THE INTERPLAY BETWEEN WORKING MEMORY
AND TOPIC FAMILIARITY IN L2 READING COMPREHENSION

By Jihye Shin

A Dissertation
Submitted in Partial Fulfillment
of the Requirement for the Degree of
Doctor of Philosophy
in Applied Linguistics

Northern Arizona University
May, 2020

Approved:
Dr. Fredricka L. Stoller (Co-chair)
Dr. Vedran Dronjic (Co-chair)
Dr. Soo Jung Youn
Dr. Yuly Asención-Delaney
Dr. William Grabe
ABSTRACT

THE INTERPLAY BETWEEN WORKING MEMORY AND TOPIC FAMILIARITY IN L2 READING COMPREHENSION

JIHYE SHIN

This study investigated whether and how working memory (WM) and topic familiarity combine to facilitate second language (L2) reading comprehension to elucidate the complex dynamics of individual differences in L2 reading. It also examined the distinctive contributions of WM subcomponents (e.g., storage, processing) to L2 reading comprehension with and without topic familiarity. A total of 158 university students in Japan ($n = 78$) and Korea ($n = 80$) studying English as an L2 participated. The current study observed participants’ performance under two conditions, reading with and without topic familiarity, which were purposefully created by the researcher. The participants completed tasks that assessed topic familiarity, WM capacity, L2 knowledge, and L2 reading comprehension in a two-phase administration of measures.

Results of mixed-effects modeling showed that there is an interaction between WM and topic familiarity in L2 reading comprehension, even when L2 knowledge was accounted for. Participants benefitted differently from the provision of topic familiarity due to constraints imposed by WM. In the presence of topic familiarity, WM amplified the effect of topic familiarity on reading, which resulted in L2 readers with high WM taking advantage of topic familiarity to a greater extent.

When closely examined, the extent to which participants were able to benefit from familiarity with the topic was linked to their processing capacity rather than storage.
capacity. The participants’ capacity for storing and maintaining information did not seem to contribute to their ability to take advantage of topic familiarity, nor was it predictive of reading comprehension. Methodologically, the results highlight the need for WM measures to tap processing because of its unique role in L2 reading comprehension.

The findings provide a nuanced understanding of the effect of topic familiarity on L2 reading comprehension and illustrate the role of working memory in L2 readers’ effective use of topic familiarity. Implications for and contributions to pedagogical practices regarding pre-reading stages of lessons are explored. Directions for future research are also proposed.
Acknowledgements

I would like to express my heartfelt gratitude to those who have contributed to the completion of easily the most rewarding journey of my life.

First and foremost, I would like to thank my dissertation committee. My co-chairs, Fredricka Stoller and Vedran Dronjic, are my twin pillars without whom none of this would have been possible. Fredricka is the most generous, encouraging, and inspiring mentor that one can ever dream of. Her thoughtful and thorough feedback has touched every single page of this dissertation. Her undying support and guidance lifted me up when I was down and brought me where I am today. I am proud to say that her insights continue to guide my perspectives on second language reading and that her dedication to teaching has shaped my identity and goals as a teacher. I aspire to emulate her devotion, passion, and kindness in all of my future endeavors.

Vedran has taken me on a wild tour of the enthralling world of psycholinguistics and has taught me how to demystify complex concepts. His encouragement and insightful comments are what made this dissertation what it is today. His constant support and unfailing belief in me have allowed me to venture out and develop my identity as a researcher. I appreciate him being on this journey with me since the inception of the pilot study almost 3 years ago. He is also the only person I know who will respond to my email at 3 a.m. in the morning, both of us being night owls.

I would like to express my gratitude to Soo Jung Youn for offering unceasing help and sharing her insights into and expertise in assessment. She is one of the kindest people I know. I am grateful to Yuly Asención-Delaney for giving me helpful comments and support in addition to sharing her wisdom with me. I would like to thank Bill Grabe for
sharing his knowledge of reading that is without parallel and for encouraging me to read widely and look for connections among areas and fields. I would like to extend my thanks to the faculty in the Applied Linguistics Ph.D. program at NAU for their support throughout the years.

I would also like to thank my parents, Yangja Choi and Sunho Shin, as well as my dear friends and Ph.D student peers: Valeriiia Bogorevich, Panjanit Chaipuapae, Shi Chen, Brett Hashimoto, Daniel Keller, Alyssa Kermad, Elnaz Kia, Garrett Larson, Mark McAndrews, and Tyler True. All of them have given me strength over the years to keep marching ahead. Having their support truly meant the world to me. My special thanks go to Jules Baumgartner who has always been there for me and fed me heavenly dinners on countless occasions. I am eternally grateful for her friendship, our racquetball games, Indian lunches, hair-dying sessions, Thanksgiving dinners, and road trips.

I would like to acknowledge the financial support from a *Language Learning* Dissertation Grant, National Federation of Modern Language Teachers Association (NFMLTA)/ *Modern Language Journal (MLJ)* Dissertation Support Grant, Northern Arizona University Support for Graduate Students Award, and Arizona TESOL.

Finally, words cannot express all that I wish to say to my husband, Jesse Conway, who has been my rock for the past five years. He has believed in me and supported me through all of the ups and all of the downs. From making bacon pancakes in the morning to bringing so much happiness and joy into my life, I appreciate all small and big things he does for me. His positive outlook on life and his unwavering support, patience, and love inspire me every day. I could not have done this without him.
# Table of Contents

Abstract ........................................................................................................................................... ii
Acknowledgements ................................................................................................................................. v
Table of Contents ................................................................................................................................. vii
List of Tables .......................................................................................................................................... x
List of Figures .......................................................................................................................................... xi

Chapter 1: Introduction ......................................................................................................................... 12
  Importance of the Dissertation .............................................................................................................. 14
  Overview of the Dissertation ............................................................................................................... 17

Chapter 2: Literature Review ................................................................................................................ 18
  Working Memory .................................................................................................................................. 18
  Working memory and academic learning ............................................................................................ 20
  Involvement of Working Memory in Reading Comprehension ............................................................. 22
  Measures of Working Memory ............................................................................................................ 25
  Reading span task as a working memory measure .............................................................................. 27
  Relationship between Working Memory and Reading Comprehension ............................................ 29
  Working memory and L2 reading comprehension .............................................................................. 30
  Inconsistent measurement of working memory and its impact on research outcomes .................... 33

  Reading Comprehension and Its Multiple-level Architecture ............................................................ 38
  Background Knowledge in Reading Comprehension ......................................................................... 39
  Combined Effects of Working Memory and Background Knowledge .............................................. 44
  Methodological Inconsistencies across Studies ................................................................................. 51

Proposed Study ...................................................................................................................................... 52

Chapter 3: Method ................................................................................................................................. 54
  Research Design ................................................................................................................................. 54
  Variables of Interest ............................................................................................................................ 55
  Participants ......................................................................................................................................... 58
  Instruments .......................................................................................................................................... 59
  Questionnaire ....................................................................................................................................... 59
  L2 Reading Span Task .......................................................................................................................... 61
  L2 Knowledge Measures .................................................................................................................... 67
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Texts in L1</td>
<td>69</td>
</tr>
<tr>
<td>Topic Familiarity Task</td>
<td>71</td>
</tr>
<tr>
<td>L2 Reading Quiz</td>
<td>72</td>
</tr>
<tr>
<td>Procedures</td>
<td>83</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>85</td>
</tr>
<tr>
<td>Pre-analysis procedures</td>
<td>85</td>
</tr>
<tr>
<td>Data-analysis for RQ 1</td>
<td>87</td>
</tr>
<tr>
<td>Data-analysis for RQ 2</td>
<td>88</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
<td>90</td>
</tr>
<tr>
<td>Study Variables and Measures</td>
<td>90</td>
</tr>
<tr>
<td>Topic Familiarity: Pre-check</td>
<td>90</td>
</tr>
<tr>
<td>Topic Familiarity: Post-check</td>
<td>91</td>
</tr>
<tr>
<td>Working memory</td>
<td>94</td>
</tr>
<tr>
<td>L2 knowledge measures</td>
<td>100</td>
</tr>
<tr>
<td>L2 reading comprehension</td>
<td>101</td>
</tr>
<tr>
<td>Research Question 1: Interaction Between Working Memory and Topic Familiarity</td>
<td>102</td>
</tr>
<tr>
<td>Research Question 2: Extent to which Working Memory Subcomponents Contribute to L2 Reading with and without Topic Familiarity</td>
<td>108</td>
</tr>
<tr>
<td>Interaction between Working Memory and Topic Familiarity in L2 Reading Comprehension: Japanese vs Korean L2 English Learners</td>
<td>110</td>
</tr>
<tr>
<td>Summary of the Chapter</td>
<td>111</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td>114</td>
</tr>
<tr>
<td>Topic Familiarity as a Type of Background Knowledge</td>
<td>114</td>
</tr>
<tr>
<td>Facilitative Effect of Topic Familiarity on L2 Reading Comprehension</td>
<td>115</td>
</tr>
<tr>
<td>Interaction between Working Memory and Topic Familiarity in L2 Reading</td>
<td>118</td>
</tr>
<tr>
<td>Contributions of Working Memory Subcomponents (Storage, Processing Accuracy, Processing Speed) to L2 Reading with and without Topic Familiarity</td>
<td>125</td>
</tr>
<tr>
<td>Pedagogical Implications</td>
<td>135</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>139</td>
</tr>
<tr>
<td>Directions for Future Research</td>
<td>142</td>
</tr>
<tr>
<td>Summary of the Chapter and Closing Remarks</td>
<td>144</td>
</tr>
<tr>
<td>References</td>
<td>146</td>
</tr>
<tr>
<td>Appendices</td>
<td>170</td>
</tr>
</tbody>
</table>
Appendix A: Questionnaire (English) ................................................................. 170
Appendix B: Questionnaire (Korean) ................................................................. 172
Appendix C: Questionnaire (Japanese) ............................................................... 174
Appendix D: Reading Span Task Sentences ..................................................... 176
Appendix E: Instructions for Reading Span Task (English) ............................... 179
Appendix F: Instructions for Reading Span Task (Korean) ............................... 181
Appendix G: Instructions for Reading Span Task (Japanese) ............................ 183
Appendix H: Reading Span Task Practice Sentences .................................... 185
Appendix I: Reading Span Task Sheet ............................................................. 186
Appendix J: C-test (Klein-Braley, 1997) ......................................................... 187
Appendix K: Background Texts (English) ....................................................... 188
Appendix L: Background Texts (Korean) .......................................................... 191
Appendix M: Background Texts (Japanese) ..................................................... 194
Appendix N: Topic Familiarity Check (English) ................................................. 197
Appendix O: Topic Familiarity Check (Korean) ............................................... 198
Appendix P: Topic Familiarity Check (Japanese) ......................................... 199
Appendix Q: Topic Familiarity Rating Protocol ............................................. 200
Appendix R: Practice Quiz .................................................................................. 201
Appendix S: Reading Quiz ................................................................................. 205
Appendix T: Answer Key .................................................................................... 221
List of Tables

Table 2.1. Scoring systems moderating the correlation between working memory and L2 reading .................................................................37
Table 2.2. Summary of studies on the interaction between working memory and background knowledge in L2 reading comprehension ................................47
Table 3.1. Table of item specifications ........................................................................................................................74
Table 3.2. Item analysis on reading comprehension test items............................................................78
Table 3.3. Flagged items.......................................................................................................................................78
Table 3.4. Comparison across passages in the revised reading quiz ..............................................81
Table 3.5. Order and duration of measures..............................................................................................85
Table 4.1. Topic familiarity pre-check ratings..........................................................................................91
Table 4.2. Comparison of topic familiarity ratings in pre- and post-check .........................92
Table 4.3. Scores on written responses at post-check..........................................................93
Table 4.4. Descriptive results of WM subcomponent scores .........................................................96
Table 4.5. Correlations among WM subcomponent scores .................................................................96
Table 4.6. Descriptive statistics of L2 knowledge ..............................................................101
Table 4.7. L2 Reading comprehension test results in two conditions ........................................101
Table 4.8. Correlations among variables of interest ..............................................................103
Table 4.9. Mixed-effects model for RQ1 .........................................................................................105
Table 4.10. Descriptive statistics of low- and high-WM groups ...............................................105
Table 4.11. Mixed-effects model for RQ1 after adding vocabulary size..............................107
Table 4.12. Mixed-effects model for RQ2 after adding vocabulary size..............................110
Table 4.13. Comparison of Japanese and Korean participants for RQ1 .......................111
Table 4.14. Comparison of Japanese and Korean participants for RQ2 .......................111
List of Figures

Figure 3.1. Illustration of the reading span task sequence .............................................65
Figure 4.1. Comparison of pre- and post-checks of topic familiarity self-ratings ........93
Figure 4.2. Scatterplot of processing speed ........................................................................98
Figure 4.3. Working memory subcomponent and composite scores (z-scores) of Japanese
and Korean participants .....................................................................................................98
Figure 4.4. Scatterplot of working memory composite scores ........................................99
Figure 4.5. The interaction between working memory and topic familiarity in L2 reading
comprehension .....................................................................................................................106
Chapter 1: Introduction

Research on predictors, or correlates, of reading comprehension has advanced our understanding of how components of language proficiency (e.g., vocabulary, grammar knowledge) and individual differences (e.g., working memory, background knowledge) influence second language (L2) reading performance (e.g., Carrell, 1987; Chang, Wang, Cai, & Wang, 2019; Jeon & Yamashita, 2014; Schmitt, Jiang, & Grabe, 2011; Shin, Dronjic, & Park, 2019; Shiotsu, 2010). Working memory (WM), a memory system involved in simultaneous storage and processing of current information, has been shown to be an integral structure for many types of learning (Dehn, 2008; Wen & Li, 2019). In particular, research on the learning of reading skills has pointed out that due to the cognitively demanding nature of reading comprehension, WM is a critical individual-difference factor that can distinguish between efficient and inefficient readers (Daneman & Carpenter, 1980; Gathercole & Alloway, 2008; Gathercole & Baddeley, 1993; Dehn, 2008).

Because WM governs cognitive control and allocation of attentional resources, it can affect first language (L1) readers who have even mastered lower-level reading skills (e.g., word recognition and decoding) and automated meaning construction processes (Kendeou, van den Broek, Helder, & Karlsson, 2014). Increasing numbers of studies on the link between WM and reading comprehension have suggested that such findings apply to L2 reading as well (e.g., Alptekin & Erçetin, 2010, 2011; Walter, 2004). L2 readers with more WM capacity have more cognitive resources available for accessing long-term memory and calling upon higher-level conceptual processes of reading, which
enable construction of a meaningful representation of text (Alptekin & Erçetin, 2009; Nassaji, 2002; Wen & Li, 2019).

With the influence of WM on cognitive resources available to readers, there has been increasing attention drawn to how WM may interact with L2 readers’ use of background knowledge in L2 reading (Alptekin & Erçetin, 2011; Leeser, 2007; Shin et al., 2019). The facilitative role of background knowledge on reading comprehension has been illustrated amply (e.g., Brantmeier, Sullivan, & Strube, 2014; Carrell, 1987; Johnson, 1982) and has a long history that dates to the late 1970s with the introduction and application of schema theory (Rumelhart, 1980) to reading comprehension processes. Background knowledge is known to affect the iterative process of the melding of new information to the existing knowledge store in reading comprehension (McNeil, 2012). Drawing on background knowledge, as Grabe (2009) states, is also part of attentional processes involved in reading comprehension. Indeed, it is another crucial individual difference associated with the successful construction of a text representation (Nassaji, 2002).

Readers’ attempts to assimilate layers of linguistic and extralinguistic information with WM regulating and facilitating the process involve both WM and background knowledge. The construction-integration (CI) model (Kintsch, 1998) illustrates that as readers undergo the process of constructing a text representation, all knowledge sources available to readers become activated and are held in WM. Accessing extralinguistic and conceptual knowledge, such as background knowledge, calls for WM as WM is

---

1 The notion of schema, a mental structure in memory acquired from past experiences, was first introduced by Bartlett (1932) in psychology. Through a schema, expectations can be set up for an upcoming event. This idea was later adapted by Rumelhart (1980), among others, to refer to the facilitative role of schemata (i.e., background knowledge) in reading comprehension.
responsible for allocating cognitive resources to long-term memory based data (Alptekin & Erçetin, 2011). As such, the description of how reading comprehension occurs in the CI model points to the cognitively demanding nature of reading as well as a possible effect of WM on readers’ ability to access relevant background knowledge.

**Importance of the Dissertation**

Discrete effects of WM and background knowledge on L2 reading have been widely examined (e.g., for WM and L2 reading, see Chang et al., 2019; Joh, 2015; Medina, Callender, Brantmeier, & Schultz, 2017; Linck et al., 2014; Spencer & Wagner, 2017; for background knowledge and L2 reading, see Auckerman, Brown, Mokhtari, Valencia, Palincsar, 2015; Brantmeier et al., 2014; Mahmoudi & Mahmmoudi, 2017; Tarchi, 2010). However, the significant involvement of WM and background knowledge in the reading process has also led to recent attempts to elucidate the joint effects of these two individual-difference factors. The current study is one of the few that shed light on the complex dynamics of individual differences in L2 reading. This study delves into such dynamics by investigating if and how L2 readers’ ability to benefit from background knowledge of the topic is influenced by their cognitive resources, namely WM. In doing so, it aims to (a) provide a nuanced understanding of the effect of background knowledge on L2 reading comprehension and (b) reveal the role of WM in L2 readers’ effective use of background knowledge.

One source of motivation for pursuing these goals stems from inconclusive findings on the interaction between WM and background knowledge that have been reported thus far. Among the small number of studies in this domain, some have shown a positive but independent relationship between WM and background knowledge (Alptekin
& Erçetin, 2011; Payne, Kalibatseva, & Jungers, 2009) while others have demonstrated an interactive relationship between them (Leeser, 2007; Shin et al., 2019). These studies show discrepancies in terms of their WM scoring procedures (e.g., using storage scores only, using composite scores) and their operationalization of background knowledge (e.g., topic familiarity, domain experience), which may be contributing to the lack of consensus in the findings.

Recognizing the methodological issues among the studies mentioned above, the current study set out to investigate the interaction between WM and background knowledge in L2 reading comprehension. That is, the study sought to address the inconsistent WM scoring procedures used in previous studies to contribute to L2 researchers’ growing efforts to ensure robust and reliable measurement practices. As such, the WM task in this study, adapted from the Daneman & Carpenter, 1980 and Waters & Caplan, 1996 reading span tasks, is designed to produce three sub-scores (storage, processing accuracy, processing speed). From a methodological standpoint, the current study aims to investigate the appropriateness of WM tasks that rely on storage scores only to address concerns that such WM tasks may lead to unreliable results and, more importantly, an inadequate representation of the construct of WM.

In addition, paying specific attention to the WM task and its subcomponents may allow the study’s findings to cast further light upon the nature of the interaction between WM and background knowledge. The current study is the first to examine the extent to which the subcomponents of WM (storage, processing accuracy, processing speed) contribute to L2 readers’ use of background knowledge. This goal of the study is motivated by recent findings: Even within the WM system, storage and processing may
be functionally independent of each other and may play different roles in reading comprehension (Sagarra, 2017).

Continuing on the methodological front, another strength of this dissertation study is that its research instruments for key variables (e.g., WM task, L2 reading test) had been piloted in an earlier version of the study (Shin et al., 2019). The instruments were examined upon completion of the pilot and were then refined for the current study. For instance, the reliability of the instruments was checked, item statistics of the L2 reading comprehension test were reviewed, leading to the revision of test items. By doing so, the study ensured the quality of the instruments and increased trustworthiness of its findings. Also, the current study expanded upon the earlier pilot with the recruitment of participants from one additional L1 group. As a result, the study involved two L1 groups (Japanese and Korean), increasing the generalizability of findings.

Finally, by exploring the complex web of cognitive processes involved in activating and utilizing background knowledge, operationalized as topic familiarity, the study aims to inform L2 teachers about how individual differences in WM may be linked to L2 readers’ efficiency in accessing their topic familiarity, which affects their success at reading comprehension. Findings from the current study, therefore, lead to pedagogical implications in regard to pre-reading instructional activities to support L2 readers’ comprehension of target texts. Ideas can be suggested based on the findings for pre-reading activities that can accommodate L2 readers with low WM. Furthermore, these findings can be used to raise teachers’ awareness of the role of WM as cognitive resources in L2 students’ reading processes when working with topic familiarity.
Overview of the Dissertation

This chapter has described the focus and importance of the research reported in this dissertation. The remainder of the dissertation includes four additional chapters. Chapter 2 (Literature Review) provides a survey of the literature pertaining to the roles of the two individual-difference constructs of interest, WM and background knowledge, operationalized as topic familiarity, in L2 reading comprehension and any possible interplay between the two. The chapter concludes by identifying an area upon which further light can be cast in this domain of research as defined by a set of research questions.

Chapter 3 (Method) describes the methods of investigation including the research design, variables of interest, participants, instruments, procedures, and data analysis. Chapter 4 (Results) reports the results of the two main research questions as well as further analyses. Chapter 5 (Discussion) summarizes and interprets the results reported in Chapter 4. This final chapter concludes with an exploration of pedagogical implications, an accounting of the limitations of the study, and directions for future research.
Chapter 2: Literature Review

This chapter provides a survey of the literature pertaining to the roles of the two individual-difference constructs of interest – working memory (WM) and topic familiarity – in L2 reading comprehension and any possible interplay between the two. The chapter first presents a brief overview of the concept of WM as well as a review of the research on the involvement of WM in reading to establish its significant role in reading comprehension. Measures of WM are discussed next with the focus on the reading span task, leading into a review of relevant studies that provide empirical support for the relationship between WM and L2 reading comprehension. Subsequently, the literature on background knowledge in L2 reading comprehension is reviewed, followed by a discussion of empirical findings on the combined effects of WM and background knowledge in L2 reading comprehension. The chapter concludes by identifying areas upon which further light can be cast in this domain of research as defined by a set of research questions.

Working Memory

Working memory (WM), known as one of the underlying cognitive processes of many kinds of learning, is the theoretical construct used in multiple disciplines in cognitive science to refer to a system responsible for active maintenance and simultaneous processing of information (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Dehn, 2008). Although a dozen WM models have been proposed and controversies surrounding them have persisted, scholars (Conway, Jarrold, Kane, Miyake, & Towse, 2007; Miyake & Shah, 1999) have suggested a view that WM models in fact “converge more than diverge on fundamental issues” (Wen, 2016; p.17). Most recently,
Wen’s (2016) close examination of commonalities and differences among the most contemporary WM models has revealed that although some inconsistencies lie in the research scope and focus, the models are more similar than different from each other in terms of the basic mechanisms that they postulate.

The highly influential WM model proposed by Baddeley and colleagues (e.g., Baddeley & Hitch, 1974; Baddeley & Logie, 1999) includes two domain-specific subsidiary systems – a phonological loop (i.e., speech-based store) and a visuospatial sketchpad (i.e., image-based store) – and one domain-general supervisory system labeled the central executive. The subsidiary systems were viewed as short-term memory components controlled by the central executive which is a supervisory attentional system that regulates and coordinates cognitive processes. Baddeley (2000) later added another subcomponent, the episodic buffer, a multimodal store that links information from the subsidiary systems with long-term memory. Baddeley’s multi-component model has been largely accepted in cognitive science research (Miyake & Shah, 1999; Wen, Mota, & McNeill, 2015). Though compatible in principle, Cowan’s (1999, 2005) embedded-process model shines more light on the role of the focus of attention and level of activation. He proposed that WM and long-term memory are mutually interdependent and that WM is essentially a part of long-term memory, only more activated and more accessible. Cowan also emphasizes that the focus of attention has limited capacity, which can restrict WM retention and processing.

More broadly, these two most influential models – Baddeley’s (2000) multi-component model and Cowan’s (1999, 2005) embedded-process model- as well as others (see Ericsson & Kintsch, 1995, for the long-term WM framework; see Kane, Conway,
Bleckley, & Engle, 2001, for the executive attention model; see Oberauer, Sub, Wilhelm, & Wittman, 2003, for the facet model) arrive at fundamentally consensual characterizations of WM in terms of its nature, structure, and function (Wen, 2016) as elaborated in the following:

- WM contains both storage and processing components and is a capacity system in which information span and duration are limited, which may lead to a trade-off between storage and processing (Daneman & Carpenter, 1980);
- WM includes non-executive functions (e.g., chunking, encoding) and higher-order executive functions (e.g., attentional control, inhibitory control), also called the central executive (Baddeley & Hitch, 1974);
- WM provides linkages between short-term memory components subsumed in WM and long-term memory, which affects performance on cognitive tasks and real-world actions (Wen, 2016).

Such an integrated view toward WM models is endorsed by the notion that WM represents a cognitive mechanism involved in “the control, regulation, and active maintenance of task-relevant information” (Miyake & Shah, 1999, p. 450) to facilitate encoding, storage, and retrieval functions that are essential for learning (Dehn, 2008).

**Working memory and academic learning.** Research on working memory has repeatedly confirmed the critical role of WM in general academic learning (Gathercole & Alloway, 2008). For learning to take place, knowledge must be constructed and modified while one interacts with long-term memory and juggles the simultaneous storage and processing of information, which is what working memory is responsible for. Thus, successful learning in terms of its rate and extent is determined by the functioning and
capacity of WM because every piece of explicit information that is remembered and learned must pass through WM (Dehn, 2008). In particular, executive functions, which govern the control and manipulation of WM resources, are at play during learning and other cognitive tasks. Individual differences in WM reflect attentional control which allows for the activation and maintenance of memory representations while alternating between tasks and suppressing retrieved information from focus as needed (Cowan, 2005; Engle, 1996). For this reason, executive WM deficits result in ineffective attentional control, which in turn leads to unsuccessful learning. Gathercole (2007) and Meltzer (2018) assert that executive WM discriminates between learners with and without learning disabilities because those with learning disabilities often struggle with simultaneous organization and coordination of multiple mental activities.

Among academic skills, the functioning of WM has been found to be directly involved in language learning such as vocabulary learning and decoding (Dehn, 2008; Dixon, LeFevre & Twilley, 1988; Gathercole & Alloway, 2008; McGrew & Woodcock, 2001; Swanson & Berninger, 1996), which are important aspects of reading. Reading comprehension is in fact one of the most researched academic skills, whose relationship with WM has been well-documented (Cirino, Miciak, Ahmed, Barnes, Taylor, Gerst, 2019; Daneman & Carpenter, 1980; Gathercole & Baddeley, 1993). The current study is situated in this very subdomain of academic learning – reading comprehension – and the role of WM as a crucial cognitive ability that underlies it. The following section discusses involvement of WM in reading comprehension based on theoretical and empirical support.
Involvement of Working Memory in Reading Comprehension

Comprehending a text goes beyond word recognition, lexical access, and word and sentence parsing. Rather, it is a model-building process with multi-level interaction and integration of information (Grabe, 2009). In the construction-integration (CI) model, Van Dijk and Kintsch (1983) postulate that readers begin this process by interacting with textual features at the surface level (e.g., decoding, accessing meaning, parsing grammatical structures) to extract information from the text itself, which gives rise to the second level, where they integrate propositions encoded in the text in their attempt to build meaning. This representation, called the textbase in Kintsch (1983), is a network of propositions in text that are connected to each other by local bridging, which refers to linking information from text currently active in WM with the immediately preceding text. After establishing text-level understanding, readers are at the third level where the propositional message is extended and refined to be integrated into a situational representation of the text, called a situation model. At this stage, readers engage in global bridging to connect ideas in the text and interpret the information in relation to their prior knowledge to generate a coherent whole (Graesser, 2007; Kintsch, 1998).

This integrative and iterative model-building process illustrated in the CI model places considerable demands on WM as the mental model of the situation presented by the text is constantly formed and updated during reading (Elleman, 2017). Just and Carpenter (1992) describe reading comprehension as a cognitive task that demands “extensive storage of partial and final products in the service of complex information processing” (p. 122). In other words, reading comprehension requires the simultaneous process of extracting pieces of information and consolidating them into larger chunks to
construct meaning. During this cognitive activity, readers must engage in the activation and use of background knowledge to holistically generate inferences and finally arrive at an integrated semantic representation of the text (Kintsch, 1998; Koda, 2005).

With its processing and storage functions, WM serves as a mental workspace that processes and holds necessary information from successive words, phrases, sentences, and propositions currently being read while keeping them available for integration, allowing a coherent representation of the text to be constructed (Daneman & Carpenter, 1980; Gathercole & Baddeley, 1993). Further, the meaning construction process as well as the metacognitive aspect of comprehension monitoring and strategy use point to the significant role of attentional control because even with an automated meaning construction process, difficulties can still arise at higher levels of comprehension (García-Madruga et al., 2013; Grabe, 2009; Kendeou et al., 2014).

The executive component of WM (i.e., Baddeley’s central executive), described as the underlying mechanisms of cognitive processes of attentional control and the regulation of the dynamics of human cognition, is indeed associated with higher-level comprehension (Daneman & Carpenter, 1980; Grabe, 2009; Miyake, Friedman, Emerson, Witzki, & Howarter, 2000). According to Cowan (2010), executive functions allow readers to allocate attention flexibly and be selective of information to be held. Also, executive functions enable continuous updates of what becomes relevant as new information enters (Palladino, Cornoldi, de Beni, & Pazzaglia, 2001; Yeari, 2017) and have consistently shown a unique contribution to reading outcomes (Cirino et al., 2019).

Correspondingly, the core executive functions as identified by Miyake and colleagues (2000) include inhibition, switching, and updating. Inhibition serves as a
restraining function that suppresses the disruptive effects of irrelevant information and controls the competition among relevant and irrelevant information activation thereby preventing erroneous interpretations of text (Chiappe, Hasher, & Siegel, 2000; Dehn, 2008). Switching is the ability to shift back and forth between multiple pieces of information and tasks. Combining information directly from text as well as connecting it with prior knowledge from long-term memory is known to be part of the switching function (García-Madruga et al., 2013). Finally, the updating function requires not only monitoring and coding incoming information of relevance but appropriately revising the items held in WM by replacing old and irrelevant information with newer and relevant pieces (Morris & Jones, 1990). Thus, it is crucial that readers keep tuning activation levels in accordance with the relevance of information until a representation of a text can be built based on the final interpretation (Palladino et al., 2001). Miyake and colleagues (2000) stress that updating represents active manipulation of information rather than simple storage, highlighting the simultaneous juggling act that occurs between processing and storage in WM.

In short, reading comprehension requires continual integration of incoming and relevant information into the evolving mental representation of the text (Kintsch, 1998) as well as the retrieval of and interaction with prior knowledge from long-term memory. This integrative computational process demands resources in WM that simultaneously process and store information to allow for meaning construction. Moreover, the executive functions of WM (i.e., inhibition, switching, updating) play a critical role in regulating and controlling attention during reading comprehension.
Measures of Working Memory

A variety of WM measures have been created and used in previous research. There are visuospatial WM tasks that require test takers to keep track of spatial orientations (see Shah & Miyake, 1996, for spatial span tasks); traditional “simple” span tasks that require remembering a series of stimuli (see Cantor, Engle, & Hamilton, 1991 for digit span and word span); and complex span tasks such as the counting span task (see Case, Kurland, & Goldberg, 1982), the operation span task (Turner & Engle, 1989) and the reading span task (Daneman & Carpenter, 1980). Because simple span tasks are storage-oriented without a concurrent processing subtask, they are now considered short-term memory span tasks (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Complex span tasks, on the other hand, have been shown to reliably predict cognitive control and executive attention; thus, they are used most commonly.

Although the counting span task, operation span task, and reading span task use different types of stimuli, the underlying structure of these tasks is the same as they tap storage and processing of information simultaneously. The counting span task requires participants to count shapes and remember the total count in each set for later recall; the operation span task requires participants to respond to an arithmetic problem to verify equations and remember a word that appears next to the problem for later recall; the original reading span task requires participants to read aloud a sentence while trying to retain the final word of each sentence for later recall. Kane and colleagues (2004) argued for a domain-general attentional aspect of WM along with Turner and Engle (1989) who created the operation span task that reflects this view. Turner and Engle stressed that WM does not depend on the nature of the processing component of the span task.
However, a domain-specific view was supported by Daneman and Carpenter (1980) who devised the reading span task, which is considered to be more relevant to WM in L2 research because it involves language processing. Shah and Miyake (1996) found support for the domain-specific view by showing that reading span was a stronger predictor of verbal Scholastic Aptitude Test (SAT) ability than spatial span; conversely, visuospatial skills were better predicted by spatial span than by reading span. Moreover, Wen (2016) advocated for the domain-specific view of WM rather than the domain-general view for research on the link between WM and L2 subskills (e.g., reading). He proposed a definition of WM for L2 research: “The limited capacity of multiple mechanisms and processes in the service of complex L2 activities/tasks” (p. 10). He advised L2 researchers to adopt WM measures that are task-specific, such as the reading span task, because the underlying construct of WM will be better tapped. This is not to discount the domain-general executive aspects of WM, but to support a domain-general executive component of WM that interacts with a domain-specific process, such as reading comprehension.

In accordance with this definition of WM specific to L2 research, proposed by Wen (2016), the current study focuses on the reading span task as a WM measure in an effort to further elucidate the role of WM in L2 reading comprehension. Because the reading span task is also the most commonly used task in research that examines the relationship between WM and reading comprehension (e.g., Alptekin & Erçetin, 2009, 2010, 2011; Bailer, Tomitch, & D’Ely, 2013; Leeser, 2007), the current study provides comparable empirical findings across relevant studies.
**Reading span task as a working memory measure.** The role of WM as an underlying cognitive ability in reading comprehension has been well established, especially with Daneman and Carpenter (1980) having paved the way by devising a verbal WM test called the reading span task. They made a strong case against using the traditional digit span, letter span, and word span tests as WM measures. These simple span tests only measure how many items an individual can store and retrieve, which goes against the architecture of the WM construct that is composed of storage and processing. Moreover, scores on simple span tasks do not correlate well with reading and listening comprehension. The reading span task, on the other hand, was designed to tap both processing and storage components and was shown to better predict reading comprehension (Daneman & Merikle, 1996). That is, WM was operationalized through a task that requires the simultaneous processing and storage of information.

In the original reading span task developed by Daneman and Carpenter (1980), 60 unrelated sentences were prepared. Participants were asked to read aloud a sentence presented to them, memorize the final word of that sentence, and report the final word of each sentence included in a set, which became progressively longer. They started with three sets of two sentences, which required participants to understand two sentences and remember the two final words in each of the sets. The next three sets included three sentences, followed by three sets of four, five, and six sentences. The highest and most demanding level was three sets of six sentences where participants would have to recall the six final words in each set. In this task, the processing function is presumably taxed by the first part of the task in which participants read sentences aloud. The recall of the final word of each sentence in the current set is what represents the storage function of
WM. Since its debut in 1980, the reading span task has been a popular choice as a WM measure because, unlike simple span tasks, it was designed to involve both functional aspects of WM (i.e., processing and storage) (Waters & Caplan, 1996; Wen, 2016).

