help this thesis would not have even been imagined. I would also like to thank the rest of my thesis committee: Assoc. Prof. Albert Ali Salah, Assist. Prof. Pavel Logačev and Assist. Prof. Duygu Özge for their valuable comments.

I would like to thank Fethiye Erbil, Mustafa Ertürk, and Rıza Memiş for their unconditional assistance. Whenever I needed help with finding participants, understanding statistics, or proof reading a text, I knew whom to call. Apart from helping me whenever I needed, they encouraged me to pursue my graduate study and do the best I can. I would also like thank my colleagues at Şehit Adem Yavuz Middle School and Mustafa Kemal Anadolu High School for their interest and assistance in various parts of my study.

Above all, I would like to thank my family for their understanding throughout the years. Firstly, my beloved wife, Betül Sivridağ, provided me with the best

environment to study and she happily sacrificed her comfort so that this thesis can be finalized. I also thank my daughter Feyza Sivridağ for her precious smile which cheered me up whenever I had a problem with my study. Lastly, I want to thank to all who have had an intentional or unintentional positive effect in my intellectual development.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: BACKGROUND LITERATURE 4

2.1 Reading in a second language 4

2.2 Working memory 14

2. 3 Inferences during reading 34

2. 4 Proficiency and automatization in L2 reading 54

2. 5 Eye-tracking 60

CHAPTER 3: METHODOLOGY 69

3.1 Participants 69

3.2 Data collection instruments 70

3. 3 Procedures 80

3. 4 Data analysis 82

CHAPTER 4: RESULTS 85

4.1 Preliminary analyses 85

4.2 Eye movements 96

4.3 Summary of EM analyses 118

CHAPTER 5: DISCUSSION 119

5. 1 Predictive inference generation 119

5.2 Time course of predictive inferencing 123

5.3 The role of WM capacity in predictive inference generation 124

5.4 The role of language proficiency in predictive inference generation 126

5.5 The interaction between WM and proficiency in inference generation 129

CHAPTER 6: LIMITATIONS AND FUTURE DIRECTIONS 132

APPENDIX A: VOCABULARY LEVELS TEST 134

APPENDIX B: PASSAGES USED IN EYE-TRACKING EXPERIMENT 136

APPENDIX C: COMPREHENSION QUESTIONS AND OPTIONS USED IN EYE-TRACKING TASK..... 146

APPENDIX D: ANOVA SUMMARY TABLES..... 152

REFERENCES..... 182

LIST OF TABLES

Table 1. Means and Standard Deviations for Vocabulary Scores as a Function of 2(WM Capacity Level) X 2(Proficiency Level) X 2(Frequency Level) 86

Table 2. Means and Standard Deviations for Comprehension Scores across Proficiency and WM Groups 88

Table 3. Means and Standard Deviation (in ms) Reading Duration of Passages as a Function of a 2(Proficiency) X 2(WM Level) Design 91

Table 4. Means and Standard Deviations for Question Response Time (in ms) as a Function of a 2(Proficiency) X 2(WM Level) Design 93

Table 5. Means and Standard Deviations for WM Scores across WM and Proficiency Groups 94

Table 6. Means and Standard Deviations for Fixation Probability, First Fixation Duration and First Run Duration of Pre-Target Region as a Function of a 2(Proficiency) X 2(WM Level) Design 99

Table 7. Means and Standard Deviations for Fixation Probability, First Fixation Duration, and First Run Duration of Target Region as A Function of a 2(Context) X 2(Proficiency) X 2(WM Level) Design 101

Table 8. Means and Standard Deviations of Fixation Probability, First Fixation Duration (in ms), and First Run Duration (in ms) for Post Target Region. 104

Table 9. Means and Standard Deviations of Fixation Probability, First Fixation Duration (in ms), and First Run Duration (in ms) For Final Region. 106

Table 10. Means and Standard Deviations of Late Processing Measures (in ms) for Pre-Target Region 110

Table 11. Means and Standard Deviations of the Late Measures (in ms) for Pre-Target Region 113

Table 12. Means and Standard Deviations of the Late Measures (in ms) for Post-Target Region 115

Table 13. Means and Standard Deviations of the Late Measures (in ms) for Final Region. 116

Table 14. Mean Facilitation (in ms) per Character 117

Table 15. p Values for Pair-wise Comparison of Facilitation per Character 118

LIST OF APPENDIX TABLES

Table D1. ANOVA Summary Table for the Effects of WM and Proficiency Level on Vocabulary Scores 152

Table D2. ANOVA Summary Table for Comprehension Scores 153

Table D3. ANOVA Summary Table for Reading Duration of Passages 154

Table D4. ANOVA Summary Table for Response Times (in ms) to Comprehension Questions in Eye-Tracking Task 155

Table D5. ANOVA Summary Table for Accuracy Scores of Complex Span Tasks 156

Table D6. ANOVA Summary Table for Complex Span Task Scores 157

Table D7. ANOVA Results for Fixation Probability on Pre-Target Region 158

Table D8. ANOVA Results for First Fixation Duration on Pre-Target Region. 159

Table D9. ANOVA Results for First Run Duration on Pre-Target Region 160

Table D10. ANOVA Results for Fixation Probability on Target Region 161

Table D11. ANOVA Results for First Fixation Duration on Target Region 162

Table D12. ANOVA Results for First Run Duration on Target Region 163

Table D13. ANOVA Results for Fixation Probability on Post-Target Word 164

Table D14. ANOVA Results for First Fixation Duration on Post-Target Region .. 165

Table D15. ANOVA Results for First Run Duration on Post-Target Region 166

Table D16. ANOVA Results for Fixation Probability on Final Region 167

Table D17. ANOVA Results for First Fixation Duration on Final Region 168

Table D18. ANOVA Results for First Run Duration on Final Region 169

Table D19. ANOVA Results for Regression Path Duration of Pre-Target Word .. 170

Table D20. ANOVA Results for Selective Regression Path Duration of Pre-Target Word 171

Table D21. ANOVA Results for Re-Reading Duration of Pre-Target Word 172

Table D22. ANOVA Results for Regression Path Duration of Target Region 173

Table D23. ANOVA Results for Selective Regression Path Duration of Target Word 174

Table D24. ANOVA Results for Re-Reading Time of Target Word 175

Table D25. ANOVA Results for Regression Path Duration of Post-Target Region 176

Table D26. ANOVA Results for Selective Regression Path Duration of Post Target Region 177

Table D27. ANOVA Results for Re-Reading Time of Post-Target Region..... 178

Table D28. ANOVA Results for Regression Path Duration of Final Region..... 179

Table D29. ANOVA Results for Selective Regression Path Duration of Final Region 180

Table D30. ANOVA Results for Re-Reading Time of Final Region 181

LIST OF FIGURES

Figure 1. Sample from Vocabulary Levels Test. 71

Figure 2. Flowchart of complex span tasks. 73

Figure 3. Example of a record showing eye-movement measures 83

Figure 4. Change in vocabulary score as a function of language proficiency and frequency level 87

Figure 5. Mean accuracy scores in complex span task as a function of WM and proficiency 95

Figure 6. Interaction of WM level and context on first fixation duration of pre target region. Exp and cont corresponds to predicting and control context, respectively 98

Figure 7. First fixation duration of target region as a function of language proficiency and context 102

Figure 8. First run duration of target region as a function of language proficiency and context 102

Figure 9. Interaction between WM and language proficiency on facilitation in regression path duration of pre-target region 108

Figure 10. Interaction between WM and language proficiency on facilitation in re-reading time of pre-target region 109

CHAPTER 1

INTRODUCTION

Reading comprehension is one of the most complex capabilities of human mind. It requires decoding visual stimuli, retrieval of word meanings and syntactic rules from long-term memory (LTM), encoding presented information, associating what is presented to existing world knowledge, and sending the results to LTM. Most of the time, comprehension goes beyond these processes and requires finding implicitly given or missing information to create a coherent text representation. These processes of temporary storage and manipulation of information take place in the realm of working memory (Baddeley & Hitch, 1974). Thus, individual differences in working memory (WM) capacity have been put forward as a major factor distinguishing "good readers" from "poor readers" (Just & Carpenter, 1992).

In any kind of text, the reader has to make inferences to connect different sentences or ideas. These inferences might be associating two items within a sentence as in anaphoric inferences, or they can be required to connect two consecutive sentences as in casual inferences. Sometimes, however, inferences go beyond the text and require the reader to predict what happens next (i.e. elaborative inferences). Although bridging and causal inferences are considered as necessary for text comprehension, elaborative inferences, including predictive inferences, have been argued not to be crucial for forming a coherent text representation (McKoon & Ratcliff, 1992). Therefore, whereas bridging and causal inferences have been proposed to occur automatically, elaborative inferences have been considered as strategic and effortful processes (Calvo & Castillo, 1998; Calvo, Meseguer, & Carreiras, 2001; O'Brien, Shank, Myers, & Rayner, 1988).

Since predictive inferences, which require comprehension of the previous text and elaboration on it, are strategic and effortful processes, WM capacity should have a sizable effect on their generation. Accordingly, in different studies it has been found that WM capacity affects whether and when predictive inferences are drawn (Budd, Whitney, & Turley, 1995; Linderholm, 2002; Long, Oppy, & Seely, 1994). Although there is plenty of evidence indicating that reading in a second language (L2) relies heavily on WM capacity, much of what we know about WM and making inferences during reading comes from studies conducted with native speakers (NS). As such, we do not know how inference generation changes when demand on WM is high due to processing of a non-native language.

Some of the processes involved in L2 reading comprehension become automatic with increasing proficiency (Segalowitz & Segalowitz, 1993). Simple processes like word recognition become automatic with training and thus require less WM resource. While initially the readers try to recognize words from shapes of letters or visual cues, with repeated exposure they start to process words as a whole (Ehri, 1996). Automatization of lower-level processes means that more working memory resources can be devoted to higher-level processes. Therefore, demand on WM is expected to decrease as the reader becomes more proficient in L2 and remaining resources can be allocated to higher order processes. This, in turn, will facilitate reading comprehension. However, still very little is known about the interaction between WM capacity and language proficiency on L2 reading comprehension.

Therefore, this study aims to investigate the interaction between WM capacity and L2 proficiency in relation to generation of predictive inferences during L2

reading. For this purpose an eye-tracking study, in which the course of continuous, natural reading is not spoiled, is conducted.

CHAPTER 2

BACKGROUND LITERATURE

2.1 Reading in a second language

Because reading is basically information processing, the models of reading comprehension have been affected by information processing theories, especially during the early times of reading research. Researchers applied steps of information processing to reading comprehension and tried to explain it by dividing it into separate phases in a bottom-up fashion. Later research indicated that reading comprehension does not follow these steps in most of the cases, and readers sometimes complete higher levels of processing before lower level ones, which suggested that reading is a top-down process. However, this top-down hypothesis came short in explaining several factors like speed of reading and determining the "top" where the readers start processing. As a result, a combination of these two hypotheses has been proposed. According to this theory, comprehension is achieved as a result of the interaction between the low and high level processes which are executed in a parallel fashion. Some of these later theories emphasized individual differences in reading comprehension ability and suggested possible mechanisms through which these differences may arise.

Bottom-up theories of reading suggest that reading starts with the fixation on the words. Then, the reader analyses the edges and the curves on the visual stimuli. After this analysis, the reader identifies the stimuli as a specific letter. After processing a few letters, the reader chunks them into systematic phonemes, which turn into lexemes when lexical access is accomplished. In the next stage, these lexemes become semantic representations. When several of these semantic

representations are combined inside the primary memory they are sent to the “Place Where Sentences Go When They Are Understood” (Gough, 1972). All these processes take place in about 700 milliseconds. Gough (1972) presents findings of research on reading durations as evidence to this theory. He indicates that letter identification takes 10 to 20 ms and when a reader makes 3 fixations per second he or she can attain reading 300 words per minute, which is around normal reading pace. Therefore, text is processed letter by letter. Another line of evidence for bottom-up processing comes from research on acquisition of reading. Gough (1972) shows that difficulties children experience while they learn how to read can be explained by a bottom-up approach in that children confuse visually similar letter pairs like b-p or b-d due to insufficient perceptual development. This shows that reading starts with the lowest step, which is perceptual analysis of letters. Another point the author presents as evidence for this model is the reading errors. The author discusses that for a reader to comprehend a text, reading must be fast and it should not be interfered with pauses. Therefore, when a child cannot decode a word he or she does not spend time trying to read it correctly or guess it as in top-down models, instead, he or she simply makes a reading mistake and continues reading.

LaBerge-Samuels (1974) proposed another bottom-up model that differentiates between automatic and controlled processes in reading. According to this model, reading starts with feature detection and follows a similar path as Gough’s model. However, most of the lower level processes are assumed to be automatic, and they do not require attention (Samuels, 2004). Automatization is defined as being able to do two tasks simultaneously. This model distinguishes between two phases of reading: Decoding and comprehension. Decoding is defined as turning printed words into spoken words, without the necessity of articulating

them. Comprehension, on the other hand, involves connecting and incorporating meanings of individual words. The model suggests that in the early phases of learning to read both decoding and comprehension require attention and as attention is a limited resource, the reader switches it between the two processes. As the reader becomes more experienced, decoding becomes automatic, and attention is reserved for comprehension. The author indicates that reading speed changes dramatically while the child is learning how to read. This change is not only a quantitative change; the reading process changes qualitatively as well by making some of low level processes automatic. He suggests that reading processes such as lexical access is very slow at the beginning because all the sub-processes require attention, but later as they become more automatic, reading gets faster.

Similar to L1 reading, bottom-up models of L2 reading also stress the direction of processing, which is from basic processes like letter identification to more complex ones such as semantic integration of newly read phrases into the existing information. Automatization of lower level processes is essential for fluent and effective reading so that limited attentional resources can be used for higher level processes (McLaughlin, Rossman, & McLeod, 1983).

A major criticisms of the bottom-up theories is that they cannot explain research findings that perception of a letter may depend on the context, both syntactic (Weber, 1970) and semantic (Meyer & Schvaneveldt, 1971), and that people read meaningful words faster than random strings of letters. In other words, they ignore the role of higher-level processes such as syntactic parsing and semantic integration in lower-level processes such as letter identification and lexical access.

On the other hand, top-down theories of reading assume that readers form hypotheses and test them during reading. Goodman (1967) defines reading as a "psycholinguistic guessing game". He highlights the readers' former reading experiences and general world knowledge. He states that the mistakes made by a fourth grader are not because of lack of knowledge or being careless. He emphasizes the pattern in the errors. For example, the child replaces "the" with "your" or "his" during reading. All these words are noun identifiers and they do not have any resemblance, either visually or aurally. Therefore, Goodman (1967) suggests that the child guesses that there will be a noun identifier at that location before he or she encounters "the". He claims that readers use graphic, syntactic, and semantic information simultaneously. After this discussion he proposes his reading model, which starts with a reader's scanning through the text line by line and selecting graphic cues based on his or her prior experiences. Then the reader forms a perceptual image, which is a combination of what he or she has seen on the text and his or her expectations. Then the reader makes a guess and stores it in WM and as he or she goes on reading, validity of this guess is checked. If it is not valid, more cues are searched in the text and in the prior knowledge. If the guess is verified, it is incorporated into what has been read. According to these models, the difference between good and poor readers arises because of the differences in how much attention they give to the visual stimuli. Good readers use their prior experiences and knowledge of the syntax to guess the written words better and faster without error, but poor readers rely more on physical features of the written words to identify them, which slows their reading and thus comprehension (Smith F. , 1971).

There are L2 versions of top-down models as well. These theories are based on a universal framework and thus suggest that reading competence is heavily

affected by conceptual processing and strategic manipulations like use of prior reading experiences and general world knowledge. Such factors do not change from language to language, they are universal and thus, these theories emphasize the role of transfer from L1 to L2 in developing L2 reading skills (Koda, 2004). A fluent L1 reader has mastered hypothesis forming and testing and when he or she learns a new language this skill will be transferred to L2. He or she will have to learn only the lowest level features of the new language. On the other hand, a poor L1 reader cannot predict what comes next in the text very accurately, so he or she will have the same problem in L2 as well (Cummins, 1979).

Top-down models have been criticised for being too vague (Stanovich, 1980). That is; they do not specify where the top and bottom are. Another point of criticism has been the ineffectiveness of these models in explaining the speed of reading. For top-down models, forming a hypothesis about the next word must be faster than decoding the word visually, otherwise, hypothesis forming becomes unnecessary. However, such a complex hypothesis forming and testing process cannot be completed in duration as short as hundred milliseconds (Stanovich, 1980). Top-down models have also been criticised for the source they suggest for individual differences. It has been shown that poor readers also benefit from prior experiences and syntactic knowledge during word identification. Studies which compared fluent and beginner readers showed that beginners use their knowledge of phonotactic constraints as much as fluent readers do (Juola, Schadler, Chabot, & McCaughey, 1978).

The connectionist models offer another approach to reading - and language in general. They explain processing as the activation and/or inhibition of certain units and the connections among them based on the stimulus and knowledge of the

processor; in this case the reader. In connectionist models there are very few or no prescriptive rules, instead, the system is allowed to generate its own rule-like behaviour. These models allow parallel processing; several steps can be carried out by different units simultaneously. Moreover, in these models, knowledge is not stored at a certain unit, it is scattered inside the system (Koda, 2004). Connections are formed and gain their value or weight with experience, and this value or weight can always change as new experiences are gained.

Rumelhart's (1977) interactive model can be an example for connectionist models because it suggested that some underlying reading processes are executed in a parallel fashion. This model was formed similar to HEARSAY II, which is a connectionist speech understanding program (Erman, Hayes-Roth, Lesser, & Reddy, 1980). In Rumelhart's model there are different modules and they operate simultaneously. Features of the visual stimuli are extracted by a module and they are fed to a "pattern synthesizer". Orthographic, lexical, semantic, and syntactic knowledge is fed to the same place by different modules, but simultaneously. Each module has different levels inside them and hypotheses about what is being read are formed and tested at each module at different levels. Stanovich (1980) criticised Rumelhart's (1977) model by indicating the hierarchical processing inherent in the model. According to Stanovich (1980), in this interactive model, higher level processes wait lower level processes for information feed. Therefore, he added a compensation mechanism to the model and named it interactive-compensatory model. In this version of the model a problem in one of the modules or knowledge sources can be compensated by other modules or knowledge sources irrespective of their place in the processing hierarchy.

Coady (1979) was one of the first researchers to apply interactive-compensatory theory to L2 reading. In Coady's (1979) model, based on Goodman's views, there were three interacting modules, and comprehension was a product of this interaction. Conceptual abilities and background knowledge were two of the modules in Coady's (1979) model. The third module, process strategies, consisted of any kind of strategy a reader benefits including syllable identification and getting contextual meaning. The efficiency of process strategies increased with proficiency. Coady (1979) also suggested that when there is a deficiency in one of the modules, it can be compensated by other modules.

Bernhardt (2011) criticized Coady's (1979) model by suggesting it had not mentioned language in any of the modules and it was not tested or improved, which made it a less scientific model. Then, she applied this interactive-compensatory model to L2 reading in a different manner. Similar to Coady's (1979) model, she suggested that reading in a non-native language requires interaction of different modules, and when one these modules fail to provide sufficient information toward comprehension, this insufficiency is compensated by other available modules. However, in L2 reading, some of the modules or knowledge sources are different from L1 reading. After examining error rates and error types made by L2 readers, she showed that there are three factors affecting L2 reading performance (Bernhardt, 1991). According to Bernhardt's model, L1 literacy is one of the knowledge sources in L2 reading. She claimed that readers transfer their L1 reading skills to L2, and readers with poor L1 reading skills will also have difficulties in L2 reading. Another source of knowledge is L2 language knowledge, which includes L2 grammar and vocabulary knowledge a reader has. She called the third knowledge source as "others", and it included other factors like strategies used by the reader, general

world knowledge and domain knowledge, reading purpose, motivation, which affect L2 reading comprehension. The author suggested that L1 skills predict 20% of the performance in L2 reading comprehension, L2 knowledge predicts 30% and other factors predict 50% of performance.

Another model which allows parallel processing has been proposed by Just and Carpenter (1980). Their model -similar to interactive-compensatory models- have different components and they interact with each other. In their model, working memory is the mediating component between reading stages and long term memory. It is also the source of individual differences in comprehension ability. In early stages of reading, the visual stimuli are processed and the output is sent to WM. In the later stages, WM content is analysed and syntactic and semantic processing are executed. The outputs of this stage are fed to WM in every step. Long term memory provides episodic domain knowledge and procedural knowledge about how to execute reading stages. The authors mention three important features of this model. First they emphasize that reading involves both serial and parallel processing. Serial processes consume limited capacity and they take some time to be executed. Parallel processes, on the other hand, are automatic and they do not consume resources. Binding of words to their meanings is given as an example of serial processes, and activation of semantic and episodic knowledge is suggested to be a parallel process. Another stressed feature of this model is the WM's having limited capacity. This has many implications including removal of some information when the processing load is heavy, therefore slowing down comprehension. The last important feature of the model is that it does not require higher order control mechanisms; the sequence of processing is adjusted by the inner dynamics of the system. In a later version of this

model the authors emphasized the role of WM. This model will be discussed in detail in the next section.

Just and Carpenter (1992) defines WM as a set of processing resources which execute symbolic computations and produce intermediate and final outputs. They suggest that comprehension, which requires processing and storage of information, is carried out by WM, and differences among individuals in terms of comprehension skills are results of differences in their WM capacity. They emphasize that WM has a storage unit which keeps interim and end products and a processing unit which carries the lexical, syntactic, and inferential processes. However, unlike early models of WM like (Baddeley & Hitch, 1974), which will be introduced in detail later, they do not include modality-specific buffers in the definition; their storage part has a broader definition. It can store lexical items, theme of the text, propositions from several sentences, and antecedents of pronouns from previous sentences. They claim that linguistic processing does not happen in distinct modules such as lexical module or syntactic module. All the processes of comprehension take place in the realm of WM and they interact with each other as much as WM capacity allows.

The theory predicts that as the burden on the WM increases, some deterioration in the comprehension will be observed as a function of capacity. Also, as the capacity diminishes, comprehension failures will be observed as a function of cognitive demand. To test these predictions they presented supporting evidence from several lines of research in which either processing burden was manipulated or capacity deficits due to aging or language impairments were taken into consideration. In most of these studies WM capacity was measured by a reading span task in which the subjects processed sentences while they were storing words in their memory.

The authors cite evidence from sentence processing studies for the relationship between WM capacity and comprehension. For instance, King and Just (1991) showed that reading time difference between high-capacity readers and low-capacity readers emerged only in the object relative clause sentences which are more difficult to process compared to subject relative sentences, suggesting that subject relative clause sentences make little demand on WM and thus do not exhaust available resources. On the other hand, object relative clause sentences make higher demands, and while high capacity readers can provide sufficient resources, low capacity readers cannot. Studies that are focused on syntactic ambiguity also show that while high-WM readers can preserve several interpretations of a syntactically ambiguous structure, low-WM readers can preserve only one interpretation until the disambiguating point of the sentence.

Another type of evidence comes from studies investigating comprehension under an extrinsic memory load. In these studies, preserving extrinsic memory load which is another task unrelated to comprehension for a certain duration consumes WM resources which otherwise would be used by comprehension processes. Extrinsic load can be doing simple mathematical operations, following moving objects, or naming words presented on the screen. This load leads to longer reading time or deteriorated comprehension. The next line of research supporting capacity theory is manipulation of distance between two related constituents like a pronoun and its antecedent. The authors cite Daneman and Carpenter (1980) which found that high-capacity readers can relate a pronoun to its antecedent over a longer distance than low- capacity readers can. This suggests that the former has sufficient resources to maintain more elements in working memory.

The authors made a computer simulation model of capacity theory. Their model, CC READER, takes items cycle by cycle, processes them, and stores interim products. WM capacity is represented by activation level which determines the number of items that can be processed in each cycle, number of items in the storage, and their activation strength, and it can be manipulated. The model shows that when syntactic parsing requires large amounts of processing, non-syntactic information is processed only if the WM capacity is set to high. By showing the similarities between their model and human data, they emphasize that as low WM readers consume their resources with syntactic parsing, they cannot benefit from semantic or pragmatic information sufficiently.

As WM is thought to have an important role in reading comprehension, it is expected that it will have an effect on many aspects of reading, ranging from lexical access to inference generation. The next section will discuss working memory in depth.

2.2 Working memory

There are several widely accepted models which explain how information is processed in WM. Despite the differences among these models in terms of the nature of WM, there is consensus on WM's simultaneous storage and processing functions as well as its close connection to LTM (Miyake & Shah, 1999). On the other hand, the models differ mainly in terms of WM's components, its capacity, the types of information it processes, and the nature of its relationship with other cognitive components such as intelligence. Despite these differences among the models, Miyake and Shah (1999) propose that they are not completely incompatible.

Baddeley's multi-component model (Baddeley & Hitch, 1974) is one of the first models of WM and it is widely accepted. In this model, WM consists of three storage modules and a central executive which manipulates attention and other storage systems. Phonological buffer and visuo-spatial sketchpad are two sub-systems which store verbal and visual input for a brief time, respectively. The episodic buffer, which was added later to the model, enables interaction of WM with LTM and integration of information from different sources (Baddeley A. , 2000). These four modules and interaction among them make complex processes, which require temporary input storage and manipulation, possible (Baddeley A. , 2017). As language processing heavily depends on brief storage, manipulation, and integration of continuous input, WM's role in language acquisition and processing has been studied widely. Especially the phonological loop and the central executive have been linked to many aspects of language processing from vocabulary acquisition to inference generation (Baddeley A. , 2003).

Another WM model proposed by Cowan (1988) emphasizes the connection between memory and attention. This model conceptualizes WM as an information processing system which consists of a sensory store, a long-term store, and a central executive. The sensory store takes input from the outside world and sends it to the long-term store. It can store information for a very brief duration and it filters the unimportant information through habituation mechanism. In the long-term store, information can exist in different levels of activation. These activation levels are controlled by attention. An activated part of LTM is considered as STM, and focus of attention can be on only a part of STM. The central executive controls what information enters the focus of attention voluntarily and attention-orienting systems control involuntary shifts in focus of attention. Apart from information which is

activated by the central executive, novel stimuli which is not habituated, and LTM content which is activated through associations can also enter the focus of attention.

According to Cowan's (1988) model, the content of focus of attention is highly activated, but this activation is capacity limited. A piece of information can stay in the focus of the attention for very long duration, but the amount of information one can attend at a particular time is limited. On the other hand, there is not a capacity limit for activated LTM content, but its activation is time limited; once activated its activation gradually fades away. In this model WM can be defined as a cognitive process which keeps information highly active and available for processing. A piece of information is maintained in WM as long as it is the focus of attention. In this model modality of information is not emphasized. There are not different modules for auditory, visual, or spatial information; they are all activated and maintained in the same manner. However, their activation is achieved through the brain regions which are responsible for processing the auditory, visual, and spatial stimuli, respectively (Cowan, 1999).

Another model which emphasizes the role of controlled attention has been devised by (Engle, Tuholski, Laughlin, & Conway, 1999). In their model, WM consists of activated LTM traces, processes to conduct these activations, and controlled attention. Similar to Cowan's model, in Engle et al.'s (1999) model, controlled attention is capacity limited and domain free. The model suggests that STM is required for maintenance of activated information and it is domain specific. For information to be processed according to current goals without interruption from distracters, controlled processing is necessary. It is capacity is limited and it carries out processing of relevant information while inhibiting or blocking interference from other sources. The model posits that individual differences in WM capacity emerge

only when the task requires use of controlled attention. Therefore, the researchers propose that it is the capacity of controlled attention, which is the analogue of central executive in other models that yields differences in WM capacity. As such, WM capacity refers to the capability of activating memory content by transferring them into the focus of attention while there is distracting stimuli.

Kintsch et al. (1999) suggest that despite the differences among models of working memory, the models are not incompatible with each other. For example, each model assigns a different role and degree of importance to the LTM and its relation to WM. However, none of them claims that LTM and WM are separate entities. Although all models approach the relationship between WM and attention from a different angle, they all emphasize the closeness between the two constructs. WM capacity is another point where despite different points of view among models, the underlying idea is compatible across all. All the models claim that WM has limited capacity. They differ in explaining the factors and mechanisms that are responsible for this limitation. None of the models mentioned above are entirely domain specific or entirely domain general. Although they propose different components of WM to be domain general or domain specific, their claims are not fundamentally incompatible. Miyake and Shah (1999) also support the view that models of WM have general agreement on the underlying factors and mechanisms governing working memory; and the differences do not make these models incompatible.

