Despite the known roles of working memory and background knowledge in the process of reading comprehension, few attempts have been made to elucidate the interaction between these two individual difference factors in second language (L2) reading comprehension. The current study investigated whether and how working memory and background knowledge combine to facilitate L2 reading comprehension in a context where L2 students encounter texts with and without relevant background knowledge. Seventy-nine adult Korean learners of English as a foreign language participated and completed tasks for working memory, L2 knowledge, and L2 reading comprehension. The results revealed that L2 readers with higher working memory capacity benefitted more from the provision of background knowledge, which led to achieving better reading comprehension than readers with low working memory. This finding highlights the role of working memory in L2 reading in terms of using existing resources to one’s advantage particularly because no significant difference on L2 measure scores was found between the high– and low–working memory groups. Pedagogical implications are discussed regarding the importance of not only providing background knowledge when it is not already present but also following up with explicit instructional support to help all readers utilize what is available to them.

doi: 10.1002/tesq.482

Reading comprehension is a multifaceted process that requires intricate interaction among multiple skills and factors. Background knowledge is a well-known factor that influences
reading comprehension as readers constantly go through an iterative process that involves the melding of new information to the existing knowledge store (McNeil, 2011). Utilizing background knowledge, readers work to connect the literal message from the text currently active in working memory with incoming information to successfully construct a mental representation of the text (Alptekin & Erçetin, 2010; Kintsch, 1988). In this sense, working memory is what constantly binds local and global understanding of a text. In fact, research has shown that working memory is a well-established predictor of reading comprehension performance (Carretti, Borella, Cornoldi, & De Beni, 2009; Daneman & Carpenter, 1980; Kintsch & van Dijk, 1978; Linck, Osthus, Koeth, & Bunting, 2014; Nouwens, Groen, & Verhoeven, 2016).

In building a mental model of a text, readers bring to the table a variety of cognitive capacities. The two constructs that are the focus of this study, working memory and background knowledge, are not equally malleable. Working memory is a relatively stable individual characteristic (Hambrick & Engle, 2002; Melby-Lervåg, Redick, & Hulme, 2016; Redick et al. 2013), whereas background knowledge is easily modifiable. Recently, there has been growth in research on the influence of individual differences in working memory capacity on reading performance (Joh, 2015). Research on the facilitative role of background knowledge on reading comprehension has a much longer history that dates to the late 1970s, when schema theory was introduced (Rumelhart, 1980). Background knowledge was also the center of the knowledge-is-power hypothesis as a prominent predictor of success in cognitive tasks (Hambrick & Engle, 2002).

Despite the significant involvement of working memory and background knowledge in the process of comprehending a text, few attempts have been made to elucidate the interplay of these two individual difference factors in L2 reading comprehension. This study set out to investigate whether and how the two combine to facilitate L2 reading comprehension with the goal of providing a better understanding of the nature of their interaction in the complex reading comprehension process.

Reading Comprehension Processes

Current theories of reading comprehension agree that reading involves interaction between the reader’s text-based and knowledge-based processes. The underlying assumption is that readers interact with the text to varying degrees (Alptekin & Erçetin, 2011). One of
the most comprehensive models that accommodates this view of reading is the construction-integration (CI) model (Kintsch, 1988), which addresses multilevel representations of a text: the surface code, the textbase, and the situation model. The CI model posits that readers first interact with textual features at the surface level (i.e., surface code) to capture the internal meaning of the text and attempt to build a semantic whole (i.e., textbase) by integrating propositions encoded in the sentences and incorporating local bridging. After establishing text-level understanding, readers extend and refine the propositional message to integrate it into a situational representation of the text (i.e., situation model). At this stage, they engage in global bridging and connect ideas in the text to generate a coherent whole (Graesser, 2007; van Dijk & Kintsch, 1983). This process takes place iteratively; mental representations of the text are constantly updated, which places considerable demands on working memory capacity (Ellemann, 2017). Because the current study is among the first to set out to untangle the complex relationship among working memory, background knowledge, and reading comprehension in a second language, it tests reading comprehension as a single construct. Our instruments aimed at the assessment of reading comprehension do not disregard the fact that comprehension occurs both at a literal and inferential level and, therefore, feature items tapping both, but they were not designed specifically with the aim of contrasting the two. Our analysis, therefore, treats reading comprehension as a unified construct reflected by a single score.