However, Daneman and Carpenter’s (1980) reading span task has raised concerns with regard to its appropriateness for measuring the processing component for the following reasons: First, the scores generated do not reflect participants’ efficiency in carrying out the task; also, it is difficult to observe whether there is a storage-processing trade-off just by recording the number of final words recalled. Basing WM solely on the number or words recalled, in turn, makes the scores difficult to interpret as they only reflect the storage function of WM (Leeser, 2007; Turner & Engle, 1989; Waters & Caplan, 1996). This original reading span task did involve a task that tapped processing by having participants read aloud. However, their processing task may not accurately measure processing as participants could simply sound out words without pay attention to them. Also, the scores of the task did not reflect the performance of the processing component. To address these issues, Waters and Caplan (1996) incorporated a sentence acceptability judgment task as well as the recording of reaction times to tap the processing component of WM explicitly. This modification of the classic reading span task allows for performance on both processing and storage to be represented in the scores (Walter, 2004).

Currently, the most commonly adopted format of a complex reading span task asks a participant to read a set of unrelated sentences, make a semantic or grammatical judgment about each sentence, and report the final word of each sentence when the set is finished. The judgment task taps the processing component of WM while the final word
recall task taps the storage component. Importantly, more and more studies are paying attention to not only the task design but also the scoring system to ensure that the resulting composite scores reflect both processing and storage scores. Complex reading span tasks have been shown to exhibit higher reliability than simple span tasks (see Conway et al., 2005). This dual-paradigm reading span task has now become a standard way of measuring WM and is widely used to date.

**Relationship between Working Memory and Reading Comprehension**

Credited with developing a complex span task to measure WM, Daneman and Carpenter (1980) showed a strong correlation between reading comprehension and WM and have prompted numerous studies to explore this relationship. Empirical findings in L1 contexts overall have converged on the fact that reading comprehension is constrained by WM; greater WM capacity allows for more successful retention and integration of information, which leads to better reading comprehension (Hannon & Daneman, 2001; Just & Carpenter, 1992).

In a meta-analysis with 77 studies, Daneman and Merikle (1996) found that individual differences in WM significantly predicted reading comprehension with weighted effect sizes \( r \) ranging from .41 to .52. Moreover, WM has been shown to explain unique variance in reading comprehension even after controlling for variables strongly associated with it, such as vocabulary and decoding skills (Cain, Oakhill, & Bryant, 2004; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; Seigneuric & Ehrlich, 2005). Studies that compared skilled and less skilled readers found that deficits in reading comprehension are associated with inefficiencies in WM especially due to unsuccessful
support from executive function and attentional control (e.g., McVay & Kane, 2012; Nouwens, Groen, & Verhoeven, 2016; Swanson & Berninger, 1996).

Of particular interest to the current study is whether these findings and claims are consistent with and applicable to L2 readers. The following discussion therefore provides a survey of the literature pertinent to the relationship WM and L2 reading comprehension.

**Working memory and L2 reading comprehension.** The overall findings on WM and L2 reading comprehension generally replicate those of L1 studies with the consensus that there is a positive association between the two (Wen & Li, 2019). For instance, Kormos and Sáfár (2008) explored the role of WM in the learning of English by Hungarian students at the novice level ($n = 122$). While phonological short-term memory (non-word span) did not appear to be an important factor in explaining overall L2 proficiency, a more resource-taxing measure of short-term memory, which bears resemblance to complex span tasks, backward digit span not only accounted for 30.2% of the variance in the overall performance on the Cambridge First Certificate Exam, but it showed a positive relationship with reading comprehension ($r = .31$). The researchers explained that holding verbal information in memory storage as well as carrying out cognitive processes simultaneously poses high attentional demands on learners, which may have contributed to the significant relationships.

Bailer, Tomitch, and D’Ely (2013) conducted a study with Brazilian high school students at the intermediate level ($n = 61$), who completed a reading comprehension test, a form recognition task while reading, and two WM tasks (reading span and operation
span task\(^2\)). They corroborated the view that individual differences in WM affect L2 reading performance by demonstrating a significant difference of reading comprehension scores between high – and low – WM students as well as a positive correlation between reading and WM \((r = .37)\). However, this was only the case when WM was represented by reading span tasks; the operation span tasks, which include arithmetic processes instead of verbal, did not generate the same results. The authors concluded that in order to achieve predictive validity when investigating WM and reading, the WM measure should incorporate a verbal storage and processing component, that is, be in the same modality.

Similar findings were reported in a study with advanced students as well. Harrington and Sawyer (1992) examined WM capacity and L1 and L2 reading skills of Japanese advanced learners of English \((n = 32)\). The participants’ WM capacity was measured through digit span (L1 and L2), word span (L1 and L2), and reading span (L1 and L2). Information regarding L1 reading skills assessment could not be found in the article. L2 reading skills were assessed through the grammar and reading sections of the TOEFL and a cloze test. It was found that digit and word span did not show a relationship with any of the L2 reading outcomes but that L2 reading span was correlated with L2 grammar \((r = .57)\) and reading comprehension scores \((r = .54)\). These findings parallel Daneman and Merikle (1986), who reported that WM measures that tap both processing and storage capacity (i.e., the reading span task) are more strongly associated with and can better predict reading comprehension.

\(^2\) A WM task developed by Turner and Engle (1989) that requires participants to read and verify an arithmetic problem and read a word that appears next to the problem.
In a series of studies with Turkish university students learning English, Alptekin and Erçetin (e.g., 2009, 2010) explored the role of WM in L2 reading comprehension by examining literal and inferential comprehension to account for the multi-level representational architecture of reading. They found that, in general, readers with higher WM outperform their counterparts with lower WM. Also, there was a significant relationship between WM (L2 reading span task) and inferential comprehension \( (r = .40 - .45) \) but not literal comprehension. The researchers explained that literal understanding of a text does not impose a heavy cognitive load on WM capacity as it is more directly influenced by L2 proficiency and lower-level processes of reading such as decoding and syntactic parsing.

Adding to the evidence from individual empirical studies are findings from meta-analytic studies that have ascertained that WM and L2 reading go hand in hand. Jeon and Yamashita (2014) conducted a meta-analysis to investigate the key variables related to L2 reading and to describe the contributions of individual differences to L2 reading comprehension. They examined reading correlates that are frequently investigated (i.e., high-evidence correlates), such as L2 vocabulary and grammar knowledge, as well as less frequently investigated ones (i.e., low-evidence correlates), such as L2 listening comprehension, metacognition, and WM. Although WM was categorized as a low-evidence correlate, the accumulated evidence revealed that WM has a medium-sized correlation with L2 passage-level reading comprehension \( (r = .42) \) (see Plonsky & Oswald, 2014, for effect size benchmarks), substantiating that WM is one of the key cognitive correlates of reading that merits more research attention.
Similarly, Linck and colleagues (2014) conducted a meta-analysis with 79 studies to provide a quantitative synthesis of the relationship between WM and L2 outcomes; they further established that WM serves as a component of the cognitive processes underlying L2 processing. They reported WM’s positive association with L2 language comprehension \( (r = .26) \) and production \( (r = .27) \) and concluded that more research on WM can enhance our understanding of both receptive and productive L2 abilities. More recently, Shin’s (under review) meta-analysis focused specifically on WM and L2 passage-level reading; Shin reported a correlation of .30, confirming the small to medium-sized relationship between WM and L2 reading comprehension. However, the findings showed that variability in the study outcomes (i.e., the range of the correlation between WM and L2 reading comprehension) was systematically affected by what type of reading span task was used in the study, calling attention to inconsistent WM measurement practices. The details of this inconsistency in measurement practices and how it may affect findings are outlined in the next section.

**Inconsistent measurement of working memory and its impact on research outcomes.** Although WM is considered a predictor of L2 reading comprehension, there exist discrepancies of correlations reported thus far, ranging from non-significant (e.g., Chun & Payne, 2004) to significant and strong \( (r = .79 \) in Walter, 2004). Such inconclusive findings likely resulted from methodological inconsistencies. In this domain of research where the most commonly used WM measure is the reading span task, several studies operationalized WM capacity as storage-based scores while disregarding processing capacity, following the traditional Daneman and Carpenter (1980) approach (e.g., Harrington & Sawyer, 1992; Ikeno, 2006). However, a number of studies used an
average of storage and processing scores to reflect WM capacity (e.g., Alptekin & Erçetin, 2009; 2010; 2011; Zhang, 2015), while a few others added one more element and used an average of storage, processing accuracy, and processing speed (e.g., Leeser, 2007; Shin et al., 2019). Considering that measurement of a construct is what essentially shapes its operationalization, Alptekin and Erçetin’s (2009) comment about there being little consensus on WM measurement holds weight. Taken together, a cautious interpretation of past research findings is warranted due to the variations in WM measures.

Incorporating three components into a WM measure is supported by Leeser (2007) and Waters and Caplan (1996). Waters and Caplan were the first to raise the issue of a speed-accuracy trade-off among the components. That is, subjects might use different strategies; they can “read slowly to increase accuracy of processing or to rehearse the final words” (p. 74). In support of this, Leeser (2007) found that reaction times were negatively correlated with semantic judgment accuracy and final word recall scores, indicating that participants may have sacrificed speed to recall more words and make more accurate judgments. With this being the case, WM scores based exclusively on the storage component will not accurately represent WM capacity and thus will be difficult to interpret (Turner & Engle, 1989). The inclusion of processing accuracy and speed components, therefore, can address this issue and reduce the trade-offs across tasks contributing to recall outcomes (Daneman & Carpenter, 1980; Leeser, 2007).

Another issue stemming from the fact that composite WM scores reflect different cognitive processes is the extent to which reading comprehension is associated with each of the components. In line with Waters and Caplan’s (1996) claim that processing rather
than storage explains most of the shared variance between WM and reading, Alptekin and Erçetin (2010) showed that L2 reading comprehension was correlated with processing scores of the reading span task ($r = .33$) but not with storage scores. There was also a lack of a significant correlation between L2 processing and storage, suggesting that the two components may be functionally independent of each other and may play different roles in reading comprehension.

Sagarra (2017) shed more light upon this matter by evaluating the impact of processing demands on maintaining information. After discovering that WM (when represented as storage scores) did not modulate the improvement of L2 grammar nor reading of L2 Spanish learners, a second experiment was conducted to determine whether the WM effect was hindered by methodological limitations. Sagarra compared the extent to which L2 reading development was predicted by two types of the reading span task: The original one (Daneman & Carpenter, 1980) in which final word recall scores were calculated to indicate storage capacity without a taxing processing measure and a more complex task that added processing accuracy and speed to the scoring system. The results showed that WM, when represented as storage-based scores, did not contribute to L2 grammar nor L2 reading. However, when processing and reaction times were accounted for, a significant main effect for WM was found in both L2 grammar ($\eta^2 = .06$) and L2 reading ($\eta^2 = .30$). Sagarra also found a negative correlation between reaction times and final word recall ($r = .16$) as did Leeser (2007).

These findings demonstrate that reading span tasks with and without a measure that places heavier cognitive demands on processing yield different outcomes in studies of reading comprehension. Further, it can be suggested that individuals with the same
storage capacity may perform differently on the processing component of the task. Each of the components may play a unique role in reading, which could not have been captured by previous studies that focus on composite WM scores. Sagarra (2017) reported that unlike recall scores, reaction times as part of processing capacity predicted longitudinal WM effects on L2 grammar and reading. Sagarra added that WM capacity can be better estimated when the metric features storage, processing accuracy (correct judgment), and processing speed (reaction times). Thus, she recommended future research include multiple performance indices beyond simply storage scores. This complex measure may in turn be able to provide more insights into the relationship between WM and L2 reading comprehension. Despite the potentially differential contributions of each of these components to L2 reading comprehension, there are only a few studies (Leeser, 2007; Sagarra, 2017; Shin et al., 2019) that followed the rigorous procedure that includes all three components in the scoring system. Delving into the individual contributions of WM components to L2 reading with a more fine-grained measure may be able to shed light on the discrepant findings on WM and L2 reading.

As previously mentioned, Shin’s meta-analysis (under review) revealed that variations of the reading span task in terms of the scoring system systematically moderate the strength of the relationship between WM and L2 reading comprehension. As summarized in Table 2.1, Shin categorized the scoring systems observed across studies of interest into three groups: (a) Storage only, (b) the average of storage and processing accuracy, and (c) the average of storage, processing accuracy, and processing speed (reaction time). Studies that exclusively relied on storage scores as WM capacity showed a mean correlation of .20, which is lower than the correlations reported in studies that
estimated WM using a more complex task that includes either processing accuracy ($r = .33$) or processing accuracy and speed ($r = .30$). This lends support for the findings of Leeser (2007), Sagarra (2017), and Alptekin and colleagues (2014) that the storage and processing components of WM may exert different influence on L2 reading comprehension, calling for further research.

To sum up, WM is a cognitive individual-difference factor that influences reading performance as a limited mental workspace in which integration and updating of information takes place. However, close examination of the literature revealed that the nature of its relationship with L2 reading has been partly obscured through an inconsistent use of methodologies, resulting in variability in study outcomes reported in the existing research. Also, empirical evidence thus far suggests that there is more to be discovered about the relationship between WM and reading comprehension with regards to differential contributions of the components of WM (storage and processing) to L2 reading comprehension.

Table 2.1

*Scoring Systems Moderating the Correlation between Working Memory and L2 Reading*

<table>
<thead>
<tr>
<th>Scoring system</th>
<th>Effect size ($r$)</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>Lower bound</td>
</tr>
<tr>
<td>Storage only</td>
<td>.20</td>
<td>.19</td>
</tr>
<tr>
<td>Average of storage and processing accuracy</td>
<td>.33</td>
<td>.32</td>
</tr>
<tr>
<td>Average of storage, processing accuracy, and reaction times (processing speed)</td>
<td>.30</td>
<td>.29</td>
</tr>
</tbody>
</table>

*Note.* Adapted from Shin (under review)
Reading Comprehension and Its Multiple-level Architecture

Reading comprehension has been shown to involve multiple-level architecture, which in turn places varying cognitive demands on the reader (Alptekin & Erçetin, 2010). There is a body of literature that emphasizes a continuum of reading comprehension levels and skills, such as literal and inferential comprehension (e.g., Carnine et al., 2010; McCormick, 1992). Although there exists a controversy on a linear progression of difficulty along the ascending level of comprehension, scholars agree that tasks or items associated with inferential comprehension, rather than literal comprehension, are more conducive to higher levels of cognitive activation and reader engagement (Alptekin, 2006; Alptekin & Erçetin, 2010; Basaraba, et al., 2013; Kintsch, 1998).

Literal comprehension is known to operate on the text-based level of understanding, which involves readers’ ability to identify the meaning of vocabulary and extract information explicitly stated in a text. For this reason, WM demands required for this level of comprehension may not be extensive, especially for proficient readers. Inferential comprehension, on the other hand, requires readers to go beyond reading for exact meaning. At this level of comprehension, readers engage in knowledge-driven processes and interact with the text to synthesize, summarize, integrate information, draw conclusions, and interpret the author’s intention, hence the expression “reading between the lines” (Alptekein, 2006; Basaraba et al., 2013). As such, inferential comprehension entails more controlled processing on the readers’ part as they connect pieces of information to form a coherent whole (Alptekin & Erçetin, 2010).

Further, the notion that reading comprehension processes involve integration of information for which WM serves as a mental workspace has been corroborated by the
construction-integration (CI) model (Kintsch, 1998). As mentioned above, the model distinguishes between two main processes: (a) a construction process in which a reader constructs propositions (i.e., the textbase) from textual information through local bridging and (b) an integration process in which integration of propositions with the reader’s global knowledge takes place through global bridging to form a coherent representation of the text (i.e., the situation model). The relatively unstructured prior knowledge then becomes more organized in accordance with the context, and the integration becomes more fine-tuned through iterative cycles as more textual input is encountered. Taken together, the operationalized construct of reading comprehension, therefore, would be best represented by a measure that taps the two levels of reading comprehension – literal and inferential comprehension.

**Background Knowledge in Reading Comprehension**

To reiterate, reading comprehension processes involve intricate interaction and integration of textual and extratextual information within and between levels of comprehension (e.g., word level, proposition level) to successfully generate inferences and ultimately form a mental representation of a text (Carrell, Devine, & Eskey, 1988; Kintsch, 1998). General knowledge relevant to the topic or content of the text at hand, is typically referred to as background knowledge, which is the all-encompassing term that will be used throughout this study. Background knowledge is known to assist in creating connections between various pieces of information, which, when functioning in concert with textual information, establishes coherence in the representation of the text in readers’ memory (Grabe 2009; Graesser, Gernsbacher, & Goldman, 2003; Kintsch, 1998; Kintsch & van Dijk, 1978; van den Broek, 1994). As such, background knowledge is a
strong contributor to reading comprehension as demonstrated in numerous studies reporting its effect across ages (e.g., for children, see Langer, 1984; Lipson, 1982; Rowe & Rayford, 1987; Tarchi, 2010; for adults, see Brantmeier, 2005; Hammadou, 2000; Kendeou & van Den Broek, 2007) and proficiency levels (e.g., for beginners, see Hammadou, 1991; for intermediate learners, see Alptekin, 2006; Liu, 2015; for advanced learners, see Brantmeier et al., 2014).

Although L2 reading research has consistently shown the significant role of background knowledge in reading comprehension, background knowledge has been operationalized in a variety of ways referred to as cultural familiarity (Alptekin, 2006; Johnson, 1982), subject knowledge (Brantmeier, 2005, Brantmeier et al., 2014), content schemata (Carrell, 1987), and topic familiarity (Chang, 2006; Pulido, 2003, 2004, 2007). Studies that used the terms subject knowledge and topic familiarity seem to be using them synonymously (Brantmeier et al., 2014; Chang, 2006) to mean content knowledge, or what Carrell (1987) called content schemata, which she defined as “knowledge relative to the content domain of the text” (p. 461). This is different from how the concept of the schema was first introduced in psychology by Barlett (1932): A mental structure in memory abstracted from past experiences that sets up expectations for an upcoming event. Influenced by Rumelhart (1980) among others, what Carrell refers to as a schema was a much narrower description of a schema that is meant to highlight reader-centered L2 reading with an emphasis on readers’ text-based knowledge and conceptually-driven background knowledge functioning in concert to achieve maximal comprehension. Cultural familiarity also seems to fall into a type of content schemata, though it refers to culture-specific content-related knowledge, also called cultural content schemata.
Nevertheless, even with different terminology, the aims of the studies above are the same: Discovering the extent to which readers’ background knowledge affects their reading comprehension. The motivation of these studies stems from the notion that reading requires knowledge of not only words but also of the world and correspondingly, these studies tend to test students on both topics familiar and unfamiliar to them to test the effect of background knowledge.

There is abundant evidence of the facilitative effect of background knowledge on L2 reading comprehension. The pioneering work in this domain includes Johnson (1982) who examined how background knowledge, operationalized as prior cultural experience, affects advanced L2 students’ \( n = 72 \) reading comprehension. Two weeks after a city-wide celebration of Halloween in which all students except one participated, Johnson gave them a reading passage that included information assumed to be familiar and unfamiliar to them: the celebration of Halloween today and the celebration of Halloween in the past. When she tested their comprehension through written recall and true or false questions, she found that readers more accurately recalled and recognized the familiar section of the passage. The students showed superior performance on recall of information not only in its quantity but in accuracy and cohesive ties between propositions in the passage designed to be familiar to them. Also, readers’ familiarity with the celebration of Halloween today had a greater impact on their reading comprehension than did their exposure to preselected vocabulary.

Similar results were reported in Horiba and Fukaya (2015) who operationalized background knowledge as topic familiarity. They compared L2 reading performance of Japanese university students \( n = 145 \) on a familiar and unfamiliar topic. Nursing and
non-nursing majors were recruited and were divided into a high and low topic-familiarity group. After reading two narrative passages on health care and a nursing situation, both of which were familiar to the nursing majors, all participants completed a written recall task and subtests of production and meaning-recognition as a vocabulary acquisition test. The written recall results showed that nursing students (i.e., high topic familiarity group) were able to generate more inferences and build more coherent representations of the text than their counterparts with lower topic familiarity, indicating that topic familiarity facilitated learning of the text content.

In search of ways to enhance readers’ comprehension of expository texts, Hammadou (2000) examined the effects of analogies and background knowledge, defined as prior content knowledge, on L2 reading. French and English native speakers were presented with two texts on the function of the eye and the chain reaction of nuclear fission in their L1 or L2. These texts were then modified again to create one with analogies and one without them. While analogies either had a debilitating effect or did not have any effect on participants’ reading comprehension assessed through a recall task, their prior content knowledge, revealed through tests that included multiple-choice questions about the topic, was found to contribute to comprehension.

This finding was repeated in Brantmeier (2005) who replicated Hammadou (2000) with the same research agenda – adding analogies to unfamiliar L1 and L2 texts to aid comprehension. While analogies had a negative effect and made “the reading process even more multifarious,” (p. 47) background knowledge, referred to as subject knowledge, had a significant effect on L2 reading comprehension regardless of students’ language backgrounds (Spanish and English), proficiency (native, intermediate,
advanced), and reading comprehension task type (recall, sentence completion, multiple choice questions) indicating the influential force of background knowledge.

Although the facilitative role of background knowledge in reading comprehension has been well established (Grabe & Stoller, 2020), some studies have reported that background knowledge does not always work in every reader’s favor. Background knowledge has been found to interact with syntactic complexity of text (Barry & Lazarte, 1995) and L2 proficiency, or linguistic ability (Carrell & Eisterhold, 1983, Hudson, 1982). As the CI model illustrates, all knowledge sources available to readers, such as syntactic and pragmatic knowledge, become activated and are held in WM in the construction stage. However, readers’ success in generating inferences efficiently and connecting information in the integration stage is affected by their differential reliance on various linguistic and extralinguistic knowledge as well (Pulido, 2009). Moreover, what is held in WM in the construction stage of reading activates content in long-term memory. In response to this signal, “the information from the discourse representation and general world knowledge becomes activated simultaneously” and influences further construction of meaning and the iterative cycles (Nassaji, 2002, p. 455). Nassaji added that greater background knowledge allows for more efficient attentional allocation during reading, which in turn enables richer analyses and interpretations of text.

Taken together, the iterative construction-integration cycle is completed by continuous assimilation of both linguistic and extralinguistic knowledge with WM regulating and facilitating the process. Yet, little is known about the potential interactive relationship between WM and background knowledge. Although inadequate language proficiency evidently leads to a propensity for lower level text-oriented processing, WM
limitations are another trait that may be at play since WM allocates cognitive resources to long-term memory-based data and facilitates access to them (Alptekin & Erçetin, 2011). Activation of background knowledge and its effect on reading comprehension, therefore, may be moderated by WM operation.

**Combined Effects of Working Memory and Background Knowledge**

Though this is still a largely unexplored domain of research, attempts have been made to investigate how WM and background knowledge contribute to performance in a range of cognitive tasks. A related viewpoint proposed in cognitive psychology is the knowledge-is-power hypothesis, which postulates that the primary source of power that determines success in cognitive endeavors is background knowledge, referred to as domain knowledge, while cognitive abilities are considered less important (Feigenbaum, 1989). Hambrick and Engle (2002) recognized that little work had been done to empirically demonstrate the issue stemming from this hypothesis, namely the interplay between background knowledge and WM; they proposed three models to describe their combined effects. First, the *compensation model* proposes that high levels of relevant background knowledge can compensate for low levels of WM. This model also suggests that an increase in adequate task-specific knowledge may lead to a decrease in the performer’s dependence on ability characteristics, such as WM capacity. Second, the *independent-influences model* predicts that WM and background knowledge have additive and independent effects on performance, predicting no interdependence. Finally, the *rich-get-richer model* posits that the facilitative effect of background knowledge can be amplified by higher levels of WM capacity; that is, people with high WM capacity tend to benefit from background knowledge to a greater extent.
In Hambrick and Engle’s (2002) study, adult participants \((n = 181)\) with a wide range of preexisting knowledge of baseball (domain knowledge), which was assessed through a questionnaire, listened to radio broadcasts of baseball games and completed a memory recall task. They found that the participants’ knowledge of baseball facilitated their memory performance, and so did their WM. The results also showed that high WM participants derived a greater benefit from their baseball knowledge than did their counterparts with lower WM, supporting the rich-get-richer model.

However, this finding was not replicated in further studies. Hambrick and Oswald (2005) slightly altered the research design of Hambrick and Engle (2002) to evaluate the three models (compensation, independent-influences, and rich-get-richer) concerning the interaction between domain knowledge and WM; instead of assessing participants’ already-existing domain knowledge through questionnaire items as Hambrick and Engle (2002) did, Hambrick and Oswald created two conditions – a domain-relevant task and a non-domain-relevant task. Then, they tested whether or not the influence of WM is different on performance in the domain-relevant baseball task and in the non-domain-relevant task where participants’ baseball knowledge would not be of use. The results showed that participants’ \((n = 318)\) domain-relevant task performance benefitted from their baseball knowledge (domain knowledge), but the effect of WM remained constant across tasks which represented background knowledge and no background knowledge conditions. Similarly, after testing 155 poker players’ WM and poker knowledge, Meinz and colleagues (2012) found that the participants’ WM and knowledge were important predictors of poker performance, but the effects were independent of each other favoring the independent-influence model, like Hambrick and Oswald.
Borrowing the three models and applying them to the context of reading comprehension, a handful of studies have been conducted to explore how WM and background knowledge might interact in L2 reading comprehension (see Table 2.2). However, findings are inconclusive across studies. In Alptekin and Erçetin’s (2011) study, Turkish participants (n = 62) read a narrative text either in its original American version or in a nativized version that was designed to raise their cultural familiarity. WM was assessed through an L2 reading span task and was represented by a composite score of storage and processing. No interaction was found between background knowledge (cultural familiarity) and WM. Similarly, Payne and colleagues (2009) showed that WM measured through a counting span task \(^3\) and domain experience (years of active study of Spanish) of Spanish learners contributed to their L2 reading comprehension but did not interact (β = .13, p = .71).

By contrast, Leeser (2007) reported that there is an interaction between WM and background knowledge (topic familiarity) and found evidence for the rich-get-richer model. He showed that L2 Spanish speakers (n = 94) with high- WM outperformed the low- WM group on a reading comprehension task if they were familiar with the topic of the texts demonstrating the interaction between WM and background knowledge (η\(^2\) = .06, p = .05). Leeser concluded that high- WM capacity aids in forming connections between ideas expressed in the text and background knowledge. In support of this finding, Joh and Plakans (2017) showed that WM was a significant predictor of L2 reading comprehension of Korean EFL learners (n = 80) when topic knowledge was provided (β = .54, p = .00).

\(^3\) A WM task developed by Case, Kurland, and Goldberg (1982) which requires participants to count shapes and remember the total count of each set for later recall.
Table 2.2

Summary of Studies on the Interaction between Working Memory and Background Knowledge in L2 Reading Comprehension

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Sample characteristics</th>
<th>Method</th>
<th>BK Operationalization &amp; treatment/assessment</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joh &amp; Plakans (2017)</td>
<td>L1: Korean L2: English Age: $M = 21$ $N = 80$ Intermediate to advanced</td>
<td>Task: L2 RST Scoring: Storage + Processing accuracy + Reaction times</td>
<td>Test: Researcher-developed Item: MCQs &amp; Short answer Qs Text: Expository (#: 8)</td>
<td>Preexisting knowledge of target language Assessment: L2 knowledge &amp; topic knowledge (topic-related vocab) WM contributed substantially to RC for high BK group ($\beta = .54$, $p &lt; .001$); less and not significant for low BK group ($\beta = .13$, $p = .41$) [Rich-get-richer model]</td>
</tr>
<tr>
<td>Study</td>
<td>L1: Language</td>
<td>L2: Language</td>
<td>Age</td>
<td>Task: Description</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
<td>-----</td>
<td>-------------------</td>
</tr>
<tr>
<td>Payne et al., (2009)</td>
<td>English</td>
<td>Spanish</td>
<td>Unknown</td>
<td>Counting span task</td>
</tr>
</tbody>
</table>

*Notes. WM = working memory, RC = reading comprehension, BK = background knowledge, RST = reading span task, MCQ = multiple choice question*
However, for the low-topic knowledge group, WM did not show a significant contribution to L2 reading comprehension ($\beta = .13, p = .41$), and L2 knowledge (grammar and vocabulary) was the only predictor of reading comprehension. Their results indicate that topic knowledge may determine the degree of WM contribution to L2 reading thus providing evidence for the interaction between the two.

Most recently, Shin and colleagues (2019), in a pilot version of the current study, found further support for WM and background knowledge operating on the interactive, rich-get-richer relationship. They examined 79 Korean EFL learners’ WM through a reading span task and gave them two passages written in their L1 the day before they took a reading comprehension test on the same two topics plus on two unfamiliar topics. The results demonstrated that background knowledge, which was provided to them by the researchers, enhanced their L2 reading comprehension ($\eta_p^2 = .48, p = .00$). However, a more complicated picture was revealed when background knowledge came into play. High-working-memory participants achieved greater L2 reading comprehension when background knowledge was present than when it was not with a very large effect size ($d = 1.34$). The same pattern was found with the low-WM group, but the effect turned out to be medium-sized ($d = .68$). In other words, not everyone was able to benefit equally from background knowledge due to the constraints of WM.

The researchers concluded that L2 readers with higher WM had more cognitive resources to put relevant background knowledge to better use more efficiently. On the flip side, their findings show that low-working-memory students need more explicit instructional support because mere presence of background knowledge does not necessarily lead to facilitation of reading comprehension if is not activated or used
effectively. In fact, being able to connect text to background knowledge is what distinguishes a strategic reader from a less strategic one (Grabe & Stoller, 2020). Shin and colleagues added that providing background knowledge should be a starting point rather than the end of instruction, reinforcing the importance of pre-reading activities that can help readers access and activate background knowledge.

What remains unknown, according to Shin and colleagues (2019), is the roles of the storage and processing functions of WM in L2 readers’ ability to utilize background knowledge available to them to enhance their reading comprehension. As previously discussed, evidence for distinct roles of the two subcomponents of WM has been shown (Sagarra, 2017) but not in the capacity of moderators that influence association between background knowledge and L2 reading comprehension. Shin and colleagues (2019) speculate that the differences between high- and low- WM readers may lie not in how much information they can hold in temporary storage but in how efficiently they utilize WM resources to access and retrieve relevant background knowledge. That is, high WM readers are better at accessing and retrieving information from long-term memory (Ericsson & Kintsch, 1995). Shin and colleagues added that high-WM readers were perhaps more strategic at directing their attentional resources. Further support can be found in Palladino and colleagues (2001) and Yeari (2017) who asserted that it is the flexible use of and attentional control that assisted readers in coordinating relevant information activation and suppression. However, more empirical evidence is needed to illustrate the roles of the WM components in interaction with background knowledge in L2 reading comprehension.
Methodological Inconsistencies across Studies

The inconclusive findings discussed above are likely due to the lack of uniformity in methods in terms of (a) WM measures utilized in existing studies and (b) the operationalization and assessment of background knowledge. A couple of different WM measures were used in this domain of research (counting span task in Payne et al., 2009; reading span task in Alptekin & Erçetin, 2011; Joh & Plakans, 2017; Leeser, 2007; Shin et al., 2019). Even within the reading span task design, task features (e.g., scoring system) differed across studies. Reading comprehension materials used in the studies were not consistent either. Shin’s (under review) meta-analysis revealed that different text types (narrative vs. expository) make a difference in the study outcome; WM tends to correlate more strongly with narrative text comprehension \((r = .33)\) than expository text comprehension \((r = .27)\). Additionally, the operationalization of background knowledge has varied from domain experience (Payne et al., 2009) to L2 knowledge (Joh & Plakans, 2017) and to cultural familiarity (Alptekin & Erçetin, 2011). Assessment methods for background knowledge have also varied from self-ratings of topic familiarity (Leeser, 2007) to self-reports of language experience (Payne et al., 2009). To create familiar and unfamiliar conditions, Alptekin and Erçetin (2011) replaced culturally unfamiliar words with words that were culturally familiar to Turkish students, and Shin and colleagues (2019) provided written texts to introduce two of the four topics on which participants were tested.

Taken together, while the above studies have undoubtedly contributed to the understanding of the roles of WM and background knowledge in L2 reading comprehension, there is insufficient evidence as to which one of the three models
(independent-influences, compensation, and the rich-get-richer best describes the interaction between WM and background knowledge. The disagreement in findings may reflect methodological discrepancies across studies, which warrants caution in the interpretation of results. Considering that the scoring system and methodological features of the reading span task as well as comprehension test materials can moderate study outcomes (Shin, under review), research with careful methodological attention can cast additional light on if and how WM and background combine to facilitate L2 reading comprehension.

**Proposed Study**

The above literature review has surveyed both established and underexplored areas pertaining to the roles of WM and background knowledge in L2 reading comprehension as well as their potential interplay. It addressed concerns with regard to methodological issues of WM measurement and overall inconsistencies observed in this domain of research. These issues likely have contributed to the difficulty of comparisons and generalization of the findings across studies. The goal of the current study is to advance our understanding of the role of WM in L2 reading comprehension in relation to its components (i.e., storage and processing) while pursuing methodological rigor. With this in mind, the researcher intends to conduct a follow-up methodological replication of the pilot study (Shin et al., 2019) on the interaction between WM and background knowledge in L2 reading comprehension while seeking to expand on their findings by examining the specific contributions of storage and processing to the interaction. By probing into how storage and processing capacity influence readers’ use of background
knowledge and L2 reading comprehension, the study aims to better explain the nature of the interaction between WM and background knowledge.

Despite the increased attention to the role of WM in L2 reading comprehension among L1 and L2 researchers, this study represents one of the few attempts to investigate the interplay between WM and another well-established predictor of reading comprehension – background knowledge, operationalized in the study as topic familiarity. Also, by simulating a classroom scenario in which L2 readers encounter familiar and unfamiliar expository texts, the study aims to lead to pedagogical implications regarding how WM capacity influences student performance. The following research questions are addressed in this investigation:

1. Is there an interaction between working memory and topic familiarity, as a specific form of background knowledge, in L2 reading comprehension? If so, which one of the models (compensation, independent-influences, and rich-get-richer) best illustrates the interaction?

2. To what extent do working memory subcomponents (storage, processing accuracy, processing speed) contribute to L2 reading with and without topic familiarity?
Chapter 3: Method

In an attempt to address the research questions stated at the end of Chapter 2, a research study was carried out in Japan and South Korea. Many of the instruments used in the current study were piloted in its earlier version (Shin et al., 2019). This chapter describes the methods of investigation including the (a) research design, (b) variables of interest, (c) participants, (d) instruments, (e) procedures, and (f) data analysis.

Research Design

This study used a quasi-experimental design to find and further examine combined effects of working memory (WM) and topic familiarity in L2 reading comprehension. To do so, quantitative data were collected to represent (a) WM capacity, (b) L2 knowledge, (c) L2 reading comprehension, and (d) the effect of topic familiarity on participants’ L2 reading comprehension. Tests for WM, L2 knowledge (via the C-test and Vocabulary Size Test), and L2 reading comprehension generated numerical scores for each of the variables. For WM and L2 knowledge, participants’ capacity and ability at the point of testing were of interest. Therefore, no intervention or manipulation was needed.

To examine the effect of topic familiarity on L2 reading comprehension, however, a repeated-measures design was followed, in which two conditions were purposefully created by the researcher. Adopting the approach used by Shin and colleagues (2019), this dissertation study was designed to observe 138 L2 participants’ performance under two conditions – reading with and without topic familiarity. In this approach, participants’ familiarity with topics was manipulated for the purpose of deliberately differentiating the conditions in which participants’ L2 reading performance is examined.
That is, the researcher provided a topic familiarity treatment to participants by distributing background texts on two of the four topics featured in L2 reading tests. Following this, participants read two passages on familiar topics and two on unfamiliar topics in the L2 reading tests. Participants’ L2 reading scores generated from the two conditions were then compared.