2. 2. 1 Measuring WM capacity

Many researchers agree that complex span tasks are better indicators of WM capacity than simple span tasks. In simple span tasks subjects are required to retrieve a list of

items without any interference from processing tasks whereas a complex span task taps both storage and processing parts of working memory. The main difference between simple span tasks and complex span tasks comes from the distinction between WM and STM (Baddeley & Hitch, 1974). Simple span tasks tap STM, which is a single module passively storing a limited amount of information for a brief time. The only function of STM is information storage. WM, on the other hand, has been suggested to involve both storage and processing of information (Baddeley A. , 1996). One of the earliest papers investigating WM span and STM span distinction was written by Klapp, Marshburn, and Lester (1983). They hypothesized that STM and WM do not share a common resource. They compared subjects' performance on simple storage tasks and storage tasks with processing components. They concluded that tasks without a processing component do not measure WM capacity. Many studies have replicated these findings. Daneman and Merikle (1996) combined 77 studies with a total of 6179 participants and examined the difference between simple storage tasks and complex storage tasks in their meta-analysis. They compared predictive powers of simple span tasks and complex span tasks on reading comprehension in L1. Their results indicated that complex span scores show greater correlation with reading comprehension than simple span tasks do. They concluded that span tasks with a processing component reflect WM capacity much better.

2. 2. 1. 1 Complex span tasks

Although various complex span tasks have been developed so far, they have several features in common. First, these tasks require simultaneous processing and storage of information. Although, they may differ in types of information used in processing and storage component, each task have both components. Another shared feature of

these tasks is that the processing part of the tasks involves processing of a single type of information. While some tasks use verbal information processing, some others use visuo-spatial information, there is not any tasks that use both types of information. Therefore, these tasks have been generally categorized as verbal complex span tasks or visuo-spatial complex tasks depending on the information they use in processing part.

One of the earliest and widely used tasks for measuring WM capacity is the reading span task (Daneman & Carpenter, 1980). In this task the subjects were asked to read two to six sentences and recall their last words. Number of sentences and words to be recalled in each set increased by one in every three sets. WM capacity of a subject was the highest number of items set where he or she was successful in at least 2 of the 3 sets. In their experiment 2, Daneman and Carpenter (1980) added verification questions after every stimulus sentence to make sure that subjects processed them. The authors showed that reading span scores correlated significantly with pronoun resolution and verbal SAT scores.

After reading span, other types of verbal and visuo-spatial complex span tasks were developed. Turner and Engle (1989) used simple mathematical operations instead of sentences in the processing component. The subjects were required to decide whether given mathematical equations were correct or false. The authors also tried words and digits in the storage component, but operation span test (OST) with digits as storage items did not correlate with verbal SAT scores, so they concluded that words should be used as storage items. Their results suggested that operation span scores correlated significantly with reading comprehension, even after individual differences in quantitative skills were removed (Turner & Engle, 1989).

In another complex span task, rotation span, developed by Shah and Miyake (1996), the processing component involved deciding whether rotated letters were normal or mirror images. After responding to processing item, the subjects were presented short or long arrows emanating from the centre of the screen to different directions. The subjects were required to retrieve both the length and the direction of presented arrows. Based on the results, the authors proposed that spatial complex span tasks predict only spatial ability whereas verbal complex span tasks measure only verbal WM capacity. This dissociation meant that there are different pools of WM resources and verbal ability cannot be measured by spatial complex span tasks and vice versa (Shah & Miyake, 1996).

Engle, Tuholski, Laughlin, & Conway (1999) developed another complex span task to investigate the relationship between WM and general fluid intelligence. In their counting span task, the participants counted the number of dark blue circles which were presented together with circles of different colours and squares with the same colour. This was the processing component of their task. The storage component required retrieval of the number of dark blue circles in serial order after two to eight processing tasks. Complex span of each participant was the sum of the numbers of dark blue circles from the trials in which the participant retrieved all the items correctly (Engle, Tuholski, Laughlin, & Conway, 1999).

Symmetry span, which was developed by Kane et al. (2004) is a visuo-spatial complex span task. In symmetry span task the participants saw 8x8 matrices on which some of the cells were painted in black. The participants had to decide whether the black painted cells were symmetric along the vertical axis of the matrices. This constituted the processing part of the task. After each matrix, a 4x4 grid with one of the cells painted in red was presented. In each trial, the participants

had to retrieve the location of red cells in correct serial order. The number of processing-storage sets ranged between two to five (Kane, et al., 2004).

In Kane et al.'s navigation span task (2004) participants mentally moved an asterisk along the edges of either the letter E or the letter H and indicated whether the asterisk is on one of the top or bottom edges of the letter or on the middle edges of the letter after each movement. After this processing part, participants were shown a moving ball and during the recall, they were asked to retrieve the paths the ball moved in the correct serial order (Kane, et al., 2004). In each trial there were two to five letter-ball pairs.

2. 2. 1. 2 Domain specific versus domain general working memory

Although many researchers agree that complex span tasks are better measures of WM capacity than simple storage tasks, there are still debated issues about the nature and functions of these tasks. One of the main debates has been whether they are domain specific or domain general. Some researchers suggested that complex span score reflects only one aspect of WM capacity; verbal or viso-spatial. For example, reading span scores have been attributed to verbal WM, or more specifically, reading skills. The other view is that WM capacity tasks are domain-general. This view suggests that there are not different capacities for verbal and visuo-spatial WM and any complex span task indicates this domain general WM capacity.

One of the first papers to propose that verbal and spatial information are processed differently was written by Brooks (1968). He made the participants memorise visual or verbal input and then make judgements about these inputs. The author manipulated the way the subjects showed their judgements. There were three different ways of giving the output: verbally articulating the answer by saying yes or

no, pointing it, or tapping on it. The results showed that when the input and output were in the same modality, response times were higher compared to giving output in a different modality than input. The author also manipulated difficulty of verbal and visual outputs and obtained similar results. When the subjects were expected to give more difficult verbal output (i.e. a polysyllabic word) their recall time for verbal stimulus increased significantly. However, when the recalled stimulus was visuo-spatial, difficulty of verbal output did not cause any significant difference on response time. Similarly, whereas changing difficulty of visual output increased response time for visuo-spatial stimuli significantly, it did not affect response time for verbal stimuli. The author concluded that there are distinct processing and recall mechanisms for verbal and spatial information.

After observing a positive correlation between reading span scores and reading comprehension, Daneman and Carpenter (1980) concluded that the RST measures only verbal WM capacity even more specifically, reading comprehension skills. In another paper where they used an RST to study the possible mechanisms through which individual differences in WM capacity are correlated with reading comprehension skills, the authors again concluded that reading span is related specifically to underlying reading skills like word encoding and retrieval (Daneman & Carpenter, 1983).

Another piece of evidence for the domain specificity of complex span tasks comes from Morrell and Park's (1993) study which examined whether adding explanatory illustrations to texts reduces WM load and thus age-related performance differences. The authors found that spatial complex span tasks predicted procedural assembly performance better than verbal and numerical complex span tasks. Procedural assembly performance was measured by analysing the errors made while

the subjects were building some novel Lego structures according to given written instructions. On the other hand, the authors also observed that verbal and numerical complex span tasks correlated with text comprehension more than spatial tasks did.

Shah and Miyake (1996) also investigated whether different types of complex span tasks tap different WM resources. They showed that rotation span task correlated with spatial ability whereas verbal span task correlated with verbal skills. They found that spatial span score explained 44% of variance in the spatial ability scores and the correlation between the two constructs was .66 whereas the correlation between verbal SAT (Scholastic Aptitude Test) scores and spatial span scores was only .07. Reading span and verbal SAT correlation was .45 and the difference between the two correlations was significant. Moreover, they could not find a significant correlation between spatial span scores and verbal span scores. In the light of these findings the authors suggested that there are distinct pools of WM resources at least for verbal and spatial processing.

Domain specificity of WM was investigated in neuro-imaging studies as well. Smith, Jonides, and Koeppe (1996) gave their subjects verbal and spatial memory tasks while they were in a PET scanner. In the spatial task the subjects were required to recall the position of dots or letters shown on the screen and in the verbal task, they were shown four letters and then a probe and they were expected to indicate if they had seen the probe letter on the previous screen. Neither of the tasks included an overt processing component, though. The results showed that for verbal task the activated areas were mostly on the left hemisphere whereas for spatial tasks right hemisphere regions were activated more than left hemisphere regions. Based on this distinct activation of brain regions, the authors suggested that verbal and spatial WM

functions are carried out by different neural networks, and thus they can be dissociated.

In another paper Friedman and Miyake (2000) also found evidence in favour of domain specific WM capacity. They manipulated the spatial or causal difficulty of texts and applied reading span and rotation span tasks. Their texts were stories that took place in a building like a mall or museum. While their spatially simple stories took place in only one floor, spatially complex stories started in the first floor, continued in the second floor, and ended in the first floor of the building the story was based on. They manipulated causal complexity by changing some sentences and making causal inferences implicit in difficult condition contrary to explicit causality in simple condition. For each story, the subjects were given six causal and six spatial questions while they were reading and they were expected to answer these questions immediately, without going back in the texts. The authors found that spatial span scores correlated with performance on spatial information related questions while reading span scores correlated with performance on causal information related questions. They also observed that increased spatial complexity reduces performance only on spatial questions whereas making causal inferences implicit reduces performance only on causal questions. Therefore, the authors concluded that verbal and visuo-spatial aspects of a situation model are maintained by verbal and visuo-spatial WM resources, respectively.

Another study reached the same conclusion by using verbal and spatial span tasks as independent measures and Tower of Hanoi task and a conditional reasoning task as dependent measures (Handley, Capon, Copp, & Harper, 2002). As independent measures the authors used two simple storage tasks and two complex span tasks. Simple storage tasks were word span and arrow span which tapped on

verbal WM and spatial WM, respectively. As complex span tasks they employed rotation span for spatial WM and reading span for verbal WM. The authors presented evidence that Tower of Hanoi task requires spatial WM resources. In their conditional reasoning task the participants were given logical problems and asked to indicate whether the conclusion is the necessary output of the premises. This task was considered as a verbal task, and thus tapping on verbal WM. The results showed that performance on Tower of Hanoi task correlated with both simple and complex spatial span tasks whereas conditional reasoning task performance correlated with only complex verbal span task. The authors also conducted a confirmatory factor analysis. The best fitting model emerged when simple verbal span, reading span, and conditional reasoning performance were loaded into one factor and simple spatial span, rotation span, and Tower of Hanoi performance were loaded into another factor. There was a moderate and significant correlation between these two factors. The authors concluded that there are distinct WM resources for verbal and spatial processing. They also proposed that this distinction is not only at the level of passive storage, but there are different executive processing resources for verbal and spatial information. They explained the correlation between the two factors in their confirmatory factor analysis by proposing a domain general attention allocation and inhibition system.

Although there are a number of papers presenting evidence in favour of distinct WM resources for verbal and visuo-spatial domains, the evidence in favour of domain generality of WM is considerably broad. In most of the papers which support a domain general WM, different WM tasks show moderate to high correlations. One of the first papers showing such a correlation was published by Turner & Engle (1986). They designed this study similar to the one conducted by

Daneman & Carpenter (1980). However, they used three different complex span tasks. Two of the tasks had reading sentences, and the other one, operation span task, had simple mathematical problems as processing component. One of the RSTs and the OST had unrelated words as storage component while the other RST task used digits as storage units. To measure reading comprehension the subjects were given several paragraphs with multiple choice questions and their verbal SAT scores were also collected. They found that the correlation between reading span and reading comprehension was very close to the correlation between operation span and reading comprehension. Therefore, they concluded that the type of processing component in a complex span task does not change the construct the task is measuring and thus it is not necessary to include a processing component similar or related to the criterion measure.

In an extended version of the first study, Turner & Engle (1989) used two different OSTs and a digit span task, which is a simple span task requiring storage of digits. In the first experiment they gave their subjects four different complex span tasks, two simple span tasks, and a reading comprehension test. They also obtained verbal and quantitative SAT scores of their participants. In two of the complex span tasks the processing part was deciding whether a given sentence made sense, and in the other two, the subjects decided whether simple mathematical equations were correct. One of the sentence spans and one of the operation span tasks had words as the retrieval items, and the other two had digits as retrieval items. The results showed that all of the four complex span tasks correlated with reading comprehension, regardless of the type of the processing task. However, for complex span tasks in which the retrieval items were words the correlation was much higher compared to tasks with digit retrieval. Moreover, only complex span tasks with a digit retrieval

part significantly correlated with verbal SAT scores. They also computed the correlation between sentence spans and operation spans and observed a similar trend. That is, tasks with the same type of storage component showed greater correlation.

Supporting evidence for this general capacity hypothesis was also found in Conway & Engle's (1996) paper. They gave their subjects a standard OST first and then the subjects were given mathematical operations with different difficulty levels. In the third part of the experiment, the subjects received three different operation span tasks with easy, moderate, and difficult levels of processing parts. The operations in different difficulty levels were determined according to each subject's performance on mathematical operations which were given in the second part of the experiment. This manipulation was made to see if the correlation between complex span scores and reading comprehension was because the subjects had different levels of ability on the processing part of the operation span. If that was the case then when all the participants were forced to use equal level of cognitive resources for processing component, the above-mentioned correlation should disappear. The authors also recorded the viewing time of the operations and retrieval time of words in operation span task. The results indicated that reading comprehension, as measured by verbal SAT, correlated with operation span scores in all three difficulty levels. Even when effects of the viewing time were removed, the correlations remained significant. The authors concluded that, the correlation between reading comprehension and complex span score is not a result of underlying processing skills or different amount of operation resources subjects have. Rather, it is because of variation in WM capacity, which is, in fact, attentional resources.

In another study (Engle, Tuholski, Laughlin, & Conway, 1999) the researchers used a latent-variable approach to evaluate the relationship between

different complex span tasks, WM capacity, STM, and general fluid intelligence. They gave their subjects a large battery of tasks including operation span, reading span, and counting span as WM tasks; forward span tasks and backward span task as STM tasks; Cattell's Culture Fair Test and Raven's Progressive Matrices as fluid intelligence tasks, and other different tasks like keeping track task, immediate free recall task, ABCD task, continuous opposites task, and random generation task. They conducted explanatory and confirmatory factor analyses on their data. Two factors emerged from this analysis. Operation span, counting span, and reading span formed one of the factors and forward span and backward span constituted the other factor. The authors suggested that the first factor represents working memory while the second one is for short term memory. The correlation between WM and STM was .68 and significant. The authors also investigated the relationship between general fluid intelligence and WM and STM. They found that while WM was significantly related to general fluid intelligence, STM was not. Even when the STM part in WM was statistically removed, WM had a significant correlation with general fluid intelligence. Therefore, the authors concluded that STM and WM are distinct but related constructs. Moreover, functions of WM go beyond storage to attention control, which does not change from domain to domain.

Another study from the same line of research used a wider variety of complex span tasks to investigate if verbal and visuo-spatial WM can be dissociated (Kane, et al., 2004). They had 260 subjects with diverse backgrounds to increase the variation in WM in the sample. Besides operation span, reading span, and counting span, the authors included rotation span, symmetry span, and navigation span tasks as measures of WM capacity. They also gave their subjects 13 different reasoning and general fluid intelligence tasks. Five of these tasks were verbal, other five were

visuo-spatial, and three tasks were for measuring general fluid intelligence. The three of the six STM tasks were verbal and the others were spatial tasks. The authors argued that the evidence in favour of the domain-specific WM reflects the domain-specificity of STM capacity through the storage component of the complex span tasks used in those studies. Therefore, the rationale in this study is that if WM is domain general, verbal and visuo-spatial complex span tasks should show greater correlation than verbal and spatial simple span tasks do. The authors made two models based on the results and compared them. One of the models had a single WM latent-variable while the other included two different WM latent variables for verbal and visuo-spatial WM capacity. In the model with two WM latent variables, correlation between them was very high. Moreover, the difference in chi-squares between two models was not significant. Therefore, the authors concluded that one factor model with a single WM construct explains the data better. As predicted, STM tasks showed a greater domain-specificity than WM tasks. Whereas the correlation between verbal and visuo-spatial WM tasks had 70% shared variance, verbal and spatial STM tasks had only 40% shared variance. Based on this evidence, the authors suggested that the domain-specific accounts of WM reflect the domain-specificity of storage, which is a part of complex span tasks. Moreover, the authors also observed that highly skilled participants showed more domain-specificity compared to less skilled participants. As most of the previous research employed highly skilled participants from top universities, prior research suggesting domain-specific WM, in fact, underestimated domain-generality of WM. In the light of these findings, the authors concluded that WM, especially the executive and attention functions of it, keeps information activated and easily accessible, regardless of the type of stimuli.

2. 2. 1. 3 Scoring complex span tasks

There are different scoring methods used in complex span tasks. In partial scoring, each correctly recalled item is given credit independent of the status of other items in the same set. For example, if a subject correctly recalls three out of seven items in a set he or she receives 3 points for the set. However, in all-or-nothing scoring, the subjects receive credit only for sets in which all the items were correctly recalled. If the task is scored such, the subject in the example above does not receive any points for three correct recalls because he or she could not retrieve all 7 items. This generates their absolute score. Partial credit scoring produces more internally consistent scores in complex span tasks compared to all-or-nothing scoring method, and it is more suitable for individual differences studies (Conway, et al., 2005).

In terms of the units there are two ways of scoring. In weighted scoring, a single unit's value depends on the set size it is in. For example, a unit in a set with 5 items contribute to the WM score more than a unit in a set with 3 items. In unit scoring, on the other hand, each item has the same value regardless of the size of the set it is presented. That is; unit in a set of 5 items has the same value as a unit in a set of 3 items. Conway et al. (2005) suggested that unit scoring is more suitable because each item in each set measures the same underlying mechanism.

2. 2. 2 Working memory in L2 reading

Considering the effortful and strategic nature of L2 processing and attention control function of WM it is not a far-fetched idea that WM plays a central role in L2 learning. Just like in L1 research, WM's role in L2 has been investigated in acquisition and processing. Although numerous studies have found a relationship between WM capacity and L2 acquisition and processing, there is still a lot of

controversy in this domain. There is a great amount of research showing WM's relationship to vocabulary acquisition (Cheung, 1996; Gathercole, Service, Hitch, Adams, & Martin, 1999; Service, 1992; Service & Kohonen, 1995), syntactic processing (Daneman & Case, 1981; Ellis, 1996; French & O'Brien, 2008) as well as receptive (Carpenter & Just, 1989; Harrington & Sawyer, 1992; Indararathne & Kormos, 2017; Tyler, 2001) and productive skills (Abu-Rabia, 2003; Kormos & Safar, 2008; Mackey, Adams, Stafford, & Winke, 2010; O'Brien, Segalowitz, Collentine, & Freed, 2006). Factors such as topic familiarity (Alptekin & Erçetin, 2011; Leeser, 2007), L1-L2 similarities (Osaka, Osaka, & Groner, 1993; Roberts, Gullberg, & Indefrey, 2008), affective factors (Rai, Loschky, Harris, Peck, & Cook, 2011), implicit/explicit processing (Erçetin & Alptekin, 2013), and proficiency (van den Noort, Bosch, & Hugdahl, 2006) have been suggested to interact with WM in L2 processing. Operationalization of working memory has also been suggested to be an important point in WM-L2 research (Erçetin, 2015; Juffs & Harrington, 2011). The discussion below presents several studies which investigated WM-L2 reading interaction from different perspectives.

One of the first studies to investigate the relationship between WM and L2 reading was conducted by Harrington & Sawyer (1992). Their subjects were English learners with L1 Japanese. The researchers used word span and digit span as simple span tasks as well as L1 and L2 reading span tasks as complex span measures. L2 reading comprehension was assessed through Grammar and Reading sections of the TOEFL. They found that L2 reading comprehension had a high correlation with L2 reading span but weak correlations with simple span measures and L1 reading span. They also found a significant correlation between L1 and L2 reading span.

Geva & Ryan (1993) investigated the role of WM in L2 reading comprehension by considering intelligence and linguistic knowledge in L1 and L2 as well. They gave their subjects two WMs task one in L1 and the other in L2. The researchers argue that WM plays a greater role in L2 comprehension than it does in L1 because when intelligence was statistically removed, WM measures showed a greater correlation to L2 comprehension than L1 comprehension (Geva & Ryan, 1993). The authors concluded that relatively simple processes such as word recognition or retrieval of meaning, which are processed automatically in L1 require strategic processing in L2 and this makes L2 comprehension more dependent on WM resources.

After Harrington & Sawyer (1992), L2 comprehension and working memory have been studied in several studies with the focus on different aspects such as proficiency or the operationalization of WM capacity. Several studies investigated the interaction of WM capacity with other factors such as topic familiarity and existing world knowledge. For instance, high-WM subjects were at a greater advantage when they were familiar with the subject of the reading passage compared to low-WM subjects (Leeser, 2007). On the other hand, it has also been shown that WM and content familiarity had independent and additive effects on inferential comprehension (Alptekin & Erçetin, 2011). Additionally, affective factors such as stress have been shown to make low-span readers focus on accuracy rather than processing time whereas high-span readers behaved this way independent of the presence of stress (Rai, Loschky, Harris, Peck, & Cook, 2011).

The relationship between WM and reading comprehension has also been investigated in neuroimaging research. These studies suggest that language processing brain structures for both L1 and L2 mostly overlap. Similar results were

found by Buchweitz, Mason, Hasegawa, & Just (2009) for reading. They compared reading comprehension of simple stories in Japanese (both in hiragana and kanji) and English. Their WM task was in subjects' L1. They also observed that reading in non-native language (English) led to elevated brain activation. They suggest that reading in a non-native language increases the need for cognitive resources and verbal working memory (Buchweitz, Mason, Hasegawa, & Just, 2009).

Linck, Osthus, Koeth, & Bunting (2014) compiled these studies in their meta-analysis. They included 79 different samples with 3707 participants in total. In their inclusion they took several criteria into consideration. For example, they distinguished WM tasks as L1 tasks or L2 tasks, simple span tasks or complex span tasks, and verbal or nonverbal tasks. They also categorized the performance measures as comprehension tasks, production tasks, or tasks that tap both. As their analysis included many different samples, it was important to determine criteria for proficiency. They divided the samples into two categories as highly-proficient learners and less-proficient learners. They included students having education in L2, masters and PhD students working on the L2, and people whose work depends entirely on L2 in their highly-proficient group. Their analysis showed that WM is significantly correlated with L2 processing (Linck, Osthus, Koeth, & Bunting, 2014). Similar to Harrington and Sawyer (1992), their results suggested that complex span tasks show higher correlation to language comprehension than simple span tasks do. They also found higher correlation between L2 comprehension and span tasks administered in L2 compared to span tasks administered in L1.However, they claim that the difference is not very powerful and L2 span tasks might bring L2 proficiency into the scene as a confounder. Therefore, it might be safer to use span tasks in L1 to evaluate WM capacity.

2. 3 Inferences during reading

There have been different models put forward to explain how, why, when, and what kind of inferences are generated while reading both in L1 and in L2. The discussion focuses on several points. One of these points is whether different types of inferences contribute to comprehension of a text differently. More specifically, are there any differences between inferences which contribute to the local comprehension of a text and those contributing to global comprehension? If so, are different types of inferences generated in different manners? This section presents three models of discourse comprehension that explain underlying mechanisms of inference generation.

One of the first models of text comprehension, Construction-Integration (CI) Model (Kintsch & van Dijk, 1978; Kintsch, 1988) puts forward that comprehension involves text-base and situation level processing. At the text base level information obtained only from the text is processed. In the situation level, this information is combined with the reader's existing world knowledge to obtain a general understanding of the text. The authors emphasize the difference between micro and macro level processes. Individual propositions and the relationship between them are processed in the micro or local level. At the macro or global level, these propositions are combined and a meaningful discourse representation is formed. Local level processing yields a text base, which is a coherent and structured compilation of propositions presented in the text. Global level processes, on the other hand, yields the situation model, which is the large, integrated structure representing the whole text.

The authors suggest that some of the processes are executed serially, while others take place in a parallel manner. For example, the first steps such as letter identification and lexical access are bottom-up. Integration of individual propositions to the situation model is suggested to take place through spread of activation, which is a parallel process. Another characteristic of the model is that processing advances in cycles, because of the limited capacity of WM. To form a meaningful text base, referential coherence should be checked. However, as the capacity is limited, this checking cannot be done on the whole text base. Therefore, the referential coherence is checked within the part of the text base which is still in the WM and an argument overlap is searched. This is an automatic process. If referential coherence is not established this way, content of the LTM is searched for possible references to the newly read information. This requires effortful processing. If it is still not established, the incoherent part of the text is re-read. As a result, a coherent text base is formed. Moreover, in each cycle concepts in LTM which are related to the ones in the text propositions are activated. If total activation of a concept reaches a certain threshold, it becomes a part of WM. During macro processing, some of the propositions in the text base are combined, generalized, or removed and the gist of the so-far-read text is formed, and a situation model is created. Different part of this situation model is connected via global inferences. Furthermore, the activated LTM concepts are used to generate knowledge-based inferences. The authors suggest that the propositions and inferences selected for situation model are partly determined by type of the text and reading purpose.

Another model of discourse comprehension is the minimalist theory proposed by McKoon & Ratcliff (1992). According to this theory, inferences are minimally encoded during reading, and they are used as cues for later generation of inferences

(McKoon & Ratcliff, 1986). These encoded cues activate related items through resonance, which is an automatic and parallel process (Ratcliff, 1978). This theory suggests that only two types of inferences can be generated automatically. First, if an inference is necessary for making a text locally coherent it can be generated online. The other types of automatically generated inferences are those which are produced from explicitly given text information or readers' easily available background knowledge. They call their model minimalist because they claim that more complex inferences can be drawn based on these basic inferences. The model makes two important distinctions. The first one is between automatic and strategic inferences. As the name suggests, automatic inferences does not require awareness whereas strategic inferences require processing similar to problem solving. The other distinction is between local coherence inferences and global coherence inferences. The authors give anaphoric resolution and casual inferences which connect adjacent clauses as examples of local coherence inferences. Global inferences, on the other hand, connect concepts and propositions scattered around the text. According to the model, local inferences are generated automatically because the necessary information is still in WM, and thus easily available, while the inference is being generated. Global inferences are also generated automatically, but only when they are necessary for the comprehension of the text. Other types of inferences, including instrumental and predictive inferences are generated neither automatically nor online.

Another constructionist model of inference generation was proposed by Graesser, Singer, & Trabasso (1994). The authors developed a model that predicts which inferences are generated on-line based on a search-after-meaning principle. According to this principle, the reader tries to form a text meaning which satisfies his or her purpose of reading, is coherent as a whole, and can explain why the text is

what it is. The authors suggest that the constructionist theory has six universal components and assumptions made by most of the theories of comprehension. These assumptions include information sources available to the reader (text, background knowledge, pragmatic context), memory storages a reader can use (STM, WM, LTM), the cognitive levels a text can be represented (surface code, textbase, situation model), increase in automaticity with repetition, attention shift among the cognitive levels of text representation. The last universal assumption is that likelihood of an inference to be encoded increases as it is activated by different information sources. In addition to the six universal assumptions, the model proposes three distinct assumptions, which are based on search-after-meaning principle. The first assumption is that readers want to satisfy their purpose of reading a certain text and they generate some inferences to this end. The second assumption is that readers need to establish both local and global coherence during reading a text, and they generate local and global inferences accordingly. The third distinct assumption is that readers try to clarify why the events and states are mentioned in the text and why the author has written a particular piece of information. Based on these assumptions, the authors suggest that inferences which are necessary to form a consistent interpretation of the text are generated on-line during reading. However, elaborative inferences can only be generated on-line in very restricted situations such as when the inference is activated by several different sources of information.