Working Memory and Reading

The comprehension process illustrated in the CI model implies that working memory plays an important role in reading comprehension. Working memory, often operationalized as a score on complex span tasks such as the reading span task (Daneman & Carpenter, 1980), is a limited-capacity information processing system that allows for the active maintenance of information while carrying out additional processing operations. It involves a system of temporary storage, the simultaneous processing of incoming information, attentional prioritizing and the inhibition of irrelevant information, and the retrieval of known information, often involving strategic searches of declarative long-term memory (Baddeley, 2000; Just & Carpenter, 1992).

The integrative processes of reading comprehension are particularly constrained by working memory because it stores necessary ideas in a text while keeping them available for integration (Daneman &
Carpenter, 1980; Gathercole & Baddeley, 1993). To illustrate, the surface level of understanding a text involves local bridging, which requires connecting information currently active in working memory with incoming information from the text being read (Kintsch, 1988). The integrative processes also rely on processing strategies such as using attention and being selective of information that needs to be held as opposed to filling the storage components of working memory with distracting and irrelevant information (Cowan, 2010). Thus, it is not maintaining information itself, but continuously updating while holding relevant information in storage that enables readers to complete the integrative processes of reading (Palladino, Cornoldi, de Beni, & Pazzaglia, 2001). Alptekin and Erçetin (2011) added that readers need to make use of logical and pragmatic means across sentences or even paragraphs to connect implicit and explicit information, which requires cognitive resources such as working memory and declarative long-term memory. Also, the inferential level of comprehension that leads to constructing a situation model requires readers to “read between the lines” as they integrate their prior knowledge with the information in the text (Alptekin, 2006; Elleman, 2017). In other words, expanding on the textual content, maintained by working memory, and drawing on relevant background knowledge become necessary especially when generating inferences to complete a coherent mental representation of the text (Daneman & Carpenter, 1980; Perfetti, 1989).

Daneman and Carpenter’s (1980) seminal work on working memory and reading has prompted numerous first language (L1) studies that showed that working memory is an important predictor of reading performance (e.g., Just & Carpenter, 1992; Waters & Caplan, 1996), paving the way for L2 research on the relationship between working memory and L2 reading. The consensus is that they are positively correlated. For instance, in a study conducted with Japanese-English bilinguals, Harrington and Sawyer (1992) reported a moderate correlation between working memory capacity measured with a reading span task (RST) and the TOEFL reading section ($r = .57$). Walter (2004) examined French speakers’ L2 English reading span and reading comprehension. The correlation found in her study was positive and strong ($r = .73$). Similarly, Leeser (2007) measured working memory and comprehension recall of L2 Spanish learners and showed that high-working memory readers outperformed their counterparts in the recall task. More recently, Alptekin and Erçetin (2009, 2010, 2011) have explored this domain in their series of studies and showed a positive correlation between working memory and inferential comprehension ($r = .40$ and .45, respectively) as well as with reading comprehension in general ($r = .31$).
The Reading Span Task (RST) as a Measure of Working Memory Capacity

One of the most common measures for assessing working memory capacity is the RST. This is an integrated task based on a dual-task paradigm designed to tax both working memory functions, processing and storage. In the original RST developed by Daneman and Carpenter (1980), the processing function is presumably taxed by the first part of the task in which participants read individual sentences aloud. At the end of each set (and the sets become progressively longer), participants must recall the final word of each sentence in the current set, which measures the storage function of working memory (i.e., phonological short-term memory). Daneman and Carpenter’s RST became a standard way of measuring verbal working memory because it involves both working memory components, unlike simple span tasks (Waters & Caplan, 1996).