It should also be noted here that the researcher included a control group \( (n = 20) \) for one variable, L2 reading comprehension. Individuals with no familiarity with any of the four topics featured in the L2 reading tests were selected to be placed in the control group. This was to ascertain comparability of the passages included in the tests. Of interest to the researcher was whether or not the passages in one condition (i.e., the topic familiarity condition for the intervention group) presented the same degree of difficulty and challenges to the participants as the other passages in the other condition (i.e., the no-topic-familiarity condition). Because the control group was not familiar with any of the topics featured, no difference in the scores in the two conditions would indicate that the passages and L2 reading tests were comparable. Thus, any difference, if found, in the intervention group’s scores in one condition compared to the other would be attributable to the effect of topic familiarity. Though the control group did not receive the background texts, they completed all the other tasks.

**Variables of Interest**

The dependent variable of the study is L2 reading comprehension, which was measured through an adapted TOEFL reading comprehension test comprising four testlets. Each testlet had a passage on a different topic and ten accompanying comprehension questions that tapped literal and inferential comprehension. L2 reading
comprehension was treated as a continuous variable as it was operationalized as a range of reading comprehension scores from 0 to 44.

The independent and predictor variables include L2 knowledge, topic familiarity, overall WM capacity, and subcomponents of the WM task. L2 knowledge was measured using a C-test (Klein-Braley, 1997) and the Vocabulary Size Test (VST) (Nation & Beglar, 2007), both of which have been often used in research to estimate L2 proficiency (e.g., Eckes & Grotjahn, 2006; Grotjahn, Klein-Braley, & Raatz, 2002; Mozgalina & Ryshina-Pankova, 2015; Shin et al., 2019; Stæhr; 2008). The scores of the C-test and VST were used as two separate indices of general L2 knowledge. Although standardized test scores (e.g., TOEFL, IELTS) would have provided an ideal estimate of participants’ L2 knowledge, it would not have been practical to administer a standardized test that takes hours to complete. Efforts were made to collect participants’ most recent standardized test scores using a questionnaire, which served as an additional control of current proficiency. Participants were also asked to report the approximate date of the test so that the researcher could gauge the currency of the test results; the researcher realized that standardized test scores more than one-year old would not be the best indicator of participants’ proficiency. Because scores reported by some of the participants were not from the same standardized test, a conversion of test scores was necessary, as was the case in Shin and colleagues (2019). L2 knowledge was treated as a continuous variable because vocabulary size determined by VST (Nation & Beglar, 2007) and C-test (Klein-Braley, 1997) results was expressed as a range of numerical scores.

Second, topic familiarity served as the operationalization of background knowledge in this study. The review of literature in the previous chapter indicated that
background knowledge is an umbrella term encompassing the idea represented by the following terms: Subject knowledge (Brantmeier, 2005), cultural familiarity (Alptekin & Erçetin, 2011), domain experience (Payne et al., 2009), and topic familiarity (Leeser, 2007), all of which have been used more or less synonymously with one another. Topic familiarity in the current study specifically refers to prior knowledge of a topic that may ease comprehension of an L2 text on the same topic. In the current study, participants’ topic familiarity was manipulated to the extent that the researcher could purposefully create two conditions: Reading with and without topic familiarity. The researcher deliberately provided participants with background knowledge on two topics with which participants had indicated little familiarity beforehand. The researcher also chose two additional passages on topics unfamiliar to participants, which were used in the reading-without-topic-familiarity condition.

Third, WM capacity was operationalized as a composite score of the reading span task used by Shin and colleagues (2019). The composite score was an average of z-scores of the subcomponents of the reading span task (storage, processing accuracy, and processing speed), which was a continuous variable. Because fast processing speed (reaction time) would indicate that the participant is able to process information without much hesitation, hence better performance, the values for processing speed were multiplied by -1 so that a higher score would indicate better performance for all three subcomponents.

Lastly, WM subcomponents were operationalized as individual performance on storage, processing accuracy, and processing speed in the reading span task, expressed as respective z-scores. Participants’ storage capacity was represented by the normalized
number of final words recalled correctly; processing accuracy was represented by the normalized number of accurate answers to semantic-acceptability task questions; and processing speed was represented by the normalized mean reaction time on the semantic judgment task. These tasks are described in more detail below, in the instrument section.

Participants

Participants from two different L1 groups (Japanese and Korean) were recruited for the current study. A total of 158 adult learners in English-as-a-foreign-language (EFL) settings in Japan (n = 78) and South Korea (n = 80) volunteered. The participants were made up of 64 males and 94 females. All of them were university students in Japan and Korea. Their ages ranged from 18 to 24 years old, with a mean age of 19.54 years old (SD = 1.39). Except for seven participants who were in their 3rd year of the university, all of the participants were in their 1st and 2nd year (n = 151). Participants were either in the exploratory program or majoring in various disciplines: English, English education, engineering, environmental science, information technology, medicine, nursing, social science, sociology, and tourism.

According to participants’ most recent standardized-test scores, their English proficiency ranged from high-beginning to advanced, with the majority of them at the intermediate level (detailed results of proficiency levels in Chapter 4). The participants reported that they had at least 10 years of experience studying English. Approximately 7% (11 out of 158 students) of the participants had spent about 6 months on average overseas studying English.
Instruments

This subsection describes a series of instruments used to measure variables of interest and materials used to raise participants’ topic familiarity. The following instruments and materials were presented to participants in the sequence specified immediately below. Each instrument is described in more detail in the subsections that follow.

(a) A questionnaire to collect demographic information and to gauge participants’ topic familiarity, which has been modified following the pilot study (Shin et al., 2019);
(b) the L2 reading span task to assess working memory;
(c) the VST (Nation & Beglar, 2007) and C-test (Klein-Braley, 1997) to measure L2 knowledge;
(d) background texts in participants’ L1s (Korean and Japanese) to provide prior knowledge of two topics (Teotihuacán, the Bantu people), followed by five comprehension-check questions;
(e) a topic familiarity task to obtain a more accurate estimate of participants’ topic familiarity after reading the background texts; and
(f) an adapted TOEFL reading test with four testlets followed by 10 questions, each featuring a different topic to assess L2 reading comprehension;

Questionnaire. A two-part questionnaire written in participants’ L1s was administered first (see Appendices A, B, and C). The purpose of the questionnaire was two-fold: To collect demographic information and to estimate participants’ familiarity with topics presented. The format of the questionnaire was essentially the same as the one
used in the pilot study (Shin et al., 2019). However, changes were made to Part 2 of the questionnaire to improve on limitations of the original pilot version. Each part of the questionnaire is elaborated below.

*Part 1 of the Questionnaire.* The purpose of Part 1 of the questionnaire was to obtain demographic information on participants. In this part of the questionnaire, participants were asked about their age, sex, major, years of English study, and standardized test scores, a reading subscore if available, and the date of the test.

*Part 2 of the Questionnaire.* The purpose of Part 2 of the questionnaire entailed determining participants’ familiarity with the four topics featured in the TOEFL reading tests (Testlet 1: the rise of Teotihuacán; Testlet 2: agriculture, iron, and the Bantu people; Testlet 3: history of the chickenpox vaccine; Testlet 4: ancient Rome and Greece). As Shin and colleagues (2019) explained, TOEFL reading passages on four obscure topics were first selected using the researchers’ best judgment. Then, they used Part 2 of the questionnaire as a confirmatory tool in anticipation of responses indicating little to no familiarity with the topics.

Similarly, in this part of the questionnaire, participants in the current study read the title of each text and were asked to self-rate their topic familiarity ranging from 1 (“I can tell you nothing about this topic.”) to 4 (“I can tell you a lot about this topic.”). The format of Part 2 of the questionnaire was adopted from Wallace (2018) who also attempted to examine topic familiarity. The pilot study (Shin et al., 2019) contained the same task in its questionnaire but used a 5-point scale and different wordings: 1 indicated *not at all familiar*, and 5 indicated *very familiar*. For the scales from 2 to 4, a blank space was left without any descriptors.
The changes in the questionnaire used in the current study were made to achieve for (a) preciseness of the descriptors and (b) ease of estimating participants’ topic familiarity while matching it with a corresponding numerical value on the scale. Because the current study aimed to create two conditions – reading with and without topic familiarity – Part 2 of the questionnaire served as a screening measure to filter out participants who were already familiar with the selected topics.

Another change that was made in Part 2 of the questionnaire, based on the pilot, entailed asking participants who indicate familiarity with one or more of the four topics to write down anything they know about the topic(s) in their L1. Such information helped the researcher (a) ascertain the extent to which participants’ topic familiarity complemented self-ratings on the 1-4 scale and (b) decide which participants to include or exclude from the study based on the indicated topic familiarity. For example, if a participant self-rated his or her topic familiarity with Testlet 3: ancient Rome and Greece as 2 (“I can tell you a little about the topic.”) with written responses on the geographical location of the two countries without specific knowledge of the relationship between them in ancient times, the participant was not excluded from the sample. On the other hand, if a participant self-rated his or her topic familiarity as 2 with written responses on Greek and/or Roman civilization and history, that participant became an exclusion candidate.

**L2 Reading Span Task.** The second measure administered to participants was an L2 reading span task for examining working memory (WM). This was a modified version of the reading span task originally developed by Daneman and Carpenter (1980). The researcher computerized the task using Paradigm by Perception Research Systems Inc.
(Tagliaferri, 2005) and piloted it in Shin and colleagues (2019). This task demonstrated high reliability in the pilot study (storage: $\alpha = .92$; processing: $\alpha = .88$) as well as in the current study (storage: $\alpha = .89$; processing accuracy: $\alpha = .88$, processing speed: $\alpha = .85$). In the sub-sections that follow are explanations of (a) the rationale for using the L2 – English – as the task language, (b) the format and procedure of the task, and (c) the scoring system.

Using the L2 as the Task Language. The decision to use the L2 as the task language was made based on empirical findings and suggestions from previous research. Although the use of L2 in the reading span task often conjures up controversy due to a potential impact of language proficiency on WM performance, Alptekin and his colleagues (2010, 2011) recommend measuring WM in the L2 when working with L2 learners of English. Alptekin and Erçetin (2010) showed that L1 and L2 WM were correlated but that L2 inferential reading comprehension was directly related to L2 WM while the same did not hold true for L1 WM. The same result was found in Ikeno (2006) as well. Furthermore, Miyake and Friedman (1998) demonstrated in their path analysis that although there is a correlation between L1 and L2 WM, L2 WM has a direct effect on L2 reading comprehension. Shin’s (under review) meta-analysis on the relationship between WM and L2 reading comprehension reports that the effect size is larger when WM is assessed in the L2 ($r = .35$) rather than in the L1 ($r = .17$). She also points out that unlike L1 WM, L2 WM has consistently shown a positive relationship with L2 reading comprehension in previous research. In the same line of logic put forth by the domain-specific view of WM, it can be expected that WM performance may be influenced by the task or modality, which, in this case, is the task language.
The concern regarding the interaction between WM and language proficiency is a valid one as demonstrated in van den Noort, Bosch, and Hugdahl (2006). However, the differences observed between L1 and L2 WM have been shown to be less significant when participants are at intermediate or higher levels (Alptekin & Erçetin, 2010; Walter, 2004). Thus, the current study took a cautious step in selecting L2 sentences for the reading span task to prevent participants’ proficiency from confounding their performance on this task. Specifically, sentences taken from graded readers (Barrell, 2000; Kershaw, 1999) overall had simple syntactic structures, mostly included words in the 1k frequency band, and did not include words that go beyond the 2k band. In this way, the difficulty of the reading span task would largely be attributable to the cognitive demands of the task, though the effects of linguistic complexity and language proficiency may not have been completely eliminated (see Chapter 5 for Limitations).

Format and Procedures. The format of the task resembled that of Leeser (2007) and Shin and colleagues (2019) in that it included two subtasks to tap storage and processing functions of WM, using final word recall and a semantic judgment task, respectively. This particular reading span task, a modified version of Daneman and Carpenter’s (1980) original reading span task, was developed and computerized by Shin and colleagues (2019) and used again in the current study.

This task contained 70 unrelated sentences with 35 semantically acceptable and 35 unacceptable sentences (e.g., After 3 years, a book left the army and studied farming) (see Appendix D). Each sentence, 10 to 12 words in length, ended with a different final word. The 70 sentences in the task were divided into two sets at different levels: Level 2 (min) to Level 8 (max). At Level 2, participants saw two sets of two sentences each
ending with a different final word. At Level 3, they saw two sets of three sentences each ending with a different final word. As the level increases up to Level 8, each set became progressively longer, with the final two sets including eight sentences.

To prepare participants for the task, the researcher walked them through the procedure by providing an oral explanation after each participant finished reading a script with instructions written in his/her L1 (see Appendices E, F, and G). After that, the participants were presented with three practice trials (Levels 2, 3, and 4) to become familiar with the task (see Appendix H). The practice trials were done with the researcher present, immediately before the participant began the actual reading span task. Questions were encouraged and answered.

The main procedure for this task is illustrated in Figure 3.1. To elaborate, participants first silently read a sentence that appeared on the screen and made a semantic judgment about the sentence after seeing the question “Is this sentence semantically acceptable?” written in their L1 (“이 문장이 말이 되나요?” ; “この文章の意味が通じますか”) by clicking Yes or No on the screen while, at the same time, trying to retain the final word of the sentence. Participants repeated this judgment task with the other sentences in a set and were asked to write the final word of each sentence in the set upon completion of the set. As previously discussed, by incorporating a semantic judgment task after each sentence, the postulation of a trade-off between storage and processing, if any, could be accounted for, as participants were encouraged to juggle, to the best of their ability, the retention and processing of information (Leeser, 2007; Wen, 2016).
The procedure for reporting the final words was adopted from Dronjic (2013): When a set was finished, participants were prompted on the computer screen, by the phrase “STOP AND WRITE,” to write down the final word of each sentence on a sheet provided by the researcher (see Appendix I). They were allowed to write the final words from the set in any order as long as they did not begin with the final word of the last sentence in the set. The plus sign (+) on the screen signaled the beginning of a new set.

Scoring. As noted in the review of reading span task studies in Chapter 2, the scoring system of the reading span task has varied from study to study. The current study
adopted the scoring system used by Leeser (2007) and Shin and colleagues (2019), which represents a more inclusive approach in that it includes three aspects of WM – storage, processing accuracy, and processing speed. Accordingly, composite WM scores were calculated based on the following three sub-scores: The number of correctly recalled final words, the number of accurate semantic judgments, and reaction times of making the semantic judgments in milliseconds. Measuring reaction time was done by using the built-in response time function in Paradigm, as was done in Shin and colleagues (2019). Because shorter reaction times indicated faster processing of information, the reaction times were multiplied by -1 so that a higher numerical value indicated better performance for all three subcomponents. These sub-scores were then transformed into z-scores and averaged to serve as a composite score.

Each correct semantic judgment and each final word were awarded one point, and each incorrect answer a zero. The researcher had set up a rule for spelling errors: If one letter is missing (e.g., acident instead of accident; parent instead of parents) or added to the word (e.g., grounds instead of ground) with the right phonological properties intact and a clear attempt to spell the word correctly, it is treated as a minor spelling error and is not penalized. However, no spelling errors occurred.

As previously mentioned, the task exhibited high internal consistency. As for reaction time that represents processing speed, it would be ideal to obtain test-retest reliability through multiple trials, which is often done in the medical field (Benesch, Pütz, Rosenbaum, & Becker, 2000; Mercer, Hankins, Spinks, & Tedder, 2009). In the current study, however, running multiple trials of the reading span task was unfortunately not feasible.
**L2 Knowledge Measures.** Although WM and topic familiarity are of primary interest to the current study, there is no doubt that L2 knowledge affects L2 reading comprehension (Yamashita & Shiotsu, 2017). Therefore, it was important to take L2 knowledge into account when investigating the interaction between WM and topic familiarity in L2 reading comprehension. Administering a standardized English proficiency test such as the TOEFL would have been ideal. However, after taking time constraints and the fatigue factor into consideration, the researcher decided to utilize more economical measures that have been used as reliable proxies for general proficiency by others as well as in the pilot study, specifically the VST and C-test.

Shin and colleagues (2019) mentioned that much of the variance in reading comprehension captured by the C-test overlapped with what the reading span task could capture, warning against using the C-test as a single language proficiency control. However, the C-test scores in the pilot were the strongest predictor of L2 reading comprehension ($r = .63$) followed by vocabulary size ($r = .58$). These two L2 knowledge measures jointly explained 69% of the variance in L2 reading comprehension (in the background knowledge condition). Therefore, the current study kept both of the measures to estimate L2 general proficiency. Below, the measures are described in detail.

*C-test.* The C in the C-test stands for the word ‘cloze’ because C-tests have a format similar to that of a cloze test. However, C-tests are different from the classical cloze test in that it uses what Klein-Braley (1997) calls the C-Principle by which “deletions are not performed at the text level but at the word level” (p. 64). The first sentence of each text remains unchanged. From the second sentence, the second half of
every second word is deleted (e.g., The mo____ interesting res____ of). For a word that includes an odd number of letters, the larger part is deleted (e.g., mix____ for “mixture”).

Eckes and Grotjahn (2006) showed that a C-test is a reliable general language proficiency measure ($\alpha = .80$ or higher) and is highly correlated with other language proficiency tests such as the TOEFL ($r = .55$ to .94) and TOEIC ($r = .62$). C-tests are not only easy to administer and score, but they also take test-takers less than 30 minutes to complete. In fact, in the current study, it took less than 20 minutes for participants to complete it. The same C-test from Klein-Braley (1997) was adopted in the current study (see Appendix J). It consisted of four short texts with 25 blanks per text to be filled in. Each correct response was awarded one point, and no partial scores were allowed. The maximum score was 100. The internal consistency was high ($\alpha = .85$).

_Vocabulary Size Test (VST)._ An additional measure used to supplement the C-test was the VST (Nation & Beglar, 2007), first developed by Nation and Beglar in 2007 and later validated by Beglar (2010). It has been acknowledged that vocabulary size is a good predictor of not only reading skills (Jeon & Yamashita, 2014) but also general proficiency (Stæhr, 2008). Milton, Wade, and Hopkins (2010), for instance, showed that vocabulary size can explain 40 to 60% of variance in the four language skills. The VST is known to be a reliable comprehensive measure that covers receptive vocabulary size of English from 1,000 to 14,000 word families based on the British National Corpus (BNC). The VST is freely available (http://my.vocabularysize.com).

In the current study, the English-English test version was used. The test consisted of 140 multiple-choice items in total, with 10 items representing each 1,000 word level. Each target word was used in a sentence in a simple non-defining context and shown to
test-takers who then chose the meaning of the target word from among four options. The VST took 20-25 minutes to complete. Because the online version of VST simply provided the final outcome (i.e., test-takers’ vocabulary size), the researcher was not able to calculate its reliability. According to Beglar (2010), however, the VST items provide a high degree of precision as evidenced by low standard errors ($SE = 1.0-1.8, M = 1.6, SD = .3$) and high reliability estimates (Rasch item reliability = .96 - .98).

**Background Texts in L1.** As stated above, the current study created two conditions – reading with and without topic familiarity. In order to do so, the researcher provided participants with background texts with the aim of purposefully raising participants’ familiarity with two unfamiliar topics (Teotihuacán, the Bantu people). Hence, the extent of participants’ familiarity with the selected topics was controlled. Note that the researcher made sure these were unfamiliar topics to the participants by checking their background knowledge in the questionnaire. The background texts were prepared in participants’ L1s (Japanese and Korean). The reason for doing so was to avoid exposure to related L2 vocabulary and to ensure that linguistic knowledge did not become a confounding factor. These texts, used in Shin and colleagues (2019) for their Korean participants, had been adapted from two Korean informational texts found in the Online Encyclopedia of Cultural Heritage and World History (terms.naver.com). Because the current study also included Japanese participants, the Korean texts were translated into Japanese. These texts introduced the two topics to both Korean and Japanese participant groups, without extensive overlap with the content of the passages in the testlets. An English version was also created not for the current study but for English-speaking readers of this study (see Appendices K, L, and M).
The Korean version of the two background texts contained 290 and 292 words. The Japanese version contained 1,235 and 1,009 characters. The two texts in Korean and Japanese were proofread and were compared to the English version by two native speakers of each language who are fluent in English. Additionally, a native Korean speaker who is fluent in Japanese checked the Korean and Japanese versions to ensure comparability.

Cognates were searched for in both Korean and Japanese background texts to prevent one particular L1 group from having a larger advantage by being exposed to key content words than the other group. The Korean texts were examined by the researcher, a native Korean speaker. The Japanese texts were examined by a native English speaker who is fluent in Japanese. The cognates found in the Korean background texts matched those in the Japanese background texts. They were mostly names of places or ethnic/cultural groups (e.g., Bantu, Africa, Congo, Tanzania, Mexico, Teotihuacán, Rome, Aztec) and a name of a food (i.e., banana).

Participants were instructed to read the texts carefully three times and answer five comprehension-check questions, also written in their L1, at the end. In the pilot study, participants were asked to read the texts twice on their own. No explicit comprehension check was carried out at any point, leaving open the question as to whether or not participants thoroughly read the texts and absorbed the information. To improve on this limitation, the background texts in the current study included five comprehension questions, results of which were later used to ensure that the participants paid enough attention to and comprehended the background texts. Those who did not correctly answer
more than three questions on each text were to be eliminated from the participant pool, though all participants got either four or all five of the questions right.

**Topic Familiarity Task.** To obtain a more accurate estimate of participants’ topic familiarity on the day of the reading comprehension test, a topic familiarity task was administered right before the reading comprehension test. It was anticipated that participants’ topic familiarity of Teotihuacán and the Bantu people increased due to the background texts that they were required to have read one day prior to the day of testing. On the other hand, their familiarity with the history of the chickenpox vaccine and ancient Rome and Greece was expected to remain the same as when they completed the first questionnaire.

As part of this topic familiarity task, participants were asked to write down as much information as possible about the four target topics in two minutes (see Appendices N, O, and P). They were allowed to write single words, phrases, or sentences in their L1. Two raters coded the number of unique details included in written responses by dividing prose into segments that represent meaningful details/ideas. Each unique detail or idea unit was awarded one point as was done in a number of studies that examined free text recall (e.g., Bernhardt, 1983; Chung-Fat-Yim, Peterson, & Mar, 2017). The scoring protocol is provided in Appendix Q.

The responses from Japanese participants written in Japanese were translated into Korean so that two raters, both of whom are native Korean speakers, could read and code the responses. Because the languages of Korean and Japanese share remarkably similar syntactic structures especially in terms of word order, word-for-word translation of Japanese responses corresponded to phrases expressing the same idea in Korean.
Pilot. To ensure reliability of the scoring scheme, the topic familiarity task was piloted with two topics (Teotihuacán and the Bantu people) for which participants’ familiarity was to be raised by the researcher. Six native Korean speakers independently read the Korean background texts on Teotihuacán and the Bantu people three times each and answered the comprehension questions at the end to check their understanding. The following day, they were asked to write as much as they know about the two topics in two minutes, mirroring the actual procedure of the current study. Because the participants were in Korea, photos of their responses were sent to the researcher via email.

The researcher, a native speaker of Korean, served as the first coder of the responses. For the second coder, a native speaker of Korean, enrolled in the MA TESL program at the same university as the researcher, was recruited. At the norming session, the coders read the Korean background texts on Teotihuacán and the Bantu people first. Then, the researcher explained the scoring protocol stated above – awarding one point for each unique detail/idea. The coders coded one participant’s recall response (6 sentences) together to find any sources of inter-rater disagreement. This led to an agreement that each content word should be given one point most of the time and a discussion of how to deal with ambiguous cases (e.g., compounds). The protocol (Appendix Q) was made based on the agreements between coders and their experiences with ambiguous cases. The rest of the coding was carried out independently. Inter-rater agreement was 96%.

L2 Reading Quiz. The researcher recognizes that reading comprehension can be better represented with a measure that taps the two levels of comprehension – literal and inferential – to match the operationalized construct (see Chapter 2 for a detailed discussion). Correspondingly, the TOEFL reading sections contain test items for
determining the meaning for vocabulary, locating information, recognizing whether specific information is stated in a passage, and paraphrasing, all of which can reflect readers’ literal comprehension (Jude & Ajayi, 2012; Saadatnia, Ketabi, & Tavakoli, 2017). For inferential comprehension, the TOEFL reading section includes items that require readers to search for relationships among pieces of information, such as interpreting the author’s intention, filling in omitted details, making inferences, and summarizing (Basaraba et al., 2013).

The L2 reading comprehension test used in the current study, labeled “Reading quiz” to reduce pressure on participants, contained four testlets found from two sources: TOEFL iBT Quick Prep provided by Educational Testing Service (ETS) (https://www.ets.org/toefl/ibt/prepare/quick_prep/) and Magoosh TOEFL reading practice questions (https://toefl.magoosh.com/), a company that provides practice for tests such as TOEFL, GRE, IELTS in printed books and online. Each of the testlets consisted of one passage (Testlet 1: the rise of Teotihuacán; Testlet 2: agriculture, iron, and the Bantu people; Testlet 3: the history of the chickenpox vaccine; Testlet 4: ancient Rome and Greece) accompanied by six literal-comprehension items and four inferential-comprehension items. The details of each testlet passage and item specifications are presented in Table 3.1. It should be noted that Table 3.1 provides information of the finalized form of the reading quiz after two rounds of modifications, which is elaborated on further below.
Table 3.1

*Table of Item Specifications*

<table>
<thead>
<tr>
<th>Testlet #</th>
<th>Passage detail</th>
<th>Literal comprehension questions</th>
<th>Inferential comprehension questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Identifying negative facts</td>
<td>Finding factual info</td>
</tr>
<tr>
<td>1</td>
<td>The rise of Teotihuacán</td>
<td>1, 3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>a. K1 (%): 78.95; K2 (%): 4.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. AWL (%): 4.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Length: 627 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Grade level: 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Agriculture, iron, and the Bantu people</td>
<td>18</td>
<td>12, 13</td>
</tr>
<tr>
<td></td>
<td>a. K1 (%): 79.22; K2 (%): 4.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. AWL (%): 4.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Length: 645 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Grade level: 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>History of the chickenpox vaccine</td>
<td>21, 24</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>a. K1 (%): 76.79; K2 (%): 4.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. AWL (%): 4.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Length: 655 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Grade level: 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Ancient Rome and Greece</td>
<td>32, 33</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>a. K1 (%): 74.04; K2 (%): 4.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. AWL (%): 4.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Length: 624 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Grade level: 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of items</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Points/Item</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Points/Item type</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Points/Level of comprehension</th>
<th>24</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>44</td>
<td></td>
</tr>
</tbody>
</table>

*Notes.* K1: 1-1000 word family; K2: 1001-2000 word family; AWL: Academic word list; Off-list words are not included.
The reading quiz incorporated literal and inferential comprehension questions were divided into four types of items for each level of comprehension. Item types for literal comprehension were (a) identifying negative facts, (b) finding factual information, (c) determining the meaning for a vocabulary word, and (d) restating a sentence. All of these item types required the participants to read for exact meaning and retrieve information explicitly stated in passages in Testlets 1 through 4. There were six of such items included in each of the testlets (1 or 2 items for identifying negative facts, 1 or 2 items for finding factual information, 2 items for vocabulary, and 1 item for sentence restatement). With respect to inferential comprehension, the four item types included in the test were (a) inferring rhetorical purpose questions, (b) making inferences using information in the passage, (c) inserting sentences into the passage, and (d) selecting sentence to complete a summary. Each of these item types was included in Testlets 1 through 4. The inferential items required readers to make connections among ideas and integrate multiple pieces of information.

First Round of Modifications. This subsection briefly describes the first round of modifications of the reading quiz made in preparation of the pilot study (Shin et al., 2019). Most of the changes were made on the passages after the instructor of pilot participants expressed concerns regarding the gap between participants’ levels of proficiency and the difficulty level of the original TOEFL reading passages. In order to increase the readability of the original passages, an ESL/EFL instructor with nine years of teaching experience modified the passages focusing on (a) rephrasing several sentences to simplify the grammatical structure, (b) shortening longer sentences, and (c) replacing low-frequency words with high-frequency ones based on his intuitions. After these
modifications, the passages were at grade levels 9 and 10, included 643 to 699 words and 29 to 34 academic words, and had a lexical density ranging from .55 to .57.

In addition to these changes, some adjustments to items were made as well. The original TOEFL test included 14 reading comprehension items per passage. Due to time constraints, however, the researchers of the pilot study (Shin et al., 2019) decided to delete four items per passage. The ten items per passage remained the same as the original, consisting of the same type of items in a multiple-choice format across all four testlets. The types of items in the reading comprehension test used in the pilot study mirror the item types included in the current study. The test exhibited overall high reliability ($\alpha = .85$).

Second Round of Modifications. To provide further evidence of test reliability, another round of modifications was made. The subsections below describe (a) item analysis using the pilot data, (b) revisions made based on the item analysis, and (c) further modifications of passages.

Item Analysis. The purpose of conducting item analysis was to ascertain performance of individual items on the reading quiz and to ensure comparability of the four testlets with respect to item difficulty and discrimination. Because the main goal of this item analysis was to ensure the sufficient spread of students along a continuum of reading comprehension skills, norm-referenced testing (NRT) item analysis was deemed appropriate. Therefore, item facility and item discrimination were examined.

Item facility (IF), the proportion of test takers who answered a particular item correctly, is often used as an index of item difficulty or ease. IF values around .50 are considered ideal in NRT testing with an acceptable range between .30 and .70 (Carr,
In non-high-stakes testing, the values can be farther from .50. IF of .30 and below indicates that the item is too difficult, and IF of .70 and above is considered too easy.

In terms of item discrimination, correlational and subtractive approaches were used to examine point-biserial and upper-lower (group) discrimination (ID_{UL}), respectively. Point-biserial values show the relationship between participants’ performance on a particular item and overall scores on the test. Point-biserial values above .30 are ideal and above .15 are considered reasonable in classroom assessment or low-stakes testing. ID_{UL} is calculated by subtracting the IF of a low performing group (approximately bottom 25%) from the IF of a high performing group (approximately top 25%) to show the extent to which items are able to discriminate strong and weak students. ID_{UL} of .20 and above is acceptable. The researcher adhered to the acceptable ranges for low-stakes testing.

As summarized in Table 3.2, item analysis using pilot data demonstrated that performance of items overall and for each testlet was fairly strong. The average IF value and IFs for all four testlets are close to .50, showing that the items are overall well centered. The IF values for Testlets 1 and 2 are slightly higher than those for Testlets 3 and 4, indicating that items in Testlets 1 and 2 were easier. This is not surprising because participants were able to raise familiarity with the topics of Testlets 1 and 2 and thus scored higher on these two than on Testlets 3 and 4. Additionally, item discrimination expressed as point-biserial and ID_{UL} showed that items of each testlet performed well in general in terms of separating participants with stronger and weaker comprehension.
Table 3.2

*Item Analysis on Reading Comprehension Test Items*

<table>
<thead>
<tr>
<th>Testlet #</th>
<th>Item facility (IF)</th>
<th>Item discrimination</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Point-biserial (Correlational)</td>
<td>ID\textsubscript{UL} (Subtractive)</td>
</tr>
<tr>
<td>1</td>
<td>.53</td>
<td>.30</td>
<td>.35</td>
</tr>
<tr>
<td>2</td>
<td>.53</td>
<td>.36</td>
<td>.44</td>
</tr>
<tr>
<td>3</td>
<td>.43</td>
<td>.24</td>
<td>.34</td>
</tr>
<tr>
<td>4</td>
<td>.44</td>
<td>.33</td>
<td>.41</td>
</tr>
<tr>
<td>Average</td>
<td>.49</td>
<td>.31</td>
<td>.39</td>
</tr>
</tbody>
</table>

However, six problematic items were identified in further analysis (see Table 3.3).

In terms of item difficulty, Item 1 was found to be too easy for the participants. The IF value of .95 indicates that 95% of the students got this item correct. On the other hand, Items 9 and 14 were too difficult for the participants (IF = .19 and .17, respectively); less than 20% of the participants got them right. Item 14 also showed low item discrimination power (Point-biserial = .14; ID\textsubscript{UL} = .11), contributing little to the reliability of the test.

The item statistics for Item 23 is particularly alarming (Point-biserial = -.07; ID\textsubscript{UL} = -.04). Because negative discrimination is an indication that the item is harming test reliability (Carr, 2011), further examination was required. Other items that showed poor discrimination are Item 22 (Point-biserial = .10; ID\textsubscript{UL} = .15) and Item 31 (Point-biserial = .13; ID\textsubscript{UL} = .11), calling for attention as well.

Table 3.3

*Flagged Items*

<table>
<thead>
<tr>
<th>Item #</th>
<th>Item facility (IF)</th>
<th>Point-biserial</th>
<th>ID\textsubscript{UL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.95*</td>
<td>.21</td>
<td>.15</td>
</tr>
<tr>
<td>9</td>
<td>.19*</td>
<td>.47</td>
<td>.41</td>
</tr>
<tr>
<td>14</td>
<td>.17*</td>
<td>.14*</td>
<td>.11</td>
</tr>
<tr>
<td>22</td>
<td>.27</td>
<td>.10*</td>
<td>.15</td>
</tr>
<tr>
<td>23</td>
<td>.22</td>
<td>-.07*</td>
<td>-.04*</td>
</tr>
<tr>
<td>31</td>
<td>.40</td>
<td>.13*</td>
<td>.11</td>
</tr>
</tbody>
</table>

*Note.* The asterisk (*) indicates that the value is not in the acceptable range.
Test Revision and Addition of a Practice Quiz. After close examination of the abovementioned problematic items, two decisions were made: (a) Revision of the test by either completely changing a couple of items or revising an existing item and (b) inclusion of a practice test (Appendix R). The researcher created a new item to replace one item in Testlet 3. The last items in Testlets 1, 2, and 4 require completing a summary. The last item in Testlet 3, however, was slightly different in that it required participants to fill in information to compare two diseases. Seeing that the rest of the items in all four testlets were consistent in terms of item type, a summary completion item for Testlet 3 was created for the purpose of maintaining consistency throughout. This new item was included as Item 30 in Testlet 3.

The researcher also replaced one item. Item 1 – the easy item – was a vocabulary question that asked the meaning of the word massive. Judging from the IF value, most of the students probably knew the meaning of this word. Its stem “mass” is at the 1000 word family level. Therefore, Item 1 was replaced with one of the vocabulary questions in the original TOEFL test, which had been deleted from the pilot study test to reduce the number of items per passage. The newly included item asked for the meaning of the word predominant, a vocabulary word listed on the academic word list (AWL). The item was added as Item 4 in Testlet 1 (see Appendix S).

As for Item 23, which had a negative discrimination value, the distractors were revised. Two of the distractors seemed to be too similar to each other: A. They took longer than expected. C. They took a long time to finish. Although neither of them was the answer, the fact that they are highly attractive distractors and that they have the same...
meaning might have affected participants’ performance on this item. Thus, Option A was changed to *They were supported by the Japanese government*.

The rest of the flagged items did not seem to have problems in themselves. However, based on the inspection of the remaining four flagged items (Items 9, 14, 22, and 31) and a consultation with an ESL instructor, these items may have presented additional challenge to students because most participants had not been exposed to the TOEFL nor those types of reading comprehension questions. For instance, Item 9 required students to find the best place to insert a sentence in the paragraph; Item 14 asked about author’s intention; Item 22, which showed near-zero discrimination, required students to find the best paraphrased sentence that did not distort the meaning or leave out essential information; Item 31 was the same type of question as Item 22.