Graesser, Singer, & Trabasso (1994) categorize the possible inferences into 13 classes. Inferences which are required to establish local coherence (i.e. referential inferences, causal antecedent inferences) and global coherence (i.e. thematic inferences, characters' emotions) are generated on-line, whereas elaborative inferences (i.e. predictive inferences, instrument inferences) are not. The authors

present evidence for this proposition from studies adopting verbal protocol, question answering latency and naming latency paradigms, in which the readers made local and global coherence inferences, but elaborative inferences were not generated automatically.

2. 3. 1 Bridging inferences

Although being given different names, it is possible to distinguish between two classes of inferences. The first class of inferences, which have been called bridging, text-connecting, or inter-sentence inferences, depend solely on the information given in the text. The other group of inferences, which have been termed as extra-textual, knowledge-based, or elaborative inferences, rely on both text-based information and readers' background knowledge. It is also possible to further divide bridging inferences into two groups; one connecting adjacent clauses, and thus establishing local coherence, and the other associating items scattered along the text, and thus establishing global coherence.

Bridging inferences connect new information obtained from a text to the previously read section of the same text and they are required for comprehension (Cain & Oakhill, 1999). Readers can connect newly encountered pronoun to its referent, different words which refer to the same entity, or newly read events to their previously mentioned causes. When a noun is mentioned second time in the text, either in the same form or by using a substitute, the original noun phrase is activated (O'Brian, Duffy, & Myers, 1986). Associating a pronoun to its antecedent is required to build a coherent model of the text. Haviland & Clark (1974) suggested that when a listener gets new information he or she first searches memory for any possible antecedent which can be associated with the new information. If the listener cannot

find an antecedent for the new information, a new antecedent is formed either by elaborating on what is already known, or by creating a new item. Accordingly, they found that when the text includes overlapping referents and antecedents, it takes less time to comprehend. Bridging inferences are necessary for comprehension. Therefore, different models of comprehension have suggested them to be generated automatically, without any strategic, effortful processing.

The Construction-Integration theory also suggests a similar, but more complex, mechanism. In this approach, when new information is received, antecedent search is done in STM and when an association is found, anaphoric inference is generated automatically. If the antecedent cannot be found in STM, LTM is searched and it is an effortful process. If there is still not an antecedent to the newly received information, the reader re-reads the text. Minimalist theory also emphasizes the importance of availability of antecedent to the reader during reading for establishing antecedent-referent association. If antecedent information is immediately available when the reader reads new information anaphoric inference is automatically generated. Similar to CI theory, minimalist theory defines immediately available information as the content of STM at a certain point.

Another important factor in building a coherent text representation is to find causes of propositions. When a reader comes across an event during reading both narratives (Graesser, Singer, & Trabasso, 1994) and expository texts (Noordman, Vonk, & Kempff, 1992), he or she tries to explain why it has happened. It has been shown in several different papers that constructing causal inferences improves comprehension and recall of texts (Singer & Ferreira, 1983). There have been different models proposed to explain how the reader finds a satisfying explanation. Van den Broek (1990) suggests that there are several text-related and reader-related

factors determining if an explanation can be found. In this view a text should provide necessary and sufficient information for two propositions to be connected causally. Reader related factors include ability to decode written stimuli, necessary world knowledge, cognitive resources, and reading purpose.

Minimalist theory differentiates between local causal inference and global causal inferences. Local causal inferences are made when the cause is close to its result. Closeness can be defined as being in the same or adjacent clauses. Global causal inferences connect two propositions between which there is a distance. Minimalist theory suggests that local causal inferences can be made automatically while global causal inferences require effortful processing. This is because; when a clause is read, information given in the same or previous clause is easily available. On the other hand, if the cause of an event has been mentioned a few clauses ago, the reader has to do an effortful LTM search. CI theory, however, states that as local causal inferences are needed for micro-level coherence and global causal inferences are necessary for building a coherent macro-level representation, both are generated automatically (Van Dijk & Kintsch, 1983).

It is possible to classify a few other types of inferences as global inferences. Graesser, Singer, & Trabasso (1994) suggest that apart from global causal inferences, emotional reactions of the characters and theme of the text are required for establishing global coherence. Thematic inferences, which are actually the gist or moral of a text, are generated on-line because they are needed to connect smaller, local propositions coherently to form an organized chunk of information, and they have been shown to facilitate reading of a text with a similar theme (Seifert, Abelson, McKoon, & Ratcliff, 1986). Evidence for online generation of character emotions comes from the study of Gernsbacher, Goldsmith, & Robertson (1992). In their study

subjects were slower in reading texts in which characters showed an inconsistent emotion compared to reading the texts where the actions and the emotions of the characters were consistent. Therefore, they suggest that the readers encoded the emotional states of the characters while they were reading. Minimalist theory, however, suggests that these kinds of global inferences require connecting pieces of information which are placed away from each other throughout the text, and thus they are not easily available to the reader. Moreover, they are not needed for establishing local coherence. Therefore, global coherence inferences are not generated on-line (McKoon & Ratcliff, 1989).

2. 3. 2 Elaborative inferences

Unlike bridging inferences, elaborative inferences are not required for understanding a text. Elaborative inferences enrich the reader's understanding, but without them, the text is still coherent. For example, in "Jane nailed the picture onto the wall." it can be inferred that she used a hammer, but without this information the sentence is sufficiently coherent. Whether such inferences are generated online has been debated in many papers. Although different theories have different explanations, it is widely agreed that elaborative inferences are not generated online.

CI Theory suggests that elaborative inferences are not encoded at text-base level, but they become a part of situation model (Van Dijk & Kintsch, 1983). In this model, after text-base propositions are formed, the reader employs his or her knowledge and discourse strategies to elaborate on these propositions. The authors support this proposal by showing that the readers process targets which are associates of words in a text faster than targets which are probable inferences in the text. They conclude that elaborative inferences are produced later than scriptal inferences,

which are highly suggested by the text (Till, Mross, & Kintsch, 1988). Based on this and other similar findings, they claim that elaborative inferences are produced during reading, but at a later stage such as during sentence wrap-up, or during the construction of situation model. As elaborative inferences require several processes like sense selection, incorporating meaning of different words, and accessing episodic memory, this delay is expected (Calvo, 2004). Graesser, Singer, & Trabasso (1994) suggest that elaborative inferences are not generated automatically unless they are supported by several sources of information. If an elaborative inference is activated by different pieces of information like reader goals, highly associative phrases, or highly constrained context it is likely that it will be generated and become a part of the situation model.

Minimalist hypothesis does not definitely state whether elaborative inferences are made automatically or not during reading. Instead, it sets a few criteria which determine the likelihood of elaborative inference generation. One of these criteria is the coherence of text. If the explicit information in the text does not have sufficient local coherence, then the reader will make elaborative inferences to establish coherence. The other criterion is the availability of necessary information. If the information required for an inference is easily available to the reader, that inference will be generated. McKoon & Ratcliff (1981) show instrumental inferences as supporting example to their hypothesis. When the instrument is highly suggested by an action and thus more available to the reader, lexical recognition was shorter compared to when the instrument was not highly suggested by the action. They also showed that when the instrument information is made available by mentioning it a few sentence prior to the action, lexical recognition duration was also shortened. They concluded that elaborative inference generation might happen depending on

how much information the reader can access easily (McKoon & Ratcliff, 1989). Another example presented to support minimalist hypothesis is predictive inferences. McKoon & Ratcliff (1992) propose that if the context does not highly suggest the to-be-predicted-outcome or if the reader does not have easily available information about the predictable event, such kind of an inference will not be encoded. They show that when there is not enough information about the predictable event, the readers' response to predictive and control contexts are the same. However, when the predictable outcome is supported and made available to the readers, they have more difficulty in deciding whether the outcome has been explicitly stated in the text (McKoon & Ratcliff, 1986).

Predictive inferences have attracted more attention in the literature compared to other types of elaborative inferences and differing findings have been reported regarding their generation and role in comprehension. Considering these facts together with the role of predictive inferences in models of comprehension (i.e. top-down processing models), the next section provides further information about the findings on predictive inferences.

2. 3. 2. 1 Predictive inferences

Predictive inferences, which are also called causal consequence inferences or forward inferences, are debated in terms of when and how they are generated during reading. There are conflicting reports in the literature on the nature of predictive inference generation (Klin, Murray, Levine, & Guzman, 1999). Some researchers have found that readers can predict the upcoming words during reading, while others have argued that readers cannot infer what will happen next online. There have been debates on the methodology to detect the occurrence of predictive inferences as well.

Many different dependent measures including naming latencies, word recognition, lexical decision, ERP components, and eye movements have been used. Some researchers have discussed the factors which affect the likelihood of predictive inference generation. For example, the context has been suggested to be an important factor. For a reader to generate predictive inference, the context should be highly suggestive of what will happen next. Another studied factor was the time required for predictive inferences. WM capacity has also been indicated to determine whether readers generate predictive inferences or not (Calvo, 2004).

One of the first studies on predictive inferences was carried out by Singer & Ferreira (1983). They compared the time subjects needed to answer questions about either explicit or implicit consequences of events in a story. They found that questions on implicitly given consequences took longer time to answer. They also compared subjects' error rates in answering questions which required either backward or forward inferences. They found that while the error rate for backward inferences was 6.5%, which was very close to the error rate of filler questions (6.2%), it was 16.2% for forward inferences. The authors concluded that forward inferences were not generated as accurately as backward inferences. McKoon & Ratcliff (1986) used different dependent measures to assess predictive inference generation. Specifically, they collected immediate recognition, cued recall, and priming data in four experiments. Immediate recognition data showed that when the subjects read a predictive sentence they were more prone to making false positive errors. Although they did not see the word to be predicted in the sentence explicitly, they claimed they did. Cued recall test showed that the subjects could remember the predicting sentence when they were cued by to-be-predicted word, which was not explicitly stated in the text. Priming data indicated that when the subjects were

primed with a word from the predicting sentence they made more mistakes in deciding if a test word explicitly appeared in the text. From these results, the authors concluded that readers encode predictive inferences during reading only minimally. By minimally, they mean the encoded inference contains only some semantic features of the actual inference. For example, if the inference is the "death" of a person, the reader can encode "something bad happened". However, this inference is not strong enough to interact with the memory of the text by itself; it can relate to the text by the help of a prime from the text, which strengthens the minimally encoded inference. The authors suggest that their results are in accordance with Singer and Ferreira (1983) because their subject did not fully encode predictive inferences and thus they could not answer the questions, which required full encoding of the target inference.

Fincher-Kiefer (1993) attempted to explain predictive inference generation from a CI theory point of view and designed three experiments to assess at which level of comprehension the readers encode predictive inferences by comparing them with bridging inferences, which are known to be produced online at the text-base level. In the first and second experiments the subjects read 6 sentence long passages and at different points they were stopped and given a word recognition task. In the first experiment they had to decide whether the probe had appeared in the text and in the second experiment they had to decide if the probe might appear in the continuation of the text. The author suggested that while the first experiment assessed content of propositional text base by focusing on explicitly given text, the second experiment pertained to situational model level because in the recognition task of this experiment the reader was more prone to focus on the meaning and think what the text was about. She found that response latencies to predictive and neutral

probes in experiment 1 were similar when the target appeared immediately after the predicting sentence while in experiment 2; the subjects had shorter response latencies for predicting probes compared to neutral probes at the same testing point. Experiment 3 gave similar results with a lexical decision task. In the light of these findings the author suggested that predictive inferences are generated online and at the situation model level that involves a more enriched and knowledge-based representation than propositional text base level. Further evidence supporting this conclusion has been obtained in several other studies.

In another study Fincher-Kiefer (1996) used a word recognition task in which target probes were explicitly stated words. She observed that bridging inferences lead to facilitation when there was one sentence delay between inference eliciting context and testing. She also observed an inhibition effect when the test probe was given immediately after inference eliciting sentence. The author explains these observations by suggesting that bridging inferences are encoded at the text base level and they remain active in working memory for a very brief duration. Therefore, when bridging inferences are tested immediately after the presentation of context their representation in the working memory leads to source confusion and the reader takes longer time and makes more errors in the word recognition task. On the other hand, in the delayed condition bridging inference is deleted from working memory as new information is received and text base is updated, and thus the reader does not experience source confusion and gives correct answers faster. On the other hand, the facilitation effect observed for bridging inferences in the delayed condition was not the case with predictive inferences. As long as the context supported predictive inference, readers showed inhibition effect in word recognition task. The author suggested that because predictive inferences, unlike bridging inferences, are not

encoded as a part of propositional text-base, but as a part of situation model they are not discarded from the working memory with the processing of new input. They become a part of situation model and stay activated as long as they are supported. This interpretation has been supported in another study by using a computer simulation as well (Schmalhofer, McDaniel, & Keefe, 2002).

Potts, Keenan, & Golding (1988) also investigated occurrence of predictive inferences in an attempt to further examine the findings of Singer & Ferreira (1983) and McKoon & Ratcliff (1986). In this study, the authors used lexical decision and word naming tasks, which are claimed to be better at detecting predictive inference generation because they do not require comparing the target word with previously presented context. The lexical decision test showed that the participants performed better in terms of response time and accuracy in deciding whether a string of letters were a word after reading a predicting text compared to a control text when the target was the to-be-predicted-word. However, word naming task showed that predictive inferences are generated only when they are required to establish coherence in the text. The authors concluded that word naming results are better indicators of inference generation and as suggested by Singer & Ferreira (1983), predictive inferences are not generated online unless they are crucial for comprehension. Potts, Keenan, & Golding (1988) suggested that, the findings of McKoon & Ratcliff (1986)did not reflect inference generation due to the nature of recognition tasks they used, which allowed context checking at the time of test.

Other researchers have also reported similar results. Magliano, Baggett, Johnson, & Graesser (1993) used lexical decision latencies after reading a context which required a backward causal inference and a context which required a forward causal inference. Their results suggested that causal antecedent inferences are

generated online because they are necessary for comprehension. Causal consequence inferences, on the other hand, are not generated because comprehension can be achieved without them. Millis & Graesser (1994) obtained compatible results in expository text. They also used lexical decision latencies as dependent measure. Their texts were modified versions of encyclopaedia entries and each text explained a scientific or technological phenomenon like nuclear energy, photosynthesis, and earthquakes. Similar to Magliano et al. (1993) they found that the subjects generated causal antecedent inferences, but not causal consequence inferences.

In another study, McKoon & Ratcliff (1989) tested Potts et al.'s (1988) context checking explanation for the elevated error rates in lexical decision after reading a predicting context. They collected compatibility ratings from their subjects by asking to what extent the predicting context and the target words were related to each other. Then, they compared these ratings with error rates in their experiment. The results indicate that compatibility ratings cannot predict error rates in a lexical decision test. Based on this finding, they suggested that their previous findings (McKoon & Ratcliff, 1986) did not emerge because of context checking, but because of minimally encoded predictive inferences.

Whitney, Ritchie, & Crane (1992) also criticized the results of Potts et al. (1988) by suggesting that the texts they used did not have sufficient context to elicit desired inferences. Whitney et al. (1992) claim that for predictive inferences to be generated, the related information should be emphasized and brought into the focus at the point where the inference is expected to be generated. The authors reviewed Potts et al.'s (1988) materials and found that in 28 out of 40 passages the key action which should lead to the inference was not foregrounded. Therefore, they formed foregrounded and backgrounded versions of Potts et al.'s (1988) passages and used

word stem completion, lexical decision, and naming tasks as dependent measures. The authors found priming effects after reading foregrounded texts compared to backgrounded and control texts in word stem completion and lexical decision tasks. However, naming data did not reveal any priming effect for any type of passage. The authors concluded that predictive inferences are generated during reading as suggested by McKoon & Ratcliff (1989) but their occurrence depends on the preceding context. They suggested that relevant information should be foregrounded so that predictive inferences can be generated. They also argued that naming task is not suitable for testing inference generation. The authors suggested these two claims as the reason why Potts et al. (1988) could not find evidence for predictive inference generation.

Murray, Klin, & Myers (1993) also examined the effects of foregrounding and contextual constraints on predictive inference generation by using naming task. Similar to the results of Whitney et al. (1992), they found that predictive inferences are generated when the context is highly suggestive and the information related to the predictable event is still in the focus at the time of testing. Their results also indicated that the length of the text and causal coherence breaks do not affect inference generation. Contrary to Whitney et al. (1992), Murray et al. (1993) argued that naming task is appropriate for assessing occurrence of predictive inferences as long as the relevant information is foregrounded and the context is sufficiently predictive.

Keefe & McDaniel (1993) also studied predictive inference generation and tried to explain the contradicting findings of McKoon & Ratcliff (1989) and Potts et al. (1988) The authors suggest that Potts et al. (1988)'s failure to find evidence in favour of predictive inference generation might be because of the fact that their subjects had to read another sentence between predicting context sentence and the

presentation of the test probes. Therefore Keefe & McDaniel (1993) presented test probe immediately after the predicting context. They found that subjects had shorter response latencies after reading a predicting context compared to after reading a control context in naming task. Moreover, response latencies were similar in predicting and explicit context, where the probe word was overtly stated in the context sentence. When they increased the duration between context passage and naming probe, this facilitation effect disappeared and naming latencies became similar in predicting and control situations. The authors concluded that forward inferences are generated during reading, but they are deactivated and lost very quickly. This conclusion is in accordance with minimal encoding view of McKoon & Ratcliff (1989) and it also explains Potts et al. (1988)'s findings.

More recent studies focused on different aspects of predictive inferences. For example Hawelka, Schuster, Gagl, & Hutzler (2015) investigated the differences in predictive inference generation between fast and slow readers. Their purpose was to investigate whether predictive inferences are required for fast reading. They analysed the eye-movements of the readers and the results suggested that fast readers are better at generating predictive inferences compared to slow readers. The authors concluded that predictive inferences are required for fluent reading.

Another study (Eichert, Peeters, & Hagoort, 2018) investigated if readers make anticipatory eye-movements in virtual reality (VR) environment. The participants saw some scenes in VR and at the same time heard several sentences while their eye-movements were being recorded. The sentences were either restrictive or unrestrictive. In restrictive sentences the verb suggested only one possible object in the scene shown on VR and in the unrestrictive sentences the verb did not specifically suggest any of the objects. The authors observed that after

hearing a restrictive verb, the participants fixated on the target object more. This finding suggested that the participants predicted the upcoming words during language processing regardless of the medium of the stimuli.

In a study (Mastrantuono, Saldana, & Rodriguez-Ortiz, 2019) conducted with deaf adolescents whether predictive inferences are generated in sign language was investigated. The authors presented their participants two types of stimuli. In the first condition the participants watched video recorded texts in sign language. In the second condition, sign language was accompanied by sign-supported speech (SSS). Then the participants were asked literal and inferential comprehension questions. The authors observed that while signers had difficulty in generating predictive inferences in only-sign language condition SSS facilitated predictive inference generation. However, they did not observe any difficulty in associative inferences for either condition. This suggested that predictive inferences are more difficult to generate than associative inferences in that they require information from a broader part of the text.

Some of the recent studies also used brain imaging technology such as electroencephalography (EEG), magnetoencephalography (MEG), and functional Magnetic Resonance Imaging (fMRI) to investigate predictive inference generation. One of these studies (Virtue, Schutzenhofer, & Tomkins, 2017) focused on the hemispheric differences in processing of predictions and found that right and left hemispheres behave differently in processing weak (low contextual constraints) and strong (high contextual constraints) predictions. While left hemisphere showed greater activation for strong predictions right hemisphere did not show any and difference between the two conditions. In another study (Wang, Hagoort, & Jensen, 2017) with MEG the authors observed alpha power suppression in several parts of

the left hemisphere while the participants were reading sentence final words following a highly suggestive context. In a similar study (Wang, Hagoort, & Jensen, 2018) the authors examined MEG oscillations when the readers came across a word violating the prediction, a word confirming the prediction or when there was no prediction elicited. The authors suggested that predicted words generated slower gamma frequency. In an fMRI study (Willems, Frank, Nijhof, Hagoort, & van den Bosch, 2016) the authors investigated the brain areas which respond to confirmation of prediction and disconfirmation of prediction, which causes surprise for the readers. The authors observed that different brain areas were activated in confirmation and surprise conditions and they based on the activated areas they concluded that prediction is generated at different levels of processing like extracting word form, lexical access, and situation level formation.

Apart from contextual constraints, WM capacity has also been suggested to affect whether predictive inferences are generated or not. The next section will provide an overview of studies that investigated WM's relation to predictive inference generation.

2. 3. 2. 2 Working memory and predictive inferences

Linderholm (2002) found that subjects with high reading span generated predictive inferences when the text is highly constrained in terms of causality. On the other hand, subjects with low reading span did not generate predictive inferences consistently; they could differentiate only between when the text confirms the predictive inference and when the text contradicts it. Moreover, while high-span readers could generate predictive inferences in speeded reading tasks, low low-span readers could show signs of predictive inferences only when they were reading at

their own pace. The author suggests that predictive inference generation is a difficult process which occupies some WM resources.

In another study the authors showed the effects of visuo-spatial processing load on predictive and bridging inference generation (Fincher-Kiefer & D'Agostino, 2004). Their hypothesis was that predictive inferences require cognitive resources, and if these resources are exhausted, predictive inferences are not generated. The participants saw a shape with dots in different locations and read a predictive or control context. Then, they were given a lexical decision task and their response latencies were collected. Right after the lexical decision task, they were shown the shape with dots and asked to decide whether the dots were at the same location as the first shape they saw. They were expected to generate predictive inferences while they were storing visual information in their working memory. The results showed no difference on the lexical decision task after reading predictive versus control passages in the case of visuo-spatial load during reading. The authors concluded that generation of predictive inferences demand visuo-spatial resources. When the readers did not have adequate resources, predictive inference generation was disrupted.

In another study Estevez & Calvo (2000) used stimulus onset asynchrony (SOA) naming task to determine the time course of predictive inference generation as a function of WM capacity. Their subjects read predicting and control sentences and were given either confirming or neutral naming probes after 500, 1000, or 1500 ms after the last word of context sentences. The authors observed that when the interval between context and probe was 500 ms neither high- nor low-WM subjects showed facilitation effect. At 1000 ms interval high-WM readers had shorter naming time for confirming probes compared to low-WM participants. At the 1500 ms interval, both groups generated predictive inferences. The authors suggested that

low-WM readers compensate their insufficient resources by spending more time on processing.

Calvo (2004) compared eye-movements (EM) of high- and low-WM participants while they were reading predictive and neutral passages in an eye-tracking experiment. The multiple regression applied on the EM data revealed that individual differences in WM capacity can explain some of the variance in gaze duration on the final word of the passage and regression from this area. In another eye-tracking study the same author compared early and late processing measures of high- and low-WM readers (Calvo, 2001). The subjects read prediction eliciting and control passages which were followed either by a confirming or unrelated word. The results showed that predictive and control passages did not cause difference in early processing measures in either WM group. However, there was a difference between predicting context and control context in late processing measures only in high-WM readers did. They read the part of the sentence following inferential word faster than low-WM readers. These results suggested that highly predictable events facilitated reading only for high-WM readers because they generated predictive inferences on-line. However, this inference generation did not take place during the early phases of reading, but during the text integration phase which happens later.

2. 4 Proficiency and automatization in L2 reading

At the first stages of second language learning comprehension and production are quite slow. As these processes involve multiple components and all these components are processed in a strategic and effortful manner, the output emerges quite slowly. For example, during reading comprehension components such as word recognition, lexical access, syntactic parsing, extraction of thematic roles, inference

generation contribute to the overall comprehension time considerably. When all these components are executed consciously, time and resources needed for comprehension increases dramatically.

It has been shown in different studies that some of the processes involved in skill acquisition become automatic upon repetition. Ackerman (1987) showed that with practice, differences in variation of reaction time across individuals disappear. After sufficient practice, within subject variation becomes highly similar. This indicates that with practice, individual differences diminish. From a perspective of information processing he explains this situation as components of skills gradually becoming independent of attentional resources. That is, the skills, partially or as a whole, become automatic (Ackerman, 1987).

Although automaticity is often used in skill learning and second language acquisition, the exact definition may change from task to task. However, there are widely accepted characteristics of automatic behaviour. One of the differences between automatic and strategic processing is the declarative versus procedural knowledge distinction. Transition of knowledge from declarative memory to procedural memory is what brings automatization. Another widely accepted feature of automatic processing is that it does not require conscious effort or strategic processing. Therefore, it is resource independent. Segalowitz (2003) makes a detailed operational definition of automaticity. He describes the characteristics of automatic processing and emphasizes that automatization is a qualitative change. That is, automatization does not mean speeding up of underlying components of a task, but elimination of some of these components due to practice. He describes automatic processing as fast, unstoppable, load-independent, effortless, and unconscious processing. He suggests that shift from algorithmic to instance processing can be a

feature of automatization in several domains. He also indicates several brain imaging studies to show neural changes in transition from strategic to automatic processing (Segalowitz N. , 2003).

Considering second language learning as skill learning, Segalowitz & Segalowitz (1993) used a similar rationale to see whether components of reading comprehension become automatic with practice. They presented the subjects two tasks. The first task was a simple visual detection task, which did not require effortful processes and thus minimized the effects of individual differences. The other task was a lexical decision task which was more demanding and effortful than the first one. Their subjects consisted of L2 English speakers with varying reading proficiency levels. Their data showed that practice leads to qualitative change in processing of complex tasks. That is, after sufficient repetition, some components of the task become automatic and do not require conscious processing. Moreover, their data also suggested that high proficiency readers showed less variation in their reaction times indicating that for them underlying processes of lexical decision had become less effortful. All in all, their results show that proficiency might bring automatization of underlying components of complex processes (Segalowitz & Segalowitz, 1993).

There have been replications and additions to the study of Segalowitz & Segalowitz (1993). For example, another study (Segalowitz, Segalowitz, & Wood, 1998) adopted both longitudinal and cross-sectional methodology and examined the effects of increasing proficiency on automatization. The study showed that language learners with advanced proficiency have automatized some underlying components of word recognition. The study also showed that beginner learners also automatized similar underlying components after two semesters of language learning, which

increased their proficiency level considerably (Segalowitz, Segalowitz, & Wood, 1998). Another study (Akamatsu, 2008) found similar results by comparing word recognition processes for high versus low frequency words after training. The results suggest that recognition of low frequency words becomes automatic after training. However, as recognition of high frequency words has already been automatic due to repeated exposure, training does not lead to any qualitative change in the way they are processed (Akamatsu, 2008).

Apart from word recognition, automatization has also been examined in the domain of acquisition of grammar. DeKeyser (1997) formed a mini agglutinative language to study whether practice of morpho-syntactical rules lead to automatization in comprehension and production. His subjects learned the lexicon and grammar of the language through explicit instruction. Then, they practiced the language in 15 sessions and at the end they were tested. The reaction times and error rates of the subjects showed a sharp and obvious decline after the initial learning, and this decline slowed over further practice. The results indicate that second language grammar acquisition is similar to skill acquisition in general. That is, initially the subjects have declarative knowledge which becomes procedural knowledge through practice (DeKeyser, 1997). The study supports the hypothesis that practice leads to automatization in comprehension processes.

Whether automatization of lower level processes like word recognition lead to enhanced reading comprehension has not been investigated thoroughly. Fukkink, Hulstijn, & Simis (2005) addressed this issue. They gave lexical access training to their subjects and measured their reading speed and reading comprehension. They observed that speeding up of lexical access did not lead to improved reading speed or reading comprehension. However, as they state in their paper their study is limited in

a number of ways to get to a certain conclusion. First, their training included only a small part of all the words in reading passages they used for measuring reading comprehension. Second, improvement in the lexical access in their experiment was not significant, which might have led to only very small effect on comprehension. Moreover, the role of lexical access alone might be too small to produce significant effects on comprehension. It is possible that automatization of other processes like syntactic parsing which comes with increasing proficiency as well might result in significant improvement in reading comprehension (Fukkink, Hulstijn, & Simis, 2005). Using more intense and diverse training tasks which improve language proficiency significantly might bring automatization and thus enhanced reading speed and comprehension. Moreover, considering the debate on the nature of predictive inferences, it is highly possible that their generation and time course are affected by language proficiency. As automatization is a qualitative change, it is possible that high and low proficiency readers may differ in the way they show the effects of predictive inferences.