However, this RST has raised concerns with regard to its appropriateness for measuring the processing component for the following reasons. The scores generated do not reflect participants’ efficiency in carrying out the task; also, it is difficult to observe whether there is a trade-off between the components just by recording the number of final words recalled. This, in turn, makes the scores difficult to interpret because they only reflect the storage function of working memory (Leeser, 2007; Turner & Engle, 1989; Waters & Caplan, 1996). To address these issues, Waters and Caplan (1996) incorporated a sentence acceptability task and recorded the reaction times to tap the processing component of working memory explicitly. This modification of the classic RST allows for performance on both processing and storage to be represented in the scores (Walter, 2004).

Background Knowledge and Reading

Because comprehension is viewed as a model-building process, extratextual information is needed to successfully generate inferences and ultimately form a mental representation of a text. As such, use of background knowledge of the topic at hand has been viewed as a robust driving force of constructing a mental model of a text (Ellemman, 2017). The facilitative role of background knowledge has been highlighted in the schema-oriented views. A schema refers to a mental structure in memory abstracted from past experiences that sets up expectations for an upcoming event. This notion of a schema was first introduced in psychology by Bartlett (1932). Rumelhart (1980), among
others, later adapted this concept to show the positive effect of background knowledge on reading with an emphasis on readers’ text-based knowledge and conceptually driven background knowledge functioning in concert to achieve maximal comprehension.

Bransford and Johnson (1972) lent support for this view by demonstrating how activating background knowledge with simply a title of a text can allow readers to easily connect the dots within a text. Acknowledging the significance of background knowledge in enriching a mental model and highlighting the pedagogical implications of activating background knowledge, Sedivy (2014) asserts that, even if readers can decode the linguistic content of a text, comprehension can suffer without background knowledge. Kintsch and Kintsch (2005) claimed that efficient activation of background knowledge can aid in forming inferences. Nassaji (2002) also showed that L2 readers’ ability to use socioculturally appropriate background knowledge improves reading comprehension.

Based on the examination of the roles of working memory and background knowledge thus far, it is evident that reading comprehension, given its interactive and integrative nature, requires intricate interaction between them to successfully build a mental model of a text. A question then arises as to the precise nature of the interplay between working memory and background knowledge.

The Joint Effects of Working Memory and Background Knowledge

Attempts have been underway to elucidate the contribution of working memory to reading, one of them being the investigation of moderating variables that may be involved (Joh, 2015). As a case in point, joint effects of working memory and background knowledge in reading have been explored by Alptekin and Erçetin (2011), Leeser (2007), Payne, Kalibatseva, and Jungers (2009), and Joh and Plakans (2017), which are discussed further below.

Hambrick and Engle (2002) proposed three models that could describe the relationship between working memory and background knowledge in reading. First, the compensation model proposes that high levels of relevant background knowledge can compensate for low levels of working memory. 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 working memory capacity. Second, the independent-influences model predicts that working memory 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 working memory capacity; that is, people with high working memory capacity tend to benefit from background knowledge to a greater extent.

Few attempts have been made to simultaneously examine the effects of working memory and background knowledge on reading comprehension. Nonetheless, valuable information can be gained by reviewing what has been achieved thus far. In Alptekin and Erçetin’s study (2011), Turkish participants were presented with a narrative text either in its original U.S. version or nativized version that provided cultural familiarity. No interaction was found between working memory and background knowledge. Similarly, Payne and colleagues (2009) showed that working memory and domain experience, operationalized as years of active Spanish study, significantly contributed to L2 comprehension. However, no evidence for an interaction was found, supporting the independent-influences model along with Alptekin and Erçetin.

By contrast, Leeser (2007) showed an interaction between working memory and background knowledge and supported the rich-get-richer model. In his study, L2 Spanish speakers with high working memory outperformed a low–working memory group on a reading comprehension task only if they were familiar with the topic of the texts. Leeser concluded that high working memory capacity aids in forming connections between ideas expressed in the text and prior knowledge. Joh and Plakans (2017) recently reported that working memory was a significant predictor of L2 reading comprehension of Korean EFL learners when topic knowledge was provided. However, for the low–topic knowledge group, L2 knowledge was the only predictor of reading comprehension. Their results indicate that topic knowledge may determine the degree of working memory’s contribution to L2 reading, thus providing evidence for an interaction between the two.