With the goal of familiarizing participants with the scope of item types on the actual reading comprehension test, a short practice test was concocted (see Appendix R). Translating the questions into participants’ L1s could have eased the process of understanding what was being asked of them. However, it would not be advisable to ignore possible confusion that could be caused by constant code-switching. Thus, the researcher decided that having participants engage in a practice test, labeled as “practice quiz,” with the same types of items that they would encounter later may best mitigate this issue.

The practice quiz includes a three-paragraph passage accompanied by seven questions representing each different item type. Although there were eight item types included in the actual reading quiz, identifying negative facts and finding factual information can be broadly categorized as retrieving factual information. Therefore, only
one question that required retrieving factual information was included, hence seven questions in total. The passage came from part of a text that Suk (2017) adopted from *Reading Power* (Mikulecky & Jeffries, 2004) to assess reading comprehension. Because the purpose of the practice quiz was to familiarize students with the test format and item types that may be new to them, the questions were devised to be relatively easy to answer. As intended, text analysis using the Vocab Profiler (lextutor.ca/vp/eng/) showed that this passage was much easier to understand than the passages in the actual reading quiz (see Table 3.4). Some of the words in the questions that may be unfamiliar to students (e.g., *infer*) were put in the margins with their definitions in students’ L1s (Korean and Japanese). Also, three options for each question were created instead of four to expedite the practice process.

This practice quiz was presented to participants right before the actual reading quiz. They were informed about the purpose of the practice quiz and were encouraged to ask any questions regarding test items and check the answers attached to the test.

Table 3.4

*Comparison across Passages in the Revised Reading Quiz*

<table>
<thead>
<tr>
<th>Testlet #</th>
<th>Grade level</th>
<th>Length</th>
<th>K1 Type (Token)</th>
<th>K2 Type (Token)</th>
<th>AWL Type (Token)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practicea</td>
<td>7</td>
<td>195</td>
<td>94 (171)</td>
<td>11 (12)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Testlet 1b</td>
<td>9</td>
<td>627</td>
<td>191 (495)</td>
<td>23 (30)</td>
<td>25 (30)</td>
</tr>
<tr>
<td>Testlet 2b</td>
<td>9</td>
<td>645</td>
<td>199 (511)</td>
<td>20 (31)</td>
<td>25 (30)</td>
</tr>
<tr>
<td>Testlet 3c</td>
<td>9</td>
<td>655</td>
<td>195 (503)</td>
<td>23 (32)</td>
<td>25 (32)</td>
</tr>
<tr>
<td>Testlet 4b</td>
<td>9</td>
<td>624</td>
<td>191 (462)</td>
<td>23 (25)</td>
<td>24 (31)</td>
</tr>
</tbody>
</table>

*Notes.* K1: 1-1000 word family; K2: 1001-2000 word family; AWL: Academic word list; Type: total number of different words; Token: the total number of words; Off-list words are not included in this table; aPassage is from Mikulecky and Jeffries (2004); bPassages are from ETS (https://www.ets.org/toefl/ibt/prepare/quick_prep/); cPassage is from Magoosh TOEFL reading practice questions (https://toefl.magoosh.com/)
Further Modifications of Passages. The passages in the reading quiz were further modified to ensure comparability across four testlets. The initial modifications in the pilot study (Shin et al., 2019) focused on splitting longer sentences to shorten them, simplifying grammatical structure, and substituting low-frequency words with high-frequency words. This time around, the focus was on ensuring that all four passages had approximately the same readability levels and similar lexical coverage with regard to proportions of K1, K2, and academic words included in the passages. First, the number of K1, K2, and academic words was identified using the Vocab Profiler (lextutor.ca/vp/eng/). Efforts were made to reduce discrepancies among passages especially by further modifying passages that had more K2 words and academic words compared to the other passages. As a result, two passages that initially indicated the grade level of 10 (Testlets 1 and 4) showed that they were at the grade level 9 after modification. Overall, the passages became more comparable; they indicated the same readability (grade level 9) and contained 624 to 655 words. K1 words constituted 74.04 to 79.22% of the passages, K2 words from 4.01 to 4.89%, and academic words from 4.65 to 4.97% (see Table 3.4).

Scoring. One point was assigned to each item, except for the summary completion item, which was worth two points, because it required selecting appropriate sentences, from among six choices, to fill in three blanks to create an accurate summary. Two points were awarded for selecting all three correct sentences, 1.5 points for selecting two correct sentences, 1 point for selecting one correct sentence, and zero for no correct sentence. The answer key is attached in Appendix T.
Procedures

The current study included a two-phase administration of measures (see Table 3.5). Each phase occurred on different days depending on volunteers’ availability. The volunteers were asked to sign up for one hour for Phase 1 on one day of their choice and for two hours for Phase 2 on a different day. The instruments administered in Phase 1 were the questionnaire, reading span task, C-test, and VST. The reading span task and VST required a computer. The researcher had access to two computers. Therefore, this phase was carried out with two students at a time. Data collection in Japan spanned four weeks mainly because the researcher scheduled to meet with each participant twice face-to-face. Data collection in Korea spanned five weeks for the same reason. After the participants completed Phase 1, they were informed that they would be receiving text messages from the researcher one day prior to their Phase 2 day and that they would be asked to read short stories in their L1s in the messages. Phase 1 was completed in approximately 1 hour. The participants were informed that they would receive compensation for their time after they completed Phase 2.

After Phase 1 and before Phase 2, a topic familiarity pre-check occurred for screening purposes, based on questionnaire responses: Participants who possessed prior knowledge of the topics of the texts were excluded from the participant pool for the current study. Without informing them of the decision, the researcher let these students continue. Their results, however, were not used for data analysis. Results reported below do not include these students who got eliminated but only include the 158 participants who passed the screening (138 in intervention 20 in control group).
Phase 2 began with participants reading two background texts on Teotihuacán and the Bantu people in their L1s, which was an independent task. The researcher contacted participants through a mobile instant messaging application that was commonly used in the country and sent them the passages with instructions. They were also asked to answer comprehension-check questions and send them to the researcher. This task took approximately 20 minutes. On the following day, participants completed the topic familiarity task and took the practice and reading quiz in a small group setting based on their availability. These two tasks took about 1 hour and 30 minutes. Phase 2 was completed in about 2 hours, and in total, the entire set of procedures (Phases 1 and 2) took approximately 3 hours.

All participants, except for 20 participants in the control group, completed all tasks. The last 10 volunteers to show up to participate in the first 10 days of the data-collection period in both Japan and Korea were placed in the control group (i.e., 10 Japanese, 10 Korean participants). As previously mentioned, these 20 participants did not receive the background texts. Their results were used only to ascertain the comparability of the L2 reading testlets. They completed all the other tasks without being aware that they were placed in a group that is different from their experimental group peers.

All participants were encouraged to ask questions regarding the tasks and procedures and to take a break in between tasks. They were assured that their performance would not affect their English course grades in any way, nor would their instructors gain access to the results. Japanese participants received compensation in the form of a gift card, and Korean participants in the form of physical gifts. The difference in the types of compensation was due to logistical constraints.
Table 3.5

Order and Duration of Measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>Setting</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Two at a time</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Reading span task</td>
<td></td>
<td>15 – 20 minutes</td>
</tr>
<tr>
<td>VST</td>
<td></td>
<td>20 – 25 minutes</td>
</tr>
<tr>
<td>C-test</td>
<td></td>
<td>15 – 20 minutes</td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>70 minutes (max)</td>
</tr>
<tr>
<td>Phase 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background texts*</td>
<td>Independent</td>
<td>20 minutes</td>
</tr>
<tr>
<td>Topic familiarity task</td>
<td>Group</td>
<td>10 minutes</td>
</tr>
<tr>
<td>Practice &amp; Reading quiz</td>
<td></td>
<td>90 minutes</td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>120 minutes (max)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>190 minutes (approximately 3 hours)</td>
</tr>
</tbody>
</table>

Note: *The control group (n = 20) did not receive the background treatment but completed all the other tasks. The intervention group (n = 138) completed all tasks.

Data Analysis

This section describes data analysis used to answer the research questions. The first part discusses the pre-analyses procedures that needed to be done. The second and third parts address data-analyses procedures for Research Questions 1 (RQ1) and 2 (RQ2), respectively.

Pre-analysis procedures

A series of statistical tests was performed to analyze the data. Before answering RQ1, the first step would be to substantiate that participants indeed performed in two different conditions – reading with and without topic familiarity as was done in Shin and colleagues (2019). To achieve this, a paired-samples t-test was run to test for a difference between participants’ self-ratings of their familiarity with two familiar and two unfamiliar topics. Written responses collected in the post-reading topic-familiarity task
were examined as well. Responses in Japanese were translated into Korean by a Japanese-Korean bilingual. This decision, instead of translating both Korean and Japanese responses into English, was made for three reasons:

(a) because Korean and Japanese share considerable similarity in their syntactic and morphological structures (e.g., word order, case-marking system), simple word-by-word translation of one language into the other could generate the same meaning;

(b) due to the linguistic structural differences between the English language and participants’ L1s (Korean and Japanese), translating participants’ written responses into English was likely to result in losing certain information; and

(c) it was more time-efficient to translate responses in one language into the other than translate two languages into English.

Two native Korean speakers (the researcher and another coder) coded the topic-familiarity written responses. As mentioned earlier, the researcher had trained the second coder by explaining the predetermined protocol and holding a norming session during which they coded sample sentences together from the data. Based on the norming session and discussion, the protocol was created (Appendix Q). After data collection, the researcher held a brief discussion session to explain the procedure. The raters coded responses independently. There were only a few discrepancies in the raters’ coding, which were resolved after discussion. The inter-rater agreement was initially 98% and 100% after discussion. In what follows, procedures for data analyses for RQ1 and RQ2 are described.
Data-analysis for RQ 1: Is there an interaction between working memory and background knowledge in L2 reading comprehension? If so, which one of the models (compensation, independent-influences, and the rich-get-richer) best explains the interaction? To answer RQ 1, a mixed-effects modeling approach was used given the repeated-measures portion of the study (reading in two conditions).

Mixed-effects models have been increasingly adopted in L2 research (e.g., Bowden et al., 2010; Tremblay et al., 2011) due to the flexibility they offer. A mixed-effects model can contain fixed and random-effect components in the same analysis, hence the name mixed-effects. Fixed effects are independent variables of primary interest, which can include categorical or continuous variables, or a mixture of the two. Random effects represent random variation that may not be of primary interest but arises in data analysis often due to random variance across the participants sampled (Cunnings, 2012). When included as random effects in a mixed-effects model, possible clusters of participants can be accounted for (Barr et al., 2013; Linck & Cunnings, 2015). By doing so, the model also recognizes that that there are multiple responses per subject (i.e., repeated measures) and accounts for this source of variation in a single model. Multiple measurements per subject can result in correlated errors, which violates the assumptions of ANOVA models. Mixed-effects-model analysis, however, provides a flexible approach that allows a wide variety of correlation patterns to be modeled (Seltman, 2013). Furthermore, mixed-effects models do not assume homoscedasticity or sphericity in the data (Linck & Cunnings, 2015). Another benefit of using a mixed-effects model, in this current study, was that the model could treat WM as a continuous variable, as opposed to a categorical variable, and still examine its interaction with topic familiarity. Dichotomizing a
continuous variable (i.e., using ANOVA) when it is correlated with another predictor variable is not recommended because it results in a loss of statistical power (Cohen, 1983; Cohen, Cohen, West, & Aiken, 2003). Therefore, using a mixed-effects model, which overcomes the limitation of ANOVA, was the most appropriate choice for the current study.

In the mixed-effects model for RQ1 in the present study, WM, topic familiarity, and the interaction between them were modeled with fixed effects to predict the dependent variable, L2 reading comprehension. Random intercepts for participants and items were included in the model because there were two kinds of responses for L2 reading comprehension per participant due to the two reading conditions created.

To present a visual illustration of the interaction pattern, particularly for the purpose of comparing the findings of the current study to previous ones (Alptekin & Erçetin 2011; Leeser, 2007; Shin et al., 2019), the researcher divided participants ($n = 138$) into three groups according to their WM scores (i.e., low, mid, high) and compared L2 reading comprehension of low- and high- WM groups ($n = 46$, respectively) with the mid-group taken out. Converting continuous data into categorical groups is often not recommended as it can result in losing some information and reducing statistical power (MacCallum et al., 2002). However, the purpose of this follow-up step was to simply present a picture of the interaction visually and to compare the resulting pattern with previous studies.

Data-analysis for RQ 2: To what extent do the subcomponents of working memory contribute to L2 reading with and without topic familiarity? Similar steps were taken to analyze data to answer RQ2. This analysis essentially sought to explore the
effects of WM subcomponents and their potential interactions with topic familiarity in L2 reading comprehension. Therefore, another mixed-effects model was fitted, using the subcomponents – storage, processing accuracy, processing speed – and topic familiarity as fixed effects and keeping the random intercept for participants in the model. The effect of L2 measures was included in an additional analysis to examine unique contributions of WM subcomponents and topic familiarity to L2 reading comprehension.

In this chapter, methods used for this dissertation study are described while providing information on the research design, variables of interest, participants, instruments, procedures, and data analysis. In the next chapter, Chapter 4, results of data analysis performed for Research Questions 1 and 2 are presented.
Chapter 4: Results

In this chapter, results of the data analysis are presented with respect to (a) an interaction between working memory (WM) and topic familiarity in L2 reading with two L1 groups of English learners (Korean, Japanese) and (b) the extent to which the subcomponents of WM contribute to L2 reading with and without topic familiarity. First, descriptive statistics for the main variables and measures are reported in each subsection. Second, the two research questions are answered. Following these findings are the results of further analyses, which were conducted to examine whether results within the two L1 groups conform to the overall picture presented.

Study Variables and Measures

The following reports descriptive results of the key variables: topic familiarity (pre- and post-check), WM, L2 knowledge, and L2 reading comprehension.

Topic Familiarity: Pre-check. At the beginning of the study, participants were asked to indicate their levels of familiarity with four topics (ancient Rome and Greece, Bantu people, history of the chickenpox vaccine, Teotihuacán) on a scale of one to four. The purpose of this measure was to determine if any potential participants would be eliminated from the participant pool. On the pre-check questionnaire, most of the participants indicated that they were unfamiliar with the four topics by marking 1 (“I can tell you nothing about this topic.”). Three participants marked 2 (“I can tell you a little about this topic.”) for history of the chickenpox vaccine and Teotihuacán, but they did not provide any elaborative written responses. Twenty-four participants marked 2 for the Ancient Rome and Greece topic. Four of the 24 respondents, however, did not provide any written responses to elaborate. Out of the 20 participants who did provide written
responses for what they knew about the topic, 12 wrote that they were familiar with Roman mythology, with 10 of them simply writing “Roman mythology” but nothing else. Responses related to Roman mythology were considered to be irrelevant because they were only about the legendary tales of the gods of the Greeks and Romans and their human qualities. Therefore, these participants were kept in the pool. Out of the remaining eight, five participants indicated minimal knowledge of the topic (e.g., “There were many philosophers”, “Famous for architectural structures”) by using no more than six content words. Considering the vagueness and limited length of the responses, these participants were kept in the pool. However, three participants who demonstrated extensive and specific knowledge of architectural development and leading philosophers of ancient Rome and Greece were eliminated from the sample. After eliminating these three, 138 participants were left in the pool (i.e., intervention group). Table 4.1 reports the mean self-ratings of the participants (n = 138) who screened in for the next and all phases of the study. As seen in the table, the ratings range from 1 to 1.17. No statistical differences were found among the means for the four topics.

Table 4.1

<table>
<thead>
<tr>
<th>Topic Familiarity Pre-check Ratings (n = 138)</th>
<th>Ancient Rome and Greece</th>
<th>Chickenpox vaccine</th>
<th>Bantu people</th>
<th>Teotihuacán</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1.17</td>
<td>1.02</td>
<td>1</td>
<td>1.02</td>
</tr>
<tr>
<td>SD</td>
<td>0.38</td>
<td>0.15</td>
<td>0</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Note. Results from a 1-4 scale questionnaire, with 1 being “I can tell you nothing about this topic” and 4 being “I can tell you a lot about this topic.”

**Topic Familiarity: Post-check.** Prior to the L2 reading quiz (the day after participants read background texts), participants were instructed to complete the same
topic familiarity task. The purpose of this step was to confirm that the participants had become familiar with two designated topics as intended and that they would be performing on the L2 reading quiz in two conditions: with and without topic familiarity.

As shown in Table 4.2, participants’ mean ratings for the Bantu people (2.01) and Teotihuacán (1.91) topics were higher than those for the other two topics featured on the L2 reading quiz (1.15 and 1, respectively), showing increased familiarity with the two designated topics. The paired-samples t-test showed that the post-check ratings were not different from the pre-check ratings for the two topics in the no-topic familiarity condition but that they significantly increased for the two topics in the other condition.

Table 4.2

<table>
<thead>
<tr>
<th></th>
<th>No-topic-familiarity condition</th>
<th>Topic familiarity condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ancient Rome and Greece</td>
<td>Chickenpox vaccine</td>
</tr>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>M</td>
<td>1.17</td>
<td>1.15</td>
</tr>
<tr>
<td>SD</td>
<td>0.38</td>
<td>0.36</td>
</tr>
<tr>
<td>t</td>
<td>1 (p = .32)</td>
<td>1.74 (p = .09)</td>
</tr>
</tbody>
</table>

Note. Results from a 1-4 scale questionnaire, with 1 being “I can tell you nothing about this topic” and 4 being “I can tell you a lot about this topic.”

In addition, the same pattern was found in participants’ written responses. Although 18 of 138 participants left the topic familiarity written-response part blank, all of them reported that they had read the background passages the day before and that they would recognize the information if it was presented again. The scores of the written responses were represented by the number of content words included in sentences. Table 4.3 summarizes the results of 120 participants’ written-response scores.
Table 4.3

Scores on Written Responses at Post-check

<table>
<thead>
<tr>
<th>No-topic-familiarity condition</th>
<th>Topic familiarity condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancient Rome and Greece</td>
<td>Bantu people</td>
</tr>
<tr>
<td>Chickenpox vaccine</td>
<td>Teotihuacán</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
</tr>
<tr>
<td>Min</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>6</td>
</tr>
<tr>
<td>Each topic</td>
<td>0</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
</tr>
<tr>
<td>Min</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>6</td>
</tr>
<tr>
<td>Each condition</td>
<td>0</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
</tr>
<tr>
<td>Min</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. Scores represent the number of content words in written responses.

No statistical analysis for comparing the scores of written responses (i.e., number of content words) between the two conditions was deemed necessary for the following reasons. First, participants’ self-ratings clearly showed that two different reading conditions had been created as intended (see Figure 4.1).

Figure 4.1. Comparison of pre- and post-checks of topic familiarity (TF) self-ratings. (The asterisk indicates a significant difference with p-values below .001.)
Second, almost no written responses were given for the two topics in the no-topic familiarity condition (ancient Rome and Greece, chickenpox vaccine), which created non-normality in the data. Medians, instead of means, are reported in Table 4.3 for this reason. Furthermore, due to the lack of variability in the data particularly within the No-topic familiarity condition, it would be inappropriate to use this information as continuous data to represent topic familiarity. It should rather be used as a secondary piece of evidence to support the paired-samples t-test results above (Table 4.2), which compared participants’ self-ratings at pre- and post-check. Therefore, topic familiarity is treated as a categorical variable, that is, the absence and presence of it.

**Working memory.** This subsection reports on participants’ working memory (WM) performance. Before creating composite WM scores, three subcomponents of the reading span task (storage, processing accuracy, and processing speed) were examined separately. Composite scores were then created by transforming the subcomponent scores into z-scores and averaging them, following Shin and colleagues (2019) and Leeser (2007). Note that processing speed was multiplied by -1. Below are the summaries of scores of WM subcomponents and composite scores. The two L1 groups’ results are also reported and compared to determine if participants’ L1 was a factor that influenced working memory capacity.

*Scores on WM subcomponents.* To reiterate the scoring method of WM, the WM subcomponents from the reading span task were calculated in the following manner. First, storage scores were obtained by tallying up the number of final words reported correctly. Processing accuracy scores were obtained by tallying up the number of correct semantic judgments. Processing speed was calculated by averaging reaction times for the
correct semantic judgments and multiplying the values by -1 so that bigger numerical values would indicate faster, better performance.

To reduce the undue influence of extreme observations in reaction times, a winsorization procedure was used. This is a robust statistical technique in which the highest and lowest reaction times are replaced with the next highest and lowest values that do not fall within the 5% margins on both ends in the distribution (Phillips et al., 2004; Price, Siegle, & Mohlman, 2012). That is, reaction times were winsorized at the 5% level, a highly conservative approach that does not change the qualitative pattern of results (Healey, Hasher, & Campbell, 2013). In this study, the seven lowest values, ranging from -2,356.79 to -2875.16, were replaced with -2,333.25, which was the next lowest value in the data. The seven highest values, ranging from -674.32 to -588.60, were replaced with -675.52, which was the next highest value in the data.

Finally, the scores for the three subcomponents were transformed into z-scores and averaged for each participant to create a composite WM score. This was to create a value that would quantitatively represent the WM capacity of each participant. Individual scores for each subcomponent were, however, kept for data analysis for RQ2 to further examine a detailed account of the interaction between WM and topic familiarity. Also, averaging the WM subcomponents essentially cancels out a source of variance (Shin, Dronjic, & Park, in preparation).

Descriptive results of the scores of the three subcomponents are presented in Table 4.4. The maximum score was 70 for both storage and processing accuracy. Processing speed was measured in milliseconds. There were three data points with particularly low storage z-scores (all -2.68) and two data points with particularly low
processing accuracy $z$-scores (-2.64, -2.84), unattached to the rest of the distributions.

However, they were not eliminated from the dataset for the following reasons. First, the processing speed of these participants with low scores on storage and processing accuracy was within a normal range, indicating that the participants were not necessarily distracted while completing the task. Also, based on a method of detecting outliers called the Median Absolute Deviation (MAD), which overcomes the limitations of the mean-plus/minus-three-standard-deviations method (see Leys, Ley, Klein, Bernard, & Licata, 2013, for details), the values fell within an acceptable range when the conservative threshold of 3 was used. No evidence of multicollinearity among the scores was found (see Table 4.5).

Table 4.4

Descriptive Results of WM Subcomponent Scores ($n = 138$)

<table>
<thead>
<tr>
<th></th>
<th>Storage ($z$-score)</th>
<th>Processing accuracy ($z$-score)</th>
<th>Processing Speed ($z$-score)</th>
<th>Winsorized processing speed ($z$-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>44.98 (0)</td>
<td>53.23 (0)</td>
<td>-1407.70 (0)</td>
<td>-1396.48 (0)</td>
</tr>
<tr>
<td>$SD$</td>
<td>6.34 (1)</td>
<td>5.14 (1)</td>
<td>-502.43 (1)</td>
<td>-468.28 (1)</td>
</tr>
<tr>
<td>Min</td>
<td>28 (-2.68)</td>
<td>39 (-2.84)</td>
<td>-2875.16 (-2.93)</td>
<td>-2333.25 (-2)</td>
</tr>
<tr>
<td>Max</td>
<td>58 (2.05)</td>
<td>63 (1.9)</td>
<td>-588.6 (1.64)</td>
<td>-675.52 (1.54)</td>
</tr>
</tbody>
</table>

*Note.* Processing speed was multiplied by -1.

Table 4.5

Correlations among WM Subcomponent Scores ($n = 138$)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.39**</td>
<td>-.20*</td>
<td>.69**</td>
</tr>
<tr>
<td>2</td>
<td>--</td>
<td>1</td>
<td>-.21*</td>
<td>.69**</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>.34**</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* $p^* < .05, p^{**} < .001$, WM = working memory
Normality of the data and comparability of the two L1 groups (Japanese and Korean) were also examined using the Levene’s test. Storage scores were normally distributed ($D(138) = .07, p = .44$). No significant difference was found between the two groups ($t = 1.09, p = .28$). Processing accuracy scores were normally distributed ($D(138) = .07, p = .43$). No significant difference was found between the two L1 groups ($t = 1.85, p = .07$). Processing speed was normally distributed ($D(138) = .06, p = .68$), but a significant difference was found between the L1 groups ($t = -12.52, p < .001$); Korean participants, overall, performed the processing accuracy task faster than Japanese participants did (see Figures 4.2. and 4.3.)

Figure 4.2 presents a scatterplot that displays a comparison of processing speed value distribution of the L1 groups. Though there is overlap of Japanese and Korean participants’ processing speed in the mid-range of the plot, it is apparent that Japanese participants’ data points are located from the middle to lower area while Korean participants’ data points are located from the middle to upper area of the plot. This illustrates that Korean participants tended to be faster than Japanese participants in making sentence judgments. This discrepancy also resulted in a significant difference in composite WM scores between the two groups. Figure 4.3 compares all WM subscores and composite scores of Japanese and Korean participants. Interpretation of results for both research questions should be made with caution, given the discrepancy between the two L1 groups.
**Figure 4.2.** Scatterplot of processing speed.

**Figure 4.3.** WM subcomponent and composite scores (z-scores) of Japanese and Korean participants. From left (x-axis), storage, processing accuracy, processing speed, composite WM scores; the asterisk indicates a statistically significant difference after the Bonferroni correction with p-values below .001.
**Composite WM scores.** Composite WM scores were obtained by averaging the z-scores of the subcomponents. The composite scores ranged from -1.57 to 1.17 with a mean of zero. The scores were normally distributed \((D(138)= .05, p = .93)\). A significant difference between the Japanese and Korean group was found, mainly due to the influence of the processing speed discrepancy between them as mentioned above \((t = -3.39, p = .001)\) (see Figure 4.4).

Although two data points were 1.5 standard deviations away from the mean (-1.52, -1.57, respectively), they were kept in the dataset for the following reasons. First, the two participants’ performance on other tasks overall appeared within a typical range. Second, no particular incidents were observed on the testing day that would have influenced the two participants’ performances. Third, using a 1.5 standard deviation as a cut-off would be too strict especially considering the fairly large sample size \((n = 138)\). Finally, all WM values fell well within an acceptable range when the MAD method was used with a conservative cut-off of 3 (Leys et al., 2013).

*Figure 4.4. Scatterplot of composite working memory scores.*
L2 knowledge measures. The Vocabulary Size Test (VST) and C-test were employed as measures of participants’ English knowledge. In addition to these scores, participants’ most recent standardized-test scores were collected. Most of the Japanese participants reported TOEFL Institutional Testing Program (ITP) scores, and most of the Korean participants reported TOEIC scores. Six Korean participants did not report any standardized-test scores. Participants’ TOEFL ITP and TOEIC scores were converted into the Common European Framework of Reference for Languages (CEFR) based on the guidelines published by Educational Testing Service (ETS). Among the 132 participants who reported their scores, 6 were at Level A2, 103 at B1, 20 at B2, and 3 at C1. The median and mode was Level B1. After comparing the six remaining participants’ vocabulary size and C-test scores with other participants within a similar score range, the missing values were substituted with B1, which was also the median and mode. These statistics are summarized in Table 4.6.

Normality of the data and comparability of the two L1 groups were checked. The VST scores were normally distributed ($D(138) = .08, p = .34$). No significant difference was found between the L1 groups ($t = -1.06, p = .29$). C-test scores were normally distributed ($D(138) = .08, p = .34$). Japanese participants ($M = 54.46$) outperformed Korean participants ($M = 47.2$) with a medium effect ($p = .001, d = .56$). When the CEFR levels (A2 coded as 1, B1 as 2, B2 as 3, C1 as 4) of the two groups were compared using a Mann-Whitney $U$-test, however, no difference was revealed ($U = 2515.5, p = .42$). Due to the difference of the mean C-test scores found between the Japanese and Korean groups, combined L2 knowledge scores were created for further examination by transforming the vocabulary size and C-test scores into $z$-scores and averaging them. A $t$-
test showed that there was no significant difference between the two L1 groups ($t = 1.27$, $p = .21$). Because of the difference on C-test scores between the two L1 groups, a separate analysis was carried out, which is presented at the end of this chapter.

Table 4.6

*Descriptive Statistics of L2 Knowledge (n = 138)*

<table>
<thead>
<tr>
<th></th>
<th>VST (word families)</th>
<th>C-test</th>
<th>CEFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>6606.52</td>
<td>50.78</td>
<td>2 (Median &amp; Mode)</td>
</tr>
<tr>
<td>$SD$</td>
<td>1149.62</td>
<td>13.36</td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>3700</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Max</td>
<td>10500</td>
<td>81</td>
<td>4</td>
</tr>
</tbody>
</table>

*Note. VST=Vocabulary Size Test, CEFR=Common European Framework of Reference (A2=1, B1=2, B2=3, C1=4)*

**L2 reading comprehension.** Table 4.7 summarizes 138 participants’ L2 reading quiz scores in the No-topic familiarity condition and topic familiarity condition. When the L2 readers’ performance in the two conditions was compared, the topic familiarity condition generated a statistically higher mean score with a medium effect size ($t = -9.34$, $p < .001$, $d = .78$) (see Plonsky & Oswald, 2014, for effect size benchmarks), which indicated that topic familiarity had a positive effect on participants’ L2 reading comprehension. The two L1 groups did not differ on their L2 reading performance in either of the conditions (Topic familiarity condition: $t = -1.39$, $p = .17$; No-topic familiarity condition: $t = -1.76$, $p = .81$).

Table 4.7

*L2 Reading Comprehension Test Results in Two Conditions (n = 138)*

<table>
<thead>
<tr>
<th></th>
<th>$M$ (%)</th>
<th>$SD$</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No topic familiarity condition</td>
<td>10.79 (49.05%)</td>
<td>2.41</td>
<td>5.5</td>
<td>19</td>
</tr>
<tr>
<td>Topic familiarity condition</td>
<td>13.22 (60.09%)</td>
<td>3.51</td>
<td>5</td>
<td>21</td>
</tr>
</tbody>
</table>
In addition to the 138 participants who completed all the tasks included in the study, 10 Korean students and 10 Japanese students took the L2 reading quiz without receiving the topic-familiarity treatment (reading background texts the day before). These 20 additional participants were unfamiliar with all four topics featured in the L2 reading quiz, thereby serving as a control group. This step was taken to ascertain comparability of the four reading tests in the quiz. The researcher compared the control group’s reading comprehension scores on (a) two topics that are in the topic-familiarity condition for the rest of the participants and (b) the other two topics that are in the no-topic-familiarity condition for the rest. No significant difference was found ($t = .27, p = .90, d = .04$).

**Research Question 1: Interaction Between Working Memory and Topic Familiarity**

This research question asked whether and how working memory (WM) and topic familiarity interact in L2 reading comprehension. First, the correlations among the variables were examined. Table 4.8 presents correlations among L2 reading comprehension scores with and without topic familiarity, WM (composite scores), vocabulary size, and C-test scores. No evidence of multicollinearity among the variables was found. Although the main interest of this study is in the interaction between WM and topic familiarity in L2 reading comprehension, L2 knowledge measures – VST and C-test scores – were also included in the analysis as they can be covariates.

As shown in Table 4.8, reading in the topic familiarity condition correlated with WM and both of the L2 knowledge measures. Reading in the No- topic familiarity condition, on the other hand, correlated with L2 measures but not with WM. WM correlated with vocabulary size but not with C-test scores. The L2 measures correlated with each other.
A mixed-effects model analysis was performed by using the *lme4* package in R (Bates, Maechler, & Bolker, 2012; R Development Core Team, 2012) to examine the interaction between WM and topic familiarity in L2 reading comprehension. WM, topic familiarity, and the interaction between them were entered as fixed effects into the model. By-subject variation was modeled as a random effect to account for the non-independence in the data. By-item variation was also modeled as a random effect to account for potential clusters. Although including items as a source of variance did not improve the model (BIC accounting for items: 23342.05, BIC not accounting for items: 23328.25), items remained as a random factor in the model. In the final and most parsimonious model, no obvious deviations from homoskedasticity or normality were found. Statistical significance was assessed by examining *p*-values from the *t* distribution and by using a likelihood ratio test.

As summarized in Table 4.9, this model revealed a significant main effect of WM (estimate = 2.58, SE = .35, *t* = 7.46, *p* < .001), with participants with higher WM

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.52**</td>
<td>.41**</td>
<td>.25**</td>
<td>.37**</td>
<td>.10</td>
<td>.46**</td>
</tr>
<tr>
<td>2</td>
<td>--</td>
<td>1</td>
<td>.16</td>
<td>.08</td>
<td>.16</td>
<td>.04</td>
<td>.40**</td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>.69**</td>
<td>.69**</td>
<td>.34**</td>
<td>.30**</td>
</tr>
<tr>
<td>4</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>.39**</td>
<td>-.20*</td>
<td>.19*</td>
</tr>
<tr>
<td>5</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>-.21*</td>
<td>.34**</td>
</tr>
<tr>
<td>6</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
<td>-.01</td>
</tr>
<tr>
<td>7</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note. p* < .05, *p** < .01; TF = topic familiarity, Proc-Acc = processing accuracy, Proc-Speed = processing speed, VST = vocabulary size test
achieving better L2 reading comprehension. There was also a significant main effect of topic familiarity (estimate = -2.59, SE = .03, t = -78.82, p < .001), indicating that familiar topics led to better comprehension. Importantly, the two main effects were qualified by a significant interaction between them (estimate = -2.19, SE = .06, t = -38.38, p < .001). A likelihood ratio test by using the anova() function was performed to compare two models directly – one with the interaction term and one without. A statistically significant difference between the models was revealed ($\chi^2(1) = 1327.9$, $p < .001$), confirming the significant interaction between WM and topic familiarity. Additionally, the Bayesian information criterion (BIC)$^4$ of the model with and without the interaction term was compared as a confirmatory step. The model with the interaction term yielded a smaller BIC (23271.87) than the one without (24586.76), indicating a better model. Using the MuMIn package, the variance explained by the fixed effects in the model (i.e., marginal $R^2$) was calculated (Nakagawa & Schielzeth, 2013). The fixed factors explained 28% of the variance in L2 reading comprehension. The addition of a fixed effect for L1 of participants did not lead to an improvement in model fit compared to the model without the effect ($\chi^2(1) = .21$, $p = .65$), which suggested that L1 did not affect participants’ L2 reading performance. Therefore, L1 was excluded to keep the model parsimonious. To conclude, results demonstrated that the effect of topic familiarity on L2 reading comprehension varied depending on the reader’s WM capacity.