Interaction of L2 proficiency and WM capacity has been investigated in few studies. Walter (2004) compared lower-intermediate and upper-intermediate English learners with L1 French. Although the groups were different in terms of L2 proficiency, age, and educational background, the author controlled the reading texts and comprehension tasks so that lower proficiency group did not have difficulty in understanding due to lack of lexical or syntactic knowledge. She measured comprehension through a summary task in L1 and L2. She also administered a pro-form resolution task in which the subjects had to find antecedents of pronouns or proverbs. The antecedents and pro-forms were either in the same or adjacent clauses (immediate condition) or there were two clauses between them (remote condition).

Based on Waters and Caplan's (1996) modified version of RST, WM capacity was operationalized through a composite score using response time, sentence processing, and recall span. The RST consisted of sentences in both languages. The language of the sentences changed between sets. Based on this single RST in two languages each participant received two RST scores, one for each language. The results showed that while WM facilitates reading comprehension for both proficiency groups, low-proficiency subjects benefit from high WM resources more. WM scores were also correlated with performance in remote pro-form condition in L2. However, this correlation was not specific to the low-proficiency group. The author suggested that low proficiency readers did more processing in reading and thus WM capacity had a greater influence on their comprehension (Walter, 2004).

In a later study, Akamatsu (2007) used word recognition training in L1 Japanese speakers with L2 English and he could not find a significant correlation between WM capacity and automatization of word recognition. However, as he stated, this can be the case because word recognition demands very little WM resource, and thus individual differences cannot influence the results in such a task (Akamatsu, 2008). Moreover, the subjects had had nearly six years of English education, which probably made word recognition a very simple task for them. In another study, however, van den Noort, Bosch, & Hugdahl (2006) found a significant correlation between proficiency level and WM capacity. They compared WM capacity of subjects in L1, L2, and L3. The subjects were native Dutch speakers who had fluent German as L2 and beginner Norwegian as L3. They observed that complex span scores were higher when the tasks were given in L1 than L2 and lowest in L3. They concluded that performance on complex span tasks is related to language proficiency level (van den Noort, Bosch, & Hugdahl, 2006). Their findings

also suggest that when complex span tasks are given in a non-native language, low proficiency participants engage their WM resources in linguistic processing and thus they perform poorly on WM capacity task.

2. 5 Eye-tracking

Eye-tracking is a non-invasive method for studying reading processes. It has been used in on-line reading heavily because, unlike other methods such as think aloud or rapid serial visual presentation, it does not interfere with natural reading. It is also possible to obtain a wide variety of data about fixations, saccades, and pupil size via eye-tracking. For example, it is possible to get first fixation duration on a word, saccade velocity between two fixations, or number of regressions to a certain point in a sentence. Moreover, eye-tracking systems have advanced considerably since the early studies conducted with tachistoscopes. One of the advantages of these eye-tracking systems is their high temporal and spatial resolution. Today a simple eye-tracker system can easily sample the gaze location 1000 times per second with a spatial precision of half a letter.

Although advanced eye-tracking systems give abundant information about eye movements, what these movements mean have been debated. The relationship between mental processes and eye-movements has been studied and several different models have been put forward. The main question in these studies has been how we decide to move our eyes, that is, which factors affect a certain fixation's duration and the length of a saccade. Each model suggests similar factors responsible for eye movements, but they assign different roles to these factors.

It is possible to categorize models of eye-movements as oculo-motor models and cognitive processing models which explain eye-movement (EM) patterns differently.

Oculo-motor models suggest that oculo-motor factors such as the physiology of the eye and surrounding muscles and visual characteristics of the text determine the eye movement patterns. Cognitive models, on the other hand, propose that factors which affect decoding (Rayner, 1998), integration, and comprehension of text have determine saccade length and fixation duration. It should be noted that neither model denies the role of the factors offered by the other. Therefore, it is possible to evaluate these models as a continuum. On the one end of this continuum there are oculo-motor models and on the other end there are processing models, with other models of eye-movements in between (Rayner, 1998).

Evidence for the oculo-motor models comes mostly from saccade length studies. It has been shown that the length of word next to the current fixation point affects the length of the saccade (O'Regan, 1990). When the next word is long the reader makes a long saccade and vice versa. This effect emerges even when the next word is a meaningless string of letters. As the image of the next word is in fovea, the reader decides where to fixate next by relying on crude information. Therefore, most of the time, the reader fixates not on the best position on the word where he or she can extract the information most efficiently. In this case, there should be a second fixation on the same word to bring the gaze on the most effective position. This effect has been observed in several studies.

This approach does not exclude the effects of cognitive processing of linguistic information on eye-movements. Supporters of this view claim that cognitive processing is not rapid enough to influence the relatively fast processes like saccade programming. For example, O'Regan (1990) suggests that considering the short duration of fixations (around 200 ms on average), both programming the

saccade by using linguistic information and extracting the meaning from the fixated point does not seem possible to take place in a single fixation.

In terms of fixation duration, O'Regan (1990) differentiates between within-word and between-word fixations. In this view, if a word receives two consecutive fixations because the first fixation was not on the optimal position where the highest amount of information can be drawn, the duration of this fixation will depend on the oculo-motor factors. Little variability in the duration of the first fixation on words which receive two consecutive fixations is presented as support for this view. Another piece of evidence for this hypothesis is the increasing probability of making a second fixation on the same word when the initial landing position moves away from the optimal viewing position (O'Regan, 1990).

If the word receives single fixation, its duration is determined by linguistic processing. Moreover, if a word receives two consecutive fixations the duration of the second fixation reflects linguistic processing, too. Evidence for this claim of oculo-motor models have been found in multiple studies. For example, fixations on verbs are longer than fixations on subjects or objects (O'Regan, 1990). Another source of evidence is the difference between fixation durations on high versus low frequency words. Fixations on high frequency words are shorter compared to low frequency words (O'Regan, 1990).

The processing models on the other end of the continuum suggest that eye-movements reflect linguistic processing occurring simultaneously. In these models, the whole processing of a word will occur only when the gaze is on that word. When the subject is looking at a particular word, he or she will encode it, retrieve its meaning, and integrate it with previous context. There are two basic assumptions of

this model. According to the eye-mind assumption readers process a certain single word while they are fixating on it. The second assumption, immediacy assumption, suggests that readers try to get as much information from a word as possible while it is fixated. The information to be extracted ranges from lexical encoding to integrating the word into the previously read text and making inferences. The model also predicts that, immediacy of processing protects comprehension against errors. That is; when the reader detects an inconsistency or error in processing, he or she spends more time on a word and the probability of making a regressive saccade increases.

The research conducted by Carpenter & Daneman (1981) found support for the predictions of eye-mind assumption and immediacy assumption. When the word was primed in the previous context, the gaze duration on it was shorter. Moreover, word frequency and gaze duration were negatively correlated. The authors suggested that gaze duration increases with the difficulty of word processing and, therefore, gaze duration reflects linguistic processing. They also found that when the context following a word contradicts its most common interpretation, probability of regression increases. The reader interprets the word incorrectly but notices this only when he or she reads the following words. When the error of interpretation is noticed, the reader moves back to the source of error and re-reads. In the light of these findings, the authors propose that encoding, retrieval, integration, and error detection occur when the gaze is on a certain word.

According to the process monitoring hypothesis, which is a model in between the two discussed above, fixation durations reflect linguistic processing, but to a certain extent. According to this model, the location of initial fixation point has only a small influence on fixation duration. Instead, fixation durations reflect cognitive

processing of the fixated words (Rayner, 1977; Rayner, Sereno, & Raney, 1996). However, whole processing of a word is not necessarily completed when the gaze is on that word. Sometimes processing of a word is completed when the gaze has moved to following words (Ehrlich & Rayner, 1983). Moreover, readers process the words in para-fovea to some extent. Therefore, fixation duration of a word also reflects the processing of the adjacent words.

Para-fovea is defined as the area between foveal vision and peripheral vision. It extends from 2° to 5° of visual angle (Rayner, 1998). Although the visual acuity is not perfect, readers can still extract some information from the content of para-fovea. Despite there is still debate on the amount of knowledge that can be obtained through para-foveal vision, it is widely accepted that the readers start to process a word when it is in para-fovea (Schotter, Angele, & Rayner, 2012).

Immediacy assumption and eye-mind assumption do not explain word skipping during reading. If a word can only be processed when it is fixated, skipped words are not processed at all. However, it has been observed that availability of para-foveal information affects word skipping probability significantly (White, Rayner, & Liversedge, 2005). Therefore, it can be inferred that processing of a word starts before it is fixated.

Several studies have examined the factors influencing fixation durations and they have found that frequency, predictability, syntactic relationships, and semantic load of a word affect how much time is spent on a word (Rayner & Duffy, 1986). These findings suggest that fixation durations indeed reflect cognitive processing (Ehrlich & Rayner, 1983). All these factors also affect lexical access. Therefore, process monitoring hypothesis suggests that fixation durations reflect lexical access

(Reichle, Pollatsek, Fisher, & Rayner, 1998). However, unlike immediacy assumption and eye-mind assumption, higher level processing of a word does not necessarily take place in a single fixation. Integration of a word to the prior context and general world knowledge happens in later fixations or in clause or sentence final positions. To test this hypothesis, Ehrlich &Rayner (1983) manipulated the distance between a pronoun and its antecedent. They observed that the larger the distance was the greater the reading time became. However, the increase in reading time was not reflected on the fixation duration of the pronoun, but fixation durations of the following words. The authors suggest that, process of pronoun assignment started when the gaze was on the pronoun, but it is not completed in the same fixation.

The studies reviewed above suggest that certain amount of processing occurs while a word is being fixated. However, during this processing relatively low level of information is extracted. Lexical access and retrieval of meaning can be achieved when a word is fixated. Higher-level processes like syntactic and semantic integration of a word to the prior text require more time. Therefore, these processes spill over and are completed in later fixations when the gaze is on next word or words. From this discussion it is possible to infer that to understand cognitive processing of a certain unit in a text, eye movement behaviours on following words should also be taken into consideration (Clifton, Staub, & Rayner, 2007).

Based on the discussion above following research questions and hypotheses were formed to be investigated in the present study.

1. Do readers show shorter reading times while they are reading a highly predictive context and a confirming continuation compared to a neutral context and continuation sentence in their L2?

Hypothesis: I hypothesize that all readers will show the effects of predictive inference generation to some extent. Experimental sentences where the continuation sentence confirms the predicting context will have shorter early and late processing durations. As long as the context is sufficiently suggestive, readers will be able to predict what is going to happen next and thus, processing will be easier.

In line with the argument that predictive inferences are minimally encoded on-line during reading, predictive inference generation will show its effect immediately when the target is read (McCoon and Ratcliff, 1986). Readers were expected to have shorter reading times for experimental sentences starting from the earliest moment when they encounter the target word (Hypothesis 1). As they will start to encode target word when they are still fixated on the pre-target region due to para-foveal processing, I expect to see shorter first pass reading durations starting from this region. This effect will be very small in early words but it will reach its maximum in the late processing of the final word because sentence wrap-up processes take place here and a predictable sentence will be added to the situational model of the reader much more easily

2. What are the separate and combined effects of WMC and L2 proficiency on predictive inference generation during L2 reading as reflected by early- and late-processing measures of eye-movements on pre-inferential, inferential, post-inferential, and final words of sentences which follow a highly suggestive context?

Given the existing research findings indicating that high-WM readers have more resources for higher-level reading processes such as inferences (Alptekin & Erçetin, 2011; Calvo, 2001; Fincher-Kiefer & D'Agostino, 2004; Just & Carpenter, 1992),

facilitation effect differences were expected to occur between the two groups of WM level. However, it was hypothesized that the differences between high and low WM readers would occur in terms of the early measures on the pre-target and target region of the continuation sentence, but not on the post-target and final words (Hypothesis 2). This can be explained by the syntactic structure of the reading passages used in the current study. Specifically, the second sentence on which the EM measures were obtained started with the subject of the first sentence. Therefore, high-WM readers were expected to make this anaphor resolution more easily and spare more resources to the processing of the word in para-fovea which is the inferential target, whereas post-target and final words were expected to be processed very fast by both groups. As such, the high-WM participants were hypothesized to generate predictive inferences earlier compared to low-WM participants.

As for late processing measures, low-WM participants were expected to show facilitation effect in late processing of the pre-target and target words since they would need some time to generate inferences. As such, low-WM readers would show a peaked facilitation effect in these regions whereas high-WM participants would show only the spill over effects from the early processing measures (Hypothesis 3).

Regarding the effect of proficiency, the high-proficiency participants were expected to show greater facilitation effects compared to low-proficiency participants in first fixation duration and gaze duration near the target word since they carry out lower-level processes like lexical access more easily. However, as the subjects of context and continuation sentences are the same and thus the words in pre-target regions are mostly prime nouns which have been explicitly mentioned at the beginning of context sentence, high proficiency participants were expected to read them very fast both in experimental and in control conditions. Therefore, it was though that there

would not be much room for facilitation effect in this region and the facilitation effect would spill over to the target region. As a result, the effect of language proficiency on predictive inference generation was expected to emerge during the early processing of the target word with the high-proficiency participants showing larger facilitation effect in this location during early processing (Hypothesis 4). On the other hand, for the post-target and final regions, no differences between the proficiency groups were expected since there would be much lower first fixation and first pass durations. Readers at the both language proficiency levels were expected to have already started making predictive inferences by post-target word and both groups would have reached the maximum facilitation which can be provided by predictive inference in these locations.

As for late processing measures, a significant but reverse effect in regression path duration and rereading time was expected. In other words, it was thought that as the high-proficiency participants would start predictive inference generation earlier and show facilitation effects in early processing measures, they would have less facilitation in late processing. Low-proficiency participants, on the other hand, would delay inference generation and facilitation effects until late processing, which would result in longer regression time in pre-target region. As a result, the effect of predictive inference generation would emerge on this region for low-proficiency participants. Therefore, in late processing measures such as regression path and rereading duration taken from pre-target region, low-proficiency participants were expected to show greater facilitation effects compared to high-proficiency participants (Hypothesis 5).

CHAPTER 3

METHODOLOGY

This study had two parts. First, a preliminary study was conducted to determine suitable passages to be used in the eye-tracking task. In the second part, subjects took Vocabulary Levels Test, complex span tasks, and eye-tracking task. None of the subjects attended both parts of the study. Further details about the methodology are provided below.

3.1 Participants

A total of 105 university participated in study. The students in the preliminary study comprised of 38 senior undergraduate students (9 male, 29 female; mean age = 22.4) enrolled in the Foreign Language Education Department of Boğaziçi University in order to determine cloze probability of target words used in the eye-tracking experiment. None of these students attended the eye-tracking task of this study. The students were volunteers and did not receive any course credit or monetary reward.

The students in the second part of the study comprised of 67 university students (46 female, 21 male; M_{age} = 20.5, SD_{age} = 4.35). Of these students, 29 were registered in the Foreign Language Teaching Department and constituted the high-proficiency group since they had all passed the university's English proficiency test and had been exposed to English-medium instruction for at least three years at the time of data collection. There were also 38 students registered in the university's English preparatory class and they constituted the lower-proficiency group in the actual study. All subjects had normal or corrected to normal vision. Subjects received either course credit or monetary reward for their participation.

The proficiency of the subjects was determined depending not only on their level of education but also on their standardized test scores from TOEFL or IELTS, if they had reported any, and their performance on the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001). As indicated above, the participants in the high-proficiency group had already passed Bogazici University English Proficiency Test (Bogazici University School of Foreign Languages, 2017), showing that their language level is sufficient for following any course which is entirely in English. The minimum passing score on the BUEPT is considered to be equivalent to minimum 79 on TOEFL IBT and minimum 6.5 on IELTS Academic (YADYOK Öğrenci El kitabı, 2019). The students in the lower-proficiency group were from advanced level classes of English preparatory school, which were formed based on an institutional placement test. However, the data of three subjects in this group were transferred to the high-proficiency group based on their standardized English test scores and performance on Vocabulary Levels Test. One of these subjects had 111 from TOEFL IBT and one subject had 7 from IELTS, and both had 30 from vocabulary levels test. The third subject had 30 out of 30 in vocabulary levels test but he did not report any standardized English test score.

3.2 Data collection instruments

3. 2. 1 The Vocabulary Levels Test

The Vocabulary Levels Test is a vocabulary size test developed by Laufer & Nation (1999) for measuring the participants' vocabulary knowledge. Although the test has five levels, each consisting of 30 items, only half of the items from the first two levels of the test were used in this study due to time limitations. The first two levels

of the test used in this study consist of the most frequent 2000 and 3000 words in English corpus, respectively. In this study a revised version of the test was used (Schmitt, Schmitt, & Clapham, 2001). The used items of the test are presented in Appendix A.

A group of items and their options are given in Figure 1 as an example. The items in the test are grouped into three. For every three items, a group of six words are given as options. Three of the options are either synonyms or short explanations for the three items. The participants are asked to find the most appropriate one among these six options for each item. The other three options are distracters. In the test, items from 2000 and 3000 frequency levels were randomly ordered.

1	business	
2	clock	_____ part of a house
3	horse	_____animal with four legs
4	pencil	_____something used for writing
5	shoe	
6	wall	

Figure 1. Sample from Vocabulary Levels Test.
Adapted from "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." by N. Schmitt, D. Schmitt and C. Clapham, 2001 *Language Testing*, 18(1). Copyright 2001 by Sage Publication.

The participants took this test on paper without any time limit. On one side of the paper there were instructions and an example test question with answers. For each correct answer the participant received 1 point and for each participant three different vocabulary scores were calculated; one for 2000 frequency level, one for 3000 frequency level, and one for the total, which is the sum of the two levels.

Vocabulary Levels Test was formed in 2001, and its validity and reliability has been investigated in several studies. The definitions and synonyms in this test were chosen so that they had higher frequency than the target words. Moreover, orthographic similarities between items in the same groups were kept at minimum. Similarities in terms of meaning among items in the same groups were also kept as low as possible so that subjects who have very little knowledge about a word can still match it correctly. However, target words were presented in alphabetical order and their definitions were in order of length so as to minimize guessing. The authors also checked the frequency in terms of word families and instead of using the base form of words, which are generally given in word counts, they used the most frequent member of each word family.

As the present study focuses on reading comprehension, a receptive vocabulary test was considered more appropriate than a productive one. One purpose of the vocabulary test in this study was to show whether high- and low-proficiency participants differed in terms of vocabulary knowledge. Since the sentences in eye-tracking experiment consisted of relatively high frequency words, it was hoped that the high- and low-proficiency participants would differ in knowledge of low-frequency words but not high-frequency words.

3. 2. 2 Complex span tasks

Foster et al.'s (2015) shortened complex span tasks were used to assess the participants' WM capacity. The authors prepared computerized versions of operation span, rotation span, and symmetry span (see Figure 2) containing fewer blocks and trials than the original ones in order to make these tasks more practical. All these

tasks were developed using E-Prime Professional 2.0 software (Psychology Software Tools, Pittsburgh, PA, 2016).

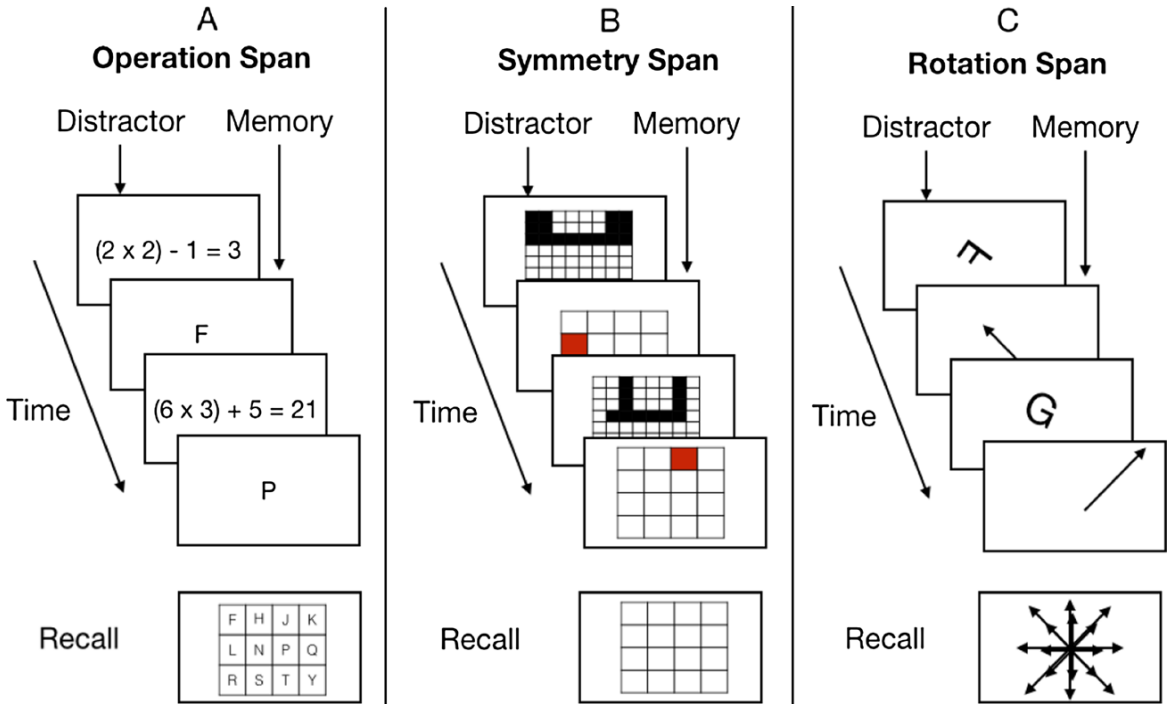

Figure 2. Flowchart of complex span tasks.
Adapted from "*Shortened complex span tasks can reliably measure working memory*." by J. L. Forster, Z. Shipstead, T. L. Harrison, K. L. Hicks, T. S. Redick, R. W. Engle, 2015 *Memory & Cognition*, 43(2). Copyright 2014 by Psychonomic Society

In operation span (see Figure 2) the storage component was letters and the processing component was simple math problems involving four basic operations. Subjects were first given math problems with an answer. When they solved the problem they clicked on the mouse and they were asked if the answer was true or false. After they made their decision, a letter appeared on the screen for 500 ms and then another math problem was given. There were 5 sets of problem-letter sequences in each block and the number of math problems and letters to be remembered in sets ranged from three to seven. After each set, a recall screen appeared and the participant was asked to choose presented letters in the correct order by clicking on the boxes next to each

letter. In recall screen there were buttons to clear the choices in case any mistakes were made and to add a blank place into the sequence so that the order of remembered letters could be entered correctly even if the participant forgot one of the letters.

In symmetry span task (see Figure 2) the storage component consisted of remembering the location of red squares which was shown on a 4x4 empty grid. The processing component was deciding whether a shown matrix was symmetrical along its vertical axis. At the end of each set, a recall screen was shown in which the participant was asked to click on the locations where red squares have been presented on a 4x4 grid. There were also buttons to clear all the choices made and to fill in the place of any forgotten location. The participants were expected to choose the locations in the order they were shown. The task had 12 sets of symmetry matrix-red square sequences divided into three blocks. In each block the number of red squares to be remembered changed between 2 and 5.

The rotation span task (see Figure 2) consisted of remembering direction and size of arrows as storage component and deciding whether rotated capital letters were facing the correct direction or not. Some of the letters were normal and some others were mirror images. As they were rotated around their centre, the subjects had to rotate them mentally before making a decision. There were sixteen possible arrows with 8 different angles and two different sizes in the storage part. After each set of arrow-letter sequence, the participants saw a recall screen in which there were sixteen possible arrows. They were expected to click on the arrow heads which they have seen in the correct order. On this screen, there were also buttons to clear all the choices made and to fill in the place of forgotten arrows. In each block there were 4 sets of sequences ranging between 2 to 5 arrow-letter pairs.

In all the tasks, the participants had beginning practice sessions in which they first did the storage part alone, then processing part alone, and finally both parts together as in the experimental trials. Their response times to processing task were recorded in these practice sessions and each participant had 2.5 standard deviation of their mean response time for answering each processing question in experimental trials. If they failed to give an answer during this time, the true/false screen was skipped and the answer is computed as a processing mistake. Moreover, in all the tasks the participants were told to keep their processing accuracy as high as possible, and they were given feedback after every sequence. This aimed at ensuring that the participants attended to the processing task and did not only rehearsed items to be recalled.

To prevent exhaustion of the subjects, Foster et al.'s 20^{th} model, which consists of one block of operation span, two blocks of symmetry span, and three blocks of rotation span, was used. This model takes relatively shorter time to complete and explains 97.8% of variance in general fluid intelligence, which could be explained if the full tests were used. The instructions, buttons, and feedback in the tasks were translated into Turkish. The translations were checked by a native Turkish speaker and two Turkish Literature teachers and any punctuation mistakes or biased wordings were corrected. Moreover, the letter "Q" in storage part of operation span was changed into letter "G" because Turkish alphabet lacks "Q" and it does not fit the letter naming system of Turkish that requires adding /e/ sound after a consonant.

The three WM capacity tasks were given to the participants in random order. They were asked to follow the instructions carefully and do the practice parts as well as they could. The experimenter was in the same room during the tasks and the participants were encouraged to ask any questions they had before the experiment

started. After each task the participant was asked if he or she wanted to have a short rest. She or he started the next task as soon as she or he was ready.

For each subject an accuracy score and a WM capacity score were obtained based on their performance on three different WM capacity tasks. Accuracy score was the percentage of correct answers to the processing parts (i.e. mathematical operation for operation span, symmetry judgements for symmetry span, and letter rotation for rotation span). WM capacity score was the total number of correctly retrieved items (i.e. number of correctly recalled letters for operation span, number of correctly recalled squares for symmetry span, and number of correctly recalled arrows for rotation span). Subjects' WM capacity level was determined according to their WM capacity task scores. For each subject a single WM capacity score was computed by summing the scores of three tasks according to unit scoring and partial scoring method, in which every correct recall counted as one point regardless of the set size each unit was in or whether all the items in the same set were recalled correctly. Subjects' accuracy scores on the processing part of the three WM capacity tasks were combined to obtain a single accuracy score. The combined accuracy score was weighted according to the number of blocks each WM capacity task had. Participants with a weighted accuracy score less than 0.75 were excluded from the analysis. Then, the subjects were ranked according to these summed scores and the ones above the median (70.5) were assigned to high-WM capacity group (N=30, min=71, max=91, mean= 78.46, sd= 5.51, median= 78.5) and the ones below the median were assigned to low WM group (N=30, min= 35, max=70, mean=59.3, sd=9.17, median= 63).

3. 2. 3 Predictive inference generation task

The participants were given an inference generation task that consisted of 64 passages and 64 questions based on these passages. The passages used in this task were developed based on the Spanish texts used in Calvo (2001). There were two groups of passages: predictive and control. The first sentences of the predictive passages established a predictive context whereas the first sentences of the control passages did not elicit any prediction.

To prevent word-based priming the words which are semantically close to the target words were used in both predictive and control sentences and thus passages were formed in pairs; for each predictive passage a similar control passage was written. As a result, on average 58.4% of the words were the same for the experimental and control pairs. The number of words in each condition was also controlled and they were made as close as possible in order to prevent the effects of sentence length on the results ($M_{\text{experimental}} = 31.47$, $M_{\text{control}} = 31.38$, $t(62) = 0.085$, $p = 0.93$).

Frequency of the target words was also controlled to prevent any effect of lack of vocabulary knowledge on eye-movement data. The median rank of target words was 671.5 according to General Service List (West, 1953). The high frequency of target words minimized the possibility that low proficiency readers did not know the meaning of these words.

A completion norm study was conducted to make sure that the target words could be predicted from the predictive context sentences but not from control sentences. For this purpose initially formed 100 pairs of predictive and control

passages were divided into 5 lists. Each list contained 20 predictive and 20 control passages which were not pairs and they were randomly ordered for each participant. Each passage consisted of the first sentence which established a context and the subject of the second sentence. Target and following words were omitted. An example is provided below.