The lack of uniformity in methods used in existing studies may have contributed to the inconclusiveness of the findings. For instance, the operationalization of background knowledge has varied from domain experience (Payne et al., 2009) to cultural familiarity (Alptekin & Erçetin, 2011). Background knowledge assessment methods include self-ratings of topic familiarity (Leeser, 2007) and on language experience (Payne et al., 2009). In Alptekin and Erçetin (2011), the participants’ familiarity with the texts was raised by replacing unfamiliar words with culturally familiar ones. Additionally, different working memory measures were used, such as a counting span task (Payne et al., 2009) and an RST (Alptekin & Erçetin, 2011; Leeser, 2007). Even within the RST design, the scoring systems differed; in addition to taking into account the scores of sentence judgments and word
recall, Leeser (2007) included reaction times for correct responses on a judgment task in calculating composite scores, addressing Waters and Caplan’s (1996) critical comment on efficiency being excluded in examining the processing function of working memory.

Although the above studies contribute to the understanding of the two individual-difference factors and their links with L2 reading, little is known about whether and how background knowledge, operationalized as topic familiarity, interacts with working memory. Also, past studies have not controlled for the degree of background knowledge that participants had prior to the reading test. Thus, the present study used a more controlled approach by clearly differentiating the conditions under which participants’ reading performance is assessed, that is, reading with and without background knowledge provided by the researchers. The study also aimed to mimic a plausible scenario where readers with differing working memory capacity and background knowledge encounter difficult L2 texts.

We investigated the interaction of working memory and background knowledge and evaluated the three models (compensation, independent-influences, and rich-get-richer models). The following research questions were addressed:

1. Is there an interaction between working memory and background knowledge in L2 reading comprehension?
2. If so, which of the models best explains the interaction?

METHOD

Participants

A total of 79 Korean EFL learners (51 females, 28 males) enrolled in local universities in South Korea participated. Their ages ranged from 18 to 26 years old ($M = 22.28$), and their academic backgrounds were varied. The participants’ reported scores on standardized English tests along with their academic standings indicated that their proficiency ranged from intermediate ($n = 75$) to advanced ($n = 4$).

Materials

Background and familiarity questionnaire. A background questionnaire was administered to obtain the participants’ demographic information. Their self-ratings of familiarity with four topics featured in the reading comprehension test were also collected. We had selected TOEFL reading passages on four obscure topics using our best
judgment and utilized the questionnaire responses as a tool to confirm our prediction. Participants were given the titles of the texts and were asked to mark their familiarity with the topics on a scale of 1 (not familiar at all) to 5 (very familiar).

**Working memory.** An RST was designed in a similar format as Waters and Caplan (1996) and Leeser (2007) in a computerized version using Paradigm by Perception Research Systems Inc. (Tagliaferri, 2005). The issue regarding the language of a working memory task is often controversial. However, Alptekin and collaborators (2011, 2014) have advised that future research on working memory in L2 reading use an L2 working memory task based on the findings that L1 and L2 working memory are positively correlated but working memory measured in the L2 is more directly related to L2 reading comprehension (Alptekin & Erçetin, 2010; Miyake & Friedman, 1998). In support of this, a recent meta-analysis on the relationship between working memory and L2 reading (Shin, under review) found that a larger effect size is produced when working memory is measured in the L2 ($r = .35$) than L1 ($r = .18$). Taking into consideration that previous similar studies also measured L2 working memory of intermediate learners (Alptekin & Erçetin, 2010; Joh & Plakans, 2017), our intention was to provide empirical findings that are comparable across studies. Therefore, we designed our RST using the L2 as the task language as well.