---

$^4$ BIC values are relative and can be computed for multiple models to be compared. A smaller BIC indicates a better fit overall, that is, a better model. A difference of over 2 shows good evidence of models being different (Seltman, 2013).
Table 4.9

**Mixed-effects Model for RQ1 (Predictors: WM, topic familiarity, interaction between WM and topic familiarity)**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>SE</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>13.24</td>
<td>.26</td>
<td>51.90*</td>
</tr>
<tr>
<td>Working memory (WM)</td>
<td>2.58</td>
<td>.35</td>
<td>7.46*</td>
</tr>
<tr>
<td>Topic familiarity (TF)</td>
<td>-2.59</td>
<td>.03</td>
<td>-78.82*</td>
</tr>
<tr>
<td>WM:TF (interaction)</td>
<td>-2.19</td>
<td>.06</td>
<td>-38.38*</td>
</tr>
</tbody>
</table>

*Note: Marginal $R^2 = .28$, $p < .001$

After confirming the significant interaction using the mixed-effects model with all data included in the analysis ($n = 138$), the best suited interaction model among the compensation, independent-influences, and rich-get-richer models was identified to provide a visual illustration of the interaction. To do so, participants were first divided into three groups according to their composite WM scores (i.e., low, mid, high), each comprising 46 participants’ data. The middle third ($n = 46$) was removed to provide a contrast between high- and low-WM groups’ performance on the reading comprehension test. These additional steps, which resemble ANOVA procedures, were taken only to create a figure which can be compared with figures in similar published studies (Leeser, 2007; Shin et al., 2019). A $t$-test was performed to compare the scores of low- and high-WM groups ($n = 46$, respectively), which revealed a significant difference with a large effect size ($t = 22.91$, $p < .001$, $d = -4.78$) (see Table 4.10).

Table 4.10

**Descriptive Statistics of Low- and High-WM groups ($n = 92$)**

<table>
<thead>
<tr>
<th></th>
<th>Composite WM score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
</tr>
<tr>
<td>Low-WM group ($n = 46$)</td>
<td>-.65</td>
</tr>
<tr>
<td>High-WM group ($n = 46$)</td>
<td>.62</td>
</tr>
</tbody>
</table>

*Note. $p < .001$; WM = working memory
As seen in Figure 4.5, the two WM groups’ L2 reading performance was plotted to visually examine the interaction pattern. L2 reading comprehension of the low- and high-WM groups did not differ in the no-topic familiarity condition ($p > .05$). Both groups significantly improved in L2 reading in the topic-familiarity condition. The degree of the improvement, however, was different between the groups, thereby illustrating an interaction between WM and topic familiarity. When familiar with the topics, the low-WM group scored higher on the L2 reading test with a small effect size ($t = -3.51$, $p < .05$, $d = .48$). The high-WM group also scored higher on the L2 reading test when familiar with the topics, but with a large effect size ($t = -8.94$, $p < .001$, $d = 1.25$). The difference of scores in the topic-familiarity condition between the two groups was also significant ($t = 4.74$, $p < .001$, $d = 0.99$). That is, the gain in the high-WM group was considerably larger than in the low-WM group.

**Figure 4.5.** The interaction between WM and TF in L2 reading comprehension (expressed as percentages). One asterisk: $p < .05$, two asterisks: $p < .001$. 

106
Further analysis was conducted to examine if the above results would hold up after accounting for L2 knowledge represented by vocabulary size and C-test scores. Therefore, another mixed-effects model was constructed with the two scores added into the model. As shown above, VST was a significant predictor of L2 reading comprehension, indicating that the larger the vocabulary size, the better L2 reading comprehension. However, C-test scores were not a significant predictor (estimate = .07, SE = .20, t = .35, p > .05). Therefore, C-test scores were removed from the model to construct a parsimonious model.

Table 4.11 presents the results of the final mixed-effects model. VST was a significant predictor (estimate = 1.10, SE = .19, t = 5.87, p < .001). The effects of the original predictors (WM, topic familiarity) held up while holding the newly added variable (vocabulary size) constant. The new model yielded a smaller BIC (23251.78) than the original model (23271.87), which showed that the new model provided a better fit overall. In fact, this set of fixed factors explained 39% of the variance in L2 reading comprehension. Compared to the first model (without VST), an additional 11% of the variance in L2 reading comprehension was accounted for by VST.

Table 4.11

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>SE</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>13.25</td>
<td>.24</td>
<td>55.19*</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>1.08</td>
<td>.19</td>
<td>5.82*</td>
</tr>
<tr>
<td>Working memory (WM)</td>
<td>2.02</td>
<td>.32</td>
<td>6.20*</td>
</tr>
<tr>
<td>Topic familiarity (TF)</td>
<td>-2.59</td>
<td>.03</td>
<td>-78.82*</td>
</tr>
<tr>
<td>WM:TF (interaction)</td>
<td>-2.19</td>
<td>.06</td>
<td>-38.38*</td>
</tr>
</tbody>
</table>

Note. Marginal $R^2 = .39, p^* < .001$
Research Question 2: Extent to which Working Memory Subcomponents Contribute to L2 Reading with and without Topic Familiarity

For RQ2, WM was decomposed into three subcomponents (storage, processing accuracy, processing speed) of the reading span task to examine their unique roles in L2 reading with and without topic familiarity. As previously mentioned, there were no signs of multicollinearity among the subcomponents. Storage positively correlated with processing accuracy ($r = .39$) but negatively with processing speed ($r = -.20$). Processing speed showed a negative correlation with processing accuracy as well ($r = -.21$). Although caution is warranted in interpreting these correlations due to the weak relationships, there are possible trade-offs, which are still relevant on a theoretical level, between (a) processing speed and processing accuracy and (b) processing speed and storage.

In terms of the relationships between the subcomponents of the reading span score and L2 knowledge variables, storage correlated with VST ($r = .19$) but not with C-test scores. Processing accuracy correlated with both VST ($r = .34$) and C-test scores ($r = .33$). Thus, the larger the storage and processing capacity, the larger the vocabulary size. Processing speed negatively correlated with C-test ($r = -.23$) but did not show a relationship with VST.

Reading in the topic familiarity condition correlated with storage ($r = .25$) and processing accuracy ($r = .37$). It correlated with C-test ($r = .22$) and showed the strongest relationship with VST ($r = .46$). Reading in the No-topic familiarity condition, on the other hand, did not correlate with any of the WM subcomponents but did with L2 knowledge; it was positively associated with VST ($r = .40$) and C-test scores ($r = .20$).
To summarize, there existed a consistent relationship between L2 knowledge and L2 reading comprehension regardless of the conditions whereas WM subcomponents were sensitive to the conditions. L2 reading comprehension in both conditions exhibited positive relationships with L2 knowledge. When it comes to the relationship with WM subcomponents, however, reading in the topic familiarity condition showed a positive relationship with two WM subcomponents, specifically storage and processing accuracy, though no relationships were found in the No- topic familiarity condition.

A mixed-effects model was fitted to examine contributions of each of the WM subcomponents to L2 reading. Similar to the steps taken for RQ1, variables of interest – storage, processing accuracy, processing speed, topic familiarity, VST, C-test scores – were entered as fixed effects and participants and items as random effects. C-test scores did not show a significant effect and were thus removed from the model. Although not all WM subcomponents were predictive of L2 reading comprehension, all were kept in the model to present the results for RQ2 (see Table 4.12). Processing accuracy was the only significant predictor of L2 reading among the WM subcomponents (estimate = .99, SE = .23, \( t = 4.22, p < .001 \)), making a unique contribution to L2 reading comprehension after accounting for VST. The interaction between processing accuracy and topic familiarity was predictive of L2 reading comprehension (estimate = -.80, SE = .25, \( t = -2134, p = .002 \)). This result was confirmed by examining the BIC of the model with and without the interaction term. The model with the interaction term yielded a smaller BIC value (1313.7) than the one without (1325.7), indicating a better model with a significant difference (\( \chi^2(1) = 17.63, p < .001 \)). Thus, the aspect of WM with which topic familiarity is interacting was processing accuracy, rather than storage or processing speed.
Table 4.12

**Mixed-effects Model for RQ2 after Adding Vocabulary Size**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>13.27</td>
<td>.22</td>
<td>61.43**</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>1.07</td>
<td>.19</td>
<td>5.68**</td>
</tr>
<tr>
<td>Storage</td>
<td>.16</td>
<td>.19</td>
<td>.81</td>
</tr>
<tr>
<td>Processing accuracy</td>
<td>.99</td>
<td>.23</td>
<td>4.22**</td>
</tr>
<tr>
<td>Processing speed</td>
<td>.31</td>
<td>.18</td>
<td>1.72</td>
</tr>
<tr>
<td>Topic familiarity (TF)</td>
<td>-2.63</td>
<td>.23</td>
<td>-11.26**</td>
</tr>
<tr>
<td>Processing accuracy: TF (interaction)</td>
<td>-1.01</td>
<td>.23</td>
<td>-4.34*</td>
</tr>
</tbody>
</table>

*Note.* Marginal $R^2 = .39$, $p^* < .05$, $p^{**} < .001$

**Interaction between Working Memory and Topic Familiarity in L2 Reading**

**Comprehension: Japanese vs Korean L2 English Learners**

Further analyses were performed due to the significant difference of processing speed between Japanese and Korean participants. Tables 4.13 and 4.14 summarize how data from Japanese and Korean participants compared in terms of the results for RQs 1 and 2. The same patterns were observed in the performance of both L1 groups, which correspond to the overall results described above. One difference that should be noted is that the set of fixed factors accounted for more variance in the L2 reading of Japanese participants (marginal $R^2 = .45$ & .46) than that of Korean participants (marginal $R^2 = .36$). In other words, 45-46% of the L2 reading comprehension performance of the Japanese participants was explained, collectively, by VST, WM, topic familiarity, and the interaction between WM and topic familiarity; 36% of the L2 reading comprehension performance of the Korean participants was explained by the same set of independent variables.
Table 4.13

Comparison of Japanese and Korean Participants for RQ1

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Japanese</td>
<td>Korean</td>
<td>Japanese</td>
<td>Korean</td>
<td>Japanese</td>
<td>Korean</td>
</tr>
<tr>
<td>Intercept</td>
<td>13.41</td>
<td>13.31</td>
<td>.31</td>
<td>.30</td>
<td>43.29*</td>
<td>44.76*</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>1.72</td>
<td>.82</td>
<td>.32</td>
<td>.22</td>
<td>5.36*</td>
<td>3.69*</td>
</tr>
<tr>
<td>Working memory (WM)</td>
<td>2.29</td>
<td>1.60</td>
<td>.60</td>
<td>.52</td>
<td>3.85*</td>
<td>3.10*</td>
</tr>
<tr>
<td>Topic familiarity (TF)</td>
<td>-2.92</td>
<td>-2.42</td>
<td>.36</td>
<td>.31</td>
<td>-8.15*</td>
<td>-7.92*</td>
</tr>
<tr>
<td>WM:TF (interaction)</td>
<td>-2.53</td>
<td>-1.97</td>
<td>.66</td>
<td>.52</td>
<td>-3.85*</td>
<td>-3.80*</td>
</tr>
</tbody>
</table>

Note. \( p^* < .001 \); Marginal \( R^2 \) from Japanese data: .45, Korean data: .36

Table 4.14

Comparison of Japanese and Korean Participants for RQ2

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
<th>Coefficient</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Japanese</td>
<td>Korean</td>
<td>Japanese</td>
<td>Korean</td>
<td>Japanese</td>
<td>Korean</td>
</tr>
<tr>
<td>Intercept</td>
<td>13.08</td>
<td>13.85</td>
<td>.40</td>
<td>.40</td>
<td>32.70**</td>
<td>34.84**</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>1.64</td>
<td>.73</td>
<td>.32</td>
<td>.23</td>
<td>5.10**</td>
<td>3.23*</td>
</tr>
<tr>
<td>Storage</td>
<td>-.12</td>
<td>.19</td>
<td>.33</td>
<td>.23</td>
<td>-.38</td>
<td>.81</td>
</tr>
<tr>
<td>Processing accuracy</td>
<td>1.52</td>
<td>.68</td>
<td>.39</td>
<td>.28</td>
<td>3.91**</td>
<td>2.40*</td>
</tr>
<tr>
<td>Processing speed</td>
<td>.35</td>
<td>-.22</td>
<td>.35</td>
<td>.39</td>
<td>.98</td>
<td>-.57</td>
</tr>
<tr>
<td>Topic familiarity (TF)</td>
<td>-2.30</td>
<td>-2.84</td>
<td>.35</td>
<td>.31</td>
<td>-6.52**</td>
<td>-9.22**</td>
</tr>
<tr>
<td>Processing accuracy:TF (interaction)</td>
<td>-1.46</td>
<td>-.82</td>
<td>.41</td>
<td>.28</td>
<td>-3.58**</td>
<td>-2.94**</td>
</tr>
</tbody>
</table>

Note. \( p^* < .05, p^{**} < .001 \); Marginal \( R^2 \) from Japanese data: .46, Korean data: .36

Summary of the Chapter

This chapter described the results of (a) the two main research questions that the study set out to answer and (b) further analyses to investigate the difference in processing speed between the two L1 groups. The major findings of the study are summarized as follows.
1. L2 reading comprehension is positively correlated with and predicted by topic familiarity, WM, and L2 knowledge.

2. As L2 knowledge variables, vocabulary size was a significant predictor of L2 reading comprehension whereas C-test scores were not.

3. There is an interaction between WM and topic familiarity; the positive effect of topic familiarity on L2 reading comprehension was amplified by readers’ WM capacity.

4. The interaction between WM and topic familiarity was significant while holding vocabulary size constant.

5. Among the compensation, independent-influences, and rich-get-richer models, findings of the study corresponded to the rich-get-richer model; the high-WM group outperformed the low-WM group in L2 reading comprehension in the presence of topic familiarity. Participants’ performance in L2 reading comprehension without topic familiarity did not differ regardless of their WM capacity.

6. Among the three subcomponents of the reading span score (storage, processing accuracy, processing speed), processing accuracy was the only significant predictor of L2 reading comprehension. Processing accuracy in the WM task reflects sentence-level processing (to be differentiated from discourse-level comprehension) under pressure.

7. Among the WM subcomponents, processing accuracy interacted with topic familiarity in L2 reading comprehension after controlling for vocabulary size.
8. Overall, the same patterns were observed in the performance of both Korean and Japanese L2 English learners; however, more variance in L2 reading of the Japanese participants, than that of the Korean participants, was accounted for by the independent variables (WM, topic familiarity, vocabulary size).

In the next chapter, the findings are further discussed with a view to informing theory, research methods, and teaching practices. In addition, the following chapter addresses limitations of the study and directions for future research.
Chapter 5: Discussion

The purpose of the study was two-fold: (a) to investigate whether and how working memory (WM) interacts with topic familiarity in L2 English reading comprehension and (b) to examine the extent to which subcomponents of WM (i.e., storage, processing accuracy, processing speed) contribute to reading in the presence and absence of topic familiarity. This chapter summarizes and interprets the results of the study with respect to the research questions answered in Chapter 4. The discussion of the results centers around how the findings add to the current literature on the roles of WM and background knowledge in L2 reading comprehension. The chapter concludes with an exploration of pedagogical implications, an accounting of the limitations of the study, and directions for future research.

Topic Familiarity as a Type of Background Knowledge

While this study operationalized background knowledge as topic familiarity, referring to readers’ pre-existing knowledge of the topic of the text at hand, it is important to note that background knowledge is the term most often used to refer to prior knowledge in general in the existing literature on reading comprehension (Carrell & Eisterhold, 1983; van den Broek, Mouw, & Kraal, 2015). Therefore, the term background knowledge is used as an all-encompassing term throughout the chapter, though clarifications are made when it is necessary to address issues that stem from different operationalizations of the construct. In fact, background knowledge has been operationalized in various ways in reading contexts: as content familiarity (Alptekin & Erçetin, 2009), cultural familiarity (Alptekin, 2006), subject knowledge (Brantmeier et al., 2014), content schemata (Carrell, 1987), language experience (Payne et al., 2009),
and topic familiarity/knowledge (Joh & Plakans, 2017; Leeser, 2009; Shin et al., 2019),
to name a few. Even among the narrow pool of comparable studies that explored the
interaction between WM and background knowledge as the current study did,
background knowledge was represented in vastly different ways (e.g., familiarity with
places and settings in the reading passage, in Alptekin & Erçetin, 2009; number of years
studying the L2, in Payne et al., 2009), possibly contributing to discrepancies in findings,
which are discussed below.

The discussion that follows focuses first on the impact of topic familiarity on L2
reading comprehension based on the analysis of data collected for the study reported
here. The discussion then focuses on (a) the interaction between WM and topic
familiarity (RQ1) and (b) the roles of the subcomponents of the reading span score in the
interaction (RQ2).

**Facilitative Effect of Topic Familiarity on L2 Reading Comprehension**

The analysis showed the facilitative effect of topic familiarity on L2 reading
comprehension, which reaffirms the literature on the positive role of topic familiarity in
reading (e.g., Hammadou, 2000; Horiba & Fukaya, 2015; Lee, 1986; McNeil, 2011).
Topic familiarity assists readers in relating textual information to the existing clusters of
information that they already have stored in their long-term memory, thereby facilitating
the creation of a mental representation presented by the text (Kendeou, Rapp, & van den
Broek, 2003; Nassaji, 2002). Interesting to note, however, is that topic familiarity had a
significant effect on L2 reading comprehension even when it did not include topic-related
vocabulary beyond what the readers had already possessed before reading. Participants’
background knowledge of the initially unfamiliar topics (Teotihuacan, Bantu people) in
the current study was, in fact, constructed in participants’ L1s. The researcher purposefully provided background texts written in participants’ L1s with the intention of not including L2 vocabulary related to the content or the topics to achieve the purpose of raising familiarity with the topics only conceptually while avoiding the potential confounding effect of L2 vocabulary learning.

This finding contradicts Chou’s (2011) discussion of what background knowledge must include for it to aid comprehension. Chou found that while providing L2 students with prior knowledge of relevant vocabulary improved their performance on a reading test, their familiarity with the topic did not have an effect. He concluded that background knowledge must include specific knowledge of relevant L2 terminology that can directly help readers comprehend the text. However, the current study showed that even a simple encounter with the topics in the L1, which increased familiarity with the topics without including or pre-teaching L2 vocabulary relevant to the content, can still improve L2 reading comprehension. This finding also supports the view that reading comprehension is not a product of text-based information alone but is achieved through an interaction between and melding of textual information and readers’ knowledge of the topic (Koda, 2007).

The differences in the effect of background knowledge on L2 reading comprehension may be attributed to readers’ understanding and grasp of background information. In Chou’s (2011) study, background information associated with the chosen topics was extracted from Wikipedia and provided to participants in their L2. However, the readability of Wikipedia entries, generally, has been described as low. The Gunning fog index of texts from Wikipedia is higher (15.8±0.4) than that of the British National
Corpus (12.1±0.5) which is considered as a reasonable approximation to “English in general” (Yasseri, Kornai, & Kertesz, 2012, p. 3). Similar concerns regarding the low readability and comprehensibility of Wikipedia articles have been expressed by others as well (e.g., Hendry & Sheepy, 2017; Jatowt & Tanaka, 2012). Therefore, it may be that L2 participants in Chou’s study, who were at the intermediate level, may not have fully grasped the information provided to build background knowledge of the topics. In the present study, on the other hand, it is unlikely that the participants, whose proficiency was also at the intermediate level, had difficulty understanding the background texts because they were written in their L1s. This method also minimized possible issues of readability, which can arise especially when working with L2 texts. The participants’ performance on the comprehension check of the background texts indicated a high level of comprehension.

In sum, the present study showed that background knowledge, operationalized as topic familiarity, successfully aided L2 readers’ comprehension despite the fact that the background knowledge did not include specific lexical knowledge relevant to the content. The current study’s findings suggest that whether or not background knowledge can enhance reading comprehension may not be determined by what kind of information the knowledge contains (e.g., topic-related vocabulary). For the effect of background knowledge to materialize, what matters more may be the extent to which the knowledge has been absorbed and is therefore available in the memory system and accessible to the reader as a conceptual resource. Readers’ capacity to activate background knowledge, therefore, becomes a factor in their success at utilizing it in reading. In fact, Grabe (2009) described reading comprehension as “a combination of text input, appropriate cognitive
processes, and the information that [readers] already know” (p.74), indicating the important role of cognitive resources in the complex and interactive processes of reading.

The following discussion centers on the interplay between WM and topic familiarity in L2 reading comprehension. The discussion sheds light on the involvement of cognitive resources in L2 readers’ use of background knowledge.

**Interaction between Working Memory and Topic Familiarity in L2 Reading**

As discussed above, the notion that topic familiarity, or background knowledge in general, has a facilitative effect on L2 reading comprehension is not new (Chang, 2006; Horiba & Fukaya 2015; Lee, 1986; Pulido, 2007). However, this current study can be differentiated from the majority of the studies on background knowledge in a number of ways. First, it incorporated an experimental intervention component to measure empirically an effect of topic familiarity on L2 reading comprehension in two distinct conditions: reading with and without topic familiarity. By doing so, the researcher aimed to simulate and compare situations in which readers encounter L2 texts having been exposed to some topics but not others. Second, and more importantly, a more controlled approach was employed in creating the two conditions by (a) selecting four topics that were proven to be unfamiliar to the participants and (b) subsequently providing background texts about two of the four topics in participants’ L1s to manipulate topic familiarity. As a result, the study was able to ensure more or less the same extent of knowledge (i.e., quantity) and its degree of interconnectedness to the actual reading (i.e., quality). By establishing more control over the quantity and quality of background knowledge, this study also addresses the limitations of previous studies that solely relied on participants’ self-ratings of their background knowledge (e.g., Joh & Plakans, 2017;
Leeser, 2007). In addition, because the participants in the current study were instructed to read the background texts the day before the reading test, they were able to work with consolidated knowledge. That is, retention of episodic memories that were formed was strengthened by sleep (Racsmany, Conway, & Demeter, 2010). Finally, this study is one of the few that recognizes the role that cognitive (processing) resources play in utilizing knowledge resources and examined the interaction between WM and topic familiarity. In what follows, discussion and interpretation of the first research question are presented.

One of the major goals of this study was to examine whether and how WM and topic familiarity interact in L2 reading comprehension in two L1 groups (Japanese and Korean). Results showed evidence for an interaction between them; L2 readers with higher WM capacity were able to derive a greater benefit from topic familiarity, which resulted in more successful reading comprehension. Also, while the role of WM in L2 reading comprehension was negligible in the absence of topic familiarity, it played a significant role above and beyond L2 knowledge when readers possessed topic familiarity.

This interactional pattern corresponds to the rich-get-richer-model, which illustrates that individuals with higher WM capacity are better able to draw upon prior knowledge in their performance on a cognitive task (Hambrick & Engle, 2002). The finding is in accordance with Leeser (2007) and the pilot of the current study (Shin et al., 2019), which supported the rich-get-richer model as a pattern for the interaction between WM and topic familiarity. In both of the studies (Leeser, 2007; Shin et al., 2019), WM did not play a significant role when L2 readers were not familiar with the topics; that is, having high-WM capacity was not necessarily a major advantage to L2 readers unless
they were familiar with the topics. In the presence of topic familiarity, WM amplified the
effect of topic familiarity on reading by aiding in forming connections between ideas
expressed in the text and prior knowledge (Leeser, 2007). Because WM was measured
through the reading span task which tapped storage and processing capacity (accuracy
and speed), it can be said that readers with a greater capacity for maintaining information
while processing additional pieces of information may be at an advantage when making
use of topic familiarity. In other words, having topic familiarity supports L2 reading
comprehension, which can be facilitated even further with sufficient WM resources. As
highlighted by Shin and colleagues (2019), this also means that the advantage of having
topic familiarity did not apply equally to all readers due to constraints imposed by WM.

By showing evidence for the interaction between WM and topic familiarity, the present
study, along with Shin and colleagues (2019) and Leeser (2007), demonstrated that not
every L2 reader can benefit from topic familiarity to the same degree.

However, findings on the interaction between WM and topic familiarity have not
been agreed upon, making it difficult to identify a firm general account of their
interaction in L2 reading comprehension at this juncture. Unlike the current study, the
findings of Alptekin and Erçetin (2011) and Payne and colleagues (2009) favored the
independent-influences model, showing that WM and topic familiarity are independent
contributors to L2 reading comprehension with no interaction. The disparity in findings
among the studies on the interaction between WM and background knowledge is likely
due to different construct operationalizations. Background knowledge was
operationalized as content familiarity in Alptekin and Erçetin (2011), where the
researchers “nativized” an American short story to enhance Turkish participants’
familiarity. To nativize the text, they replaced the names of settings and locations (e.g., New York City → Istanbul) and modified the story to cater to participants’ customs of the time of the story (e.g., couples who are dating → engaged couple). In Payne and colleagues (2009), language experience was used as an operationalization of background knowledge, which was represented by the number of years that participants spent studying L2 and the number of L2 classes that they took. In both of these studies, what the researchers conceptualized as background knowledge may not have been a crucial factor for participants’ comprehension of target texts.

Leeser (2007) and Shin and colleagues (2019), on the other hand, operationalized background knowledge as topic familiarity, as the current study did. In all three studies, two distinct conditions (i.e., reading on familiar and unfamiliar topics) were created to explicitly target differing degrees of topic familiarity while controlling for textual variables (e.g., passage length, readability). Therefore, L2 readers may have benefitted from being familiar with the key concepts in the texts, which is the type of knowledge known to enhance inference generation and memory for details (Lupo et al., 2017; Pulido, 2007). Because meaning construction entails relating information in the textual input to information stored in long-term memory (Kintsch, 2013), actively searching in long-term memory to find relevant knowledge would have been necessary for the participants.

The results of the current study highlight the importance of WM resources in mobilizing topic familiarity in reading comprehension, adding another layer to the complex dynamics of reading comprehension. Schema theory (Oller, 1995; Rumelhart, 1980) postulates that preexisting knowledge structures stored in the mind (i.e., schemata)
allow for more efficient performance on a given task. This applies to a reading task as well because reading entails mapping textual information onto preexisting knowledge structures. However, schema theory, though often evoked to describe the role of background knowledge in reading, seems to take for granted readers’ ability to utilize background knowledge. As pointed out in Nassaji (2002), schemata are static and inflexible due to the assumption that schemata can be accessible whenever needed and are resistant to modification. Schema theory also claims that the reading comprehension process proceeds in a top-down manner. For these reasons, schema theory is unable to explain how readers navigate their way in comprehending new information without background knowledge.

The current study suggests that flexibility may actually be key as readers attempt to (a) build background knowledge when it is not present and (b) comprehend texts as they draw on background knowledge without the guarantee of putting it to use successfully. This view is in line with Nassaji’s stance that retrieval and use of background knowledge may be a more effortful cognitive process than schema theory implies. By showing the interaction between WM and topic familiarity, the current study also suggests that the process of using background knowledge occurs through cognitive operations that rely on WM.

Theoretical support for this finding can be found in the construction-integration (CI) model in terms of the networks of knowledge being called up from textual input active in WM (Kintsch, 1998, 2013). This perspective aligns with the interactive nature of reading, which underscores the interaction between text- and reader-driven information. The CI model acknowledges that the cycle involved in reading processes –
extracting meanings embedded in the sentences, accumulating the propositions through successive processing of textual information, achieving discourse coherence, and retrieving relevant information from long-term memory – is underpinned by WM. The integration aspect of the CI model, in particular, describes how relevant background knowledge activated and processed in WM serves as support for comprehension, ultimately being integrated into the mental representation of the text.

In this respect, the current study reinforces the idea of the complexity of reading comprehension while underscoring the limited-capacity nature of WM as well as its role in materializing the interaction between text- and reader-based knowledge. This study’s findings showed that not all L2 readers were able to benefit equally from the provision of background knowledge; instead, readers with higher WM capacity outperformed their lower WM counterparts in L2 reading comprehension. As Grabe (2009) stated, the integration processes of reading comprehension require communication of ongoing mental processes with long-term memory, facilitated by the mental workspace that WM affords readers. More specifically, the juggling of processing textual information and integrating it with preexisting knowledge in an activated and accessible state calls upon WM resources. In fact, individuals with high WM have been found to be better at accessing and retrieving information from long-term memory than those with low WM (Kaakinen, Hyönä, & Keenan, 2003). Thus, high-WM readers’ ability to “communicate” better with information stored in long-term memory resulted in more efficient retrieval of background knowledge as well as the merging of textual information with background knowledge, which may have been key to taking advantage of background knowledge and, ultimately, achieving better L2 reading comprehension.
In this study, the composite WM scores, obtained through the reading span task, represented participants’ ability to maintain verbal information (correct final word recall) while simultaneously processing additional information as accurately as possible (correct semantic judgments) and recording reaction times. Therefore, individuals who had high WM scores were those who showed better performance on the three subcomponents (storage, processing accuracy, processing speed) of the WM task. These readers with high WM, equipped with greater ability to alternate between storage and processing tasks under pressure, were able to utilize what they knew about the topic more effectively than their lower WM counterparts. This difference is likely because readers with high WM had more cognitive resources available for higher-level processes during reading, such as activating and using background knowledge and integrating information (Shin et al., 2019).

The findings can also be explained by the notion that readers’ ability to strategically use their attention is critical (Friedman & Miyake, 2004). Readers must select information to maintain an active and retrievable state while controlling attention to avoid distraction and suppress irrelevant material (Engle, 2002). As shown in Kaakinen and colleagues (2003), individual differences in how readers process relevant and non-salient information with controlled attention are reflected in individual differences in WM. Attentional control, governed by WM’s central executive, contributes to the orchestration of skills and abilities in higher-order cognitive processes such as reading comprehension (Butterfuss & Kendeou, 2018; Miyake et al., 2000). In other words, successful reading comprehension is dependent on the flexible use of attentional control, which allows readers to coordinate activation and suppression of information
(Friedman & Miyake, 2004; Palladino et al., 2001; Yeari, 2017) as well as update information in WM (Carretti, Cornoldi, De Beni, & Romanò, 2005; Miyake et al., 2000). Although direct measures of executive functions (e.g., updating, inhibition) were not employed in the current study, greater WM capacity is an indication of greater ability to effectively manipulate attention (García-Madruga et al., 2013). Thus, it is likely that readers with high WM had more cognitive resources for attentional control as they engaged in the iterative process of updating information held in WM, allowing them to benefit more from topic familiarity.

The current study adds to the body of work on the complex web of cognitive processes involved in L2 reading comprehension as it demonstrates what the role of WM is in reading and how it interacts with another reader variable, topic familiarity. In doing so, it adds nuance to the long-standing literature on the effect of background knowledge on reading comprehension by revealing that the ability to use background knowledge to one’s advantage is affected by WM capacity. The following section delves further into the interaction between WM and topic familiarity by focusing on the roles of WM subcomponents (storage, processing accuracy, and processing speed) in L2 reading with and without topic familiarity.

**Contributions of Working Memory Subcomponents (Storage, Processing Accuracy, Processing Speed) to L2 Reading with and without Topic Familiarity**

The second research question (RQ2) set out to investigate the contributions of each of the WM subcomponents (storage, processing accuracy, processing speed) to L2 reading comprehension with and without topic familiarity. It should be noted that the three subcomponents represent performance on the reading span task with respect to the
storage (final word recall) and processing aspects (semantic judgments and the speed at which the judgments are made) of the task. By doing so, the current study aimed to (a) provide a detailed account of the interplay between WM and topic familiarity in L2 reading and (b) contribute to the discussion of appropriate WM measurement practices, which evidently affect how the construct of WM and its relevance in reading are interpreted.

The results showed that WM as a whole predicted L2 reading comprehension only in the presence of topic familiarity. When closely examined, storage (final word recall) and processing accuracy (semantic judgment) scores positively and moderately correlated with L2 reading comprehension with topic familiarity. That is, participants who showed better performance on final word recall and sentence meaning judgment in the concurrent WM task (reading span task) were more likely to comprehend L2 texts with familiar topics better. Processing speed did not show a significant relationship with L2 reading comprehension regardless of background knowledge condition. Storage and processing accuracy scores showed a positive, moderate correlation, indicating that storage capacity and processing capacity go hand in hand: The more information readers can hold in memory, the more information they can process concurrently. Processing speed showed a negative relationship with both storage and processing accuracy, which suggests that participants sacrificed speed in their attempts to perform well in the processing and storage component of the WM task. A more detailed discussion of this trade-off is provided further below.

While both storage and processing were related to the participants’ use of topic familiarity during L2 reading, processing accuracy was the only significant predictor
when L2 knowledge (measured by the Vocabulary Size Test and C-test) was accounted for. In other words, processing capacity had a unique contribution to L2 readers’ ability to benefit from topic familiarity. By probing into the degree of specificity that the WM subcomponents contribute to the interaction between WM and topic familiarity, the current study extends the findings of Shin and colleagues (2019) and Leeser (2007). Not only do this study’s results confirm the rich-get-richer model with regard to the interaction between WM and topic familiarity, but they also demonstrate that it is the processing capacity of WM that is responsible for making use of topic familiarity during L2 reading. Processing capacity in the reading span task can be described as an ability to make accurate judgments on semantic plausibility of a sentence while holding a varying number of to-be-remembered items in short-term memory, which adds the pressure of this concurrent task. As explained by Palladino and colleagues (2001), effective maintenance of information (i.e., storage) does not necessarily lead to successful comprehension without coordinating information activation through the flexible use of attentional control. In this respect, the participants in the current study most likely drew upon their processing capacity of WM rather than storage to (a) access their background knowledge of the topics during reading and (b) integrate it into the textual information as they underwent the iterative process of updating information held in WM.

The discovery that processing capacity, rather than storage, provides unique information for utilizing background knowledge in reading joins previous findings that support the processing-oriented hypotheses described in Yeari (2017). The crux of processing-oriented hypotheses is that what differentiates readers with high WM capacity from those with low WM capacity is their ability to (re)activate information during
reading. Storage-oriented hypotheses, on the other hand, focus on differences in readers’ short-term retention of information and consider maintenance of information as key to constructing a coherent text representation (Just & Carpenter, 1996; Yuill, Oakhill, & Parkin, 1989). While readers’ capacity to retain and maintain information in L2 reading with topic familiarity is important as evidenced in the significant correlation between storage and L2 reading with topic familiarity, its effect is limited, as it was subsumed by L2 knowledge.

As elaborated in Yeari (2017), high-WM readers demonstrate better performance on what seems to resemble the integration process of the construction-integration model (Kintsch, 1998, 2013). That is, they are more successful at performing a controlled intentional search of background knowledge, (re)activating relevant information and connecting textual ideas and information to a single coherent representation (Linderholm, 2002; Yeari & van den Broek, 2015). The limited WM resources, which are by definition deployed effortfully, independent of vocabulary knowledge and any other automatic language- and reading-related processes, have been shown to contribute to making inferences by assisting readers in drawing elaborations during the integrative process in reading comprehension (Calvo, 2004; Estevez & Calvo, 2000). The complex higher-level processes involved in this integration part of constructing a text representation go beyond maintenance of information, which explains the current finding that the unique contribution of WM to reading comprehension is derived from processing capacity but not from storage capacity.

Furthermore, this study’s findings are in accord with the view corroborated by the processing-oriented hypotheses in that WM resources allow an enhanced access to long-
term memory and efficient retrieval of information (McNamara & Kintsch, 1996; McNamara & Scott, 2001). Perhaps, the individual differences in participants’ ability to process information (while holding previously entered information) are also associated with the efficiency with which they access relevant background knowledge to facilitate L2 reading comprehension. Although the researcher in this study did not record how long it took for participants to complete the L2 reading comprehension tests, all participants were instructed to finish them in one hour, and their tests were collected at the one-hour mark. Considering the same timeframe given to the participants, it is likely that more efficient search of and access to relevant background knowledge led to better performance in comprehension. The finding that processing capacity rather than storage did most of the heavy lifting in the interaction with topic familiarity may also be an indication that L2 readers’ strategic and flexible allocation of attention is tapped by the processing capacity component of the reading span task used in this study.