Next Monday was their fifth anniversary so William was planning a fine dinner with his wife, but he had to make a reservation. William …

38 university students were given one of the 5 lists. The subjects were asked to read the context sentences and complete the second sentence by answering the questions "What happened next?" They were told to use the first idea that came to their mind and use a few words only. The task was administered on paper and the participants did not have time limit. Based on the participants' answers a cloze probability score for each target word was calculated. For this calculation, each answer was evaluated by two independent judges. Each judge determined whether the given answer could be regarded as the intended target. For example if the target was "called" the answer "talked on the phone" was also counted as target. In addition to predictive passages, answers for control passages were also evaluated. If similar answers other than the target words were given to the same passage by different subjects they were marked as unanticipated targets by judges. Only answers which were marked as target by both judges were counted as target in cloze probability calculation. Percent agreement between the judges was 0.82 and Cohen's kappa was 0.70. For each passage, cloze probabilities for pre-determined targets and unanticipated targets were calculated.

For inclusion in the eye-tracking experiment, predictive and control sentence pairs were evaluated together. For a pair to be used, its predictive passage was expected to elicit the target with at least 0.55 probability. If the control passage of the same pair elicited the pre-determined target or an unanticipated target word with a probability of 0.30 or higher, that pair was excluded. After this exclusion, 64 passages, 32 with predictive context and 32 with control context were used in the eye-tracking experiment along with 6 practice passages. For each passage a simple two-choice question was written to make sure that the participants read the passages for comprehension. An example of a predictive and control context pair is given below. A full list of practice, predictive, and control passages are presented in Appendix B and comprehension questions are presented in Appendix C.

Control context + inferential word:

Charles had planned a fine dinner with his wife for their fifth anniversary which was next Monday, and had made a reservation. Charles called his favourite restaurant.

Predicting context + inferential word:

Next Monday was their fifth anniversary so William was planning a fine dinner with his wife, but he had to make a reservation. William called his favourite restaurant.

For each passage four areas of interest were specified for the eye-tracking experiment. One or two words before the target word were the pre-target region. Target word was another region of interest. The words between the target and sentence final word was the post-target and the last word of each passage was specified as the final region of interest. For the first example above “Charles” is pre-

target, “called” is target, “his favourite” is post-target, and “restaurant” is the final areas of interest. Both early and late measures of processing were computed for each target. Early processing measures were first-fixation duration, first pass duration, and probability of fixation. Late processing measures were regression-path duration, selective regression path duration, and re-reading time.

The 64 passages were divided into 2 lists. Each list contained only one passage from each pair in order to maximize the interval between predictive and control pairs during eye-tracking task. The order of the lists and the order of the passages in the lists were randomized across participants.

3. 3 Procedures

Before starting the experimental tasks, the participants were asked to read and sign a consent form. First, they answered a short questionnaire on their demographic information, sight related problems if there were any, and language background. After the questionnaire they were given the Vocabulary Levels Test on paper.

After Vocabulary Levels Test, the participants received automated complex span tasks. Order of the WM capacity tasks was randomized across participants. The stimuli were presented on a desktop computer with 15.6 inch screen. The subjects were seated in front of a table approximately 70 cm. away from the computer screen. They used the mouse for entering their choices. Subjects took the tasks individually.

The last task was predictive inference generation task. To record eye-movements during this task, EyeLink® 1000 Plus Desktop Mount was used. In this setup, the participant read the stimuli on a display screen without any attachment to his or her body. A chin and forehead rest was used to minimize head movements. The distance between participant’s eyes and the screen was 70 cm. Reading was

binocular but data from only one eye were recorded. Whether to record left or right eye was determined based on subject's preference. If the subject did not state any preference, movements of right eye were recorded. The eye-movement data had a temporal resolution of 1000 Hertz.

The task was prepared by using SR Research Experiment Builder software (SR Research Experiment Builder 2.1.1, 2017). The passages were presented on a 17 inch LCD screen with a resolution of 1280x1024 and refresh rate of 60 Hertz. The eye-tracker was connected to a host PC, which the subject could not see during the task. After the instructions were given a nine-point calibration was conducted.

In this task, the subjects read short passages and answered simple, two-choice questions about these passages while their eye-movements were being recorded. The stimuli were presented in Times New Roman 14 font size in black colour on white background. The participants did not have any time limit on reading the passages and answering the questions. They were given written instructions first and then they saw 6 practice passages one by one. They were encouraged to ask any questions they had during this time. After reading each passage, they clicked on the mouse to see the related question. They entered their responses by clicking on either of the options. After every two passage-question pairs a fixation dot appeared on the screen where the first letter of the next passage would be. The experimenter checked if there were any problems with calibration in this drift check phase. If not, the next passage was presented. Eye-tracker was calibrated again when drifts exceeded one degree of visual angle, and then the experiment continued from the same point.

3. 4 Data analysis

The participants were rank-ordered according to their storage scores on the complex span tasks and divided into high- and low-WM groups using a median-split method. Thus, for each proficiency level there were two groups of WM. Seven subjects were removed from the analysis either because they did not satisfy required accuracy on WM capacity task (N= 4) or the eye-tracking data were not usable due to artefacts (N=2). One subject did not take the eye-tracking part of the experiment. This left 60 subjects with 30 high-proficiency readers and 30 low-proficiency readers. As a result, 4 groups were formed. These were high proficiency-high WM (n=15), high proficiency- low WM (n=15), low proficiency-high WM (n=15), low proficiency-low WM (n=15).

Before analysing EM data, preliminary analyses using factorial ANOVA were conducted on vocabulary scores, reading comprehension scores, and reading duration as well as the processing and storage scores of the WM task in order to examine whether the initial assumptions were met and whether proficiency and WM groups were formed reliably. These initial assumptions were that both groups of language proficiency would not have difficulty in understanding passages but differ in terms of their proficiency level, language proficiency and WM were independent, and WM groups were different in terms of WM capacity. Moreover, it was thought that these preliminary analyses would reveal if vocabulary knowledge and difficulty had any confounding effects in comprehension. For these analyses, WM level and proficiency levels were between subject variables and context was within subject variable. Subjects were treated as random effect.

To analyse the EM data, analyses were conducted on early and late measures for each region of interest, namely pre-target region, target region, post-target region, and final region. The early measures were fixation probability, first fixation duration, and first run duration. Fixation probability was calculated as the percentage of all fixations on a passage which landed on the area of interest. It is calculated by dividing the number of fixation on an area by total number fixations during the passage was on the screen. As its name suggests, first fixation duration is the time spent on the first fixation falling on the area of interest. First run duration is the sum of all fixations on an area of interest from the start of first fixation until the gaze leaves the area to left or right. Figure 3 shows an example of hypothetical recording.

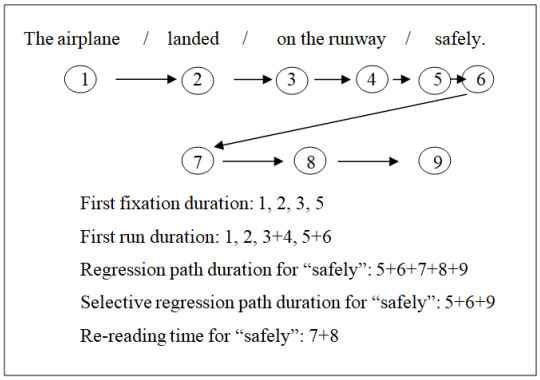

Figure 3. Example of a record showing eye-movement measures

As for the late processing measures, regression path duration, selective regression path duration, and re-reading time were used as dependent variables. Regression path duration is the sum of all fixations starting when the gaze first enters an area of interest until another area on the right is fixated. It includes all the fixations on the area of interest and regressions made from this area. Selective regression path duration is the total duration of fixations on an area of interest. This measure includes both first run duration and re-fixations on the same area. Re-reading duration is the time spent during regressions from an interest area.

For each region of interest and eye-movement measure, 2 (WM level) x 2(Proficiency Level) x 2 (Context) ANOVA with subjects as random effect was performed. In these analyses, WM level and proficiency were between subject variables and context was within subject variable.

As ANOVA is a parametric test, its assumptions of normality of the scores and homogeneity of variance across groups were checked before each analysis. Specifically, score distributions were examined through histograms for each comparison groups prior to the analysis. None of the histograms indicated non-normal distributions. The assumption of homogeneity of variances (i.e. homoscedasticity) was checked via Levene's Test of Homogeneity of Variance. In all the analyses reported in the results section, the significance level of Levene's Test was above .05 indicating group variances do not differ significantly.

CHAPTER 4

RESULTS

4.1 Preliminary analyses

Before analyzing eye-movements, data collected through other measures were examined to test the difference between WM and proficiency groups as well as to see if any confounding variables were affecting the results. To this end, vocabulary scores, reading comprehension scores, reading and question response durations, and WM task scores were analyzed and compared. The statistics for the significant effects are indicated within the text and the complete ANOVA summary tables are provided in Appendix D.

4.1.1 Vocabulary scores

Table 1 shows means and standard deviations of vocabulary scores as a function of WM capacity and language proficiency in two levels of vocabulary frequency as well as overall vocabulary scores. It should be noted that although there is difference between scores of high and low proficiency participants, both groups have quite high scores from both levels of vocabulary, given that the maximum possible score on the test is 30. Moreover, as the frequency of words decrease, the difference between scores of high and low proficiency readers increase.

Table 1. Means and Standard Deviations for Vocabulary Scores as a Function of 2(WM Capacity Level) X 2(Proficiency Level) X 2(Frequency Level)

	2000 frequency proficiency level				3000 frequency proficiency level			
	High		Low		high		Low	
WM	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
high	14.93	0.26	14.07	1.58	14.47	1.06	11.93	2.12
low	14.73	0.59	13.07	2.19	14.13	1.25	12.13	3.60
Total	14.83	0.46	13.57	1.94	14.30	1.15	12.03	2.91

Note. M and *SD* represent mean and standard deviation, respectively.

In order to examine whether the proficiency or WM groups were different in terms of their vocabulary scores, a 2 (low vs. high proficiency) x 2 (low vs. high WM) x 2 (2000 vs. 3000 frequency levels of vocabulary) mixed ANOVA was conducted with WM capacity and proficiency level as between subject variables and vocabulary frequency as within subject variable. Subjects were treated as random effect variable. The ANOVA results revealed a significant main effect of language proficiency, $F(1,56)=16.87$, $p < 0.01$ and vocabulary frequency, $F(1,56) = 22.71$, $p < 0.01$ while WM capacity did not have a significant main effects. Thus, high-proficiency participants ($M = 29.13$, $SD = 1.38$) had significantly higher overall vocabulary scores than low-proficiency participants ($M = 25.60$, $SD = 4.45$). The significant effect of frequency level suggests that participants had higher scores in 2000 frequency level than in 3000 frequency level.

The only significant interaction effect (see Figure 4) was between language proficiency and frequency level of words, $F(1,56) = 5.32$, $p < 0.025$.

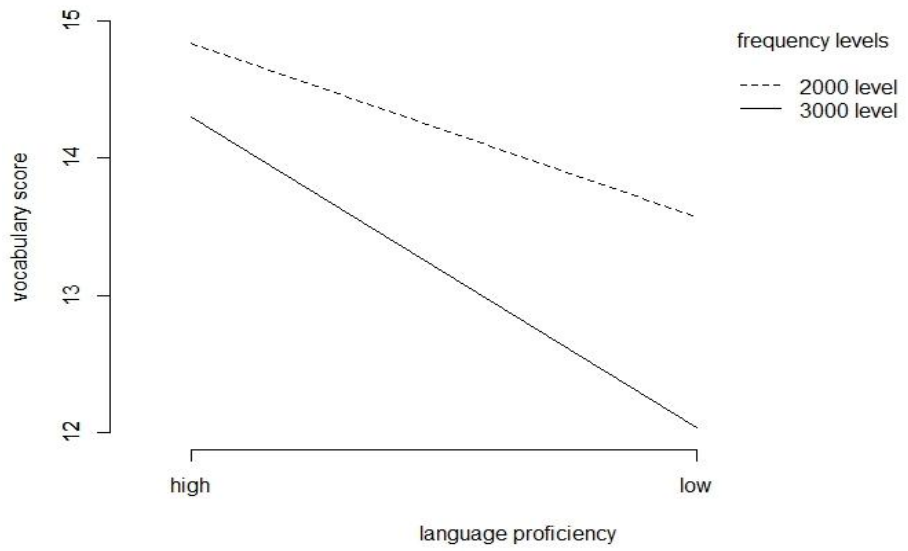

Figure 4. Change in vocabulary score as a function of language proficiency and frequency level

Figure 4 shows that as the frequency of words decrease the difference in vocabulary scores of high and low proficiency readers increase. This is in line with the initial assumption that while high proficiency readers would do similarly in both levels of vocabulary frequency, low proficiency readers' vocabulary scores would decline as the frequency of words decrease. Moreover, this also indicates that as the frequency of words increases, the difference in vocabulary knowledge of high and low proficiency readers decreases. Considering the high frequency levels of target words used in the passages, these results suggest that the effect of vocabulary knowledge on comprehension and eye-movement data had been minimized. Also, despite the lack of a standardized proficiency test the participants in this study were assigned to high and low proficiency groups quite efficiently.

4.1.2 Reading comprehension

Table 2 shows means and standard deviations of reading scores as a function of language proficiency, WM capacity level, and context. These reading comprehension scores were collected during eye-tracking task from the participants' answers to questions on passages. Similar to vocabulary scores, comprehension scores were also quite high, regardless of proficiency, WM level, or context considering the possible maximum scores was 64; 32 for control and 32 for predicting context questions.

Table 2. Means and Standard Deviations for Comprehension Scores across Proficiency and WM Groups

	WM capacity level				Context			
	High		Low		Control		Predictive	
proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	29.13	1.91	29.07	1.74	28.17	2.00	30.03	0.93
Low	28.37	1.97	29.00	1.74	27.60	1.81	29.77	1.19
Total	28.75	1.96	29.03	1.73	27.88	1.91	29.90	1.07

Note. M and SD represent mean and standard deviation, respectively.

A 2 (low vs. high proficiency) x 2 (low vs. high WM) x 2 (control vs. predictive context) mixed ANOVA on reading comprehension scores was conducted to see if the participants differed on their comprehension of the stimuli due to their language proficiency and WM capacity scores. Proficiency and WM capacity level were between subject variable while context was within subject variable. Subjects were treated as random effect variable. The results did not indicate a significant main effect of either proficiency or WM. However, context had a significant main effect, $F(1,56) = 59.93$, $p < 0.01$, suggesting that control passages received significantly less

correct answers than predictive sentences. None of the interaction effects reached significance.

These results indicate that there was not a difference between readers with different levels of language proficiency in terms of their comprehension of the passages and the differences in eye-movements data is not because of the difference in their overall comprehension of the passages. The significant difference in comprehension scores between predictive and control sentences shows that control sentences were relatively more difficult to understand. This was an expected result because prediction facilitates comprehension. However, comprehension scores of control passages were also quite high (27.88 on average out of 32), which indicates that the readers did not have much difficulty in understanding these passages.

4.1.3 Reading duration

For each subject the time it took to read the presented passages and to answer the comprehension questions given after each passage had been recorded (see Table 3). The effects of language proficiency, WM, and context on the subjects' reading times were examined through 2 (low vs. high proficiency) x 2 (low vs. high WM) x 2 (predictive vs. control context) mixed ANOVA. Proficiency and WM capacity level were between subject variable while context was within subject variable. Subjects were treated as random effect variable. The results showed that WM did not have a significant effect on reading times. None of the interaction effects approached significance. Language proficiency, $F(1,112) = 11.86$, $p < 0.01$ and context, $F(1,112)= 8.67$, $p < .01$ had significant effects on reading duration. The main effect of context had been expected considering the facilitative effect of prediction in reading (Goodman, 1967). The effect of language proficiency showed that high

proficiency participants read the sentences faster than low proficiency participants. This is in line with general course of language learning, in which reading speed increases with experience and language proficiency.

Table 3. Means and Standard Deviation (in ms) Reading Duration of Passages as a Function of a 2(Proficiency) X 2(WM Level) Design

	WM level				Context			
	High		Low		Control		Predictive	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
high	9748.76	3139.61	8661.04	2007.20	9951.92	2941.67	8457.89	2165.63
low	11012.51	2779.22	10979.25	3489.32	11780.81	3283.78	10210.95	2800.37
Total	10380.64	3007.93	9820.15	3054.68	10866.36	3225.52	9334.42	2634.61

Note. M and SD represent mean and standard deviation, respectively.

4.1.4 Response time for questions

Table 4 shows means and standard deviations of response time for questions in eye-tracking task of the study. The effects of language proficiency, WM capacity, and context on the subjects' reading times were examined through 2 (low vs. high proficiency) x 2 (low vs. high WM) x 2 (predictive vs. control context) mixed ANOVA. Proficiency and WM capacity level were between subject variable while context was within subject variable. Subjects were treated as random effect variable. This analysis also revealed a similar pattern with WM having no significant effect. None of the interaction effects reached significance. Similar to reading time results, language proficiency, $F(1,112)= 10.43$, $p < 0.01$ and context, $F(1,112)= 4.41$, $p = 0.04$ had significant effects and as discussed above these observations were expected because of the facilitative effect of prediction and gain in reading speed with increasing proficiency.

Table 4. Means and Standard Deviations for Question Response Time (in ms) as a Function of a 2(Proficiency) X 2(WM Level) Design

	WM				Context			
	High		low		control		predictive	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	2462.83	420.61	2570.61	393.82	2593.18	390.24	2440.27	416.70
Low	2750.66	535.10	2907.78	715.29	2956.01	630.05	2702.43	616.87
Total	2606.75	498.76	2739.19	597.18	2774.59	550.86	2571.35	538.38

Note. M and SD represent mean and standard deviation, respectively.

4.1.5 Performance on WM tasks

The participants' performance on the processing and storage the complex span tasks were compared across MW and proficiency groups. For the processing task, the accuracy scores were obtained for each participant based on the percentage of correct answers to processing questions. As for the storage scores, the total number of accurately recalled items constituted the storage measure. As high- and low-WM participants were grouped according to the storage scores by median-split method, the difference between the high- and low-WM subjects' scores were, as expected, statistically significant, $t(58)= 9.8$, $p< 0.001$. Table 5 shows means and standard deviations of both processing and storage scores as a function of WM level and proficiency.

Table 5. Means and Standard Deviations for WM Scores across WM and Proficiency Groups

	Proficiency					
	High		low		Marginal	
	M	*SD*	*M*	*SD*	M	SD
Processing						
High-WM	0.94	0.05	0.92	0.05	0.93	0.05
Low-WM	0.89	0.07	0.93	0.04	0.91	0.05
Total	0.92	0.06	0.93	0.05	0.92	0.05
Storage						
High-WM	77.87	5.73	79.07	5.42	78.47	5.51
Low-WM	57.07	10.02	61.53	7.96	59.30	9.18
Total	67.47	13.27	70.30	11.15	68.89	7.35

Note. *M* and *SD* represent mean and standard deviation, respectively.

Table 5 shows that while the difference between high and low WM participants' accuracy scores were quite high; the participants from both groups of language proficiency did almost equally well. This also supports the independency between language proficiency and WM.

A 2 (proficiency) x 2 (WM level) ANOVA with subjects as random effect was conducted on the processing and storage scores separately. For the processing scores, the results indicated no significant main effect. However, the interaction effect (see Figure 5) was significant, $F(1,56)= 5.16$, $p= 0.03$. Post-hoc comparisons showed that the only significant difference in accuracy scores was between high proficiency-high WM and high proficiency-low WM groups, $t(28) = 2.44$, $p = 0.02$. The non-significant main effect of proficiency level suggests that it did not affect the participants' performance on the WM task. This observation is important in that it shows independency between proficiency and WM for the scope of this study.

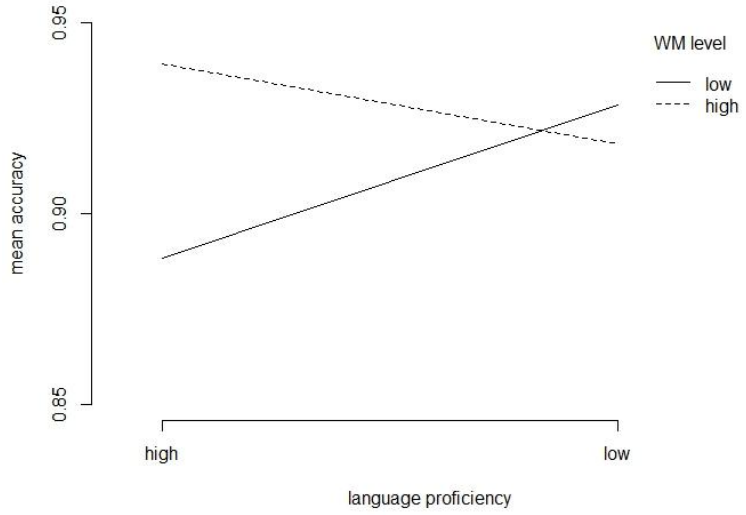

Figure 5. Mean accuracy scores in complex span task as a function of WM and proficiency

As for storage scores, the ANOVA results showed neither main effect of language nor any interaction effect. As expected, WM level had a highly significant main effect, $F(1,56)=97.55$, $p < 0.01$.

4.1.6 Summary of preliminary analyses

These analyses have some important implications for the integrity of this study. First, they show that despite the lack of a standardized test of proficiency, the participants were assigned to the correct proficiency groups representing their language level. Secondly, these results indicate that all the participants understood the passages, and any difference in eye-movement data is not because some participants did not understand the passages used in the eye-tracking task. Lastly, the results show that WM and language proficiency are independent constructs for the scope of this study. These will enable more confidence in discussion of the eye-movement data.

4.2 Eye movements

For each measure and location, I conducted a 2 (language proficiency) x 2 (WM level) x 2 (context) ANOVA with subjects as random effect. Language proficiency and WM level were between subject variable while context was within subject variable. When context showed a significant or close to significant interaction effect with either of the other variables, I carried out further analyses to see how the differences in reading times are affected by language proficiency or WM capacity level. For this aim, I calculated facilitation effects by subtracting value of predicting context from the value of control context and analysed the effects of independent variables on this derived facilitation value.

4.2.1 Early measures

For each area of interest the effects of language proficiency, WM level, and context on fixation probability, first fixation duration, and first run duration were analysed.

4.2.1.1 Pre-target region

Table 6 shows the means and standard deviations of fixation probability on pre target region as a function of language proficiency, WM level, and context. As can be seen from the ANOVA conducted on these variables, the only significant effect was of context, $F(1,56) = 4.09$, $p = 0.05$. Predictive passages helped readers to skip first word of the continuation sentence. This shows the facilitative effect of predictive inferences in that readers could predict upcoming word after reading a suggestive context and did not need to fixate on it as much.

Means and standard deviations for first fixation duration as a function of proficiency, WM level, and context are presented in Table 6. As can be seen from the ANOVA table, for first fixation duration language proficiency had a significant main effect $F(1,56) = 8.13$, $p = 0.006$. High proficiency readers spent much less time on this region compared to low proficiency readers. Moreover, WM level and context showed a small interaction effect, $F(1,56) = 3.61$, $p = 0.062$. While high WM readers could show facilitation effect here, low WM readers could not. Figure 6 shows this interaction.

For first run duration, language proficiency and context had significant main effects (language proficiency: $F(1,56) = 8.18$, $p = 0.006$; for context: $F(1,56) = 4.03$, $p = 0.05$). Similar to first fixation duration, high proficiency readers dwelled on this region less than low proficiency readers did. The significant effect of context

indicates that predictive inference facilitated reading pre-target region. The lack of this effect of context in first fixation duration suggests that this facilitation was caused not by first fixation, but by following fixations on the same word. Implications of this result in terms of para-foveal processing were discussed together with facilitation in regression duration in discussion section.

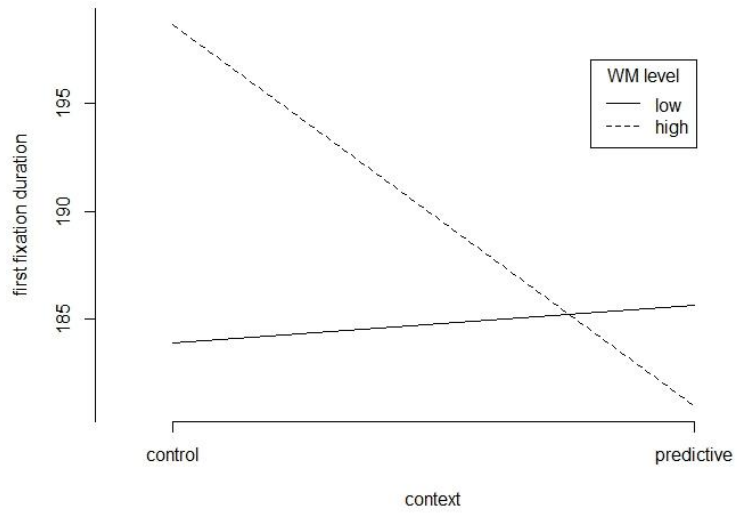

Figure 6. Interaction of WM level and context on first fixation duration of pre target region. Exp and cont corresponds to predicting and control context, respectively

Table 6. Means and Standard Deviations for Fixation Probability, First Fixation Duration and First Run Duration of Pre-Target Region as a Function of a 2(Proficiency) X 2(WM Level) Design

Context	Proficiency	WM level	Fixation probability		First fixation duration		First run duration	
			Mean	SD	Mean	SD	Mean	SD
Control	High	high	0.03	0.03	180.59	147.55	211.67	190.13
		low	0.04	0.03	171.05	127.01	192.63	154.35
	Low	high	0.04	0.03	216.72	126.03	248.29	162.9
		low	0.03	0.03	196.78	148.83	239.06	197.49
Predictive	High	high	0.03	0.03	160.69	132.98	185.74	175.79
		low	0.03	0.03	164.3	123.21	184.73	147.51
	Low	high	0.04	0.03	201.3	126.04	241.71	172.4
		low	0.03	0.03	206.97	210.48	236.9	229.32

4.2.1.2 Target

Table 7 shows means and standard deviations for fixation probability, first fixation duration, and first run duration of target words as a function of context, language proficiency, and WM level. For all three measures, target words of predictive context sentences received shorter fixations compared to control context sentences. This emerged as significant main effect of context on fixation probability, F (1,56) = 21.17, $p < 0.001$; first fixation duration, $F(1,56) = 4.68$, $p = 0.034$; and first run duration $F(1,56) = 9.42$, $p = 0.003$.

There were small interaction effects in first fixation duration (see figure 7) and first pass reading time (see Figure 8), but these effect did not reach significance. Comparison of high and low language proficiency groups' facilitation means in first fixation duration and first pass reading time indicated that readers with high language proficiency showed greater facilitation than low proficiency readers (for first fixation duration: $M_{high\ proficiency} = 14.31$, $M_{low\ proficiency} = 1.76$; $t(58) = 1.72$, $p = 0.091$; for first pass reading time: $M_{high\ proficiency} = 20.68$, $M_{low\ proficiency} = 6.03$; $t(58) = 1.71$, $p = 0.093$). This suggests that high proficiency readers showed slightly bigger facilitation in early processing of the target word than low proficiency readers did. WM did not have any significant effect in early processing of target word.

Table 7. Means and Standard Deviations for Fixation Probability, First Fixation Duration, and First Run Duration of Target Region as A Function of a 2(Context) X 2(Proficiency) X 2(WM Level) Design

Context	Proficiency	WM level	Fixation probability		First fixation duration		First run duration	
			Mean	SD	Mean	SD	Mean	SD
Control	High	high	0,04	0.03	213.47	120.28	244.73	162.68
		low	0,05	0.03	219.32	104.29	251.3	140.63
	Low	high	0,04	0.03	216.06	103.06	244.85	140.91
		low	0,04	0.03	217.38	135.3	263.27	183.68
Predictive	High	high	0,04	0.03	198.64	129.59	222.21	159.56
		low	0,05	0.03	205.53	106.97	232.45	131.11
	Low	high	0,04	0.03	211.19	118.25	241.19	150.49
		low	0,04	0.02	218.74	129.58	254.87	161.2

.