A total of 70 sentences were included and were divided into two sets of two to eight levels. The sentences were taken from graded readers (Barrell, 2000; Kershaw, 1999), all of which indicated the same level, lower intermediate. This is equivalent to the Common European Framework of Reference for Languages (CEFR) Level B1. Care was taken to avoid a confound between lexical proficiency and working memory performance as much as possible; the selected sentences included words from the 1k to 2k frequency band, which is far lower than the average vocabulary size of the participants that was later calculated (8k families). Each sentence, 10–12 words in length, ended with a different final word. Three practice trials were provided to help participants familiarize themselves with the task procedure.

Participants read a sentence silently and decided if it was semantically acceptable immediately after reading by clicking on either a *yes* or *no* button. Out of the total pool of 70 sentences, half were acceptable and the other half were unacceptable. This judgment task was intended to ensure that the participants were not solely focusing on memorizing the last words but were processing the information in the sentences. Reaction times for answering the semantic judgment questions were recorded to take processing speed into account (Waters & Caplan, 1996). By targeting the simultaneous processing of sentences
presented and the storage of final words, this study adopted a dual-task assessment procedure to account for the postulation of a trade-off between the two components of working memory, storage and processing (Wen, 2016).

After the judgment task, the next sentence in the set appeared. When each set was finished, participants were prompted by the phrase “STOP and REPORT” on the screen to report the final words of all the sentences in the set by writing them on a sheet provided. As outlined in Dronjic (2013), the participants were instructed to write the words in any order except that they could not begin with the final word of the last sentence. They were notified that a score of 0 would be assigned if the final word of the last sentence was reported first. One point was awarded for each correct word and 0 for each incorrect one. Minor spelling errors were ignored.

The RST composite scores were based on mean reaction times in milliseconds for correct judgments, the number of correct judgments, and the number of correct final words recalled. They were transformed into z-scores, and the reaction times were multiplied by −1 because shorter reaction times meant better performance on processing information. The z-scores of the three tasks were then averaged to represent each participant’s reading span score (Leeser, 2007). Internal consistency was examined using Cronbach’s alpha. Both storage and processing tasks were found to be highly reliable (.92 and .88, respectively).

**Language measures.** Two measures indexing L2 knowledge included the Vocabulary Size Test (VST) (Coxhead, Nation, & Sim, 2014; Nation & Beglar, 2007) where 100 word families are represented by each item (http://my.vocabularysize.com) and a C-test (Klein-Braley, 1997). VST scores have often been used as a proxy for L2 proficiency levels (Nation, 2006) and have been shown to be highly correlated with general proficiency test scores (Alderson, 2005).

To supplement the VST result, a C-test was administered. In addition to its appealing features regarding administration time (less than 30 minutes) and objective, easy scoring, C-tests exhibit high reliability and a high correlation with general proficiency tests (Eckes & Grotjahn, 2006). We adopted Klein-Braley’s (1997) C-test, which included four texts with 25 blanks per text to be filled in (see the Appendix). The deletion technique used in the test as described in Klein-Braley’s article is as follows: The first sentence remains as is, and the deletion starts with the second word in the second sentence. The second half of every second word was deleted. When there was a word with an odd number of letters, the larger part was deleted. One point was assigned to each correct response and zero to each incorrect response.
Background texts in L1. The day before taking a reading comprehension test, participants received short texts on Teotihuacán and the Bantu people, which are two of the four topics of passages featured in the test (see Table 1). They were written in the participants’ L1, Korean, to prevent exposure to any related L2 vocabulary. The texts were adapted from two Korean passages found in the Online Encyclopedia of Cultural Heritage and World History (terms.naver.com). Care was taken to ensure that these texts introduced the topics without extensive overlap with the content of the L2 passages on the test. The background texts were similar in length (290 and 292 words, respectively). The participants were asked to read the background texts twice carefully on their own. They were informed that the texts were intended to provide them with background knowledge of the topics and that the information in the texts would assist in their comprehension of the English texts that would be on the reading test.