It should be noted that subscribing to the processing-oriented hypotheses is, however, by no means tantamount to suggesting that the ability to store information is of no use in reading with topic familiarity. What the current findings indicate is that there is little evidence that the extent to which L2 readers were able to benefit from familiarity with the topic is linked to their storage capacity. This conclusion is based on the following. First, the quantity and quality of background knowledge were controlled by the researcher and thus were equal across participants. Second, storage scores did not predict L2 reading comprehension above and beyond L2 knowledge while processing accuracy did, demonstrating that processing capacity is the primary WM subcomponent that interacted with topic familiarity.
Taken together, the results of the current study yielded evidence that processing capacity lies at the heart of the interplay between WM and topic familiarity in L2 reading comprehension. The processing-oriented view bolsters the interpretation of the findings that processing accuracy may embody the capacity that enables readers to utilize topic familiarity with efficiency during reading. The participants’ capacity for storing and maintaining information, on the other hand, did not seem to contribute to their ability to take advantage of topic familiarity, nor was it predictive of reading comprehension after L2 knowledge was accounted for. The discussion of RQ2 reinforces the assertion made above that the process of accessing and using topic familiarity calls on WM resources. The findings pertinent to RQ2 provide nuanced insights into the resources upon which L2 readers draw when reading with topic familiarity: It is not enough to be able to store background knowledge in long-term memory for it to enhance L2 reading comprehension. Rather, it is what readers can do with that knowledge (e.g., activating it, using it to make inferences) in a limited amount of time that uniquely contributes to the degree to which readers can benefit from background knowledge to facilitate their comprehension. L2 readers’ ability to do so is likely to be reflected in the processing capacity measured in the WM task.

In addition to shedding light on the unique roles of WM subcomponents in reading with and without topic familiarity, the results obtained in this study can also contribute to the discussion of WM measurement practices in L2 reading research. Because WM is defined as the capacity to simultaneously store and process information, both storage and processing components should be measured in a concurrent task in association with the criterion construct of interest. However, this is often not the case as
many studies tend to equate storage capacity (recall scores) with WM capacity, which has been criticized with respect to the appropriateness of the WM tasks that rely solely on storage (Sagarra, 2017; Unsworth et al., 2009). Because the WM task design employed in this study allows measuring participants’ simultaneous performance on storage, processing accuracy, and processing speed, the current study is able to address this methodological issue.

Decomposing reading span performance into three subcomponents – storage, processing accuracy, processing speed – and measuring each of them in the time-sensitive complex reading span task made it possible for the current study to observe the interrelationships among them. To reiterate the results, processing speed (the speed at which semantic judgments were made) yielded a weak, negative correlation with storage (final word recall) as well as with processing capacity (semantic acceptability judgment under the pressure of a concurrent task). In other words, it is possible that the participants sacrificed speed for higher levels of sentence-judgment accuracy while holding information in memory, alluding to a potential speed-accuracy tradeoff. A moderate, positive relationship between processing capacity and storage was found, which shows that individuals who held more information in memory were also able to process more information accurately under pressure. The interrelationships among the subcomponents, with one being positive and the others inverse albeit weak, hold theoretical importance though more empirical work is warranted. Still, because of the nature of the reading span task, we can assume that the participants were dividing their attentional resources. That is, participants traded off on one component of the WM task for another, echoing the
growing body of research that endorses an inclusive scoring system as opposed to relying on storage scores.

For instance, Waters and Caplan (1996) questioned the appropriateness and reliability of WM measures that take only storage scores into account because in their study, too, processing speed and storage yielded a negative correlation, as did processing speed and processing accuracy. Shin and colleagues (in preparation) and Leeser (2007) also reported the same patterns for the interrelationships among WM subcomponents and highlighted that a WM score based solely on storage performance is difficult to interpret, warranting a composite WM score that reflects both storage and processing. Sagarra (2017) asserted that the presence of a taxing processing component is one of the contributing factors of the discrepant findings in WM literature. Further, Unsworth and colleagues (2007) argued against the earlier notion that processing accuracy and processing speed are redundant and that either can be used as an index of processing efficiency (Daneman & Tardif, 1987). This was because, just like this current study, their results showed that processing accuracy and speed were negatively related and indexed different constructs. They concluded that inspection of both processing accuracy and speed can generate more insights into investigating the underlying constructs within WM capacity.

In addition, the way WM is measured influences the resulting relationship between WM and L2 reading. According to Shin’s meta-analysis (under review) on WM and L2 reading, the correlation between WM and L2 reading varies depending on what subcomponents constituted WM scores. Shin reported that the relationship was the weakest when WM scores reflected only storage but significantly increased when both
storage and processing accuracy were represented in the composite scores. When processing speed was included in the composite WM score in addition to storage and processing accuracy, the correlation approximated the overall aggregated effect size of the relationship the best ($r = .30$). At the same time, Shin pointed out the limited amount of research that addressed processing speed in a WM task and stated that the results should be interpreted with caution. In the current study, only the scores based on processing accuracy accounted for variance in L2 reading with topic familiarity. This shows that measurement of storage alone may not generate informative findings that can further the understanding of how WM interacts with other learner variables in cognitive performance such as reading comprehension.

It is worth mentioning here that there was a considerable discrepancy of processing speed between Japanese and Korean participants. However, L1 was not a significant predictor of participants’ L2 reading comprehension, nor did this difference affect the patterns of results when further analyses were conducted with the two L1 groups separately. This difference in processing speed, based purely on the researcher’s observation, may be attributed to Japanese students’ tendency to be extremely cautious and thorough to provide accurate answers to the semantic judgment task. Korean students, however, did not seem to focus on one specific subcomponent of the WM task.

From a methodological point of view, obtaining composite WM scores that reflect both storage and processing components is an improvement from relying solely on storage. Depending on the focus of L2 research, however, use of the composite WM scores may make it difficult to unearth the distinctive contributions that each of the WM task subcomponents make to L2 performance (e.g., reading comprehension). This is
because averaging the standardized scores (z-scores) for the subcomponents essentially cancels out a source of variance. It has become increasingly clear that performance on WM task subcomponents provides unique information about (a) the underlying constructs of WM and (b) individual differences in each of them that are independent of the others (Sagarra, 2017; Unsworth et al., 2009). Therefore, a cautious interpretation of past findings is warranted particularly because using only storage scores may not accurately capture one’s WM capacity. Because providing a fine-grained breakdown of the roles of WM subcomponents in relation to the construct of WM is beyond the scope of the current study, interested readers are referred to further readings (e.g., Friedman & Miyake, 2004; Unsworth et al., 2009; Vergauwe et al., 2014; Waters & Caplan, 1996).

The present study represents one of the few attempts aimed at furthering our understanding of the role of WM in L2 reading comprehension by investigating how WM and topic familiarity interact. By examining the pattern of the interplay between WM and topic familiarity, the study illustrates that (a) using topic familiarity to one’s advantage is an action that requires WM resources and (b) the facilitative effect of topic familiarity is amplified by high-WM capacity: the rich get richer.

Additionally, the study is the first to provide nuanced insights into which WM subcomponent (measured through the reading span task) contributes to L2 readers’ efforts to access information and benefit from topic familiarity. To reiterate the results, storage did not significantly predict L2 reading with topic familiarity, but processing accuracy did even when L2 knowledge was accounted for. Based on this finding, the study concludes that it is not enough to be able to simply store background knowledge of the topics in memory for it to enhance L2 reading comprehension. What drives the
individual differences in the effective use of topic familiarity in L2 reading turns out to be L2 readers’ ability to efficiently access and activate relevant knowledge; it is such ability that may be linked to their processing capacity.

The current study also suggests methodological implications and points out the necessity to include a measure that can tap processing accuracy because it plays a unique role in L2 reading comprehension. Thus, using WM scores that do not reflect processing accuracy may not be able to (a) capture the unique aspect of WM and (b) generate informative findings that can contribute to our understanding of how WM interacts with other learner variables in reading comprehension.

**Pedagogical Implications**

The current study suggests implications for L2 reading instruction in regard to what teachers should be aware of when designing activities in the pre-reading stages of their lessons. The findings not only reinforce the facilitative effect of topic familiarity on L2 reading comprehension but also reveal the potential challenges faced by L2 readers in accessing their prior knowledge and using it to their advantage due to the constraints imposed by WM. The literature on background knowledge generally converges on the notion that the more a reader knows, the easier it is to understand a text (e.g., Neuman et al., 2014). In fact, background knowledge has been shown to empower readers’ comprehension achievement by helping them compensate for cohesion gaps in the texts that they are reading and make inferences during reading (e.g., Aukerman et al., 2015). Informational texts, in particular, tend to include concepts that are often directly related to students’ topic familiarity (Horiba & Fukaya, 2015). The current study demonstrated that even a simple pre-reading activity – during which students read a short text in their
L1s that simply introduces the topic – can significantly improve reading comprehension. While this activity may not be feasible in a linguistically heterogeneous class, this can also be done in learners’ L2. The objective of this activity is to introduce the topic prior to reading the target text so that students can be exposed to the new topic and build relevant knowledge, which ultimately facilitates their comprehension of a text on the same topic.

There is no denying that topic familiarity, or background knowledge in general, enhances reading comprehension. However, L2 teachers should not assume that students can make use of their familiarity with the topic without training. As addressed in Chapter 5, not every reader benefitted equally from topic familiarity. Teachers should be aware that the ability to take advantage of background knowledge is often taken for granted and that the process of accessing and retrieving background knowledge during the cognitively demanding task of L2 reading takes the orchestration of cognitive resources; students with low WM capacity are less likely to put their existing knowledge sources to use effectively. Grabe and Stoller (2020) list connecting text to background knowledge as one of numerous empirically validated reading strategies, which implies that it is a deliberate activity undertaken by readers and that the strategy is teachable. In a similar vein, Carrell (1988) emphasized that being equipped with prior knowledge is not the same as being able to activate and use that knowledge by integrating it with information being extracted from the text. Carrell added that pre-reading activities can fulfill their purpose when they help students (a) build new background knowledge and (b) activate existing background knowledge. Considering that a classroom will likely include students with a wide range of WM capacities, the benefits of pre-reading instruction can be maximized if
background-knowledge provision is not the end of a set of instructional activities but a starting point (Shin et al., 2019). Therefore, teachers should not only provide background knowledge if assigning unfamiliar topics but also employ explicit instruction to help students activate background knowledge, especially to accommodate low-WM L2 readers.

Examples of pre-reading activities that might help in this regard include (a) text previewing to help students think about what they already know about the topic and (b) the use of a set of related content resources (e.g., short texts, videos) prior to asking students to read the target text. (See Anderson, 2009; Grabe & Stoller, 2020; Hedgcock & Ferris, 2018, for additional pre-reading activities that tap students’ background knowledge.) In a text-previewing activity, the teacher can guide students in familiarizing themselves with the target text by bringing their attention to the title, subheadings, visuals, and other outstanding text characteristics. Teachers can then ask students to predict the topic of the text, prompting students to activate what they know about the topic. This task can be extended to a full-class discussion during which teachers can elicit relevant information and students can expand their topic-related knowledge by sharing their thoughts with peers.

Another pre-reading activity, which entails the use of a set of related content resources, involves guiding students through a series of texts with an aim to build background knowledge of the topic that students will soon encounter. A well-planned implementation of this method – using a series of texts – is discussed in detail in Lupo and colleagues (2017). They developed a pre-reading protocol, called a quad text set framework, in which four different types of texts are included: the target text, which is
the challenging one, and three other text types (texts with visuals, informational texts, accessible texts) that are intentionally less challenging and sequenced with the purpose of building and activating background knowledge. In a set, the texts are put in the following order: visual, informational, accessible, and target text. While adaptations and modifications can be made according to students’ specific needs, the goals of each of the texts included in the set are as follows: (a) introduce the topic through a visual text; (b) provide more specific information about the topic and build additional background knowledge through an informational text; (c) help readers make connections to and find relevance in the topic through an accessible text, such as young adult fiction or a text about popular culture; and (d) achieve instructional objectives through the target text. This sequence prompts students to use their background knowledge as they continue to read about the same topic; the four-text routine has been shown to help readers identify the main theme in the target text, which was more challenging than the previous texts.

In terms of the timing of background knowledge provision in the pre-reading stages of a lesson, it may be advisable for teachers to assign short readings to students the day prior to reading the target text(s). Sleep has been shown to consolidate memories and learning as it strengthens memory representations (Fenn & Hambrick, 2011; Gais, Lucas, & Born, 2006). Thus, giving students time to sleep on their newly instilled background knowledge may help the process of declarative memories becoming consolidated as well as their retrieval later when students encounter the target text on the same topic. Whether or not there is a positive effect of sleep specifically on the use of background knowledge in L2 reading contexts is an empirical question. Nonetheless, because memory consolidation during sleep will lead to more successful and efficient retrieval of
information, experimenting with the timing of background knowledge provision may be a worthwhile endeavor.

One common theme that runs through the pre-reading activities described above is that they serve the goal of providing opportunities for students to build relevant knowledge and use the knowledge (e.g., in prediction activities, in full-class discussions, with multiple texts on the same topic as the target text). The benefits that result from effective pre-reading activities such as these support L2 readers, particularly those with low WM, who need additional support for taking advantage of their background knowledge to facilitate reading comprehension.

Limitations of the Study

There are limitations regarding the methods employed in this study that should be addressed. The measures were used to provide quantitative information of participants’ performance on the following: WM via the reading span task (RST); L2 knowledge via the Vocabulary Size Test (VST) and C-test; and L2 reading comprehension via adapted TOEFL tests.

First, the current study used the L2 RST as a WM measure. According to a domain-specific view on WM, the RST is more directly relevant for L2 research because it is associated with language processing whereas the use of non-linguistic tasks such as the operation span task (Turner & Engle, 1989), for instance, is a product of a domain-general view on WM. Wen (2012) argued that L2 researchers who wish to investigate the effects of WM on specific L2 sub-skills (e.g., reading) should construct complex span tasks based on the domain-specific view because complex span tasks that are associated with language processing (e.g., RST) can tax the underlying construct more closely (see
Juffs & Harrington, 2011 and Wen, 2016 for more detail on domain-specificity and domain-generality of WM).

An additional issue concerning the L2 RST comes from the fact that because the RST requires knowledge of the language used in the task, which, in this case was participants’ L2, proficiency may be a confounding factor. While there is the possibility that less proficient students may experience challenges with this type of RST due to working in their L2, more advanced students have been reported to perform comparably in L1 and L2 RSTs (Alptekin & Erçetin, 2011). The current study mitigated this potential issue by incorporating in its RST English sentences consisting of words below the 2K band from graded readers. This cut-off – the 2K band – is far below the average English vocabulary size (6K word families) of the participants. Also, the correlation between participants’ performance on the RST and L2 reading comprehension test was non-significant in the reading-without-topic-familiarity condition and was significant in the reading-with-topic-familiarity condition. That is, variance in WM and reading comprehension did not overlap in the absence of topic familiarity, which shows that the targeted ability in WM is independent of general reading ability, unless called upon by an external variable (e.g., topic familiarity). In addition, because studies that focus on the interaction between WM and L2 reading have also used the L2 RST (e.g., Alptekin & Erçetin, 2011; Shin et al., 2019), the current study aimed to provide comparable results to allow for generalizations of the findings. Nonetheless, the potential issues stemming from using the L2 RST are acknowledged, and empirical work on examining the interplay between WM and L2 reading by using a non-linguistic WM task, such as the operation span task (Engle, 2002; Turner & Engle, 1989) or an L1 reading span task, would shed
light on this matter. However, it should be noted that L1 reading span has shown a weaker correlation with L2 reading comprehension than L2 reading span (Miyake & Friedman, 1998) and that L2 WM has shown a stronger link with L2 reading comprehension (Erçetin, 2015).

Also related to the RST is the discrepancy observed between Japanese and Korean participants, with regard to the processing speed subcomponent of the task. As further analysis revealed, the L1 made no difference to the mixed-effects model as a fixed factor, and the two L1 groups showed the same patterns of results. However, the fact that one particular group of participants used more time in their attempts to perform well on the semantic judgment task (while holding a set of verbal items in memory) raises a question of the possibility that some participants may have been able to use strategies during the task. Imposing a time limit on the semantic judgment task would be a way to minimize participants’ reliance on strategies.

Another limitation of the study concerns the L2 knowledge measures. Although a more comprehensive test such as TOEFL would have provided a more accurate index of L2 proficiency, the study used the VST and C-test due to time constraints and cost. While the VST was a significant predictor of L2 reading comprehension, the C-test was not, which indicates that the C-test may not have been an effective measure of L2 knowledge. Although vocabulary size serves as a predictor of general proficiency (Stæhr, 2008), the study’s findings pertinent to L2 knowledge should be interpreted with caution. Future studies are advised to use multiple and more rigorous measures to index participants’ L2 knowledge (e.g., TOEFL).
Involving two L1 participant groups (80 Korean speakers and 78 Japanese speakers) in this study in its attempt to increase generalizability of the findings was no small undertaking. Also, the results from the two L1 groups turned out to be comparable. However, culture-specific behavior can exist in other L2 speakers of English (e.g., Japanese participants showing much slower processing speed than Koreans), which might affect their performance and the overall picture of the findings as a result. It should be noted that Korean and Japanese speakers learning English as an L2 are in no way representative of the entire L2-learner population. Additionally, the researcher’s intention is not to assume that the current study’s findings apply to all L2 learners but to add to the growing body of research on the interplay between WM and background knowledge in L2 reading. The researcher attempted to do so by providing reliable results generated from a relatively large-scale study that examined two L1 groups using piloted and refined measures. Nonetheless, future studies will, I hope, expand the L2 learner group that has been involved in research on this topic (L1 Turkish – L2 English in Alptekin & Erçetin, 2011; L1 Korean – L2 English in Shin et al., 2019; L1 English – L2 Spanish in Leeser, 2007; L1 English – L2 Spanish in Payne et al., 2009; and L1 Japanese, Korean – L2 English in this study).

**Directions for Future Research**

The results of the study open up new avenues for future research. First, the study provided empirical evidence that the ability to take advantage of topic familiarity in L2 reading comprehension is amplified by WM resources, hence favoring the rich-get-richer model. However, participants were provided with a very limited amount of information to be merely exposed to the topic rather than to build knowledge prior to reading the target
texts. Perhaps, the provision of more content information in the pre-reading stage of an intervention would result in different interactional patterns. For instance, expert or extensive knowledge on a topic could compensate for low WM. Future research might consider investigating the pattern in which WM interacts with the degree of topic familiarity across participants. Findings from such research might yield evidence-based informative implications for language instructors as well as material writers. If it does turn out that expert knowledge can compensate for low WM, it would mean that L2 readers can benefit from instructional materials that incorporate multiple texts on the same topic (e.g., quad text set framework), which might help them establish extensive topic knowledge before reading the target text.

Second, the current study is one of the few that examined WM by using a time-sensitive concurrent span task that generated scores for storage, processing accuracy, and processing speed. This scoring system allowed for the examination of interrelationships among the three subcomponents, which demonstrated that processing speed negatively correlated with both storage and processing capacity. These inverse relationships were reflected in the composite WM scores as the scores on the three WM task subcomponents were standardized and averaged (note that processing speed was multiplied by -1). More research on this scoring procedure in terms of validity and reliability is called for to ensure an accurate conceptualization of the construct of WM and to illustrate the disadvantages of relying solely on storage scores. With more future research, this inclusive scoring system may also provide a more informative account of the role of WM in L2 performance with respect to the distinctive contributions of each of the subcomponents. As previously mentioned, a time limit on the semantic judgment task
may be an important feature to add to task procedures to ensure that all participants perform within a comparable time frame while also eliminating the possibility of some participants taking more time in their attempts to use strategies.

Finally, future studies that investigate the interplay between WM and background knowledge might consider including measures of the components of executive function such as updating, shifting, and inhibition (see Miyake et al., 2000; Miyake & Friedman, 2012 for more discussion on executive function). For instance, perhaps readers’ ability to switch their attention flexibly is reflected in their processing capacity and is also related to L2 reading comprehension. Or is it perhaps updating ability that is associated with processing capacity as well as with L2 reading comprehension? Examining which control mechanisms are linked to L2 learners’ ability to use background knowledge in reading would offer more insights into learners’ behavior with respect to self-regulation during L2 reading.

**Summary of the Chapter and Closing Remarks**

This last chapter of the dissertation opened with a brief review of major findings from the investigation of the interplay between working memory (WM) and topic familiarity in L2 reading comprehension. The review was followed by a more in-depth interpretation of those results; when pertinent, current findings were connected to previous research. This discussion led to unique insights into (a) the role of WM as cognitive resources upon which L2 readers draw and (b) L2 learners’ behavior regarding how they divide attentional resources during the WM task. In addition, implications for and contributions to pedagogical practices were explored. Furthermore, the study’s limitations were discussed and directions for future research were proposed.
The research reported in this dissertation contributes to the ongoing discussion of the interplay between WM and background knowledge (operationalized as topic familiarity in this study) in L2 reading comprehension. Among the three possible models (compensation, independent-influences, richer-get-richer) originally proposed by Hambrick and Engle (2002), the findings in the current study corresponded to the rich-get-richer model. This model indicates that, on the one hand, WM resources can amplify the facilitative effect of background knowledge and, on the other hand, not everyone can take advantage of what they know due to the constraints imposed by WM. In this regard, the current study highlights the dynamics of L2 readers’ individual differences and provides a more nuanced understanding of the impact of background knowledge on L2 readers’ performance. It is my hope that this study will serve as an impetus for future investigations into the significance of WM in L2 reading comprehension and WM’s interaction with other reader-based variables, such as background knowledge. Such studies will likely continue to shed light on the complex dynamics of individual differences in the contexts of L2 reading and will contribute to promoting instructional approaches and strategies that recognize L2 readers’ different needs.
References


Bates, D.M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.


Kintsch W. (2013). Revisiting the construction-integration model of text comprehension
Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 807–

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and


University Press.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second
language reading development. In K. Koda (Ed.), *Reading and language learning*
(pp. 1–44). Special issue of *Language Learning Supplement, 57*, 1–44.


*Reading Research Quarterly, 19*, 468-481.


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


Shin, J. (under review). A meta-analysis of the relationship between working memory and L2 reading comprehension: Does task type matter?


Appendices

Appendix A: Questionnaire (English)

Part 1.

Thank you for participating in this study. Please answer the following questions.

1. Name: _____________________
2. Age: ______________________
3. Sex (Please circle one): Male / Female
4. Major: ______________________
5. Years of studying English: ______________________________________
6. What was your total score (approximately) of the most recent English test (e.g., TOEFL, TOEIC, IELTS)? If you remember, please provide your reading subscore and the approximate date of the test.
   Test ____________ Total score _________________ Reading score ________________
   Date of test ______________________________________

Part 2.

I would like to know what you know about the topics below. Please respond to the question. Then, write down as much as you know about the topics using your first language.

<table>
<thead>
<tr>
<th>How familiar are you with the following topics?</th>
<th>I can tell you ( ) about this topic.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nothing</td>
</tr>
<tr>
<td>1. Ancient Rome and Greece</td>
<td>1</td>
</tr>
<tr>
<td>2. Agriculture, iron, and the Bantu people</td>
<td>1</td>
</tr>
<tr>
<td>3. History of the chickenpox vaccine</td>
<td>1</td>
</tr>
<tr>
<td>4. The rise of Teotihuacán</td>
<td>1</td>
</tr>
</tbody>
</table>
1. What I know about **ancient Rome and Greece**:

2. What I know about **agriculture, iron and the Bantu people**:

3. What I know about **the history of the chickenpox vaccine**:

4. What I know about **the rise of Teotihuacán**:
Appendix B: Questionnaire (Korean)

Part 1. 연구에 참여해주셔서 감사합니다. 아래의 질문에 답해주세요.

1. 이름: _____________________
2. 나이: _____________________
3. 성별 (둘 중 하나를 골라주세요): 남 / 여
4. 전공: _____________________
5. 영어 공부 기간: _____________________
6. 가장 최근에 받은 영어테스트 점수 (TOEFL, TOEIC, IELTS 등)는 약 몇점 인가요? 테스트의 읽기부분 점수와 테스트 날짜를 대략 알고 계시다면 적어주세요.

테스트 이름 __________ 총점수 __________ 읽기점수 __________
테스트 날짜 ___________________

Part 2. 아래 주제 관련한 배경지식을 얼마나 가지고 계신지 알아보려 합니다. 아래 질문에 답한 후 첫장에 각 주제에 대해 얼마나 알고 있는지 한글로 적어주세요.

<table>
<thead>
<tr>
<th>아래 주제에 관해 얼마나 많이 알고 있나요?</th>
<th>이 주제에 관해 ( )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>아무것도 모른다</td>
</tr>
<tr>
<td>1. 고대 로마와 그리스</td>
<td>1</td>
</tr>
<tr>
<td>2. 반투족의 농사와 철</td>
<td>1</td>
</tr>
<tr>
<td>3. 수두백신의 역사</td>
<td>1</td>
</tr>
<tr>
<td>4. 테오티우아칸의 성장</td>
<td>1</td>
</tr>
</tbody>
</table>
1. 고대 로마와 그리스에 대해 아는 것:

2. 반투족의 농사와 철에 대해 아는 것:

3. 수두백신의 역사에 대해 아는 것:

4. 테오티우아칸의 성장에 대해 아는 것:
Appendix C: Questionnaire (Japanese)

Part 1. この研究にご参加していただいてありがとうございます。続いて、以下の質問を答えてください

1. 名前: _____________________
2. 年齢: _____________________
3. 性別 (二つのうち一つを選んでください): 男 / 女
4. 専攻: _____________________
5. 英語の勉強期間: _____________________
6. 一番最近になった英語の試験の点数は何ですか (TOEFL, TOEIC, IELTS 等)。読解点数とテストの日付を大体知っているなら書いてください。

テストの名前____________ 総点数____________ 読解点数____________
試験日 _______________________

Part 2. 背景知識がどれだけいるか調べてみましょう。下記の質問に答えてから次のページに各トピックに関して知るほど日本語で書いてください。短い単語や文章を少なくとも大丈夫です。

以下のトピックについてどの程度知っていますか

<table>
<thead>
<tr>
<th>トピック</th>
<th>全く知らない</th>
<th>ほんの少し</th>
<th>多少は</th>
<th>たくさん</th>
</tr>
</thead>
<tbody>
<tr>
<td>古代ローマギリシャ</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>パンツー族の農業と鉄</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>水疱瘡ワクチンの歴史</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>テオティワカンの成長</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
1. 古代ローマギリシャについて知ってること:

2. バンツー族の農業と鉄について知ってること:

3. 水疱瘡ワクチンの歴史について知ってること:

4. デオティワカンの成長について知ってること:
## Appendix D: Reading Span Task Sentences

<table>
<thead>
<tr>
<th>Level</th>
<th>Order</th>
<th>Sentences</th>
<th>Final word</th>
<th>Semantic judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>1</td>
<td>A pencil died in August in a horrible car accident.</td>
<td>accident</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>She fell in love when she was seventeen years old.</td>
<td>old</td>
<td>YES</td>
</tr>
<tr>
<td>2-2</td>
<td>3</td>
<td>After 3 years, a book left the army and studied farming.</td>
<td>farming</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>When Jenna’s father died, he left her a lot of money.</td>
<td>money</td>
<td>YES</td>
</tr>
<tr>
<td>3-1</td>
<td>5</td>
<td>She enjoyed swimming and going for picnics on the beach.</td>
<td>beach</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Sarah ran home as fast as she could because she was late.</td>
<td>late</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>The diamonds were a present to the cake from the king.</td>
<td>king</td>
<td>NO</td>
</tr>
<tr>
<td>3-2</td>
<td>8</td>
<td>A table went to a summer camp with its sister last year.</td>
<td>year</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>People from all over the world like the taste of chocolate.</td>
<td>chocolate</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>She was not interested in singing but was good at dancing.</td>
<td>dancing</td>
<td>YES</td>
</tr>
<tr>
<td>4-1</td>
<td>11</td>
<td>It was a beautiful house, but it was also very cold.</td>
<td>cold</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>John asked his box if it would make his favorite meal.</td>
<td>meal</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>When Sally came back, Stephen was ready to leave the town.</td>
<td>town</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>Jack started talking, and the truck listened carefully and nodded often.</td>
<td>often</td>
<td>NO</td>
</tr>
<tr>
<td>4-2</td>
<td>15</td>
<td>The three birds saw little men flying in the sky.</td>
<td>sky</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Bill’s dad was going to drive Kevin to his computer’s house.</td>
<td>house</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Jenna was surprised when she saw the inside of a watch.</td>
<td>watch</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>Anna put her hand over her eyes to hide her tears.</td>
<td>tears</td>
<td>YES</td>
</tr>
<tr>
<td>5-1</td>
<td>19</td>
<td>People were eating ice cream although it was quite angry outside.</td>
<td>outside</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>Michael needed some money and wanted to sell his sofa.</td>
<td>sofa</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>After 5 years, Jack made enough money to support his father.</td>
<td>father</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>Healthy muscles help the mountain enjoy fun activities like swimming.</td>
<td>swimming</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>The calendar must go again and catch another fish for his dinner.</td>
<td>dinner</td>
<td>NO</td>
</tr>
<tr>
<td>5-2</td>
<td>24</td>
<td>Four men were riding on cell phones along a country road.</td>
<td>road</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>Nick saw a few grey hairs after a shower that night.</td>
<td>night</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>A little microwave with wide eyes was holding onto his hand.</td>
<td>hand</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>The armchair stood by the window waving until Stephen disappeared.</td>
<td>disappeared</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Jack saw many paintings on the wall by the chair.</td>
<td>chair</td>
<td>YES</td>
</tr>
<tr>
<td>6-1</td>
<td>29</td>
<td>The garden was not big, but it was a wonderful place.</td>
<td>place</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>30</td>
<td>Stephen thought the living room sounded a bit bored yesterday.</td>
<td></td>
<td>yesterday</td>
<td>NO</td>
</tr>
<tr>
<td>31</td>
<td>Sam talked to his cat when the door was lonely.</td>
<td></td>
<td>lonely</td>
<td>NO</td>
</tr>
<tr>
<td>32</td>
<td>The sun was burning hot and the air was getting warm.</td>
<td></td>
<td>warm</td>
<td>YES</td>
</tr>
<tr>
<td>33</td>
<td>When the wallet opened, the moon fell upon the ground.</td>
<td></td>
<td>ground</td>
<td>NO</td>
</tr>
<tr>
<td>34</td>
<td>It took hours to get there, but the wall was happy.</td>
<td></td>
<td>happy</td>
<td>NO</td>
</tr>
<tr>
<td>35</td>
<td>Nicole heard Sam talking to buildings on the phone.</td>
<td></td>
<td>phone</td>
<td>NO</td>
</tr>
<tr>
<td>36</td>
<td>Now, the highway is ready to start its own business.</td>
<td></td>
<td>business</td>
<td>NO</td>
</tr>
<tr>
<td>37</td>
<td>They walked down the narrow street to get to the market.</td>
<td></td>
<td>market</td>
<td>YES</td>
</tr>
<tr>
<td>38</td>
<td>I came to have drinks with the swimming pool’s parents.</td>
<td></td>
<td>parents</td>
<td>NO</td>
</tr>
<tr>
<td>39</td>
<td>Everyone was very polite, and I thanked them for the invite.</td>
<td></td>
<td>invite</td>
<td>YES</td>
</tr>
<tr>
<td>40</td>
<td>I spoke to Sally, and trees walked to the mall.</td>
<td></td>
<td>mall</td>
<td>NO</td>
</tr>
<tr>
<td>41</td>
<td>Nicole worked very hard but was worried about her exam.</td>
<td></td>
<td>exam</td>
<td>YES</td>
</tr>
<tr>
<td>42</td>
<td>It was only 4 o’clock, but the Chinese restaurant was closed.</td>
<td></td>
<td>closed</td>
<td>YES</td>
</tr>
<tr>
<td>43</td>
<td>The shop was full of green beans in their blue uniforms.</td>
<td></td>
<td>uniforms</td>
<td>NO</td>
</tr>
<tr>
<td>44</td>
<td>They spent 2 hours sitting in the traffic in the morning.</td>
<td></td>
<td>morning</td>
<td>NO</td>
</tr>
<tr>
<td>45</td>
<td>John was surprisingly happy to be studying again after the holidays.</td>
<td></td>
<td>holidays</td>
<td>YES</td>
</tr>
<tr>
<td>46</td>
<td>Eliot sat down and watched the rain running down his window.</td>
<td></td>
<td>window</td>
<td>YES</td>
</tr>
<tr>
<td>47</td>
<td>Michael took his country to his parents’ house for two days.</td>
<td></td>
<td>days</td>
<td>NO</td>
</tr>
<tr>
<td>48</td>
<td>Mary tried to call several times, but there was no answer.</td>
<td></td>
<td>answer</td>
<td>YES</td>
</tr>
<tr>
<td>49</td>
<td>Nick soon realized that the field was tired after the run.</td>
<td></td>
<td>run</td>
<td>NO</td>
</tr>
<tr>
<td>50</td>
<td>Nicole walked to the market and bought a lot of vegetables.</td>
<td></td>
<td>vegetables</td>
<td>YES</td>
</tr>
<tr>
<td>51</td>
<td>Dogs tried to build many apartments in a lot of cities.</td>
<td></td>
<td>cities</td>
<td>NO</td>
</tr>
<tr>
<td>52</td>
<td>After we got out, the lake started to go up a hill.</td>
<td></td>
<td>hill</td>
<td>NO</td>
</tr>
<tr>
<td>53</td>
<td>The balloon booked a table for 8 o’clock for two people.</td>
<td></td>
<td>people</td>
<td>NO</td>
</tr>
<tr>
<td>54</td>
<td>It started raining, and there was a huge crash of thunder.</td>
<td></td>
<td>thunder</td>
<td>YES</td>
</tr>
<tr>
<td>55</td>
<td>He sent ships to the hat to trade with foreign lands.</td>
<td></td>
<td>lands</td>
<td>NO</td>
</tr>
<tr>
<td>56</td>
<td>The storm had passed but the sky was still grey.</td>
<td></td>
<td>grey</td>
<td>YES</td>
</tr>
<tr>
<td>57</td>
<td>Milton kindly gave all of his money to the school.</td>
<td></td>
<td>school</td>
<td>YES</td>
</tr>
<tr>
<td>58</td>
<td>The history of music goes back to the beginning of culture.</td>
<td></td>
<td>culture</td>
<td>YES</td>
</tr>
<tr>
<td>59</td>
<td>The police station was crowded with people when the ocean arrived.</td>
<td></td>
<td>arrived</td>
<td>NO</td>
</tr>
<tr>
<td>60</td>
<td>You should be careful with the sun in the hot season.</td>
<td></td>
<td>season</td>
<td>YES</td>
</tr>
<tr>
<td>61</td>
<td>Jack tried to get up, but he could not move his legs.</td>
<td></td>
<td>legs</td>
<td>YES</td>
</tr>
<tr>
<td>62</td>
<td>A woman came in when the stone was still cleaning.</td>
<td></td>
<td>cleaning</td>
<td>NO</td>
</tr>
<tr>
<td>63</td>
<td>The tree stood in the yard, ready for slippers to climb.</td>
<td></td>
<td>climb</td>
<td>NO</td>
</tr>
<tr>
<td>64</td>
<td>I walked along the cow before anyone else was awake.</td>
<td></td>
<td>awake</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>There was nothing to do except wait, as the snow continued.</td>
<td>continued</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>Anna suggested going for a walk around the harbor after lunch.</td>
<td>lunch</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>There were books all over the floor in Bernard’s room.</td>
<td>room</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>John finally opened his closet and pulled out the blue suitcase.</td>
<td>suitcase</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>A group of young tables was playing basketball at the park.</td>
<td>park</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Jack just went into his juice to watch television and rest.</td>
<td>rest</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>
Appendix E: Instructions for Reading Span Task (English)

[Slide 1]

In this test, you will be asked to silently read English sentences and do your best to remember the last word in each sentence. After reading each sentence, you will be asked to decide whether or not the sentence semantically makes sense. So, when you see the question “Does this make sense?” you will respond by clicking on either YES or NO on the screen.