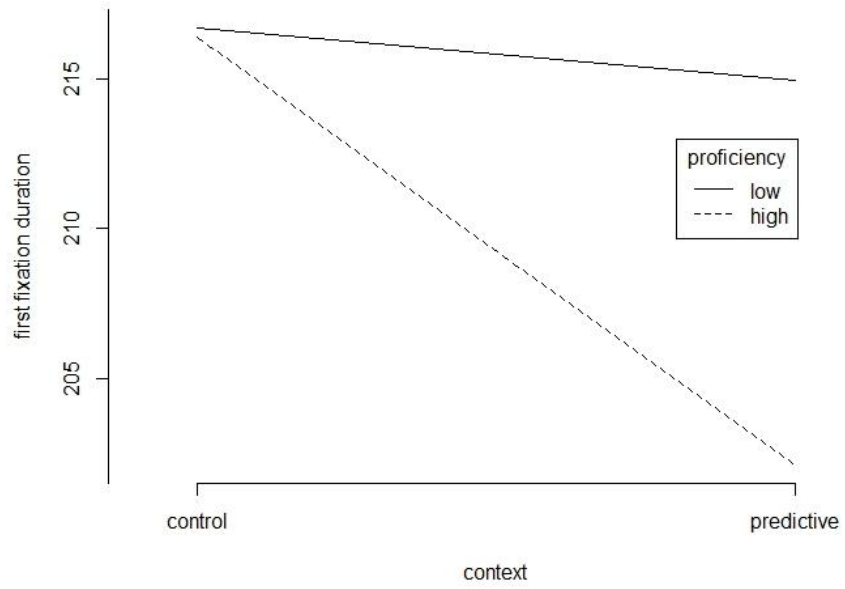

Figure 7. First fixation duration of target region as a function of language proficiency and context

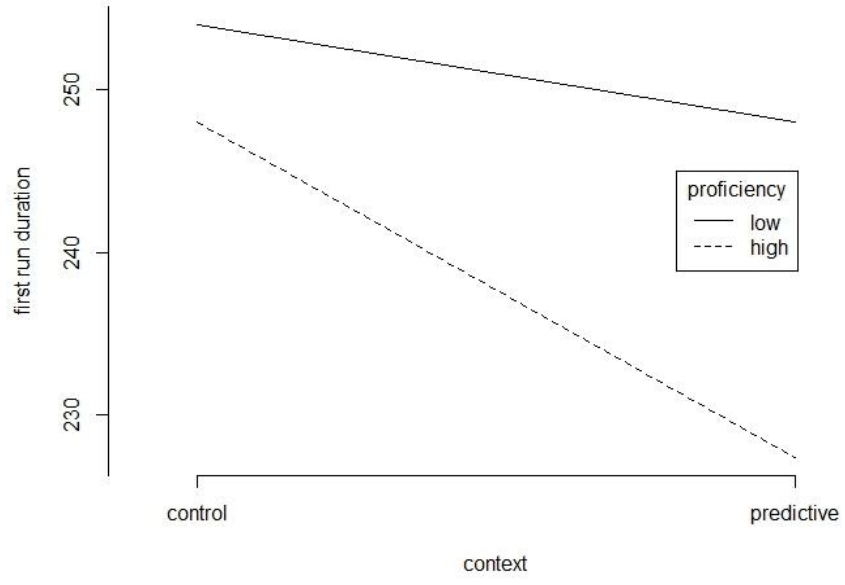

Figure 8. First run duration of target region as a function of language proficiency and context

4.2.1.3 Post-target

Table 8 summarizes the means and standard deviations of fixation probability, first fixation duration, and first run duration as a function of context, language proficiency and WM level for post-target region. High proficiency readers skipped this area more often than low proficiency readers did. On the post-target region context and language proficiency had main effects on fixation probability (context: $F(1,56) = 27.95$, $p < 0.001$; proficiency: $F(1,56) = 5.64$, $p = 0.02$). There were not any significant difference between high and low proficiency readers on first fixation duration but high proficiency ones spent significantly less time in first run duration; $F(1,56) = 4.31$, $p = 0.04$. WM level did not show any significant effect on any of the measures. None of the interaction effects reached significance. This was expected because after pre-target and target word, variance among readers in terms of facilitation would disappear due to ceiling effect. Implications of this observation in terms of prevalence and time course of predictive inferences were further discussed in discussion section.

Table 8. Means and Standard Deviations of Fixation Probability, First Fixation Duration (in ms), and First Run Duration (in ms) for Post Target Region

Context	Proficiency	WM level	Fixation probability		First fixation duration		First run duration	
			Mean	SD	Mean	SD	Mean	SD
Control	High	high	0.09	0.05	249.93	106	496.17	323.62
		low	0.09	0.05	229.31	92.8	463.13	317.76
	Low	high	0.08	0.04	256.6	99.38	537.61	321.84
		low	0.08	0.04	248.32	112.42	538.53	340.15
Predictive	High	high	0.08	0.04	243.42	100.19	468.25	276.41
		low	0.08	0.05	233.74	94.96	420.03	247.87
	Low	high	0.08	0.04	246.83	101.01	486.83	294.01
		low	0.07	0.04	249.54	140.06	517.84	327.2

4.2.1.4 Final region

Table 9 provides means and standard deviations for early eye movement measures on final region. On this location, predictive context had lower values for all three measures. This was supported by the significant main effect of the context in ANOVA. For the final region context had a significant effect on all three measures (fixation probability: $F(1,56) = 10.0$, $p = 0.009$; first fixation duration: $F(1,56) = 24.66$, $p < 0.001$; first pass reading time: $F(1,56) = 16.75$, $p < 0.001$). As final location is where sentence wrap-up processes take place, this effect was expected because prediction makes integration of incoming information to existing structure of the text easier. However, as most of the sentence wrap-up processes happen during late processing, the difference between durations of control and predictive context sentence were not very high at this point.

Table 9. Means and Standard Deviations of Fixation Probability, First Fixation Duration (in ms), and First Run Duration (in ms) For Final Region

Context	Proficiency	WM level	Fixation probability		First fixation duration		First run duration	
			Mean	SD	Mean	SD	Mean	SD
Control	High	High	0.04	0.04	185.76	144.93	243.99	228.09
		Low	0.03	0.03	150.19	158.43	179.7	194.74
	Low	High	0.03	0.03	149.16	226	193.81	404.15
		Low	0.03	0.03	168.11	151.79	226.93	235.37
Predictive	High	High	0.04	0.03	160.42	138.21	203.82	209.77
		Low	0.03	0.03	119.48	133.25	150.18	194.57
	Low	High	0.02	0.02	135.13	136.89	164.66	185.34
		Low	0.03	0.03	154.83	151.96	205.64	227.74

4.2.2 Late measures

For each area of interest, the effects of language proficiency, WM level, and context on regression path duration, selective regression path duration, and re-reading time were analysed.

4.2.2.1 Pre-target

Table 10 shows means and standard deviations of regression path duration, selective regression path duration and re-reading time on pre-target region. In pre-target region context showed a significant main effect in all three measures of late processing (regression path duration: $F(1,56) = 5.95$, $p = 0.017$; selective regression path duration: $F(1,56) = 4.72$, $p = 0.034$, rereading time: $F(1,56) = 4.00$, $p = 0.05$). Readers spent longer time in regressions from pre-target words of control sentences than the ones of predictive sentences. This shows that readers started to demonstrate the effects of predictive inference in late processing of pre-target region.

Main effect of language proficiency emerged in regression path duration, $F(1,56) = 4.58$, $p = 0.036$ and selective regression path duration, $F(1,56)= 8.36$, $p= 0.005$. High proficiency readers had shorter durations than low proficiency readers for these two measures.

In regression path duration there was a significant three-way interaction; $F(1,56) = 4.36$, $p = 0.04$. I conducted further analysis on facilitation values to see how the groups differ (see figure 9). In regression path duration there was a significant difference between high proficiency-low WM and low proficiency-low WM groups ($M_{high\ proficiency\text{-}low\ WM} = -11.41$, $M_{low\ proficiency\text{-}\ low\ WM} = 91.01$; $t(28)= 2.19$, $p= 0.036$).

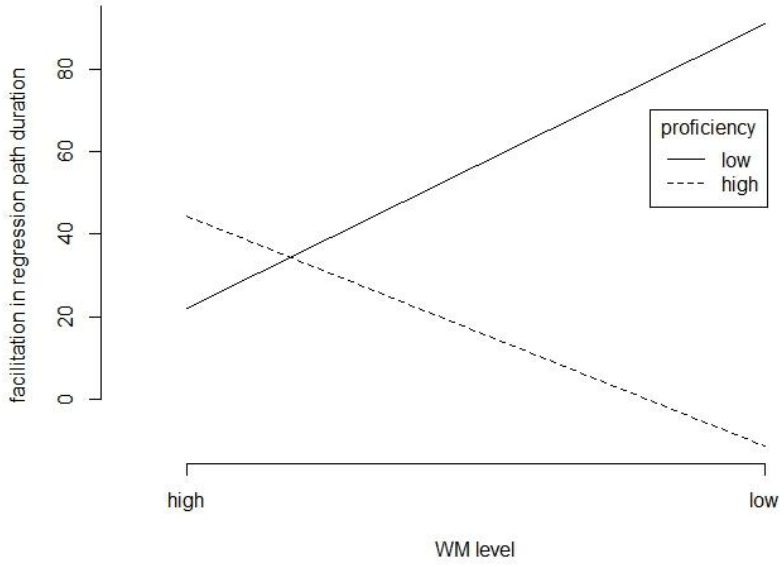

Figure 9. Interaction between WM and language proficiency on facilitation in regression path duration of pre-target region

The same three way interaction was also significant in re-reading time; $F(1,56) = 5.00$, $p = 0.029$. Post-hoc analysis on facilitation values of rereading time of pre-target region showed the same pattern in which low proficiency-low WM readers had greater facilitation than high proficiency-low WM readers; $M_{high\ proficiency\text{-}low\ WM} = -16.15$, $M_{low\ proficiency\text{-}\ low\ WM} = 85.47$; $t(28) = 3.0$, $p = 0.006$ (see figure 10). These findings were interesting in that it showed that proficiency and WM affected predictive inference generation and reading comprehension in general, through different mechanisms.

Moreover, in re-reading time there was a significant two way interaction between context and proficiency; $F(1,56) = 4.22$, $p = 0.044$. Here, low proficiency readers had greater facilitation than high proficiency readers. The relationship between this result and the difference in reading speed between proficiency groups were further discussed in discussion section.

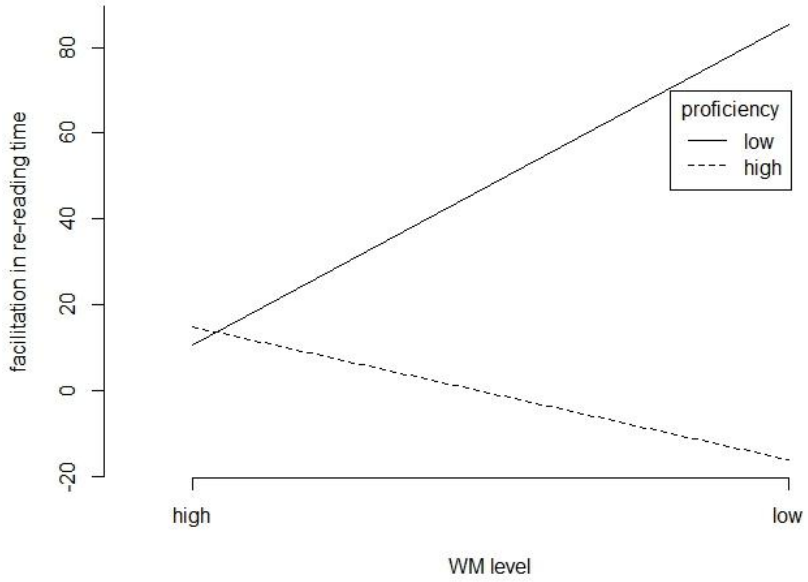

Figure 10. Interaction between WM and language proficiency on facilitation in re-reading time of pre-target region

The opposite trend of facilitation between selective regression path duration and other two late processing measures suggested that high and low proficiency readers showed the effects of predictive inference generation through different mechanisms. That is; while low proficiency readers read the predictive passages faster because they shorter regressions, high proficiency readers' advantage was due to para-foveal processing. Implications of this finding, together with the results obtained in early measures, were further discussed in discussion section.

Table 10. Means and Standard Deviations of Late Processing Measures (in ms) for Pre-Target Region

Context	Proficiency	WM level	Regression path duration		Selective regression path		Rereading time	
			Mean	SD	Mean	SD	Mean	SD
Control	High	high	338.55	708.45	225.99	205.92	112.57	628.5
		low	247.98	343.3	201.59	165.86	46.38	260.67
	Low	high	381.23	650.54	267.04	194.12	114.19	595.61
		low	388.21	847.65	256.59	221.39	131.62	790.63
Predictive	High	high	294.22	609.09	196.52	194.86	97.7	554.16
		low	259.39	444.98	196.85	168.11	62.54	372.59
	Low	high	359.35	730.52	255.72	184.67	103.63	680.08
		low	297.2	468.52	251.05	245.59	46.15	365.56

4.2.2.2 Target

Table 11 shows means and standard deviations for the late measures on target words. The ANOVA results revealed a significant main effect of context in regression path duration, $F(1,56) = 4.19$, $p = 0.045$ and selective regression path duration, $F(1,56) = 16.26$, $p < 0.01$. As control sentences caused longer regressions, it can be concluded that the readers were showing the facilitative effects of predictive inferences.

Furthermore, the interaction effect of language proficiency by context was significant for regression path duration, $F(1,56) = 4.84$, $p = 0.032$ and rereading time, $F(1,56) = 5.35$, $p = 0.024$. For both measures, high proficiency readers showed greater facilitation effect than low proficiency readers [regression path duration:, $M_{\text{high proficiency}} = 76.73$ $M_{\text{low proficiency}} = -2.74$, $t(58)= 2.17$, $p = 0.033$; rereading duration: $M_{\text{high proficiency}}= 52.74$, $M_{\text{low proficiency}} = -19.84$, $t(58)=2.28$, $p = 0.026$] This indicates that while high proficiency readers could show the effect of predictive inference generation at this point, low proficiency readers could not. What this finding suggests in terms of the time course of predictive inference generation is further discussed in the discussion section along with findings in other regions.

There was a small WM by context interaction in regression path duration, $F(1,56) = 3.24$, $p = 0.077$ as well as re-reading time, $F(1,56) = 3.53$, $p = 0.065$. Pair-wise comparisons for facilitation effects showed that both in regression path duration and rereading time low-WM participants showed greater facilitation when they were reading a predicting context [regression path duration: $M_{\text{high WMC}} = 4.45$, $M_{\text{low-WM}} = 69.54$, $t(58)= 1.76$, $p = 0.084$; rereading time: $M_{\text{high-WM}} =-13.03$, $M_{\text{low-WM}} = 45.9$, $t(58)= 1.83$, $p = 0.073$] This is in line with Hypothesis 2 which stated that high WM readers employ para-foveal processing more, in that they showed facilitation effect

of target word while they were fixated on pre-target word because they could process the target word para-foveally then.

Table 11. Means and Standard Deviations of the Late Measures (in ms) for Pre-Target Region

Context	Proficiency	WM level	Regression path duration		Selective regression path		Rereading time	
			Mean	SD	Mean	SD	Mean	SD
Control	High	high	389.31	723.99	264.21	175.22	125.09	681.42
		low	447.47	813.4	273.93	164.82	173.54	748.15
	Low	high	349.9	309.35	279.65	162.64	70.26	222.65
		low	363.35	352.71	294.61	215.35	68.74	238.37
Predictive	High	high	351.39	614.19	243.64	181.78	107.76	553.43
		low	331.91	431.86	246.52	142.93	85.39	377.76
	Low	high	378.92	710.47	265.26	162.89	113.66	673.37
		low	339.81	373.1	274.79	169.9	65.03	308.46

4.2.2.3 Post-target and final regions

Table 12 and table 13 show means and standard deviations for post-target and final regions, respectively. For these regions, the only difference was between predicting and control contexts. Readers with different WM and proficiency levels showed similar facilitation effects. Only context had a significant main effect (all *p*s< 0.001 for all three measures of both regions). None of the other main or interaction effects reached significance.

This implies that during the late processing of post-target and final regions all the readers were benefiting from facilitative effect of predictive inferences, and this facilitation peaked at these regions, where sentence wrap-up processes take place.

Table 12. Means and Standard Deviations of the Late Measures (in ms) for Post-Target Region

Context	Proficiency	WM level	Regression path duration		Selective regression path		Rereading time	
			Mean	Sd	Mean	SD	Mean	SD
Control	High	High	1439.29	2393.04	650.22	504.5	789.07	2070.23
		Low	1393.11	2001.42	602.37	435.07	790.74	1746.29
	Low	High	1477.79	1990.31	656.17	429.16	821.62	1763.1
		Low	1264.85	1753.44	656.71	418.73	608.14	1531.81
Predictive	High	High	963.9	1408.74	536.94	345.16	426.96	1215.68
		Low	1072.23	1348.98	521.41	350.69	550.82	1166.94
	Low	High	1149.72	1705.48	587.53	437.06	562.19	1467.35
		Low	860.48	947.91	577.29	407.68	283.18	752.89

Table 13. Means and Standard Deviations of the Late Measures (in ms) for Final Region

Context	Proficiency	WM level	Regression path duration		Selective regression path		Rereading time	
			Mean	Sd	Mean	SD	Mean	SD
Control	High	High	1747.97	2908.68	322.19	377.8	1425.78	2706.55
		Low	1325.42	2231.62	230.15	269.66	1095.28	2121.83
	Low	High	1956.47	3001.53	271.74	483.15	1684.73	2762.9
		Low	2059.22	3387.42	319.61	361.46	1739.61	3233.61
Predictive	High	High	1048.64	2027.56	239.47	255.07	812.26	1905.03
		Low	759.76	1521.46	186.15	291.56	586.65	1396.99
	Low	High	1281.17	2101.48	201.14	265.06	1060.73	1983.21
		Low	1333.17	2422.95	247.46	293.02	1070.18	2313.2

4. 2. 3 Location-wise comparisons

As fixation duration is highly dependent on the word length, I compared the regions by dividing the mean facilitation effect for each region by the mean number of characters. Table 14 presents mean facilitation for six measures across the four locations. It should be noted that late processing measures of final region are drastically longer than the other measures and locations. This shows the sentence wrap-up processes that take place at the end of sentences.

For all the measures the differences in facilitation among the four locations were significantly different for all the measures. The results of pair-wise comparisons (Table 15) indicated that the facilitation effect was significantly larger in the final region, especially during late processing. Mean facilitation effect per character and pair-wise comparison results for each location are given in the tables 14 and 15, respectively.

Table 14. Mean Facilitation (in ms) per Character

	FFD	FPD	RPD	SRPD	RRT
Pre-target	1.13	1.47	5.39	1.70	3.68
Target	1.54	2.44	7.52	3.66	3.86
Post-target	-1.00	1.00	22.54	4.25	18.29
Final	3.28	4.58	86.66	9.4	77.93

Note. FFD = First fixation duration, FPD = First pass duration, RPD = Regression path duration, SRPD = Selective regression path duration, RRT = Rereading time

Table 15. p Values for Pair-wise Comparison of Facilitation per Character

	Pre-target –Target	Pre-target - Post-target	Pre-target - Final	Target - Post-target	Target : Final	Post-target - Final
FFD	0.98	0.31	0.31	0.19	0.48	0.01
FPD	0.8	0.97	0.06	0.56	0.29	0.03
RPD	0.98	0.04	<0.001	0.08	<0.001	<0.001
SRPD	0.45	0.24	<0.001	0.97	0.01	0.01
RRT	0.99	0.07	<0.001	0.07	<0.001	<0.001

Note. FFD = First fixation duration, FPD = First pass duration, RPD = Regression path duration, SRPD = Selective regression path duration, RRT = Rereading time

4.3 Summary of EM analyses

The results of this study clearly show that all readers make predictive inferences regardless of their WM capacity level and proficiency. This finding confirms Hypothesis 1. WM capacity level and language proficiency on the other hand have effects on the time course of predictive inference generation. However, these two variables exert their effects through different mechanisms. As suggested by Hypothesis 2, high WM capacity readers started making predictive inferences on early processing of the pre-target word while low WM capacity readers started showing the effects of predictive inference later, during the late processing of the same word. This observation is in accordance with the predictions of Hypothesis 3. Although language proficiency did not determine whether predictive inferences are drawn, it determined when the readers started to benefit the facilitative effect of these inferences. High proficiency readers read the pre-target word very fast and did not show any facilitation effect here. As predicted by Hypothesis 4, their facilitation effect spilled over to the early processing of the target word. Low proficiency readers, on the other hand, could show the effect of predictive inferences during the late processing of pre-target region, which had been suggested by Hypothesis 5.

CHAPTER 5

DISCUSSION

The findings presented in the previous section are discussed in terms of their implications for predictive inference generation and its time course. Moreover, effects of WM capacity, language proficiency and their interaction on predictive inference generation are further examined.

5. 1 Predictive inference generation

One of the main findings of this study is that sentences following a predictive context were read faster than sentences following a neutral context by all the participants. This indicates that, predictive inferences are generated easily and neither language proficiency nor WM capacity determine whether or not they are drawn. Regardless of language proficiency and WM capacity, all participants in the current study showed facilitation effect while they were reading a sentence which followed a highly predictive context. These two factors, however, determine at which point in the sentence the facilitation effect of predictive inferences starts to emerge by varying available processing resources and defining how much of the processes become automatic. Thus, similar to previous findings the effects of WM and processing efficiency manifested their effects not at generation, but in the time course of predictive inferences (Murray & Burke, 2003).

Previous literature on the nature of forward inferences and the extent to which they are generated during reading has presented mixed results. While several studies did not find any evidence for online predictive inference generation (Magliano, Baggett, Johnson, & Graesser, 1993; Potts, Keenan, & Golding, 1988; Singer &

Ferreira, 1983), in others reading a predictive context led to facilitation in the measurement task (Calvo, 2001; Fincher-Kiefer, 1993; Klin, Murray, Levine, & Guzman, 1999). The sources of this discrepancy have been mostly attributed to the passages and measurement tasks used in different studies. For predictive inferences to be generated the context passages should be sufficiently constrained and the predicting context should be in the focus when the to-be-predicted target is presented. That is, there should not be long intervals between prediction eliciting context and target. Moreover, the measurement task should also be devised carefully. As the time course of predictive inferences is complex, the target should be presented just at the right moment so that effects of predictive inference can be observed. The results of the current study are in line with those of previous studies that used highly constrained texts and tasks which were close to normal reading in nature. As the cloze probabilities of the passages used in the current study were quite high and there was not any intervening time or task between context and target sentences, the passages were expected to elicit predictive inferences. Moreover, the eye-tracking technique did not alter the nature of normal reading at all. The readers were actively reading during the whole task and they did not have any opportunity to engage in context checking. Most of the papers which examine generation of predictive inferences using eye-movement technique have also reported similar facilitation effects (Calvo, 2001; Calvo & Castillo, 1998; Calvo, Meseguer, & Carreiras, 2001).

In the current study the facilitation effect of predictive inferences was observed for all locations and for both early and late measures. However, the greatest facilitation effect was observed during the late processing of the final region. These results can be considered as evidence for the proposal that sentence integration process takes place at the sentence final location suggested by the CI model of

comprehension. Thus, in the current study, the effect of predictive inferences on building situation model emerged as much larger facilitation at the sentence final word.

In the macro-level processing, the reader has to form a text-base representation, which is a representation of the global text. If the reader has all the necessary propositions to form the text-base, it is generated relatively easily. If not, a series of strategic processing such as LTM search or re-reading the text are employed. This is the source of large facilitation effects observed in the final locations. As the to-be-predicted proposition was active in WM due to predictive inference generation and as it is confirmed by the explicit text, integration process might have become less effortful in the experimental passages. For the control sentences, on the other hand, as the context sentence did not elicit any predictions, the readers may have employed strategic and effortful processes to form a situational model of the passage.

The relatively small facilitation effects observed in pre-target and target regions show that after reading a predictive context the target words were more easily available to the readers. While the readers were processing the context sentence, the target word was receiving activations in each cycle of text-base formation process. As the context sentences in experimental passages are highly constrained, activation of the target word exceeds the threshold and it becomes a part of the WM. Thus, in the current study, the spreading activation was observed as the facilitative effect emerged during early measures. Several studies suggested that this activation is a result of word-based priming instead of inferential processing (Elman, 1990; Keenan & Jennings, 1995). However, the facilitation observed in the current

study cannot stem from word-based priming because both experimental and control passages had the same content words to a great extent.

The finding that L2 readers were able to generate predictive inferences easily regardless of proficiency level, needs to be discussed also in relation L2 reading processes. Intuitively, reading in an L2 is harder because it requires more conscious processing compared to L1 reading, in which most low-level processes are automatic. As discussed before, simple processes like lexical access may require conscious effort in L2 reading. This is expected to consume processing resources of L2 readers and restrict higher level processes like elaborative inferences. However, the fact that even low proficiency level readers could show evidence of predictive inference generation contradicts this intuition. One of the possible reasons for this can be the use of top-down processing by L2 readers. As they have difficulty in lower-level processes, it can be more productive for them to first make predictions about the text and then test them with new input as suggested by top-down models of reading. In this case predictive inferences are expected to be observed in L2 reading more prevalently than they are in L1 reading (Horiba & van den Broek, 1993). Another possible explanation for this observation can be the proposition that as the text becomes more difficult, the reader tries to employ elaborative inferences more efficiently to make comprehension as complete as possible. This was also suggested by Keefe and McDaniel (1993), who demonstrated that when the text was difficult the priming effect was greater and it lasted longer compared to normal text. As reading in L2 introduces some level of difficulty through increased strategic processing, facilitation effect in L2 reading is expected because the readers are trying to compensate for the deficiency in their lower level processing with the help of elaboration processes.

5.2 Time course of predictive inferences

In inference generation research "when" question has been examined heavily. Although some studies have suggested that predictive inferences are generated immediately following a predicting context and their effects can be observed early in processing, others have claimed that inference generation takes time and its effects do not emerge in early processing. The proposed time required for generation of predictive inferences ranges from around 200 ms to more than a second. The results of the present study supports both types of time course and indicates that this discrepancy might be caused by the differences in methods and samples employed in different studies.

The results of the current study show that facilitation effects of predictive inferences appear immediately after reading the predicting context; on the first word of the target sentence. Mean regression path duration of the final word of the context sentence was 380.06 ms when averaged for all the participants. As regression path duration is the time from first fixation until the eyes left the word to the right, it can be concluded that the effects of predictive inference generation took around this amount of time, which agrees with the finding of prior research, which used techniques like naming latency or word recognition (Keefe & McDaniel, 1993).

Several studies which used eye-tracking to examine predictive inference generation did not find early facilitation effects (Calvo, 2001; Calvo & Castillo, 1998). In these studies the effects emerged on the late processing of the final word of the target sentence. The results of the current study show a similar pattern; facilitation in the late processing of the final location is significantly larger than facilitation in earlier locations. However, the earlier facilitation effects found in the

present study are not in accordance with previous eye-tracking findings. This can be caused by reading in L2 and might have several contributing factors.

First, readers with higher language proficiency read the first words of target sentences faster than readers with low proficiency. As a result, high proficiency readers could not show facilitation effect in the early processing of pre-target region. The speed of reading might cause the L1 readers' lack of facilitation in early words of the sentences. As they can read both experimental and control sentences very fast, there is not any room for facilitation effect to emerge here. It is possible that although L1 readers generate predictive inferences much earlier, they reach the maximum reading speed even on control sentences because of their native proficiency. Only during the sentence wrap-up processes L1 readers spend enough time on a word for effects of predictive inference to emerge. For L2 readers in the current study, on the other hand, reading was relatively slow due to some processes being effortful, and thus there was a small room for variation between experimental and control sentences even when the pre-target and target words were being processed. Considering the fact that low-proficiency readers started showing facilitation effects earlier than the high proficiency readers, this explanation indicates that predictive inferences are generated early, but their effects emerge only when reading speed limit allows.

5.3 The role of WM capacity in predictive inference generation

In the present study, the participants with both high- and low-WM showed facilitation effect so WM was not a determining factor. However, the effects of WM emerged on the time course of predictive inference generation. The participants with high-WM started inference generation earlier than those with low-WM. While high-

WM participants showed significant facilitation effects on early processing of pre-target word, for the low-WM participants the difference between the experimental and control sentences reached significance on the late processing of the pre-target and target word depending on their language proficiency. There can be two mechanisms involved in this observation: easier referent-antecedent association and faster para-foveal processing. These mechanisms are working together to achieve the mentioned advantage. That is, WM resources that are spared as a result of easier referent-antecedent association are used for para-foveal processing.