Reading comprehension test. Four passages featuring four topics were adapted from TOEFL practice tests (see Table 1). The grade levels of the original passages were at Grades 12–13. Upon consulting the participants’ instructor, the texts were modified to lower their grade levels to 9–10. The modified texts were comparable in terms of readability, length, number of academic words, and lexical density. Each passage was accompanied by 10 comprehension questions in a multiple-choice format. The types of items were consistent throughout the four passages, consisting of five literal and five inferential comprehension questions. For questions 1–9, one point was given for each correct answer and zero for each incorrect one. Question 10 (summary completion) was worth two points, and partial scores were allowed. We prepared two types of test packet where the order of passages was counterbalanced. In Packet A, the passages on Teotihuacán and the Bantu People appeared first. In Packet B, they appeared after the other two unfamiliar passages. Approximately half of the participants received Packet A, and approximately half received Packet B.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Grade level</th>
<th>Length (words)</th>
<th>Academic words</th>
<th>Lexical density</th>
</tr>
</thead>
<tbody>
<tr>
<td>The rise of Teotihuacán</td>
<td>10</td>
<td>652</td>
<td>34</td>
<td>0.57</td>
</tr>
<tr>
<td>Agriculture, iron, and the Bantu people</td>
<td>9</td>
<td>699</td>
<td>29</td>
<td>0.57</td>
</tr>
<tr>
<td>History of the chickenpox vaccine</td>
<td>10</td>
<td>676</td>
<td>28</td>
<td>0.57</td>
</tr>
<tr>
<td>Ancient Rome and Greece</td>
<td>9</td>
<td>643</td>
<td>30</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Participants were instructed to read the passages in the order they were presented. The test was found to be highly reliable ($\alpha = .85$).

**Postreading topic familiarity questionnaire.** After the reading test, we asked the participants to rate their familiarity with the four topics again to confirm that they performed on the test with some background knowledge on two topics and with little prior knowledge on the other two.

**Procedures**

The study was administered in two phases. Phase 1 included four tasks: the background and familiarity questionnaire, the RST, the VST, and the C-test. Short breaks were allowed, and it took 15–20 minutes to complete each of the tasks except for the questionnaire, which took 2–3 minutes. In total, Phase 1 took 50–60 minutes per participant. Phase 2 included participants reading two background texts on their own and taking the modified TOEFL reading test the following day. The test took approximately one hour to complete.

**RESULTS**

Due to missing data points, two out of the original 79 participants were excluded from analysis. Of the 77 remaining participants, three were subsequently removed from analysis due to extreme working memory scores (see below). The results reported here are for the remaining sample of 74 participants.

**Topic Familiarity**

Several participants indicated some prior knowledge of the history of the chickenpox vaccine and ancient Rome and Greece on the questionnaire. After the reading comprehension test, however, almost all of them reported that they were in fact unfamiliar with the content of the passages on the history of the chickenpox vaccine ($M = 1.16, SD = 0.41$) and ancient Rome and Greece ($M = 1.09, SD = 0.29$) but somewhat familiar with the Bantu explosion ($M = 2.41, SD = 0.49$) and Teotihuacán ($M = 2.32, SD = 0.47$) due to the background passages they had read the day before. It should be noted that the purpose of including the topic familiarity intervention was to create two conditions—reading with and without background knowledge—by providing
a sense of familiarity without extensively exposing the participants to the content of the actual L2 reading texts. Our further analysis confirmed that the difference between the two conditions was substantial \((t(74) = 21.18, p < .001)\) with a very large effect size \((d = 3.40)\).

**Working Memory**

Composite working memory scores were calculated as described above for the initial sample of 77 participants. Represented in \(z\)-scores, they ranged from \(-2.01\) to \(1.02\), with a mean of 0, a standard deviation of 0.60, and a range of 3.04. We detected two outliers, defined as values between 1.5 and 3 interquartile ranges below the 25th percentile or above the 75th percentile, and one extreme value, defined as more than 3 interquartile ranges below the 25th percentile or above the 75th percentile. One of the outliers was also the only participant to score 0 on any of the reading measures. These three participants were removed from further analysis, and the results below apply to the remaining 74 participants.