The number of answers you get right is important. And how quickly you can answer the questions is also important.

Here is an example of an acceptable and unacceptable sentence:

- Acceptable: Jack took a walk this morning.
- Unacceptable: A winter coat built a very tall tower.

When a set is finished, you will see “STOP AND WRITE” on the screen. When you do, please write down the last word of each sentence you read in the set. Please use the sheet in front of you. You may write them down in any order, but you cannot begin with the last word of the last sentence. You will get a zero if you do. Don’t worry too much about spelling errors.

Do you have any questions?

You will practice with some sentences to become familiar with the task.

<Press SPACE to continue.>
Before we start the practice, let me remind you what the tasks are:

1. Read English sentences silently.
2. Try to remember the last word in each sentence.
3. Decide if the sentence makes sense or not by clicking on YES or NO.
4. When you see “STOP AND WRITE,” write down the final word of each sentence you read on the sheet in front of you. Remember, you can write them in any order. But you cannot begin with the last word of the last sentence.

Let’s begin the practice test.

<Press SPACE to continue.>

Now, we will begin the test. Please try your best to remember final words AND answer the judgment questions as accurately and quickly as possible. The number of answers you get right is important. And how quickly you can answer the questions is also important.

<Press SPACE to begin the test..>
Appendix F: Instructions for Reading Span Task (Korean)

[Slide 1]

이 테스트에서 하실일은 다음과 같습니다. 영어 문장을 속으로 읽고 각 문장의 마지막 단어를 기억하시면 됩니다. 그 마지막 단어를 머리속에 계속 기억하면서, 한 문장을 읽을 때마다 그 문장이 의미상 말이 되는지 답하시세요. “이 문장이 말이 되나요?” 라는 질문이 나오면, 스크린에 YES 와 NO 버튼 중 하나를 클릭하시면 됩니다.

몇개의 단어를 정확하게 기억하는지도 중요하지만, 얼마나 빨리 그 문장이 의미상 말이 되는지 답하는 것도 중요합니다.

의미상 말이 되는 문장과 안되는 문장의 예시:
- 의미상 말이 됨: Jack took a walk this morning.
- 의미상 말이 안됨: A winter coat built a very tall tower.

한 세트가 끝날 때마다 스크린에 “STOP AND WRITE” 지시가 나옵니다. 그 때마다 기억하신 각 문장의 마지막 단어를 최대한 많이 주어진 종이 적으시면 됩니다. 어떤 순서로 적어도 상관이 없지만, 그 세트의 마지막 문장에 있는 마지막 단어부터 시작하시면 안됩니다. 이 경우 0점을 받게됩니다. 스펜링에 관해서는 크게 걱정하지 않으셔도 됩니다.

질문이 있으신가요?

몇가지 문장을 가지고 연습해보도록 하겠습니다.

<스페이스바를 누르면 다음으로 넘어갑니다>
연습을 시작하기 전에 다시 한번 더 테스트에서 하실 일을 확인하겠습니다.

1. 속으로 영어문장을 읽는다.
2. 각 문장의 마지막 단어를 머리속에 기억한다.
3. 각 문장의 의미가 맞는지 결정하고 YES 나 NO 를 눌러 답한다.
4. 스크린에 STOP AND WRITE 이라는 지시가 떠면 그 세트에서 읽은 문장의 마지막 단어들을 주어진 종이에 최대한 많이 적는다. 그 세트에서 마지막으로 본 문장의 마지막 단어로는 시작할 수 없다.

연습을 시작하겠습니다.

<스페이스바를 누르면 다음으로 넘어갑니다.>

[Instructions after practice]

이제 실제 테스트를 시작하겠습니다. 문장의 마지막 단어를 할 수 있는 한 최대한 기억하려고 애쓰면서 문장이 맞는지 또한 최대한 정확하고 빠르게 답해주세요.

몇개의 단어를 정확하게 기억하는지도 중요하지만, 얼마나 빠르게 그 문장이 의미상 맞이 되는지 답하는지도 중요합니다.

<스페이스바를 누르면 테스트가 시작합니다.>
このテストでは、英語の文章を黙読していただき、各文の最後の単語を出来る限り暗記していただきます。その最後の単語を覚えながら、各文章を読むたびにその文章の意味が通っていたか、を答えていただきます。“この文章の意味が通じますか？”という質問が出ると画面上の“YES”か“NO”をクリックしてください。

何個かの単語を正確に覚えるのも重要ですが、どれほど早くその文章の意味が通じているのか把握するのも重要です。

意味が通っているかどうかの例はこちらです。
• 意味が通っている: Jack took a walk this morning.
• 意味が通っていない: A winter coat built a very tall tower.

各セットが終わると、画面に“STOP AND WRITE”と表示されます。その時に覚えた各文章の最後の単語を、出来るだけあなたの前にある紙を使って書いてください。どの順番で書いていただいても構いませんが、そのセットの最後の文書にあった最後の単語から始めてはいけません。この場合は0点になります。単語のスペルをミスする事(つづりのミス)に関しては気にしなくても大丈夫です。

何か質問はありますか？

テストの内容に慣れていただくために、何問か練習を行っていただきます。

<”SPACE”を押して進んでください>
練習を始める前に、もう一度テストの方式について確認させて頂きます。

1. 英語の文章を黙読してください。
2. 各文章の最後の単語を暗記してください。
3. その文章の意味が通っていたかどうかを判断していただき、“YES”か”NO”をクリックしてください。
4. “STOP AND WRITE”と画面に表示されたら、そのセットで読んだ各文章の最後の単語を、あなたの前にある紙を使って書いてください。どの順番で書いていただいても構いませんが、そのセットの最後の文章にあった最後の単語から始めてはいけません。

それでは練習を始めましょう。

＜”SPACE”を押して進んでください＞

[Instructions after practice]

では、テストを始めます。出来る限り最後の単語を暗記していただき、そして意味が通っていたかどうかの質問にも出来る限り正確で素早く回答してください。何個かの単語を正確に覚えるのも重要ですが、どれほど早くその文章の意味が通じているのか把握するのも重要です。

＜”SPACE”を押して進んでください＞
# Appendix H: Reading Span Task Practice Sentences

<table>
<thead>
<tr>
<th>Level</th>
<th>Practice sentences</th>
<th>Final word</th>
<th>Semantic acceptability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1 The boy ran home and hid under the desk.</td>
<td>desk</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>2 The truck was hungry and ate 3 melons.</td>
<td>melons</td>
<td>NO</td>
</tr>
<tr>
<td>3</td>
<td>3 John tried to fix the coffee maker but failed.</td>
<td>failed</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>4 There was a plane crash in Japan in August.</td>
<td>August</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>5 Before the movie started, the juice went to get some popcorn.</td>
<td>popcorn</td>
<td>NO</td>
</tr>
<tr>
<td>4</td>
<td>6 Because it was cold, Jim wanted to wear a sweater.</td>
<td>sweater</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>7 An orange walked around the campus in the afternoon.</td>
<td>afternoon</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>8 My backpack was very heavy because of the books.</td>
<td>books</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>9 The movie theater was very quiet last Friday.</td>
<td>Friday</td>
<td>YES</td>
</tr>
</tbody>
</table>
Appendix I: Reading Span Task Sheet

Write down the last word in each sentence in a set.

각 세트에서 읽은 문장의 마지막 단어들을 적어주세요.

各文の最後の単語を書いてください。

<table>
<thead>
<tr>
<th>Practice set 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice set 2</td>
<td></td>
</tr>
<tr>
<td>Practice set 3</td>
<td></td>
</tr>
<tr>
<td>Set 1</td>
<td></td>
</tr>
<tr>
<td>Set 2</td>
<td></td>
</tr>
<tr>
<td>Set 3</td>
<td></td>
</tr>
<tr>
<td>Set 4</td>
<td></td>
</tr>
<tr>
<td>Set 5</td>
<td></td>
</tr>
<tr>
<td>Set 6</td>
<td></td>
</tr>
<tr>
<td>Set 7</td>
<td></td>
</tr>
<tr>
<td>Set 8</td>
<td></td>
</tr>
<tr>
<td>Set 9</td>
<td></td>
</tr>
<tr>
<td>Set 10</td>
<td></td>
</tr>
<tr>
<td>Set 11</td>
<td></td>
</tr>
<tr>
<td>Set 12</td>
<td></td>
</tr>
<tr>
<td>Set 13</td>
<td></td>
</tr>
<tr>
<td>Set 14</td>
<td></td>
</tr>
<tr>
<td>Set 15</td>
<td></td>
</tr>
<tr>
<td>Set 16</td>
<td></td>
</tr>
</tbody>
</table>

Last words
Appendix J: C-test (Klein-Braley, 1997)

Please fill in the blanks with the most appropriate word. As you can see, first part of the word is provided for you. You have 20 minutes to complete this task.

빈칸에 가장 알맞는 단어를 써주세요. 각 단어의 첫째 부분은 주어져 있습니다. 지금부터 20분이 주어집니다.

空欄に最適な単語を書いてください。単語の初めの部分は示されています。このテストは20分間で完了してください。

Text 1

Within the last twenty to thirty years the blood groups of peoples in all parts of the world have been studied. The mo____ interesting res____ of th____ studies h____ been th____, with f____ exceptions, nea____ every hu____ group exam____ has be____ found t____ consist o____ a mix____ of t____ same fo____ blood gro____; human ra____ differ i____ the rela____ numbers o____ persons wi____ them wh____ fall in____ each o____ the fo____ groups. Universal donors, group O, are found in every race and are generally the commonest type; group A is also common, while group B, and especially AB, is less common.

Text 2

If we go back to the Norman Conquest, we find nothing in this country which we could properly call a legal system. Indignant citi____ asserted th____ rights, a____ complained th____ their neig____ had bro____ his oblig____, and th____ spoke o____ taking h____ to co____. But i____ all exc____ the mo____ important mat____ between t____ king’s mo____ powerful subj____, the co____ to wh____ they ref____ was a lo____ court, a____ the rig____ which th____ claimed were those recognized by the custom of the neighborhood.

Text 3

For many years I have studied psychological processes entailed by our linguistic skills in communicating with one another. Since m____ interest i____ the psycho____ aspects o____ communication i____ even ol____ than auto____ computers I c____ remember wh____ those da____ before comp____ were li____. When I t____ to com____ them wi____ the pre____ I can th____ of n____ summary stat____ more appro____ than th____ made b____ a famous Amer____ athlete w____ said, “I’____ been rich, and I’ve been poor, and believe me, rich is better.” Believe me, computers are better.

Text 4

Many students of society – historians, political scientists, philosophers – find the study of works of literature useful and readily say so. They d____ not fe____ threatened b____ a different ki____ of disci____ or tem____ to ov____-stress th____ own subj____ special myst____. The hi____ degree o____ imagination nece____ for distin____ work i____ the human____ or soc____ science ens____ that m____ with th____ power d____ not mis____ the tech____ boundaries bet____ academic disci____ for div____ within human experience.
Appendix K: Background Texts (English)

Please read the following two passages 3 times and answer the comprehension questions at the end. The information in these passages will help you understand the TOEFL reading passages better that you will see on the test tomorrow.

1. The Bantu people

Bantu is one of the black tribes in Africa. The Bantu peoples are the proto-Bantu language speaking group, which includes Swahili, known as the most widely used in Africa, as well as Zulu, spoken by the biggest tribe in the Republic of South Africa. The Bantu peoples consist of about 300 tribes and each has its own language or dialect. “Bantu” is the term for “people” in the Bantu language. From 800 B.C. until 500 B.C., many Bantu tribes moved from the Congo Basin to the west and east and to south afterwards. The Bantu people migrated in small communities rather than as a unit of large families.

The migrations of the Bantu people to East Africa show 4 major dispersions, reaching the Great Lakes, the middle and Northern Tanzania, the coast and highlands, and Southern Tanzania. From the Congo forest in search of fertile land, they settled down in northern and western parts of Victoria Lake. From 1350 to 1500, the Bantu migrants who relied on agriculture were colonized by the Chwezi who relied on livestock. The Chwezi, a tribe that migrated south from the northern part of Africa, was skillful in handling iron and built protective fortress using soil and rocks.

It is believed that the population of Bantu increased rapidly around A.D. 1000. This expansion was probably caused by the introduction of yam, banana, and the use of iron. Some researchers think that this change spread out from the coast through the entry to Zambezi and big lakes. East Africa has the most diverse banana trees, and the bananas from there are served as basic food in Baganda. Banana plants are suckers that grow from the base of the banana plant and cannot be grown from seeds. Because of this, the Bantu, who uses bananas as a main food source, stopped their nomadic lifestyle and settled down in the forest that was humid but not dense. And they started making bronze farming tools to grow new crops.

Unfortunately, there is not enough archaeological evidence to find how iron metallurgy reached East Africa. It could have come from in the 4th century B.C. from Meroe which was flourishing with iron industry or from Indonesia through the Indian Sea or from the southwest of Africa, or from all three places. The only archaeological evidence we can rely on are ashes from blast furnaces and bellows. But considering the fact that in East Africa there was a use of bellows with dish-shaped blast, it is estimated that iron technology was probably spread from Sudan through Meroe.
Comprehension check

Please answer the questions below.

1) What does the term “Bantu” mean?

2) Why did the Bantu people migrate from the Congo forest?

3) What is the name of the tribe that colonized the Bantu people?

4) Why did the Bantu people stop moving around and settle down?

5) Why do people think that iron technology came from Sudan?

2. Teotihuacán

Teotihuacán is the biggest ancient historic site in the Americas and is located 50 kilometers northeast of Mexico City, the capital of Mexico. The term Teotihuacán means “birthplace of the gods.” This is because the Aztecs believed that the gods created the universe at that site. According to research, people started living in Teotihuacán before Christ and the city is thought to have been established around 100 B.C. The major monuments, irrigation facilities, roads, and so on were completed around the 1st A.D century, but many buildings were already constructed before A.D. Teotihuacán had a significant effect on civilizations in the ancient Americas.

Teotihuacán was the first metropolitan city of the Americas with a probable population of 125,000 during the fifth and sixth centuries A.D. The city declined between the seventh and tenth centuries A.D. until it was finally abandoned. Although archeologists can document the actual abandonment of the city, there is little evidence pointing to why it may have been abandoned.

The “Calzada de los Muertos,” or “Avenue of the Dead,” is the main path that connects the major historical sites at Teotihuacán from south to north. The Aztecs first called this road the Avenue of the Dead because they believed the platforms there were tombs, inspiring the name of the avenue. Scholars have now established that these were ceremonial platforms that were topped with temples. In the region far from the Avenue of the Dead, there is a residential area and farmlands for artisans, farmers, and merchants. Before Teotihuacán declined, it is said that 150,000 people lived in the city. Its population was similar to that of Athens and its size was bigger than Rome.
The Temple of the Feathered Serpent Quetzalcoatl, the irrigation canal, farmlands, and a residential area are placed at the southern end of the avenue. This area included a large plaza surrounded by temples that formed the religious and political center of the city. People used to do rituals here for feathered serpents that used be considered sacred. The outside edges of each level of the temples are decorated with feathered serpent heads. The irrigation canal which crosses the Avenue of the Dead is pretty wide and shows how large the city used to be. On the exterior, there are only remnants of big residential and farming areas. Excavations are still in process at the historical site of Teotihuacán where the civilization and culture of the ancient Americas is revealed.

Comprehension check
Please answer the questions below.

1) Where is Teotihuacán located?

2) When did people start to live in Teotihuacán?

3) Where did the street name “Avenue of the Dead” come from?

4) Why was the Feathered Serpent Quetzalcoatl important in the old days?

5) What is one piece of evidence showing that Teotihuacán was a large city?
다음 두 지문을 각각 세번째 읽고 질문에 답해주세요. 이 두 지문의 내용은 내일 영어 토플 읽기 테스트에서 읽게 될 지문과 관련있는 주제이므로 테스트에 도움이 될 내용을 포함하고 있습니다.

1. 반투족

반투족은 아프리카 흑인 종족 가운데 하나로, 반투어군에 속하는 언어들을 사용하는 민족들을 가리키는 말이며, 아프리카 최대 언어라고 일컬어지는 스와힐리어와 남아프리카 공화국 최대 민족인 줄루어의 사용자를 모두 포함하고 있다. 반투족은 약 300 개의 부족으로 이루어져 있으며, 모두 고유 언어나 방언이 있다. 반투족에서 반투는 반투어족 언어에서 공통적으로 인간이라는 뜻을 가지고 있다. 기원전 800 년부터 500 년 사이 반투어족의 여러 민족이 콩고 분지에서 서쪽과 동쪽으로, 후에는 남쪽으로 이동했다. 이들은 대가족 단위보다 작은 단위의 공동체로 이동하였다.

반투족의 동아프리카 이주는 크게 네 가지 이동 형태를 나타내고 있다. 이들은 호수 지역의 반투, 중부와 북부 탄자니아의 반투, 해안과 고지대의 반투, 남부 탄자니아의 반투 등으로 분류된다. 콩고 삼림 지대로부터 비옥한 땅을 찾아 이동한 이들은 빅토리아호의 북부와 서부 지대에 정착하게 된다. 농경 생활을 위주로 하는 반투족은 1350 년부터 1500 년까지 목축을 하는 츠웨지족의 지배를 받게 되는데, 츠웨지족은 북부 아프리카에서 남하해 온 부족으로, 이미 제철 기술을 숙지하고 있었으며, 흙과 돌을 이용하여 방어용 요새를 측성하기도 했다.

대략 서기 1000 년경에 반투족 인구의 급증이 있었다. 이것은 암(참마), 바나나의 전체와 절의 사용에 의해 야기되었을 것이라고 알려져 있다. 어떤 학자들은 이 변화들이 잠배지 인구와 큰 호수들에 경제적 해안으로부터 펼진 듯하다고 생각한다. 동아프리카는 세계에서 가장 많은 종류의 바나나나무가 있는데, 그것은 오늘날 바간다의 기본 식량로 제공되고 있다. 바나다는 희지로부터 자라야만 하고, 씨로는 키울 수가 없다. 이 때문에 주된 식량원으로 바나나를 이용하는 반투족은 이동 생활을 그만두고, 습하지만 너무 밀접된 숲이 아닌 곳에 정착하였고 새로운 작물 개발을 위해 철제 농기구를 만들기 시작했다.

그러나 불행하게도 제철 기술의 지식이 동아프리카에 도착하게 된 경로를 추적해 볼 만한 고고학적인 증거가 불충분하다. 그것은 기원전 4 세기경에 제철 공업의 번성지였던 메로에로부터 들어왔을 수도 있고, 인도네시아로부터 인도양을 경유하여 들어왔을 수도 있으며, 남서쪽에서 들어왔을 수도 있고, 또는 세 군데 모두로부터 들어왔을 수도 있다. 대체로 우리가 동아프리카에서 의존해야만 하는 고고학적인 증거라고는 용광로 재와 풀무뿐이다. 그러나 접시 모양의 용광로를

<table>
<thead>
<tr>
<th>Appendix L: Background Texts (Korean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>다음 두 지문을 각각 세번째 읽고 질문에 답해주세요. 이 두 지문의 내용은 내일 영어 토플 읽기 테스트에서 읽게 될 지문과 관련있는 주제이므로 테스트에 도움이 될 내용을 포함하고 있습니다.</td>
</tr>
<tr>
<td>1. 반투족</td>
</tr>
<tr>
<td>반투족은 아프리카 흑인 종족 가운데 하나로, 반투어군에 속하는 언어들을 사용하는 민족들을 가리키는 말이며, 아프리카 최대 언어라고 일컬어지는 스와힐리어와 남아프리카 공화국 최대 민족인 줄루어의 사용자를 모두 포함하고 있다. 반투족은 약 300 개의 부족으로 이루어져 있으며, 모두 고유 언어나 방언이 있다. 반투족에서 반투는 반투어족 언어에서 공통적으로 인간이라는 뜻을 가지고 있다. 기원전 800 년부터 500 년 사이 반투어족의 여러 민족이 콩고 분지에서 서쪽과 동쪽으로, 후에는 남쪽으로 이동했다. 이들은 대가족 단위보다 작은 단위의 공동체로 이동하였다.</td>
</tr>
<tr>
<td>반투족의 동아프리카 이주는 크게 네 가지 이동 형태를 나타내고 있다. 이들은 호수 지역의 반투, 중부와 북부 탄자니아의 반투, 해안과 고지대의 반투, 남부 탄자니아의 반투 등으로 분류된다. 콩고 삼림 지대로부터 비옥한 땅을 찾아 이동한 이들은 빅토리아호의 북부와 서부 지대에 정착하게 된다. 농경 생활을 위주로 하는 반투족은 1350 년부터 1500 년까지 목축을 하는 츠웨지족의 지배를 받게 되는데, 츠웨지족은 북부 아프리카에서 남하해 온 부족으로, 이미 제철 기술을 숙지하고 있었으며, 흙과 돌을 이용하여 방어용 요새를 측성하기도 했다.</td>
</tr>
<tr>
<td>대략 서기 1000 년경에 반투족 인구의 급증이 있었다. 이것은 암(참마), 바나나의 전체와 절의 사용에 의해 야기되었을 것이라고 알려져 있다. 어떤 학자들은 이 변화들이 잠배지 인구와 큰 호수들에 경제적 해안으로부터 펼진 듯하다고 생각한다. 동아프리카는 세계에서 가장 많은 종류의 바나나나무가 있는데, 그것은 오늘날 바간다의 기본 식량으로 제공되고 있다. 바나다는 희지로부터 자라야만 하고, 씨로는 키울 수가 없다. 이 때문에 주된 식량원으로 바나나를 이용하는 반투족은 이동 생활을 그만두고, 습하지만 너무 밀접된 숲이 아닌 곳에 정착하였고 새로운 작물 개발을 위해 철제 농기구를 만들기 시작했다.</td>
</tr>
<tr>
<td>그러나 불행하게도 제철 기술의 지식이 동아프리카에 도착하게 된 경로를 추적해 볼 만한 고고학적인 증거가 불충분하다. 그것은 기원전 4 세기경에 제철 공업의 번성지였던 메로에로부터 들어왔을 수도 있고, 인도네시아로부터 인도양을 경유하여 들어왔을 수도 있으며, 남서쪽에서 들어왔을 수도 있고, 또는 세 군데 모두로부터 들어왔을 수도 있다. 대체로 우리가 동아프리카에서 의존해야만 하는 고고학적인 증거라고는 용광로 재와 풀무뿐이다. 그러나 접시 모양의 용광로를</td>
</tr>
</tbody>
</table>
가진 풀무를 동아프리카에서 최근 사용한 것으로 미루어 그 제철 기술이 메로에로부터 수단을 경유하여 퍼진 것으로 추정된다.

지문 이해 체크

다음 질문에 답해주세요.

1) 반투족에서 반투가 의미하는 것은 무엇인가요?

2) 반투족의 이동의 목적은 무엇인가요?

3) 반투족을 지배한 부족의 이름은 무엇인가요?

4) 반투족이 이동 생활을 그만 둔 이유는 무엇인가요?

5) 제철 기술이 수단으로부터 들어왔다고 여겨지는 이유는 무엇인가요?

2. 테오티우아칸

멕시코의 수도 멕시코시티 북동쪽 50km 떨어진 곳에는 아메리카 대륙에서 가장 큰 고대 유적지 테오티우아칸이 있다. 테오티우아칸은 신들의 도시라는 뜻인데, 이것은 아즈텍 사람들이 신들이 이곳에서 우주를 창조했다고 믿었기 때문이다. 학자들의 연구에 따르면 테오티우아칸에는 기원전부터 사람들이 삶기 시작하고, 기원전 100년경부터 도시의 틀을 갖추게 되었다고 한다. 지금 남아 있는 주요 건축물과 관개 시설, 도로 등은 기원후 1세기경에 완성된 것이지만 이미 기원전에 상당히 많은 건축물들이 건설되었다. 테오티우아칸은 아메리카 대륙에서 탄생한 여러 문명에 커다란 영향을 미쳤다.

테오티우아칸은 아메리카 대륙 첫번째 메트로폴리탄 도시로, 기원 후 5, 6 세기에 12만 5천명의 인구가 살았다고 전해진다. 그러나 테오티우아칸은 기원후 7-10세기에 소멸하였고 결국 폐허가 되었다. 고고학자들이 저마다 테오티우아칸이 언제 폐허가 되어있는지 기록을 내놓고 있지만 왜 폐허가 되었는지를 보여주는 증거는 거의 남아있지 않다.
테오티우아칸의 주요 유적들은 유적지를 남북으로 가로지르는 ‘죽은 자의 거리’를 따라 모두 연결되어 있다. 이 길을 ‘죽은 자의 거리’라고 부르기 시작한 것은 아스텍 사람들이다. 이곳을 발견한 아스텍 사람들이 길 양쪽에 세워진 플랫폼을 무덤이라고 잘못 생각해 죽은 자의 거리라는 이름을 붙이게 되었다. 현재 학자들에 의해 밝혀진 바에 따르면 이것은 신전 밑에 있던 의식을 행하던 연단이라고 한다. 죽은 자의 거리에서 조금 떨어진 지역에는 장인과 농부, 상인들이 살았던 주거지와 농경지가 있다. 테오티우아칸이 멸망하기 전까지 이곳에 살았던 사람은 15 만 명이나 되었다고 전해지는데, 당시 인구는 아테네 인구와 비슷한 정도였고, 면적은 로마보다 넓었다.

죽은 자의 거리 남쪽 끝에서는 케찰코아틀 신전과 관개 수로 유적, 대광장, 주거지와 농지 등을 볼 수 있다. 케찰코아틀 신전은 주변은 큰 광장이었고 종교적 정치적 중심지를 이룬 신전들로 둘러쌓여있었다. 또한 성스러운 동물로 받들었던 깃털 달린 범을 위한 의식을 치르던 곳으로, 깃털 달린 범 머리 조각상이 신전 벽 바깥 부분에 충렬로 장식되어 있다. 죽은 자의 거리를 동서로 가로질러 흐르는 관개 수로는 재법 넓어, 당시 테오티우아칸이 얼마나 커다란 도시였는지 잘 보여 준다. 그리고 외곽에는 지금은 잔해만 남아 있는 넓은 주거지와 경작지가 있다.

고대 아메리카 대륙의 도시와 문화를 엿볼 수 있는 테오티우아칸 유적지는 지금도 발굴 작업이 진행되고 있다.

지문 이해 체크

다음 질문에 답해주세요.

1) 테오티우아칸은 어디에 위치해 있나요?

2) 테오티우아칸에 사람들이 언제부터 살기 시작했나요?

3) 테오티우아칸의 거리 이름인 ‘죽은 자의 거리’는 어떻게 지어진 이름인가요?

4) 케찰코아틀 신전은 어떤 역할을 하던 곳인가요?

5) 테오티우아칸이 거대한 도시였다는 것을 보여주는 증거를 하나만 적으세요.
以下の2つの文章を3回読み、質問に答えなさい。これらの文章に含まれている情報は明日のテストに出題されるTOEFLリーディングセクションの練習になります。

1. バントゥー

バントゥー族はアフリカの黒人系民族のひとつである。バントゥー族は、アフリカで最も広く使われていることで知られるスワヒリ語や、南アフリカ共和国のもっとも大きな族で話されているズル語を含む原始のバントゥー語を話す。およそ300の部族がこのバントゥー系に属し、それぞれ固有の言語または方言を使用している。バントゥー語族の言語でバントゥーという単語は人間を意味する。紀元前800年から500年までバントゥー語族に属する多数の民族がコンゴ盆地から西と東方面に、後には南方面に移住した。バントゥーは大家族より小規模な共同体をなして移住した。

バントゥー族の東アフリカへの移住は、湖地域、タンザニア中部及び北部、海岸や高地帯、タンザニア南部の大きく四つのパターンに分かれる。コンゴの山林帯から肥沃な土地を求め、彼らはビクトリア湖の北部と西部に定着する。農耕に従事していたバントゥー族は1350年から1500年まで牧畜を生業とするチュウェジ族の支配下に入った。チュウェジ族はアフリカ北部から南下してきた部族で、すでに製鉄技術を熟知しており、泥と石を活用して防御用の要塞を築城していた。

およそ西暦1000年頃バントゥー族の人口が急増したとされている。このような現象はヤムやバナナの伝来、鉄の利用によって引き起こされたのだろう。一部の研究者はこれらの変化はザンベジ川の入口と広い湖を経由し海岸地域から広まったと考えている。東アフリカで栽培されているバナナの品種の数は世界一で、バナナは今日バカンダ族の主食になっている。バナナは種子によって繁殖できず吸枝から育てるしかない。主要食料源となったバナナを栽培するためにバントゥー族は移動生活をあきらめ、木々が密集した森を避け湿った土地に定着し始めた。そして彼らは新たな作物を育てるために鉄製の農耕器具を作り始めた。

しかし残念ながら製鉄技術の知識が東アフリカに辿り着いた経緯を追跡できる考古学的証拠はまだ十分揃えていない。これは紀元前4世紀頃に製鉄工業で栄えたメロエから導入してきたとも、インドネシアからインド洋を経由し流入されたとも、南西部から伝わったとも、もしくはこの三
195

つの方面すべてから入ってきたとも言える。そもそも、私たちが東アフリカにおいて拠り所とする考古学的証拠とは高炉の灰とふいごだけだ。ただ、皿の形をした高炉つきのふいごが東アフリカで使用されていたことから、その製鉄技術はメロエからスーダンを経由して広まったと推定される。

質問
質問に答えなさい

1) 「バントゥー」とは何ですか。

2) なぜバントゥー族はコンゴの森林地帯から移住したのですか。

3) バントゥー族を占領した部族の名前は何ですか。

4) なぜバントゥー族は移動をやめて、定住したのですか。

5) なぜ製鉄技術はメロエからスーダンに広がったと考えられていますか。

2. テオティワカン

メキシコの首都メキシコシティの50km北東にはアメリカ大陸最大の規模を誇る古代遺跡地テオティワカンがある。テオティワカンという言葉の意味は「神の生まれた土地」である。これはアステカの人々は神がその場所で宇宙を造ったと信じているからである。研究によると、人々は紀元前にテオティワカンに住み始め、その町は紀元前1000年に造られたと考えられている。現在残っている主な建造物や灌漑設備、道路などは紀元後1世紀頃に完成されたものだが、建造物の多くはすでに紀元前に建設されている。テオティワカンはアメリカ大陸から誕生した文明に多大な影響を及ぼした。

テオティワカンは紀元後5から6世紀の間に約12万5千名の人口を持つアメリカ大陸で最初のメトロポリタンシティであった。この町は紀元後7世紀から10世紀の間に衰退し結局崩壊した。考古学者たちはその町が崩壊された事実を示すことは出来るが、なぜ崩壊されたかを示す証拠は殆ど残っていない。
テオティワカンの主な遺跡は遺跡地の北と南をつなぐ「死者の大通り」に沿って配置されている。この道路を「死者の大通り」と称し始めたのはアステカ人たちだ。なぜなら、ここを発見したアステカ人たちは道路の両側に建っている建造物を墓だと勘違いし「ミカオトリ」と呼んだ。古代アステカ語で「死者の大通り」という意味だ。研究者達は今、これらは神殿の下に置いて儀式的な建造物だと結論付けた。死者の大通りから少し離れた場所に職人や農夫、商人たちが暮らしていた住居地区と農地がある。テオティワカンが滅びる前までここで暮らしを営んでいた人の数は15万名以上だったと伝えられているが、当時の人口はアテネの人口と肩を並べるくらいで、土地の広さはローマより広大だった。

死者の大通りの南端ではケツァルコアトルの神殿や灌溉水路の遺跡、大広場、住居地区、農地などを見ることができる。この地区は、町の宗教や政治の中心であった神社に囲まれた大きな広場であった。ケツァルコアトルの神殿は、アメリカ大陸で栄えた古代文明の人たちが聖獣としていた翼の生えた蛇を崇める祭儀を執り行った場所で、翼の生えた蛇の頭像が神殿の壁に飾られている。死者の大通りを東西に横切る広い灌溉水路は当時テオティワカンがいかに巨大な都市であったかを目の前に突きつけてくる。そして外郭には住居地区や農地があったが、今はその姿は消え、残骸だけが残っている。

古代アメリカ大陸の都市や文化を垣間見られるテオティワカン遺跡地では今なお発掘作業が続いている。テオティワカン遺跡地はまだ究明されていない疑問があるだけに人々の好奇心を掻き立てる興味深い文化遺産だ。

質問

1) テオティワカンはどこにありますか。
2) テオティワカンに人が住み始めたのはいつですか。
3) 「死者の大通り」という通りの名前の由来は何ですか。
4) どうして、当時、羽の生えた蛇神・ケツァルコアトルは大事な存在でしたか。
5) かつてテオティワカンが大きい街だったということを示す証拠の一つは何ですか。

196
# Appendix N: Topic Familiarity Check (English)

In **2 minutes**, please write as much as you know about each topic **using your first language**. You may write single words, short/long phrases, or sentences.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1. | What I know about **Ancient Rome and Greece**:
|   |   |
| 2. | What I know about **the history of the chickenpox vaccine**:
|   |   |
| 3. | What I know about **the rise of Teotihuacán**:
|   |   |
| 4. | What I know about **the agriculture, iron, and the Bantu people**:  |
Appendix O: Topic Familiarity Check (Korean)

주제 당 각 2 분 안에 아래 주제에 관해 아는 것을 한글로 최대한 많이 적어주세요. 짧은 단어로 적으면도 되고 문장도 괜찮습니다.

1. 고대 로마와 그리스에 대해 아는 것:

2. 반투족의 농사와 철에 대해 아는 것:

3. 수두백신의 역사에 대해 아는 것:

4. 테오티우아칸의 성장에 대해 아는 것:
Appendix P: Topic Familiarity Check (Japanese)

各テーマ党2分以内に、以下のテーマに関して知るほど日本語で書いてください。短い単語や文章を少なくとも大丈夫です。

<table>
<thead>
<tr>
<th>1. 古代ローマギリシャについて知ってること:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. バンツー族の農業と鉄について知ってること:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. 水痘ワクチンの歴史について知ってること:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. テオティワカンの成長について知ってること:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
Appendix Q: Topic Familiarity Rating Protocol

1. Two raters – the researcher and another native Korean speaker – first read the background text written in Korean carefully.

2. Raters hold a norming session in which they rate three participants’ responses together.
   - Each unique idea or detail should be given one point. In the case of the Korean language, this usually means that each content word is awarded one point.
   - When ambiguity arises regarding what constitutes a unique idea or detail, raters first check to see if it is a case of a compound word. If it is, the compound word is counted as one word and thus receives one point.
   - In other cases, raters decide if removing the word would change the meaning of the phrase/sentence. If yes, the word should be given one point.
   - If ambiguity still remains, raters discuss the case and come to an agreement so that they can maintain consistency for a similar case that might occur.