In the reading passages used in the current study, the subjects of first and second sentences were the same and second sentence always started with the subject. Although it was not a pronoun, the reader had to find the antecedent of this subject when he or she started reading the second sentence (Clark & Sengul, 1979). The advantage of having high-WM in making reference-antecedent association has been reported in several studies (Carpenter, Miyake, & Just, 1994; Just & Carpenter, 1992). As high-WM readers have more capacity, they can store more information in their WM and therefore they will not have difficulty finding this antecedent. They can spare more resources for other processes. Considering the span of para-foveal processing (McConkie & Rayner, 1975), it is highly possible that they started processing the target word which was in the para-fovea at this point. As the target word was a highly predictable one in experimental passages, they showed facilitation effects even when it was in para-fovea. Meanwhile, low-WM readers were trying to associate the subject of the second sentence with its antecedent so that what they read made sense. Therefore, they could not start processing the target word, which is in the para-fovea at this time. Readers with low-WM could not start processing the para-foveal target word during the early processing of pre-target, which delayed the

emergence of facilitation effect until late processing of the pre-target and target words.

This observation is further supported by the pattern emerged in late processing of target region. At this point, there was a marginally significant difference between high- and low- WM participants' facilitation effect. The low-WM participants showed greater facilitation effect compared to high-WM ones. However, the exact measure where this significant difference stemmed was re-reading duration, which is the sum of all the regressions to previous words. This measure excludes the fixations and re-fixations on the current interest area. However, there was not a significant difference between the WM groups in selective regression path duration, which is the sum of all fixations on the current interest area; the target word. This clearly shows that low-WM readers showed facilitation effect because they needed fewer regressions while reading a highly predictable sentence whereas high-WM readers showed facilitation because they could process para-foveal word more efficiently while reading a highly predictable sentence.

In the post target and final regions, readers with high- and low-WM showed significant facilitation effect but the magnitude of this effect did not differ between groups because all readers had reached maximum facilitation and because of this ceiling effect, differences between groups were not significant. This finding suggests that the effect of WMC on predictive inference generation is quite small and it exerts its effect through indirect mechanisms.

5.4 The role of language proficiency in predictive inference generation

Similar to WM, language proficiency was not a determining factor because readers with both levels of language proficiency showed facilitation effect while they were

reading highly predictable sentences. However, the location where this facilitation effect started to emerge was related to language proficiency. The low-proficiency participants started to show facilitation earlier than the high-proficiency participants. This can be explained by the differences in speed of reading and efficiency of processing lower-level features between the proficiency groups.

In the present study pre-target words were mostly prime nouns which were mentioned in the context sentences of the passages and the time spent on these words were dependent on the readers' language proficiency. The high-proficiency participants (M = 285.03 ms) spent much less time on these words compared to the low proficiency participants (M = 356.5 ms), regardless of whether the passage was experimental or control, $t(58) = 2.14$, p =0.04. The mean regression path duration of the high-proficiency participants on the pre-target region was 276 ms. in experimental passages and 293 ms. in control passages. Prior research indicates that this duration is almost the minimum time needed to encode a word (Just, Carpenter, & Wolley, 1982). Moreover, high-proficiency participants skipped the pre-target region more often than low-proficiency readers did. The difference between the two groups of readers in skipping rate in the pre-target region was large, although not significant ($M_{high\ prof}$ = 42.96, $M_{low\ prof}$ = 34.79), $t(58) = 1.75$, p = 0.08). Although high-proficiency high-WM readers could show facilitation here, it was mostly due to the effect of WM, not language proficiency. For high proficiency low-WM readers this duration was just enough to complete referent-antecedent association. As low-WM readers could not benefit from para-foveal processing as much as high-WM readers, they did not show any facilitation effects. Low proficiency readers, on the other hand, were slow enough on late processing of pre-target word to show facilitation.

The differences in reading durations decreased on the target word and thus the high-proficiency participants could start to show the effects of predictive inference at this location. Their facilitation effect at this point was significantly greater than the facilitation effect shown by the low-proficiency participants. Interestingly, the low-proficiency participants showed almost no facilitation effect on the target word. Considered together with the facilitation effect observed in the pre-target region, this can be an argument for automatization of processing with increasing proficiency. As stated before, the pre-target regions were mostly prime nouns appeared in the first sentences of the passages. However, the target region consisted of verbs encountered for the first time in the passages. It can be argued that the low-proficiency participants showed facilitation in the pre-target region because they did not have to engage in resource consuming processes such as lexical access.

However, while low-proficiency participants were fixating on the target word they had to access its meaning and integrate it to the text read so far. Because processes like lexical access and syntactic parsing were less automatic for these readers, they probably had to employ effortful processes. As a result, they did not have enough resources left for the effects of predictive inferences to emerge. The effect of prediction spilled over to the post target region instead. The same pattern emerged in final region, too. High-proficiency participants showed significantly greater facilitation effect than the low-proficiency participants on the early processing of the final region. This observation suggests that the high-proficiency participants were faster in completing lower-level processing, which mostly occur during first pass reading, and thus they could show the effects of predictive inferences earlier.

On the post-target region, language proficiency did not have any significant effects, either in early or in late processing measures. The post target region in the passages used in the current study consisted of more than one word. At this region the participants had already started benefiting from the predictive inference and its effect was reaching the peak. The facilitation effect in the late processing of the post-target region is much higher for both groups compared to facilitation effects in previous words. As this effect is at its maximum for both groups of readers, the effects of language proficiency disappeared here. The same happened in the late processing of the final word, where the facilitation effect was significantly greater than that observed in all the other locations and measures. This ceiling effect concealed the effect of language proficiency, which was relatively small compared to the effect of predictive inference in sentence wrap-up processes.

5.5 The interaction between WM and proficiency in inference generation

During early processing, WM and language proficiency started exerting their effects at different points. Therefore, no interaction effect was observed in early processing of any of the locations. However, language proficiency and WM capacity had reverse effects during the late processing of pre-target word probably because they contributed to the facilitation effect through different mechanisms.

At late processing of pre-target word, the participants with low language proficiency and low-WM showed the greatest facilitation effect because they could start making referent-antecedent association here. As predictive passages were easier in this respect, low-proficiency low-WM readers read pre-target regions of predictive passages faster. However, due to their low-proficiency they could not employ this facilitation in early processing. Another group which showed facilitation here was

high-proficiency participants with high-WM. However, their facilitation was because of the advantage in para-foveal processing, instead of faster referent resolution. Because they had high WM capacity, they did not have much difficulty in associating the referent to its antecedent during early processing of pre-target region. Besides, they could benefit from para-foveal reading more, so they probably started processing the target word while they were fixating on the pre-target word. Thanks to their high language proficiency, they could show the facilitation effect of predictive inference immediately because their lower level processes were automatic. This enabled them to access the meaning of the target word earlier and start integration processes faster. As a result, the facilitation effect of predicting the target word could emerge even before it was directly fixated.

The argument that different mechanisms were involved in the observed effects can be evidenced by the detailed examination of these facilitation effects. For low-proficiency participants with low-WM, the greater part of the facilitation effect emerged in rereading time, which is the duration of regressions made from the area of interest. This shows that their facilitation was due to shorter regressions in predictive passages because they were not looking for the antecedent in these sentences as much as they did in control passages. On the other hand, high-proficiency participants with high-WM showed the greater facilitation in selective regression path duration, which is the sum of all fixations on the area of the interest. This indicates that their facilitation effect emerged while they were fixating directly on the pre-target word. As they did not have difficulty in antecedent-referent association, the source of this facilitation can only be para-foveal processing of the target word.

The low proficiency participants with high-WM and the high proficiency participants with low-WM did not show facilitation here because the former had already shown it during early processing due to their high-WM capacity and although they could have started processing the target word para-foveally here, probably due to their low language proficiency they spent their time in low-level processing, which is not affected by prediction much (Ito, Martin, & Nieuwland, 2017). The latter group, on the other hand, cannot benefit from para-foveal processing due to their low WM. Also, as they read the pre-target region very fast, they did not show any facilitation effect on early processing of this region. As a result, their facilitation effect spilled over to the target region and they showed a very large facilitation effect on the target word; especially during late processing.

CHAPTER 6

LIMITATIONS AND FUTURE DIRECTIONS

The biggest limitation of this study involves operationalization of language proficiency. Due to limited resources, it was not possible to use a standardized test of English to determine the participants' language proficiency. Instead, their education status and vocabulary size scores were used, which enabled eliminating most of the mis-categorization of the participants. Although these measures discriminated high- and low-proficiency participants well, using the score of a standardized test of English could have differentiated the groups better. There were several participants studying English preparatory class although they had very high level of English because they wanted to spend a year before starting their academic program. It was possible to catch these participants through the vocabulary test and a personal questionnaire they filled and to correctly place them in the high proficiency group.

Another limitation is the total sample size of 60 and 15 for each cell. With more participants, the effects of WM and language proficiency and their interactions could have been determined more reliably.

Another limitation involves the passages used in the study. Although the experimental and predictive passages were meticulously written and tested, there were not any filler passages in the experiment; the passages were either predictive or their control ones. Including several neutral filler passages could have made the reading task more reliable. However, this would add many new passages to the task and thus introduce the effects of fatigue. Considering the participants took all tasks in one session, the fatigue could have been an important confounder. Moreover, the order of experimental and control passages was randomized across participants and

no two items in a pair immediately followed each other in the reading task. This eliminated the effects of guessing and fatigue on the results. Moreover, I did not check the plausibility judgements of the passages I used. It is possible that especially in some of the control passages the event mentioned in the continuation sentence was not a plausible event for the previous context. This might have introduced an element of surprise to the eye-movement data. However, for such kind of studies it is highly difficult to write sentences which are neither predictable nor surprising. Plausibility judgements can be included in a future study with a broader scope.

This study can be broadened to include L1 readers to better understand how L2 inference generation processes differ from L1 reading. This can clarify the differences between the results presented here and the ones found in L1 reading studies with eye-tracking technique. Moreover, such a design can show the interaction between language proficiency and WMC better. Also, to examine the para-foveal processing and spill over effects better, eye-tracking can be combined with simultaneous ERP measures. This way it will be possible to obtain better temporal resolution and better insights on the exact time course of predictive inference generation.

APPENDIX A

VOCABULARY LEVELS TEST

Participant Number: _____

A vocabulary levels test: Version 1

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning. Here is an example.

l business	
2 clock	__ part of a house
3 horse	__ animal with four legs
4 pencil	__ something used for writing
5 shoe	
6 wall	

You answer it in the following way.

l business	
2 clock	6 part of a house
3 horse	3 animal with four legs
4 pencil	4 something used for writing
5 shoe	
6 wall	

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are business, clock, shoe. Try to do every part of the test.

1 birth
2 dust _____ game
3 operation _____ winning
4 row _____ being born
5 sport
6 victories

1 acid
2 bishop _____ cold feeling
3 chill _____ farm animal
4 ox _____ organization or framework
5 ridge
6 structure

1 cap
2 education _____ teaching and learning
3 journey _____ numbers to measure with
4 parent _____ going to a far place
5 scale
6 trick

1 boot
2 device _____ army officer
3 lieutenant _____ a kind of stone
4 marble _____ tube through which blood flows
5 phrase
6 vein

1 cream
2 factory _____ part of milk
3 nail _____ a lot of money
4 pupil _____ person who is studying
5 sacrifice
6 wealth

1 betray
2 dispose _____ frighten
3 embrace _____ say publicly
4 injure _____ hurt seriously
5 proclaim
6 scare

1 bake
2 connect _____ join together
3 inquire _____ walk without purpose
4 limit _____ keep within a certain size
5 recognize
6 wander

1 assist
2 bother _____ help
3 condemn _____ cut neatly
4 erect _____ spin around quickly
5 trim
6 whirl
1 original
2 private _____ first
3 royal _____ not public
4 slow _____ all added together
5 sorry
6 total

1 dim
2 junior _____ strange
3 magnificent _____ wonderful
4 maternal _____ not clearly lit
5 odd
6 weary

APPENDIX B

PASSAGES USED IN EYE-TRACKING EXPERIMENT

Odd numbered ones are predictive context sentences and even numbered ones are control context sentences. The passages P1 to P6 are practice passages.

P1 Gabriel had been sitting on the deck for almost an hour when he finally felt a movement on his fishing rod, so he pulled it immediately. Gabriel caught a very big fish.

P3 The deadline for the project was very close so Ryan went to the library, found a table and opened his laptop. Ryan typed his project.

P4 When her phone fell on the floor, Lauren took it and checked if it was still working and then she took a wet towel from her bag. Lauren wiped her phone.

P5 After Oliver left the office, he went to the car park, found his car, and sat on the driver seat. Oliver started his car.

P6 After eating a toast for breakfast Aria washed her hands and opened the new toothpaste tube she bought the day before. Aria brushed her teeth.

1. On the day of his interview for the new job, Philip got up with sunrise, had breakfast, and wore his suit. Philip left the house early.

2. Arthur got up with the sunrise, wore his suit which he bought for the interview of his new job and decided to have breakfast. Arthur left the house early.

3. Jackob was waiting at the bus stop when the bus appeared on the corner of the street, approached right and opened its doors. Jackob got on the bus immediately.

4. When the bus approached the bus stop where Harvey was waiting and opened its doors, he noticed that it was not his bus. Harvey got on the bus immediately.

5. While Sophia was reading her book in the bus, she felt a vibration in her bag so she took her phone out and saw it was her husband. Sophia answered the phone in two seconds.

6. While Amelia was reading her book in the bus, she felt a vibration and saw her husband on the next seat reaching his bag. Amelia answered the phone in two seconds.

7. Next Monday was their fifth anniversary so William was planning a fine dinner with his wife, but he had to make a reservation. William called his favorite restaurant.

8. Charles had planned a fine dinner with his wife for their fifth anniversary which was next Monday, and had made a reservation. Charles called his favorite restaurant.

9. Ethan was very hungry when he came home from work so he went to the kitchen, saw empty fridge and took his phone. Ethan ordered some food online.

10. Dylan left his phone for repair and ate some fast food before coming home because he knew that the fridge in the kitchen was empty. Dylan ordered some food online.

11. After jogging for about 30 minutes Jane was exhausted and thirsty so she sat on a bank and opened her bag. Jane drank water from her bottle.

12. Before leaving home, Luna opened her bag to check if she had everything ready for jogging, which would make her exhausted and thirsty. Luna drank water from her bottle.

13. A cold wind was blowing when Emily arrived home, but she didn't feel warm inside because a window was open. Emily closed the window immediately.

14. When Molly looked out of the window she realized a cold wind was blowing, so she wore her thick coat to feel warm. Molly closed the window immediately.

15. Alex shut the door, stood in front of it, put his hand in his pocket and grabbed his keys. Alex locked the door with his key.

16. Leon grabbed the keys before he left home, shut the door, tied his shoes and checked his pockets for the last time. Leon locked the door with his key.

17. Mia climbed up the stairs, stopped in front of the door, opened her bag and grabbed her keys. Mia unlocked the black metal door.

18. Eva climbed up the stairs, stopped in front of the door, but she couldn't find her keys in her bag. Eva unlocked the black metal door.

19. After opening the refrigerator, Lily saw the brownish apples which she bought almost a month ago. Lily threw the apples into trash.

20. Upon opening the refrigerator, Iris saw brown eggs and green apples that she bought three days ago. Iris threw the apples into trash.

21. Because Sofia had felt very sleepy the night before, she made a mistake and set her alarm to four o'clock. Sofia woke up very early that morning.

22. Clara bought an alarm clock for four dollars because she was feeling very sleepy lately, but realized that it wasn't working. Clara woke up very early that morning.

23. After flying over the ocean the plane was finally approaching the airport, so it was descending and getting slower and slower. The plane landed safely on the runway.

24. After departing from the airport the plane was finally approaching the ocean, so the wind was descending and getting slower and slower. The plane landed safely on the runway.

25. Daniel went to the kitchen and saw the pile of dishes that remained from the dinner he gave for his friends, so he put on gloves and turned the tap on. Daniel washed the smelly dishes.

26. Austin went to the kitchen, loaded the dishwasher, took off his gloves, turned the tap off, and ate some leftovers from the dinner he gave for his friends. Austin washed the smelly dishes.

27. Liam was very confused with the explanation of the teacher so he raised his hand and when the teacher allowed he stood up. Liam asked a question about the subject.

28. Toby was very confused with the explanation of the teacher so he raised his hand but when the teacher allowed him the bell rang. Toby asked a question about the subject.

29. The postman climbed up the stairs with the last package of the day, rang the bell, heard footsteps from inside and waited. The postman delivered the heavy package.

30. The postman entered the building with the last package of the day and heard footsteps and someone ringing a bell from another floor. The postman delivered the heavy package.

31. Mathew wanted to read the document before signing but the font was too small for him, so he opened his bag and took out his glasses. Mathew wore his old glasses.

32. Edward wanted to buy new glasses because he had difficulty in reading small fonts, but he had already spent lots of money on a new bag. Edward wore his old glasses.

33. Hannah was walking to her class when she noticed some money on the floor and she decided to hand it to the school administration, so she knelt slowly. Hannah picked up the money from floor.

34. Evelyn was walking to her class when she noticed that she had dropped some money when she knelt to take her books from her locker. Evelyn picked up the money from floor.

35. Amelia had no plans for the weekend and she noticed that the fence looked very rusty, so she went to the hardware store and bought brushes and dye. Amelia painted the fence in a bright color.

36. Emilia had no plans for the weekend and she noticed that the fence looked very rusty, but she didn't have any energy to go to the hardware store for brushes and dye. Emilia painted the fence in a bright color.

37. Andrew was about to enter the library to study for his upcoming exam but the receptionist stopped him, so he took out his wallet. Andrew showed his ID to the receptionist.

38. Harley was about to enter the library to study for his upcoming exam but the receptionist stopped him because a wallet was stolen in the library. Harley showed his ID to the receptionist.

39. The old and abandoned building was empty for years and it already had many cracks on its walls when the earthquake started. The building collapsed in two minutes.

40. The old and abandoned building was empty for years but it didn't have any cracks on its walls when the earthquake ended. The building collapsed in two minutes.

41. James left home only a few minutes late, rushed to the bus stop, but saw the bus leaving so he started to run but the bus was too fast. James missed the bus that morning.

42. Logan left home only a few minutes late, rushed to the bus stop, and saw the bus had already gone, so he hailed a taxi. Logan missed the bus that morning.

43. It was a nice weekend and David decided to do some sports and take some fresh air, so he packed his lunch, checked the pressure of tires and wore his helmet. David rode his bike to the forest.

44. It was a nice weekend and Felix decided to do some sports and then go shopping to buy new tires and helmet for his brother as a birthday present. Felix rode his bike to the forest.

45. Joseph was studying for an important exam in the crowded library but the music coming from the student next to him was very loud, so Joseph touched his shoulder gently. Joseph warned the student to be careful.

46. Albert was listening to loud music in front of the crowded library, when another student touched his shoulder gently while passing through. Albert warned the student to be careful.

47. Anna was enjoying a lot in the concert where she was dancing and she wanted to have a memory of it so she took out her phone. Anna recorded a video with her phone.

48. Rose started a song of her favourite singer on her phone, danced, and enjoyed the memories she had in the concert. Rose recorded a video with her phone.

49. The cat was drinking water when it heard a dog barking, so it started to run towards the apple trees which were nearby. The cat climbed the nearest tree.

50. The cat was wandering inside the house when it heard a dog barking outside near the apple tree. The cat climbed the nearest table.

51. Sarah was drinking coffee while she was reading her book which was about two people in love; she took a sip of coffee and realized that she finished the page. Sarah turned the page over.

52. Nancy was about to finish reading the book which was about two people in love; she looked away from the page to take a sip of coffee. Nancy turned the page over.

53. Ashley was very excited for her travel the next day, so she brought her suitcase from storage, cleaned it and opened her wardrobe. Ashley packed her suitcase with clothes.

54. Martha was looking for her passport for her travel the next day, so she went through her wardrobe, opened her suitcase, and searched the storage and finally found it. Martha packed her suitcase with clothes.

55. Claire was bored at home and she decided to go to the park for a short walk, but she noticed that it was about to rain hard so she changed her mind. Claire stayed home for the rest of the day.

56. Jessie was bored at home and she decided to go to the park for a short walk, but she noticed that it was about to rain hard, so she thought about just opening the window. Jessie stayed home for the rest of the day.

57. Lola was eager to know the end of the novel, so she lay down comfortably and opened it at the page she had reached the last time. Lola read the pages that remained.

58. Emma knew the author of the novel whose photo appeared on the first page of the newspaper, so he phoned to congratulate her on her success. Emma read the pages that remained.

59. While Maria walked barefoot over the rocks, she put her foot down, without realizing, on a piece of glass which had been left on the floor. Maria cut herself with pain at that moment.

60. In order to avoid putting her dirty shoes down on the floor, Eliza walked barefoot up to the glass display cabinet to place present in it. Eliza cut herself with pain at that moment.

61. The woman went into the church, spoke with the priest for a few minutes and afterwards knelt down in front of the altar. The woman prayed a prayer with devotion.

62. After having spoken with the priest for a few minutes in front of the church's altar, the woman knelt down to do her shoe up. The woman prayed a prayer with devotion.

63. When the party was over, there were bags and papers all over the floor, so Susana picked up the broom. Susana swept the floor thoroughly.

64. In order to decorate the party Jasmine hung up the coloured papers with the broom that was on the floor. Jasmine swept the floor thoroughly.

APPENDIX C

COMPREHENSION QUESTIONS AND OPTIONS USED IN EYE-TRACKING TASK

P1 to P6 are the questions and options for practice passages. Odd numbered questions are for predictive context passages and even numbered passages are for control passages.

Question	Option 1	Option 2
P1 Where was Gabriel sitting?	on the deck	on the boat
P2 What color was the cat?	grey	black
P3 Where did Ryan go?	library	school
P4What did Lauren take from her bag?	water	wet towel
P5 Who did Jack take to the park?	his daughter	his son
P6 Could Mia plug her earphone?	yes	no
1. What did Philip wear?	t-shirt	suit
2. Did Arthur get up early?	yes	no
3. Where was Jackob waiting?	bus stop	school

4.	Did the bus stop?	yes	no
5.	Who answered the phone?	Sophia's husband	Sophia
6.	What was Amelia doing in the bus?	listening to music	reading book
7.	Which day was their anniversary?	Monday	Sunday
8.	What did Charles plan for the anniversary?	dinner	brunch
9.	What did Ethan order?	food	water
10.	How did Dylan order his food?	by phone	online
11.	What sport did Jane do?	jogging	biking
12.	Was Luna going swimming?	yes	no
13.	What was the weather like?	rainy	windy
14.	What did Molly wear?	coat	sweater
15.	Did Alex close the door?	yes	no
16.	Did Leon check his bag?	yes	no
17.	Where were Mia's keys?	in her bag	in her pocket

18. What color was the door?	blue	black
19. When did Lily buy the apples?	almost a month ago	three days ago
20. Where were the eggs?	on the table	in the refrigerator
21. Did Sofia wake up late?	yes	no
22. How much did the clock cost?	5 dollars	4 dollars
23. Where did the plane land?	runway	ocean
24. Did the plane have an accident?	yes	no
25. Who did Daniel give dinner for?	his family	his friends
26. Where did Austin put the dishes?	sink	dishwasher
27. How did Liam feel about teacher's explanation?	confused	satisfied
28. Did Toby raise his hand before asking question?	yes	no
29. Was it the last package of the day?	yes	no
30. Was the package light?	yes	no

31. Did Mathew wear new glasses?	yes	no
32. What did Edward spend his money on?	bag	glasses
33. Where was Hannah walking to?	her class	her home
34. Where were Evelyn's books?	on the floor	in her locker
35. What did Amelia paint?	the fence	the wall
36. Did Emilia go to the hardware store?	yes	no
37. Where was Andrew about to enter?	library	school
38. What was stolen in the library?	a computer	a wallet
39. Were there anybody in the building?	yes	no
40. Were there any cracks on the walls after the earthquake?	yes	no
41. Did James miss the bus?	yes	no
42. Did Logan leave home on time?	yes	no
43. Where did David ride his bike to?	school	forest
44. How did Felix go to the forest?	by bike	on foot

45. Was the library crowded?	yes	no
46. Who touched Albert's shoulder?	a teacher	a student
47. What did Anna use to record the video?	a camera	a phone
48. Did Rose dance?	yes	no
49. What did the cat hear?	a dog	a bird
50. Was the cat outside?	yes	no
51. What was Sarah drinking?	coffee	tea
52. Was Nancy watching a film?	yes	no
53. What did Ashley put in her suitcase?	clothes	books
54. What was Martha looking for?	her clothes	her passport
55. Did Claire go out?	yes	no
56. How was Jessie feeling?	bored	excited
57. What was Lola reading?	a novel	a poem
58. Whose photo did Emma see?	an author	a director

59. Was Maria wearing shoes?	yes	no
60. Where did Eliza put the present?	in a cabinet	on a table
61. Where was the woman?	church	school
62. Did the woman speak with the priest?	yes	no
63. What did Susana use to sweep the floor?	brush	broom
64. Why was Jasmine decorating the room?	for a wedding	for a party

APPENDIX D

ANOVA SUMMARY TABLES FOR PRELIMINARY ANALYSES

Table D1. ANOVA Summary Table for the Effects of WM and Proficiency Level on Vocabulary Scores

Effects	*df*	*SS*	*MS*	*F*	*p*
WM level	1	3.33	3.33	0.60	.442
proficiency	1	93.63	93.63	16.87	. < .001
frequency level	1	32.03	32.03	22.71	. < .001
WM level x proficiency	1	0.13	0.13	0.02	.877
WM level x frequency level	1	2.13	2.13	1.51	.224
proficiency x frequency level	1	7.50	7.50	5.32	.025
WM level x proficiency x frequency level	1	3.33	3.33	2.36	.130
Error / Between	56	310.87	5.55		
Error / Within	56	79.00	1.41		

Table D2. ANOVA Summary Table for Comprehension Scores

Predictor	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	5.21	5.21	1.91	.173
WM capacity level	1	2.41	2.41	0.88	.352
Context	1	122.01	122.01	59.93	< .001
Proficiency x WM capacity level	1	3.67	3.67	1.35	.251
Proficiency x Context	1	0.68	0.68	0.33	.567
WM capacity level x Context	1	2.41	2.41	1.18	.281
Proficiency x WM capacity level x Context	1	2.41	2.41	1.18	.281
Error / Between	56	152.80	2.73		
Error / Within	56	114.00	2.06		

Table D3. ANOVA Summary Table for Reading Duration of Passages

Effects	*df*	*SS*	*MS*	*F*	*p*
(Intercept)	1	12242149706.64	12242149706.64	792.24	< .001
Proficiency	1	96228235.38	96228235.38	6.23	.016
WM level	1	9424558.06	9424558.06	0.61	.438
Context	1	70405653.91	70405653.91	90.11	< .001
Proficiency x WM level	1	8339062.41	8339062.41	0.54	.466
Proficiency x Context	1	43130.13	43130.13	0.06	.815
WM level x Context	1	94252.13	94252.13	0.12	.730
Proficiency x WM level x Context	1	135058.40	135058.40	0.17	.679
Error / Between	56	865342807.30	15452550		
Error / Within	56	43755998.99	781357.10		

Table D4. ANOVA Summary Table for Response Times (in ms) to Comprehension Questions in Eye-Tracking Task

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	2929694.63	2929694.63	5.66	.021
WM level	1	526267.18	526267.18	1.02	.318
Context	1	1239274.42	1239274.42	28.31	< .001
Proficiency x WM level	1	18255.51	18255.51	0.04	.852
Proficiency x Context	1	76011.69	76011.69	1.74	.193
WM level x Context	1	1277.56	1277.56	0.03	.865
Proficiency x WM level x Context	1	443.02	443.02	0.01	.920
Error / Between	56	29001127.17	517877.3		
Error / Within	56	2451570.41	43778.04		

Table D5. ANOVA Summary Table for Accuracy Scores of Complex Span Tasks

Effects	*df*	*SS*	*MS*	*F*	*p*
(Intercept)	1	50.62	50.62	18694.90	< .001
WM level	1	0.01	0.01	2.31	.134
proficiency	1	0.00	0.00	0.51	.478
WM level x proficiency	1	0.01	0.01	5.16	.027
Error	56	0.15	0.01		

Table D6. ANOVA Summary Table for Complex Span Task Scores

Effects	*df*	*SS*	*MS*	*F*	*p*
(Intercept)	1	284694.82	284694.82	5039.91	< .001
proficiency	1	120.42	120.42	2.13	.150
WM level	1	5510.42	5510.42	97.55	< .001
proficiency x WM level	1	40.02	40.02	0.71	.404
Error	56	3163.33	56.49		

Table D7. ANOVA Results for Fixation Probability on Pre-Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	< .01	< .01	0.87	0.354
WM level	1	< .01	< .01	0.06	0.812
Context	1	< .01	< .01	4.09	0.048
Proficiency x WM level	1	< .01	< .01	2.82	0.099
Proficiency x Context	1	< .01	< .01	0.72	0.399
WM level x Context	1	< .01	< .01	1.39	0.243
Proficiency x WM level x Context	1	< .01	< .01	0.68	0.411
Error / Between	56	0.01	< .01		
Error / Within	56	< .01	< .01		

Table D8. ANOVA Results for First Fixation Duration on Pre-Target Region.