The composite working memory scores for the final sample ranged from \(-1.25\) to \(1.18\), with a mean of 0, a standard deviation of 0.56, and a range of 2.43. The scores were normally distributed \((D(74) = .075, p = .20)\).

**Grouping Participants by Working Memory Scores**

Following comparable L2 studies (Alptekin & Erçetin, 2011; Leeser, 2007), we categorized participants into low- and high–working memory groups based on working memory scores. We justify our decision to do so below. A \(k\)-means cluster analysis of the working memory scores revealed two or three clusters of uneven sizes. Grouping participants according to these clusters would have jeopardized the reliability of the analysis of variance (ANOVA) on reading scores due to uneven numbers of participants in the individual cells of the mixed factorial design. We therefore decided to pursue the option of dividing the sample into two or three groups of equal or approximately equal sizes.

In a series of analyses, the reading comprehension data were examined on whether they would meet the assumptions of a mixed ANOVA, with working memory as a between-participant factor (two or three levels) and background knowledge as the within-participant factor. The data were examined for departures from normality in their raw form, in square-root-transformed form, in log-transformed form, and as logit-transformed proportions both under a \(2 \times 2\) and a \(3 \times 2\)
factorial design. These analyses revealed that square-root-transformed data in a $2 \times 2$ factorial design departed from ANOVA assumptions the least: The Shapiro-Wilk test indicated no departures from normality across the cells, and the Kolmogorov-Smirnov test was significant only for participants with high working memory in the condition in which background knowledge had been provided ($D(37) = .16$, $p = .018$), indicating only a mild departure from normality. In the $2 \times 2$ design, the error variances were equal with Levene’s tests not reaching significance.

We therefore opted to divide participants into two groups according to their working memory scores, and they differed significantly on these scores ($t(72) = 10.70$, $p < .001$; see Table 2). It is important to note that our participants’ working memory scores and overall reading comprehension scores (combining the background and no-background conditions) were not correlated ($r = .21$, $p = .07$). Under these conditions, dichotomization of a continuous variable has been shown to have no negative effects on the reliability of the conclusions drawn from a factorial ANOVA model (Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015).

### Language Measures

Table 3 shows the two groups’ scores on the two measures of English proficiency, the VST and the C-test. There were no significant between-group differences on either the VST ($t(72) = .08$, $p = .41$) or the C-test ($t(72) = 1.95$, $p = .06$). Although the groups were not significantly different on C-test scores, the fact that this difference

### Table 2

<table>
<thead>
<tr>
<th>Working Memory Scores in the High– and Low–Working Memory (WM) Groups</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>High WM</td>
<td>-.43</td>
<td>.39–1.25</td>
<td>.06</td>
<td>1.31</td>
<td></td>
</tr>
<tr>
<td>Low WM</td>
<td>.43</td>
<td>.30</td>
<td>.07</td>
<td>1.18</td>
<td>1.10</td>
</tr>
</tbody>
</table>

### Table 3

Scores on the Vocabulary Size Test (VST) and C-Test

<table>
<thead>
<tr>
<th>VST: Estimated vocabulary size (word families)</th>
<th>SD</th>
<th>C-Test</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>High WM</td>
<td>8713.51</td>
<td>1180.67</td>
<td>57.65</td>
</tr>
<tr>
<td>Low WM</td>
<td>8416.22</td>
<td>1834.77</td>
<td>51.05</td>
</tr>
</tbody>
</table>

*Note.* WM = working memory.
approaches significance and that working memory and C-test scores were correlated ($r = .32, p < .01$) calls for caution.

To further examine the two groups’ comparability on English proficiency, we converted participants’ TOEIC or TOEFL scores into CEFR levels using guidelines published by Educational Testing Service (ETS). Among the participants, 37 were at Level B1, 28 at B2, and 4 at C1.