3. Raters complete the rating task independently and compile the ratings once finished.
Appendix R: Practice Quiz

Before we begin the actual reading quiz, please complete this practice quiz to familiarize yourself with the question types. Please read the passage below. Then, answer the questions. When you finish, you can check the answers on the next page. Definitions of some words in the questions that might not be familiar to you are given in the box on the right side. Give yourself 10 minutes to complete this practice quiz.

읽기테스트를 시작하기 전에 질문타입에 익숙해지도록 연습문제를 먼저 풀어보겠습니다. 아래 글을 읽고 질문에 대답해주세요. 질문은 익숙하지 않을 수 있는 단어는 오른쪽 박스에 단어 뜻이 주어져 있습니다. 10분에 연습문제를 풀어보세요.

1. Which one of the following sentences below best expresses the important information in the highlighted sentence in Paragraph 1? Incorrect choices will change the meaning in important ways or leave out important information.

   (A) People want to know why and how they can communicate with animals.
   (B) People know how to communicate with each other but don’t know how to do the same with animals.
   (C) People know that animals know how to express what they want to each other but don’t know why it is not possible to do the same with humans.
2. The author mentions Dr. Dolittle in Paragraph 1 to
   (A) give an example of how much people have wanted to speak with animals
   (B) introduce a good movie to children who want to speak with animals
   (C) show that people can communicate with animals

3. Paragraph 2 mentions all of the following as animals’ ways of communicating **EXCEPT**
   (A) using sounds
   (B) using food
   (C) using body parts

4. It can be inferred from Paragraph 2 that
   (A) scientists know how to speak with animals
   (B) animals do not like to express themselves
   (C) there has been a lot of research on how animals communicate

5. The word “wag” in Paragraph 3 is closest in meaning to
   (A) lightly shake
   (B) drop
   (C) touch

6. In Paragraph 3 of the passage, there is a missing sentence. The paragraph is repeated below and shows four letters (A, B, C, and D) that show where the following sentence could be added.

```
Parrots and other birds, for example, can learn to say words.
```

**Where would the sentence best fit?**

- **People used to think it was possible to teach human language to animals.**
- **(A) But the birds just repeat the words. They don’t understand them.**
- **(B) Some American scientists tried to teach English to chimpanzees, close relatives of human beings. However, this was impossible.**
- **(C) Chimpanzees can’t move their mouths and throats the right way. They could never learn to speak like people.**

   (A) Option A
   (B) Option B
   (C) Option C
7. *Directions*: The first sentence of a short summary of the passage is given below. Complete the summary by selecting the THREE answer choices that express the most important ideas in the passage. Some sentences do not belong in the summary because they express ideas that are not in the passage or are minor ideas in the passage.

Write your answer choices in the spaces where they belong. You can either write the letter of your answer choice or you can copy the sentence.

<table>
<thead>
<tr>
<th>People have always tried to communicate with animals.</th>
</tr>
</thead>
<tbody>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
</tbody>
</table>

(A) People have been curious about how animals communicate with each other, so they have studied the ways they do.
(B) People made movies about how animals can communicate with each other.
(C) People tried to teach animals to talk like humans.
(D) People learned that it is impossible for animals to speak like they do.
Now, check your answers.

1. B
2. A
3. B
4. C
5. A
6. A
7. A, C, D
Now, let’s begin the reading quiz. This packet has 4 passages and 10 questions per passage. Read the passages carefully. Then, answer the questions. At the end of each set, you will be asked to indicate if there are any questions that you were able to answer just by using your background knowledge. Give yourself 20 minutes to complete each set. In total, you have 80 minutes to complete the entire quiz.

The city of Teotihuacán is located about 50 kilometers northeast of modern-day Mexico City. It began its growth by 200–100 B.C. Between about A.D. 150 and 700, it probably had a population of more than 125,000 people and covered at least 20 square kilometers. It had over 2,000 apartment complexes, a great market, a lot of industrial and commercial workshops, an administrative center, a number of massive religious buildings, and a regular pattern of streets and buildings. Clearly, much planning and central control went into the development of this great city. Moreover, the city probably had economic and religious relationships with most parts of Mesoamerica (modern Central America and Mexico).

How did this huge development happen? Why did it happen in the Teotihuacán Valley? The first reason is Teotihuacán’s excellent location: It is on a natural trade route to the south and east of the Valley of Mexico. The second reason is obsidian resources in the Teotihuacán Valley. Third, the valley had potential for extensive irrigation. The exact role of other reasons is difficult to pinpoint. For instance, perhaps it is Teotihuacán’s religious importance as a holy place. Maybe it is the historical situation in and around the Valley of Mexico toward the end of the first millennium B.C. Perhaps it is the ability of Teotihuacán’s elite to plan for the future. Finally, the impact of natural disasters, such as the volcanic explosions of the late first millennium B.C. may also be another reason.

This last reason is indirectly related to Teotihuacán’s growth. Prior to 200 B.C., there were many centers in and near the Valley of Mexico. Around this time, the largest of these centers, Cuicuilco, was seriously affected by a volcanic explosion. A lot of the agricultural land in Cuicuilco was destroyed. Therefore, Cuicuilco was not a potential rival anymore. Because of this, any relatively modest centers might have become leading economic and political powers in Central Mexico. However, the archaeological evidence clearly shows that Teotihuacán became the predominant force in the area by the first century A.D.
It seems likely that Teotihuacán’s natural resources along with the city elite’s ability to see their potential gave the city a competitive edge over its neighbors. The valley was rich in obsidian. This hard volcanic stone had been popular for many years, at least since the rise of the Olmecs (an early civilization in Mexico). Obsidian apparently had a steady and stable market. Moreover, recent research has shown that some collected by the Olmecs was originally from near Teotihuacán. Teotihuacán obsidian was probably considered as a valuable material for many centuries before Teotihuacán’s growth.

Long-distance trade in obsidian helped the elite people of Teotihuacán to get a lot of exotic products and have a relatively comfortable life. Their success probably brought people from other towns to Teotihuacán. It is also possible that as early as 200 B.C., Teotihuacán may have achieved religious significance. And its holy places may have been another population magnet. Finally, people in Teotihuacán probably had enough food for the growing population because of the increased number and size of watered fields.

The picture of the rise of Teotihuacán is a classic picture of positive feedback among obsidian mining and crafting, trade, population growth, irrigation, and religious tourism. The thriving obsidian operation, for example, would need more miners, more makers of obsidian tools, and more traders to carry the products to new markets. It all led to increased wealth. And the wealth attracted more people to Teotihuacán. The growing power of the elite, who controlled the economy, could also force people to move to Teotihuacán for additional labor work. More irrigation works would have to be built to feed the rapidly growing population. And this resulted in more power and wealth for the elite.

1. In Paragraph 1, each of the following is mentioned as a feature of the city of Teotihuacán between A.D. 150 and 700 EXCEPT

(A) regularly arranged streets
(B) several administrative centers spread across the city
(C) many manufacturing workshops
(D) apartment complexes

2. The word “pinpoint” in Paragraph 2 is closest in meaning to

(A) identify precisely
(B) make an argument for
(C) describe
(D) understand

3. Which of the following is NOT mentioned in Paragraph 2 as a main reason for the development of Teotihuacán?

(A) The presence of obsidian in the Teotihuacán Valley
(B) The potential for extensive irrigation of Teotihuacán Valley lands
(C) A long period of volcanic inactivity in the Teotihuacán Valley
(D) Teotihuacán’s location on a natural trade route
4. The word “predominant” in Paragraph 3 is closest in meaning to

(A) most aggressive  
(B) most productive  
(C) principal  
(D) earliest

5. What can be inferred from Paragraph 3 about Cuicuilco prior to 200 B.C.?

(A) It was a fairly small city until that date.  
(B) It was located outside the Valley of Mexico.  
(C) It emerged rapidly as an economic and political center.  
(D) Its economy relied heavily on agriculture.

6. Which of the following allowed Teotihuacán to have “a competitive edge over its neighbors”?

(A) Well-exploited and readily available natural resources  
(B) The presence of a highly stable elite class  
(C) Knowledge derived directly from the Olmecs about the art of toolmaking  
(D) Scarce natural resources in nearby areas such as those located in what are now the Guatemalan and Mexican highlands

7. According to Paragraph 4, what has recent research shown?

(A) Obsidian’s value was understood only when Teotihuacán became an important city.  
(B) The residents of Teotihuacán were sophisticated toolmakers.  
(C) The residents of Teotihuacán traded obsidian with the Olmecs as early as 400 B.C.  
(D) Some of the obsidian used by the Olmecs came from the area around Teotihuacán.

8. In Paragraph 6, the author discusses “The thriving obsidian operation” in order to

(A) explain why manufacturing was the main industry of Teotihuacán  
(B) give an example of an industry that took very little time to develop in Teotihuacán  
(C) illustrate how several factors influenced each other to make Teotihuacán a powerful and wealthy city  
(D) explain how a successful industry can be a source of wealth and a source of conflict at the same time

9. In Paragraph 1 of the passage, there is a missing sentence. The paragraph is repeated below and shows four letters (A, B, C, and D) that indicate where the following sentence could be added.

In fact, artifacts and pottery from Teotihuacán have been discovered in sites as far away as the Mayan lowlands, the Guatemalan highlands, northern Mexico, and the gulf Coast of Mexico.

Where would the sentence best fit?

The city of Teotihuacán is located about 50 kilometers northeast of modern-day Mexico City. It began its growth by 200–100 B.C. At its height, between about A.D. 150 and 700, it probably had a population of more than 125,000 people and covered at least 20 square
Teotihuacán was a highly developed city in Mesoamerica that reached its peak between about A.D. 150 and 700.

- The number of the architectural, administrative, commercial, and religious features of Teotihuacán show their centralized planning and control.
- Teotihuacán may have developed its own specific local religion as a result of the cultural advances made possible by the city’s great prosperity.
- There are many reasons for Teotihuacán’s extraordinary development, including its location, rich natural resources, irrigation potential, intelligent elite, and the misfortune of rival communities.
- As a result of its large number of religious shrines, by the first century A.D., Teotihuacán became the most influential religious center in all of Mesoamerica.
- In many important areas, from the obsidian industry to religious tourism, Teotihuacán’s success and prosperity typified the classic positive feedback cycle.
- Although many immigrants settled in Teotihuacán between A.D. 150 and 700, the increasing threat of coerced labor discouraged further settlement and limited Teotihuacán’s population growth.

The reading quiz continues on the next page.
AGRICULTURE, IRON, AND THE BANTU PEOPLE

1. There is evidence of agriculture in Africa before 3000 B.C. Many scholars believe that the spread of agriculture and iron throughout Africa is connected to the major centers of the Near East and Mediterranean areas. Because of the drying up of what is now the Sahara Desert, many peoples had to move to sub-Saharan Africa. At first, they settled in hunting- and-gathering areas. But those who fished lived near lakes and rivers. Because they had a more secure food supply, they lived in larger population communities. Agriculture perhaps started from the Near East because the first crops were not originally African but West Asian. When the idea of planting spread, Africans began to develop their own crops like rice. They were also receptive to the idea of imports. The farming areas of African crops ranged from Ethiopia across southern Sudan to West Africa. Other crops like bananas were introduced from Southeast Asia.

2. Livestock came from outside Africa, too. Cattle, domestic sheep, and goats were introduced from Asia. Horses were introduced by the Hyksos invaders of Egypt and then spread across the Sudan to Africa. Rock paintings in the Sahara indicate that horses and chariots were used to cross the desert. By 300–200 B.C., there were trade routes across the Sahara. Horses were adopted by peoples of the West African savanna. Later, their powerful cavalry forces allowed them to carve out large empires. Finally, the camel was introduced around the first century A.D. This was an important innovation. The camel’s ability to do well in harsh desert conditions and to carry large amounts cheaply made it an effective means of transportation. The camel changed the desert into a still difficult but more accessible route of trade and communication.

Iron came from West Asia. But the routes of the spread of iron were different from those of agriculture. Most of Africa presents an interesting case: Many African societies moved directly from a technology of stone to iron without passing through the middle stage of copper or bronze metallurgy. But some early copper-working sites have been found in West Africa. Knowledge of iron making spread into the forests of West Africa when it was reaching Europe. Evidence of iron making has been found in Nigeria and Ghana.

This technological change caused significant changes in African societies. Iron meant power. In West Africa, blacksmiths who made tools and weapons had an important place in society. They often had special religious powers. Tools like iron hoes made the land more productive. Iron weapons made the soldiers more powerful. This was important in many of African societies. People who knew how to make iron gained ritual and political power.

Unlike in the Americas, where metal-making was a late and limited development, Africans had iron from a relatively early date. So they developed an ingenious heater to produce the high heat to control the amount of air that reached the carbon and iron ore that was for making iron. Much of Africa moved right into the Iron Age by taking the technology and changing it to make it more appropriate for local conditions and resources.

The spread of agriculture and later of iron was followed by a movement of people who may have carried these innovations. They came from eastern Nigeria probably because of peoples fleeing the drying up of the Sahara. They spoke a language, Proto-Bantu (“bantu” means “the people”). It was the first language among the large number of Bantu languages, spoken in sub-Saharan Africa. Why and how these people spread out into central and southern Africa is a mystery. However, research has shown that they used their iron
weapons to fight and conquer people who were using stones to hunt and gather food. Still, the process is not clear, and peaceful migration—or simply rapid population growth—may have led to the Bantu explosion.

11. In Paragraph 1, what is implied about changes in the African environment during this time period?

(A) The climate was becoming milder, allowing for a greater variety of crops to be grown.
(B) Although periods of drying forced people south, they returned once their food supply was secure.
(C) Population growth along rivers and lakes was dramatically decreasing the availability of fish.
(D) A region that had once supported many people was becoming a desert where few could survive.

12. According to Paragraph 2, camels were important because they

(A) were the first domesticated animal to be introduced to Africa
(B) allowed the people of the West African savannahs to carve out large empires
(C) helped African peoples defend themselves against Egyptian invaders
(D) made it cheaper and easier to cross the Sahara

13. According to Paragraph 2, which of the following were subjects of rock paintings in the Sahara?

(A) Horses and chariots
(B) Sheep and goats
(C) Hyksos invaders from Egypt
(D) Camels and cattle

14. The author talks about how iron came from West Asia to Africa in Paragraph 3 to

(A) contrast the development of iron technology in West Asia and West Africa.
(B) discuss a non-agricultural contribution from Asia to Africa.
(C) introduce evidence that a knowledge of copper-working reached Africa and Europe at the same time.
(D) compare the rates at which iron technology developed in different parts of Africa.

15. The word “ritual” in Paragraph 4 is closest in meaning to

(A) military
(B) physical
(C) ceremonial
(D) permanent
16. Which one of the following sentences below best expresses the essential information in the highlighted sentence in Paragraph 5? Incorrect choices change the meaning in important ways or leave out essential information.

(A) While American iron makers developed the latest furnaces, African iron makers continued using earlier techniques.
(B) Africans produced iron much earlier than Americans, inventing technologically sophisticated heating systems.
(C) Iron making developed earlier in Africa than in the Americas because of the ready availability of carbon and iron ore.
(D) Both Africa and the Americas developed the capacity for making iron early, but African metallurgy developed at a slower rate.

17. The word “fleeing” in Paragraph 6 is closest in meaning to

(A) afraid of
(B) displaced by
(C) running away from
(D) responding to

18. Paragraph 6 mentions all of the following as possible causes of the “Bantu explosion” EXCEPT

(A) superior weapons
(B) better hunting skills
(C) peaceful migration
(D) increased population

19. In paragraph 6 of the passage, there is a missing sentence. The paragraph is repeated below and shows four letters (A, B, C, and D) that indicate where the following sentence could be added.

These people had a significant linguistic impact on the continent as well.

Where would the sentence best fit?

The spread of agriculture and later of iron was followed by a movement of people who may have carried these innovations. ■ (A) They came from eastern Nigeria probably because of peoples fleeing the drying up of the Sahara. ■ (B) They spoke a language, Proto-Bantu ("bantu" means "the people"). It was the first language among the large number of Bantu languages, spoken in sub-Saharan Africa. Why and how these people spread out into central and southern Africa is a mystery. However, research has shown that they used their iron weapons to fight and conquer people who were using stones to hunt and gather food. ■ (C) Still, the process is not clear, and peaceful migration—or simply rapid population growth—may have led to the Bantu explosion. ■ (D)
Agriculture and iron working probably spread to Africa from neighboring regions.

Answer Choices

(A) Once Africans developed their own native crops, they no longer borrowed from other regions.

(B) The harshness of the African climate meant that agriculture could not develop until after the introduction of iron tools.

(C) The use of livestock improved transportation and trade and allowed for new forms of political control.

(D) As the Sahara expanded, the camel gained in importance, eventually coming to have religious significance.

(E) The spread of iron working had far-reaching effects on social, economic, and political organization in Africa.

(F) Today’s Bantu-speaking peoples are descended from a technologically advanced people who spread throughout Africa.

The reading quiz continues on the next page.
THE HISTORY OF THE CHICKENPOX VACCINE

1 Chickenpox is a highly contagious disease. It is caused by the Varicella zoster virus. People who suffer from Chickenpox get an itchy rash that can spread all over the body. The disease can last for up to 14 days and can happen to both children and adults, but young children are more easily infected. People with chickenpox can experience a high but tolerable level of pain and a fever. Chickenpox was once considered to be a “rite of passage” by parents in the U.S. and thought to provide children with greater and improved immunity to other sicknesses later in life. However, this view changed after more research. Scientists found dangers that can be caused by the virus. Over time, the results of this research have changed attitudes toward the chickenpox and the utility of seeking ways to prevent it.

2 A vaccine against chickenpox was first created by Michiaki Takahashi, a Japanese doctor and research scientist, in the mid-1960s. Dr. Takahashi began his work to isolate and grow the virus in 1965. In 1972, he began clinical trials with a live but weak virus that caused the human body to create antibodies. Japan and several other countries began widespread chickenpox vaccination programs in 1974. However, it took over 20 years for the chickenpox vaccine to be approved by the U.S. Food and Drug Administration (FDA). The U.S. government gave approval for widespread use in 1995. Yet, even after the chickenpox vaccine was recommended by the FDA, parents did not immediately vaccinate their children against this chickenpox. They thought that it was not serious enough.

3 This attitude began to change when scientists discovered the link between the virus that causes chickenpox and shingles. Shingles is a far more serious, harmful, and longer-lasting disease in older adults that impacts the nervous system. They concluded that the chickenpox virus can stay inactive in the body and make it more likely for someone to develop shingles. As a result, doctors in the U.S. encouraged the development, adoption, and use of a vaccine against chickenpox to the public. Shingles can appear within one person many years after chickenpox—generally many decades. However, the increased risk in developing shingles as a younger adult was enough to convince U.S. doctors and medical communities that chickenpox immunization and vaccination are important.

4 Another reason that the chickenpox vaccine was not immediately accepted by parents in the U.S. was because the vaccine simply did not last long enough. In other words, scientists thought the benefits of the vaccine were temporary when it is given to young children. They also feared that it increased the chances for a person to become infected with chickenpox later as a young adult. The rash at that age can be more painful and prevalent and can last up to three or four weeks. Therefore, allowing young children to develop chickenpox rather than take a vaccine against it was believed to be the “lesser of two evils.” This idea changed over time as booster shots of the vaccine improved immunity and countered the perceived limits on the strength of the vaccine itself.

5 Today, use of the chickenpox vaccine is common throughout the world. Doctors suggest a first vaccination shot after a child turns one year old and recommend booster shots after the child turns eight. The vaccine is expected to be up to 90% effective. It has reduced worldwide cases of chickenpox infection to 400,000 cases per year from over 4,000,000 cases before vaccination became widespread. In light of such numbers, most doctors believe that the potential risks of developing shingles outweigh the benefits of avoiding rare problems associated with the vaccine. Of course, many parents still do not think the
chickenpox is very harmful and do not take steps against it. Even this trend among parents did not stop the decline of chickenpox among the most vulnerable populations. As increasing numbers of students are vaccinated and the virus has become increasingly rarer.

21. According to Paragraph 1, which of the following is true of the chickenpox virus?

(A) It leads to a potentially deadly disease in adults.
(B) It is associated with a possibly permanent rash.
(C) It is easily transmittable by an infected individual.
(D) It has been almost eradicated in the modern world.

22. Which one of the following sentences best expresses the essential information in the highlighted sentence in Paragraph 1? Incorrect answer choices change the meaning in important ways or leave out essential information.

(A) U.S. parents believed that having chickenpox benefited their children.
(B) U.S. parents believed that chickenpox led to immunity against most sickness.
(C) U.S. parents wanted to make sure that their children developed chickenpox.
(D) U.S. parents did not think that other vaccinations were needed after chickenpox.

23. Which of the following can be inferred from Paragraph 2 about the clinical trials for the chickenpox vaccine? Choose one that best describes the clinical trials as suggested by Paragraph 2.

(A) They were supported by the Japanese government.
(B) They cost a lot of money to complete.
(C) They took a long time to finish.
(D) They were ultimately successful.

24. According to Paragraph 3, which of the following is true of Varicella Zoster?

(A) It typically attacks adults who are over 60 years old.
(B) It is linked to a serious disease that occurs more commonly in adults.
(C) It likely is not a serious enough threat to human health to require a vaccine.
(D) It is completely eradicated from the body after chickenpox occurs.

25. The author uses “booster shots” as an example of

(A) a scientifically approved medicine to eliminate chickenpox
(B) a preferred method of chickenpox rash and fever treatment
(C) a way to increase the effectiveness of the chickenpox vaccine
(D) a strategy for parents to avoid vaccinating their child altogether
26. The word “countered” in Paragraph 4 is closest in meaning to

(A) affirmed
(B) refuted
(C) supported
(D) defied

27. According to Paragraph 4, many parents did not choose the chickenpox vaccine because

(A) they believed that the virus was weak and not especially harmful
(B) they thought that scientists did not have enough data to reach a conclusion
(C) they were unsure about the utility of the vaccine given its expected duration
(D) they were convinced it was potentially very toxic, particularly for older children

28. The word “prevalent” in Paragraph 4 is closest in meaning to

(A) dangerous
(B) widespread
(C) infectious
(D) contaminated

29. In Paragraph 5 of the passage, there is a missing sentence. The paragraph is repeated below and shows four letters (A, B, C, and D) that indicate where the following sentence could be added.

Meanwhile, some people continue to be unconvinced, citing a supposed potential of the vaccine to do harm.

Where would the sentence fit best?

Today, use of the chickenpox vaccine is common throughout the world. Doctors suggest an initial vaccination shot after a child turns one year old and recommend booster shots after the child turns eight. The vaccine is expected to be up to 90% effective. It has reduced worldwide cases of chickenpox infection to 400,000 cases per year from over 4,000,000 cases before vaccination became widespread. ■ (A) In light of such numbers, most doctors believe that the potential risks of developing shingles outweigh the benefits of avoiding rare problems associated with the vaccine. ■ (B) Of course, many parents still do not take steps against it. ■ (C) As increasing numbers of students are vaccinated and the virus has become increasingly rarer, however, even this trend among parents did not stop the decline of chickenpox among the most vulnerable populations. ■ (D)

(A) Option A
(B) Option B
(C) Option C
(D) Option D
30. Directions: The first sentence of a short summary of the passage is given below. Complete the summary by selecting the THREE answer choices that express the most important ideas in the passage. Some sentences do not belong in the summary because they express ideas that are not in the passage or are minor ideas in the passage.

Write your answer choices in the spaces where they belong. You can either write the letter of your answer choice or you can copy the sentence.

The notion of vaccinating children against chickenpox, an irksome but relatively harmless ailment, has been controversial.

- 
- 
- 

Answer choices

(A) People used to seek preemptive measures against chickenpox because they believed that every child who experiences chickenpox develops stronger immunity.

(B) After scientists discovered that chickenpox can lead to shingles, a more serious and lingering illness, medical communities began to prefer immunization to the traditional alternative.

(C) Scientists found the link between chickenpox and shingles and insisted vaccination against shingles because shingles afflicts adults by affecting their nervous system.

(D) Because observations were made by scientists that the chickenpox vaccine may not confer a lifetime of immunity and could cause an occurrence later in life, parents did not accept the idea of the chickenpox vaccination right away.

(E) With the discovery of booster shots that elongate immunity, vaccination against chickenpox has become widespread resulting in a significant decrease in cases of chickenpox infection.

(F) Infection primarily occurs as a result of close contact with infected itchy rashes which cause fever and intolerable pain.

The reading quiz continues on the next page.
<table>
<thead>
<tr>
<th>The Roman world had a quality of cohesiveness that neither Greece nor any other civilization had in ancient or modern times. Like the stones of a Roman wall, which were held together by cohesive design and powerful Roman cement, many parts of the Roman realm were bonded into a massive entity by physical, organizational, and psychological controls. The physical bonds included the network of armies and the stone-built roads that connected the provinces with Rome. The organizational bonds were based on the common ideas of law and administration; an army of officials enforced common standards of conduct. The psychological controls were built on fear and punishment. If anyone or anything threatened the authority of Rome, it would be completely destroyed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Roman togetherness came from the pattern of Rome’s early growth. While Greece had grown from many cities that were spread out, Rome grew from one single place. The Greek world had expanded along the Mediterranean Sea lanes, but the Roman world was put together by the gaining territory. Of course, the difference is not so large: in Alexander the Great, the Greeks had found the greatest territorial conqueror of all time. The Romans learned the importance of sea power when they left Italy. But there is one main difference: The key to the Greek world was in its powerful ships, but to Romans, it was their armies. The Greeks used the sea, but the Romans used the land. The Greek was a sailor at heart; the Roman, a landsman.</td>
</tr>
<tr>
<td>To explain the Roman world, one would have to understand the importance of gaining land. Romans valued the organization, use, and defense of their land. It is likely that it was the fertile plain of Latium, where the Latins who started Rome are from, that developed the skills important to Romans. These skills were landed settlement, landed property, landed economy, landed administration, and a land-based society. From this came the Roman skill for military organization and orderly government. In turn, a deep attachment to the land and the stable rural life fostered the Roman values. These were gravitas: a sense of responsibility, pietas: a sense of dedication to family and country, and iustitia: a sense of the natural order.</td>
</tr>
<tr>
<td>People show different attitudes toward Roman civilization. Some are impressed by Roman civilization, and others are not. As always, there are people who worship power, especially among historians, who always admire power. They are attracted to and prefer the might of Rome to the subtlety of Greece. At the same time, there is a group that dislikes Rome. Many argue that Rome is at best the imitator and the continuator of Greece on a larger scale. Greek civilization had quality; Rome had mere quantity. Greece was original; Rome tried to copy. Greece had style; Rome had money. Greece was the inventor; Rome was the research and development division. This was the opinion of some of the more intellectual Romans. “Had the Greeks held novelty in such disdain as we,” asked Horace in his Epistles, “what work of ancient date would now exist?”</td>
</tr>
<tr>
<td>Greece had a huge influence on Rome. The Romans took Greek religion and moral philosophy. In literature, Greek writers were used as models. The Romans thought that an educated Roman should be fluent in Greek. In philosophy and the sciences, the Romans made almost no advance on early achievements.</td>
</tr>
<tr>
<td>But it would be wrong to suggest that Rome was only a junior partner in Greco-Roman civilization. The Roman genius spread into new spheres—especially into those of law, military organization, administration, and engineering. Moreover, the tensions within the Roman state produced great literary and artistic sensibilities. It was not surprising that many leading Roman soldiers and statesmen were great writers.</td>
</tr>
</tbody>
</table>
31. Which one of the following sentences below best expresses the important information in the highlighted sentence in Paragraph 1? Incorrect choices change the meaning in important ways or leave out important information.

(A) The regularity and power of stone walls inspired Romans attempting to unify the parts of their realm.
(B) Although the Romans used different types of designs when building their walls, they used regular controls to maintain their realm.
(C) Several types of control united the Roman realm, just as design and cement held Roman walls together.
(D) Romans built walls to unite the various parts of their realm into a single entity, which was controlled by powerful laws.

32. According to Paragraph 1, all of the following are controls that held together the Roman world EXCEPT

(A) administrative and legal systems
(B) the presence of the military
(C) a common language
(D) transportation networks

33. According to Paragraph 2, which of the following was NOT characteristic of Rome’s early development?

(A) Expansion by sea invasion
(B) Territorial expansion
(C) Expansion from one original settlement
(D) Expansion through invading armies

34. Why does the author mention “Alexander the Great” in the passage?

(A) To acknowledge that Greek civilization also expanded by land conquest
(B) To compare Greek leaders to Roman leaders
(C) To give an example of a Greek leader whom Romans studied
(D) To indicate the superior organization of the Greek military

35. The word “fostered” in Paragraph 3 is closest in meaning to

(A) accepted
(B) combined
(C) introduced
(D) encouraged
36. Paragraph 3 suggests which of the following about the people of Latium?

(A) Their economy was based on trade relations with other settlements.
(B) They held different values than the people of Rome.
(C) Agriculture played a significant role in their society.
(D) They possessed unusual knowledge of animal instincts.

37. The word “spheres” in Paragraph 6 is closest in meaning to

(A) abilities
(B) areas
(C) combinations
(D) models

38. Which of the following statements about leading Roman soldiers is supported by Paragraphs 5 and 6?

(A) They could read and write the Greek language.
(B) They frequently wrote poetry and plays.
(C) They focused their writing on military matters.
(D) They wrote according to the philosophical laws of the Greeks.

39. In Paragraph 4 of the passage, there is a missing sentence. The paragraph is repeated below and shows four letters (A, B, C, and D) that indicate where the following sentence could be added.

They esteem symbols of Roman power, such as the massive Colosseum.

Where would the sentence best fit?

People show different attitudes toward Roman civilization. Some are impressed by Roman civilization, and others are not. ■ (A) As always, there are people who worship power, especially among historians, who always admire power. They prefer the might of Rome than to the subtlety of Greece. ■ (B) At the same time, there is a group that dislikes Rome. ■ (C) For many, Rome is just the copier and the continuator of Greece on a larger scale. ■ (D) Greek civilization had quality; Rome had mere quantity. Greece was original; Rome tried to copy. Greece had style; Rome had money. Greece was the inventor; Rome was the research and development division. This was the opinion of some of the more intellectual Romans. “Had the Greeks held novelty in such disdain as we,” asked Horace in his Epistles, “what work of ancient date would now exist?”

(A) Option A
(B) Option B
(C) Option C
(D) Option D
40. **Directions**: The first sentence of a short summary of the passage is given below. Complete the summary by selecting the THREE answer choices that express the most important ideas in the passage. Some sentences do not belong in the summary because they express ideas that are not in the passage or are minor ideas in the passage.

Write your answer choices in the spaces where they belong. You can either write the letter of your answer choice or you can copy the sentence.

<table>
<thead>
<tr>
<th>The Roman world drew its strength from many important sources.</th>
</tr>
</thead>
<tbody>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
<tr>
<td>•</td>
</tr>
</tbody>
</table>

**Answer Choices**

(A) Numerous controls imposed by Roman rulers held its territory together.

(B) The Roman military was organized differently from older military organizations.

(C) Romans valued sea power as did the Latins, the original inhabitants of Rome.

(D) Roman values were rooted in a strong attachment to the land and the stability of rural life.

(E) Educated Romans modeled their own literature and philosophy on the ancient Greeks.

(F) Rome combined aspects of ancient Greek civilization with its own contributions in new areas.

*This is the end of the reading quiz. Thank you for your participation.*
Appendix T: Answer Key

C-test answers

Text 1
The most interesting result of the studies has been that, with few exceptions, nearly every human group examined has been found to consist of a mixture of the same four blood groups; human races differ in the relative numbers of persons within them who fall into each of the four groups.

Text 2
Indignant citizens asserted their rights, and complained that their neighbor had broken his obligation, and then spoke of taking him to court. But it all excluded the most important matters between the king’s most powerful subjects, the court to which they referred was a local court, and the rights which they claimed were those recognized by the custom of the neighborhood.

Text 3
Since my interest in the psychological aspects of communication is even older than automatic computers I can remember what those days before computers were like. When I try to compare them with the present I can think of no summary statement more appropriate than that made by a famous American athlete who said “I’ve been rich, and I’ve been poor, and believe me, rich is better.” Believe me, computers are better.

Text 4
They do not feel threatened by a different kind of discipline or tempted to over-stress their own subject’s special mystique/mysteries. The high degree of imagination necessary for distinguishing work in humanities or social science ensures that men with these powers do not mistake the technical boundaries between academic disciplines for division within human experience.

Background test comprehension check answers

<table>
<thead>
<tr>
<th>The Bantu people</th>
<th>Teotihuacán</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) People</td>
<td>(1) 40 kilometers northeast of Mexico City</td>
</tr>
<tr>
<td>(2) To find fertile land</td>
<td>(2) Before Christ</td>
</tr>
<tr>
<td>(3) The Chwezi</td>
<td>(3) The Aztecs called the street “Avenue of the Dead” because they believed that the platforms there were tombs.</td>
</tr>
<tr>
<td>(4) Because they found a place where they can grow bananas, their major food source.</td>
<td>(4) Feathered serpents were considered sacred.</td>
</tr>
<tr>
<td>(5) Because there was a use of bellows with dish-shaped blast</td>
<td>(5) The size/width of the irrigation canal that crosses the Avenue of the Dead.</td>
</tr>
</tbody>
</table>
# Reading comprehension test answers

<table>
<thead>
<tr>
<th>Practice</th>
<th>Testlet 1 (Teotihuacán)</th>
<th>Testlet 2 (Agriculture, Iron, and the Bantu people)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) B</td>
<td>(1) B</td>
<td>(11) D</td>
</tr>
<tr>
<td>(2) A</td>
<td>(2) A</td>
<td>(12) D</td>
</tr>
<tr>
<td>(3) B</td>
<td>(3) C</td>
<td>(13) A</td>
</tr>
<tr>
<td>(4) C</td>
<td>(4) C</td>
<td>(14) B</td>
</tr>
<tr>
<td>(5) A</td>
<td>(5) D</td>
<td>(15) C</td>
</tr>
<tr>
<td>(6) A</td>
<td>(6) A</td>
<td>(16) B</td>
</tr>
<tr>
<td>(7) A, C, D</td>
<td>(7) D</td>
<td>(17) C</td>
</tr>
<tr>
<td></td>
<td>(8) C</td>
<td>(18) B</td>
</tr>
<tr>
<td></td>
<td>(9) D</td>
<td>(19) B</td>
</tr>
<tr>
<td></td>
<td>(10) A, C, E</td>
<td>(20) C, E, F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Testlet 3 (History of the chickenpox vaccine)</th>
<th>Testlet 4 (Ancient Rome and Greece)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(21) C</td>
<td>(31) C</td>
</tr>
<tr>
<td>(22) A</td>
<td>(32) C</td>
</tr>
<tr>
<td>(23) D</td>
<td>(33) A</td>
</tr>
<tr>
<td>(24) B</td>
<td>(34) A</td>
</tr>
<tr>
<td>(25) C</td>
<td>(35) D</td>
</tr>
<tr>
<td>(26) B</td>
<td>(36) C</td>
</tr>
<tr>
<td>(27) C</td>
<td>(37) B</td>
</tr>
<tr>
<td>(28) B</td>
<td>(38) A</td>
</tr>
<tr>
<td>(29) B</td>
<td>(39) B</td>
</tr>
<tr>
<td>(30) B, D, E</td>
<td>(40) A, D, F</td>
</tr>
</tbody>
</table>