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	39489.87	39489.87	8.12	0.006
WM level	1	763.81	763.81	0.16	0.693
Context	1	1906.03	1906.03	2.44	0.124
Proficiency x WM level	1	130.6	130.6	0.03	0.870
Proficiency x Context	1	859.68	859.68	1.10	0.298
WM level x Context	1	2816.04	2816.04	3.61	0.063
Proficiency x WM level x Context	1	291.02	291.02	0.37	0.544
Error / Between	56	272203.6	4860.779		
Error / Within	56	43695.79	780.282		

Table D9. ANOVA Results for First Run Duration on Pre-Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	68533.25	68533.25	8.18	0.006
WM level	1	2180.8	2180.8	0.26	0.612
Context	1	3398.02	3398.02	4.03	0.050
Proficiency x WM level	1	67.69	67.69	0.01	0.929
Proficiency x Context	1	1180.48	1180.48	1.4	0.242
WM level x Context	1	944.65	944.65	1.12	0.295
Proficiency x WM level x Context	1	346.8	346.8	0.41	0.524
Error / Between	56	469170.9	8378.05		
Error / Within	56	47269.8	844.10		

Table D10. ANOVA Results for Fixation Probability on Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	< .01	< .01	2.01	0.162
WM level	1	< .01	< .01	1.69	0.199
Context	1	< .01	< .01	21.17	< .001
Proficiency x WM level	1	< .01	< .01	4.33	0.042
Proficiency x Context	1	< .01	< .01	0.06	0.801
WM level x Context	1	< .01	< .01	0.30	0.587
Proficiency x WM level x Context	1	< .01	< .01	2.16	0.147
Error / Between	56	< .01	< .01		
Error / Within	56	< .01	< .01		

Table D11. ANOVA Results for First Fixation Duration on Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	1307.01	1307.01	0.55	0.462
WM level	1	875.31	875.31	0.37	0.547
Context	1	1935.28	1935.28	4.68	0.035
Proficiency x WM level	1	28.31	28.31	0.01	0.914
Proficiency x Context	1	1181.07	1181.07	2.86	0.096
WM level x Context	1	99.07	99.07	0.24	0.626
Proficiency x WM level x Context	1	50.74	50.74	0.12	0.727
Error / Between	56	133551.9	2384.85		
Error / Within	56	23135.86	413.14		

Table D12. ANOVA Results for First Run Duration on Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	5365.05	5365.05	1.14	0.291
WM level	1	4487.34	4487.34	0.95	0.333
Context	1	5350.85	5350.85	9.42	0.003
Proficiency x WM level	1	439.4	439.4	0.09	0.761
Proficiency x Context	1	1609.67	1609.67	2.83	0.098
WM level x Context	1	2.18	2.18	< .001	0.951
Proficiency x WM level x Context	1	132.3	132.3	0.23	0.631
Error / Between	56	263957	4713.51		
Error / Within	56	31817.38	568.16		

Table D13. ANOVA Results for Fixation Probability on Post-Target Word

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	< .01	< .01	5.64	0.021
WM level	1	< .01	< .01	0.42	0.519
Context	1	< .01	< .01	27.95	< .001
Proficiency x WM level	1	< .01	< .01	0.26	0.615
Proficiency x Context	1	< .01	< .01	0.07	0.789
WM level x Context	1	< .01	< .01	0.09	0.768
Proficiency x WM level x Context	1	< .01	< .01	0.43	0.516
Error / Between	56	0.01	0.000179		
Error / Within	56	< .01	< .01		

Table D14. ANOVA Results for First Fixation Duration on Post-Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	3776.86	3776.86	1.56	0.216
WM level	1	2410.63	2410.63	1.00	0.322
Context	1	211.59	211.59	0.37	0.547
Proficiency x WM level	1	1145.66	1145.66	0.47	0.494
Proficiency x Context	1	78.36	78.36	0.14	0.714
WM level x Context	1	901.15	901.15	1.56	0.217
Proficiency x WM level x Context	1	0.01	0.01	< .001	0.997
Error / Between	56	135243.5	2415.06		
Error / Within	56	32343.94	577.57		

Table D15. ANOVA Results for First Run Duration on Post-Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	101986.7	101986.7	4.31	0.043
WM level	1	4561.41	4561.41	0.19	0.662
Context	1	38068.65	38068.65	14.38	< .001
Proficiency x WM level	1	24027.58	24027.58	1.01	0.318
Proficiency x Context	1	0.39	0.39	< .001	0.99
WM level x Context	1	417.08	417.08	0.16	0.693
Proficiency x WM level x Context	1	3843.07	3843.07	1.45	0.233
Error / Between	56	1325716	23673.5		
Error / Within	56	148253.1	2647.37		

Table D16. ANOVA Results for Fixation Probability on Final Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	< .01	< .01	1.67	0.201
WM level	1	< .01	< .01	0.25	0.622
Context	1	< .01	< .01	10.00	0.003
Proficiency x WM level	1	< .01	< .01	2.79	0.100
Proficiency x Context	1	< .01	< .01	0.63	0.431
WM level x Context	1	< .01	< .01	0.08	0.783
Proficiency x WM level x Context	1	< .01	< .01	0.41	0.524
Error / Between	56	0.03	0.000536		
Error / Within	56	< .01	< .01		

Table D17. ANOVA Results for First Fixation Duration on Final Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	139.35	139.35	0.02	0.898
WM level	1	2687.94	2687.94	0.32	0.575
Context	1	13032.55	13032.55	24.66	< .001
Proficiency x WM level	1	24868.8	24868.8	2.95	0.092
Proficiency x Context	1	1548.01	1548.01	2.93	0.093
WM level x Context	1	39.96	39.96	0.08	0.784
Proficiency x WM level x Context	1	70.25	70.25	0.13	0.717
Error / Between	56	472545.6	8438.31		
Error / Within	56	29601.41	528.59		

Table D18. ANOVA Results for First Run Duration on Final Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	333.85	333.85	0.01	0.905
WM level	1	3602.89	3602.89	0.16	0.695
Context	1	27062.85	27062.85	16.75	< .001
Proficiency x WM level	1	69133.5	69133.5	2.99	0.089
Proficiency x Context	1	694.05	694.05	0.43	0.515
WM level x Context	1	643.02	643.02	0.40	0.531
Proficiency x WM level x Context	1	14.63	14.63	0.01	0.925
Error / Between	56	1295633	23136.3		
Error / Within	56	90460.15	1615.36		

Table D19. ANOVA Results for Regression Path Duration of Pre-Target Word

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	153208.9	153208.9	4.58	0.037
WM level	1	61140.16	61140.16	1.83	0.182
Context	1	39863.77	39863.77	5.95	0.018
Proficiency x WM level	1	9250.5	9250.5	0.28	0.601
Proficiency x Context	1	11990.63	11990.63	1.79	0.186
WM level x Context	1	336.15	336.15	0.05	0.824
Proficiency x WM level x Context	1	29241.24	29241.24	4.37	0.041
Error / Between	56	1871489	33419.45		
Error / Within	56	375059.3	6697.48		

Table D20. ANOVA Results for Selective Regression Path Duration of Pre-Target Word

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	82253.31	82253.31	8.36	0.005
WM level	1	2877.83	2877.83	0.29	0.591
Context	1	4890.03	4890.03	4.72	0.034
Proficiency x WM level	1	150.12	150.12	0.02	0.902
Proficiency x Context	1	564.28	564.28	0.54	0.464
WM level x Context	1	1743.98	1743.98	1.68	0.200
Proficiency x WM level x Context	1	673.47	673.47	0.65	0.424
Error / Between	56	551222.8	9843.26		
Error / Within	56	58021.26	1036.09		

Table D21. ANOVA Results for Re-Reading Duration of Pre-Target Word

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	10945.49	10945.49	0.66	0.422
WM level	1	37488.68	37488.68	2.25	0.140
Context	1	16829.97	16829.97	4.00	0.050
Proficiency x WM level	1	7043.75	7043.75	0.42	0.519
Proficiency x Context	1	17757.25	17757.25	4.22	0.045
WM level x Context	1	3611.46	3611.46	0.86	0.358
Proficiency x WM level x Context	1	21039.35	21039.35	5.00	0.029
Error / Between	56	934922.4	16695.04		
Error / Within	56	235479	4204.98		

Table D22. ANOVA Results for Regression Path Duration of Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	14550.27	14550.27	0.46	0.499
WM level	1	317.89	317.89	0.01	0.920
Context	1	41060.75	41060.75	4.19	0.045
Proficiency x WM level	1	7762.22	7762.22	0.25	0.621
Proficiency x Context	1	47374.55	47374.55	4.84	0.032
WM level x Context	1	31781.01	31781.01	3.24	0.077
Proficiency x WM level x Context	1	1180.09	1180.09	0.12	0.73
Error / Between	56	1761855	31461.70		
Error / Within	56	548565.6	9795.81		

Table D23. ANOVA Results for Selective Regression Path Duration of Target Word

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	13868.84	13868.84	2.05	0.157
WM level	1	2579.61	2579.61	0.38	0.539
Context	1	12663.94	12663.94	16.26	< .001
Proficiency x WM level	1	264.59	264.59	0.04	0.844
Proficiency x Context	1	356.43	356.43	0.46	0.501
WM level x Context	1	282.13	282.13	0.36	0.55
Proficiency x WM level x Context	1	3.65	3.65	0.12	0.946
Error / Between	56	378214.8	6753.83		
Error / Within	56	43607.44	778.70		

Table D24. ANOVA Results for Re-Reading Time of Target Word

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	56830.05	56830.05	2.44	0.124
WM level	1	1086.38	1086.38	0.05	0.830
Context	1	8118.08	8118.08	1.10	0.299
Proficiency x WM level	1	10893.03	10893.03	0.47	0.497
Proficiency x Context	1	39512.55	39512.55	5.35	0.024
WM level x Context	1	26074.32	26074.32	3.53	0.066
Proficiency x WM level x Context	1	1052.43	1052.43	0.14	0.707
Error / Between	56	1306934	23338.11		
Error / Within	56	413863.6	7390.42		

Table D25. ANOVA Results for Regression Path Duration of Post-Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	25095.15	25095.15	0.02	0.885
WM level	1	363058.4	363058.4	0.30	0.584
Context	1	4381863	4381863	33.52	< .001
Proficiency x WM level	1	597113.2	597113.2	0.50	0.483
Proficiency x Context	1	7637.56	7637.56	0.06	0.81
WM level x Context	1	11466.69	11466.69	0.09	0.768
Proficiency x WM level x Context	1	99887.72	99887.72	0.76	0.386
Error / Between	56	66940384	1195364.00		
Error / Within	56	7319828	130711.2		

Table D26. ANOVA Results for Selective Regression Path Duration of Post Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	52139.34	52139.34	1.22	0.273
WM level	1	10011.85	10011.85	0.23	0.63
Context	1	219684.4	219684.4	33.6	< .001
Proficiency x WM level	1	5403.98	5403.98	0.13	0.723
Proficiency x Context	1	3998.83	3998.83	0.61	0.437
WM level x Context	1	869.58	869.58	0.13	0.717
Proficiency x WM level x Context	1	3481.34	3481.34	0.53	0.469
Error / Between	56	2387333	42630.94		
Error / Within	56	366141.1	6538.23		

Table D27. ANOVA Results for Re-Reading Time of Post-Target Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	149579.3	149579.3	0.16	0.688
WM level	1	252490.3	252490.3	0.27	0.602
Context	1	2639277	2639277	290	< .001
Proficiency x WM level	1	716126.8	716126.8	0.78	0.381
Proficiency x Context	1	583.55	583.55	0.01	0.936
WM level x Context	1	6020.83	6020.83	0.07	0.798
Proficiency x WM level x Context	1	66073.33	66073.33	0.73	0.398
Error / Between	56	51418153	918181.30		
Error / Within	56	5097332	91023.78		

Table D28. ANOVA Results for Regression Path Duration of Final Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	5730640.58	5730640.58	2.13	0.15
WM level	1	581033.88	581033.88	0.22	0.644
Context	1	13330021.04	13330021.04	47.44	< .001
Proficiency x WM level	1	1406702.04	1406702.04	0.52	0.473
Proficiency x Context	1	34861.93	34861.93	0.12	0.726
WM level x Context	1	12895.49	12895.49	0.05	0.831
Proficiency x WM level x Context	1	63772.15	63772.15	0.23	0.636
Error / Between	56	150887782.1	2694424.68		
Error / Within	56	15736250.12	281004.46		

Table D29. ANOVA Results for Selective Regression Path Duration of Final Region

Effects	*df*	*SS*	*MS*	*F*		*p*
Proficiency	1		7207.5	7207.5	0.16	0.693
WM level	1		4911.2	4911.2	0.11	0.744
Context	1		136143.82	136143.82	38.52	< .001
Proficiency x WM level	1		107591.64	107591.64	2.35	0.131
Proficiency x Context	1		481.25	481.25	0.14	0.714
WM level x Context	1		2593.54	2593.54	0.73	0.395
Proficiency x WM level x Context	1		3041.39	3041.39	0.86	0.358
Error / Between	56		2559474.58	45704.90		
Error / Within	56		197945.45	3534.74		

Table D30. ANOVA Results for Re-Reading Time of Final Region

Effects	*df*	*SS*	*MS*	*F*	*p*
Proficiency	1	5013995.89	5013995.89	2.17	0.146
WM level	1	453443.44	453443.44	0.20	0.659
Context	1	10940535.48	10940535.48	42.67	< .001
Proficiency x WM level	1	721791.78	721791.78	0.31	0.578
Proficiency x Context	1	55018.08	55018.08	0.21	0.645
WM level x Context	1	6630.07	6630.07	0.03	0.873
Proficiency x WM level x Context	1	42366.99	42366.99	0.17	0.686
Error / Between	56	129370367.1	2310185.12		
Error / Within	56	14359136.32	256413.14		

REFERENCES

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.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3-27.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics, 29*, 175-193.

Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly, 45*(2), 235-266.

Alptekin, C., Erçetin, G., & Özdemir, O. (2014). Effects of variations in reading span task design on the relationship between working memory capacity and second language reading. *The Modern Langauge Journal, 98*(2), pp. 536-552.

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences, 93*(24), 13468-13472.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communicative Disorders, 36*, 189-208.

Baddeley, A. (2017). Modularity, working memory and language acquisition. *Second Language Research, 33*(3), 299-311.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower, *The psychology of learning and motivation* (pp. 47-89). San Diego, California: Academic Press.

Bernhardt, E. B. (1991). *Reading development in a second language: Theorethical, empirical, & classroom perspectives.* Norwood, NJ: Ablex Publishing Corporation.

Bernhardt, E. B. (2011). *Understanding advanced second language reading.* New York: Routledge.

Bogazici University School of Foreign Languages. (2017). *YADYOK Ogrenci El Kitabi*. Retrieved 2017- 01-11 from YADYOK: http://www.yadyok.boun.edu.tr/birim/ogrenci-el-kitabi.htm#12

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology/Revue canadienne de psychologie, 22*(5), 349-368.

Buchweitz, A., Mason, R. A., Hasegawa, M., & Just, M. A. (2009). Japanese and English sentence reading comprehension and writing systems: An fMRI study of first and second language effects on brian activation. *Bilingualism: Language and Cognition, 12*(2).

Budd, D., Whitney, P., & Turley, K. J. (1995). Individual differences in working memory strategies for reading expository text. *Memory and Coginition, 23*(6), pp. 735-748.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal, 11*, 489-503.

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory, 9*(4/5/6), pp. 365-381.

Calvo, M. G. (2004). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learning and Individual Differences, 15*, 53-65.

Calvo, M. G., & Castillo, M. D. (1998). Predictive inferences take time to develop. *Psychological Research, 61*(4), pp. 249-260.

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: Processing time during versus after the priming context. *Memory and Cognition, 27*(5), 834-843.

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research, 65*(3), pp. 158-169.

Carpenter, P. A., & Daneman, M. (1981). Lexical retrieval and error recovery in reading: A model based on eye fixaions. *Journal of verbal learning and verbal behavior, 20*, 137-160.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr, & K. Kotovsky, *Complex information processing: The impact of Herbert A. Simon* (pp. 31-68). N.J.: Erlbaum Associates.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension. In M. A. Gernsbacher, *Handbook of Psycholinguistics* (pp. 1075-1122). San Diego: Academic Press.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*(5), 867-873.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition, 7*(1), 35-41.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill, *Eye Movements: A Window on Mind and Brain* (pp. 341-371). Boston: Elsevier.

Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, & R. R. Jordan, *Reading in a second language* (pp. 5-12). Rowley, MA: Newbury House.

Conway, A. R., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4*(6), 577-590.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Conway, A. R., 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.

Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language, 44*(2), 220-234.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraint withing the human information-processing system. *Psychological Bulletin, 104*(2), 163-191.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake, & P. Shah, *Models of Working Memory* (pp. 62-101). Cambridge, UK: Cambridge University Press.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior, 19*(4), 450-466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 561-584.

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology, 17*(4), 367-378.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

DeKeyser, R. M. (1997). Beyond explicit rule learning. *Studies in Second Language Acquisition, 19*, 195-221.

Ehri, L. C. (1996). Development of the ability to read words. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson, *Handbook of Reading Research, Vol. 2* (pp. 383-418). New Jersey: Lawrence Erlbaum Associates.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 75-87.

Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye-movements in virtual reality. *Behavioral Research, 50*(3), 1102-1115.

Ellis, N. C. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49A*(1), 234-250.

Elman, J. L. (1990). Representation and structure in connectionist models. In G. T. Altmann, *Cognitive models of speech processing: Psycholoinguistic and computational perspectives* (pp. 345-382). Cambridge, MA: MIT Press.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences şn working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972-992.

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 prefrontal cortex. In A. Miyake, & P. Shah, *Models of Working Memory* (pp. 102-134). Cambridge, UK: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Erçetin, G. (2015). Working memory and L2 reading: Theorethical and methodological issues. *ELT Research Journal, 4*(2), 101-110.

Erçetin, G., & Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied Psycholinguistics, 34*(4), 727-753.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., & Reddy, D. R. (1980). The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys, 12*(2), 213-253.

Estevez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. *Memory, 8*(1), 51-61.

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16*(1-2), 99-124.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), pp. 981-995.

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predicitive inferences. *Discourse Processes, 22*(3), 225-246.

Fincher-Kiefer, R., & D'Agostino, P. R. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes, 37*(3), 205-224.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition, 43*(2), 226-236.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*, 463-487.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General, 129*(1), 61-83.

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *The Modern Langauge Journal, 89*(1), 54-75.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a casual connection? *Journal of Memory and Language, 29*, 336-360.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Furher

evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*, 65-77.

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. (1992). Do readers mentally represent characters' emotional states? *Cognition & Emotion, 6*(2), 89-111.

Geva, E., & Ryan, E. B. (1993). Linguistic and cognitive correlates of academic skills in first and second language. *Languge Learning, 43*(1), 5-42.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Literacy Research and Instruction, 6*(4), 126-135.

Gough, P. B. (1972). One second of reading. *Visible Language, 6*(4), pp. 291-320.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), pp. 371-395.

Handley, S. J., Capon, A., Copp, C., & Harper, C. (2002). Conditional reasoning and the Tower of Hanoi: The role of spatial and verbal working memory. *British Journal of Psychology, 93*, 501-518.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Haviland, S. E., & 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.

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2015). On forward inferences of fast and slow readers. An eye-movement study. *Scientific Reports, 5*, 8432-.

Horiba, Y., & van den Broek, P. W. (1993). Second language readers' memory for narrative texts: Evidence for structure-preserving tod-down processing. *Language Learning, 43*(3), 345-372.

Indararathne, B., & Kormos, J. (2017). The role of working memory in L2 input: Insights from eye-tracking. *Bilingualism: Language and Cognition*, 1-20.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Pscyhology: Learning, Memory, and Cognition, 43*(4), 635-652.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137-166.

Juola, J. F., Schadler, M., Chabot, R. J., & McCaughey, M. W. (1978). The development of visual information processing skills related to reading. *Journal of Experimental Child Psychology, 25*(3), 459-476.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), pp. 329-354.

Just, M. A., & Carpenter, P. A. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. *Psychological Review, 99*(1), pp. 122-149.

Just, M. A., Carpenter, P. A., & Wolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228-238.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32*(4), pp. 446-463.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32*(4), 446-463.

Keenan, J. M., & Jennings, T. M. (1995). The role of word-based priming in inference research. In R. F. Lorch, & J. O. O'Brien, *Sources of coherence in reading* (pp. 37-50). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F., & Salthouse, T. A. (1999). Eight questions and some general issues. In A. Miyake, & P. Shah, *Models of Working Memory* (pp. 412-441). Cambridge, UK: Cambridge University Press.

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General, 112*(2), 240-264.

Klin, C. M., Murray, J. D., Levine, W. H., & Guzman, A. E. (1999). Forward inferences: From activation to long-term memory. *Discourse Processes, 27*(3), 241-260.

Koda, K. (2004). *Insights into second language reading.* New York: Cambridge University Press.

Kormos, J., & Safar, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261-271.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229-270.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working Memory and Second Language Comprehension and Production: A meta-analysis. *Psychonomic Bulletin & Review,, 21*(4), pp. 861-883.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and casual text constraints. *Discourse Processes, 34*(3), pp. 259-280.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual Differences in the Time Course of Inferential Processing. *Journal of Experimental Psychology: Lerning, Memory, and Cognition, 20*(6), pp. 1456-1470.

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.

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discoure Processes, 16*(1-2), 35-53.

Mastrantuono, E., Saldana, D., & Rodriguez-Ortiz, I. R. (2019). Inferencing in deaf adolescents during sign-supported speech comprehension. *Discourse Processes, 56*(4), 363-383.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578-586.

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior, 20*, 671-682.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(1), 82-91.

McKoon, G., & Ratcliff, R. (1989). Assesssing the occurence of elaborative inference with recognition: Compatibility checking versus compound cue theory. *Journal of Memory and Language, 28*(5), 547-563.

McKoon, G., & Ratcliff, R. (1989). Semantic associations and elaborative inference. *Journal of Experimental Psychology, 15*(2), 326-338.

McKoon, G., & Ratcliff, R. (1992). Inferences During Reading. *Psychological Review, 99*(3), pp. 440-466.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33*(2), 135-158.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

Millis, K. K., & Graesser, A. C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. *Journal of Memory and Language, 33*(5), 583-599.

Miyake, A., & Shah, P. (1999). Emerging general consensus, unresolved theorethical issues, and future research directions. In A. Miyake, & P. Shah, *Models of Working Memory* (pp. 442-481). New York, NY: Cambridge University Press.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33*(2), 175-202.

Morrell, R. W., & Park, D. C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging, 8*(3), 389-399.

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes, 35*(2), 81-102.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language, 32*(4), 464-473.

Noordman, L. G., Vonk, W., & Kempff, H. J. (1992). Causal inferences during reading of expository texts. *Journal of Memory and Language, 31*, 573-590.

O'Brian, E. J., Duffy, S. A., & Myers, J. L. (1986). Anaphoric inferences during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 346-352.

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative Inferences During Reading: Do They Occur On-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), pp. 410-420.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*, 377-402.

O'Regan, K. J. (1990). Eye movements and reading. In E. Kowler, *Eye movements and their role in visual and cognitive processes* (Vol. 4, pp. 395-453). Elsevier Science Limited.

Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society, 31*(2), 117-118.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27*(4), 399-415.

Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning, 61*(1), 187-218.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition, 5*(4), 443-448.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity, and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*(3), 191-201.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1188-1200.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30*(3), 333-357.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic, *Attention and Performance 6* (pp. 573-603). Hillsdale, NJ: Erlbaum.

Samuels, J. S. (2004). Toward a theory of automatic information processing in reading, revisited. In R. B. Ruddell, & N. J. Unrau, *Theorethical Models and Processes of Reading* (pp. 1127-1148). Newark, USA: International Reading Association.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33*(2), 105-132.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Atention Perception Psychophysics, 74*, 5-35.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty, & M. H. Long, *The Handbook of Second Language Acquisition* (pp. 382-408). Malden, MA: Blackwell Publishing.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 269-385.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled preformance, practice, the differentiation of speed-up automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369-385.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll, & A. M. DeGroot, *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). New York: Oxford University Press.

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.

Seifert, C. M., Abelson, R. P., McKoon, G., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(2), 220-231.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly journal of experimental psychology, 45A*(1), 21-50.

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. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior, 22*, 437-448.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*(1), 11-20.

Smith, F. (1971). *Understanding Reading.* New York: Holt, Rinehart and Winston.

SR Research Experiment Builder 2.1.1. (2017). Mississauga, Ontario, Canada: SR Research Ltd.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32-71.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition, 16*(4), 283-298.

Tools, P. S. (2016). *E-Prime*. (Psychology Software Tools, Inc. [E-Prime 3.0]) Retrieved from Psychology Software Tools: http://www.pstnet.com

Turner, M. L., & Engle, R. W. (1986). Working memeroy capacity. *Proceedings of the Human Factor Society Annual Meeting, 30*(13), 1273-1277.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent. *Journal of Memory and Language, 28*, 127-154.

Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning, 51*(2), 257-280.

van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner, *Comprehension Processes in Reading* (pp. 423-446). NY: Lawrence Erlbaum Associates.

van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*(4), 289-296.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Virtue, S., Schutzenhofer, M., & Tomkins, B. (2017). Hemispheric processing of predictive inferences during reading: The influence of negatively emotional valenced stimuli. *Laterality: Asymmetries of Body, Brain and Cognition, 22*(4), 455-472.

Walter, C. (2004). Transfer of reading comprehension skills to L2 in linked to mental representations of text and to L2 working memory. *Applied Linguistics, 25*(3), 315-339.

Wang, L., Hagoort, P., & Jensen, O. (2017). Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience, 30*(3), 432-447.

Wang, L., Hagoort, P., & Jensen, O. (2018). Gamma oscillatory activity related to language prediction. *Journal of Cognitive Neuroscience, 30*(8), 1075-1085.

Weber, R.-M. (1970). First graders' use of grammatical context in reading. In H. Levin, & J. T. Williams, *Basic Studies in Reading.* New York: Basic.

West, M. P. (1953). *A general service list of English words with semantic frequencies and a supplementary word-list for the writing of popular sicence and technology.* Addison-Wesley Longman Limited.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). The influence of parafoveal word length and contextual constraints on fixation durations and word skipping in reading. *Psychonomic Bulletin & Review, 12*(3), 466-471.

Whitney, P., Ritchie, B. G., & Crane, R. S. (1992). The effect of foregrounding on readers' use of predictive inferences. *Memory and Cognition, 20*(4), 424-432.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex, 26*, 2506-2516.

YADYOK Öğrenci El kitabı. (2019). Retrieved from School of Foreing Languages Bogazici University: http://yadyok.boun.edu.tr/birim/ogrenci-el-kitabi.htm#12