Data for five participants were missing (four in the low–working memory group). We replaced the missing values with the code for Level B1, the lowest level, because this was both the median and the mode. This substitution was conservative; it increased the likelihood of finding a difference in proficiency between the groups, because four additional cases were coded as the lowest proficiency level in the low–working memory group and only one in the high–working memory group. A Mann-Whitney $U$-test revealed no difference between the groups based on CEFR levels ($U = 754.5, z = -.75, p = .45$). Thus, we conclude that they were comparable on English proficiency.

**Reading Comprehension**

Table 4 summarizes the two groups’ reading comprehension scores in the background-knowledge (henceforth background) and no-background-knowledge conditions (henceforth no background). The results of a $2 \times 2$ mixed ANOVA indicate a significant main effect of background ($F(1, 71) = 67.37, p < .001, \eta_p^2 = .48$), a significant main effect of working memory capacity ($F(1, 71) = 4.49, p = .04, \eta_p^2 = .06$), and a significant interaction between background and working memory capacity ($F(1,71) = 4.09, p = .047, \eta_p^2 = .05$).

Bonferroni-corrected post hoc tests indicated that high–working memory participants achieved higher reading comprehension scores when they were provided with background knowledge than when they were not ($t(36) = 8.17, p < .001$), with a very large effect size ($d = .1.34$). This pattern was also evident among the low–working memory participants, but they benefitted from the provision of background less ($t(36) = 3.97, p < .001$) with a medium effect size ($d = .65$). The

<table>
<thead>
<tr>
<th></th>
<th>Mean RC score (no background)</th>
<th>SD</th>
<th>Mean RC score (background)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>High WM</td>
<td>51.35</td>
<td>16.94</td>
<td>67.32</td>
<td>14.45</td>
</tr>
<tr>
<td>Low WM</td>
<td>47.17</td>
<td>18.48</td>
<td>56.39</td>
<td>18.66</td>
</tr>
</tbody>
</table>

*Note.* WM = working memory.
groups performed identically when no background was provided \((t(72) = 1.09, p = .28)\); however, the high-working memory participants achieved significantly higher reading comprehension scores relative to their low-working memory counterparts when they had access to background knowledge \((t(72) = 2.96, p = .004)\) with a medium effect size \((d = .68)\). Figure 1 summarizes this pattern of results.

We ran a hierarchical regression analysis to examine the relative contributions of the VST, C-test, and working memory scores to reading comprehension. The VST and C-test scores were entered in the first step and working memory scores in the second step to examine whether interindividual variation in working memory significantly predicted reading comprehension once lexical knowledge and performance on the C-test had been accounted for. Because the effect of working memory was not significant in the no-background condition, and the correlation between working memory and overall reading comprehension scores did not reach significance, we limited this analysis to the background condition. The results reveal that C-test scores were the strongest predictor of reading comprehension in the background condition, followed by VST scores, and working memory scores (see Table 5).

The results of the regression are reported in Table 6. In Step 1, both VST and C-test scores were significant predictors of reading comprehension in the background condition. This model accounts for 69\% of the variance in reading comprehension scores. Entering

![Figure 1. The interaction between working memory (WM) and background knowledge. The reading comprehension scores are expressed as percentages. Asterisks denote statistically significant contrasts.](image-url)
working memory scores in Step 2 did not yield a significantly improved model, implying that working memory was not a unique predictor of reading comprehension in the background condition.

Thus, the test of working memory and the C-test captured an overlapping portion of the variance in reading scores in the background condition. VST scores, on the other hand, were not correlated with working memory capacity. A mixed 2 × 2 analysis of covariance (ANCOVA) model with C-test scores as a covariate pointed to a significant main effect of background ($F(1,71) = 9.48, p = .003, \eta_p^2 = .12$), a significant main effect of C-test ($F(1,71) = 55.03, p < .001, \eta_p^2 = .44$), no significant main effect of working memory, a significant interaction of background and working memory ($F(1,71) = 4.84, p = .03, \eta_p^2 = .06$), but no significant interaction of background and C-test. This additional analysis supports the conclusion that differences in working memory between the two groups were responsible for high–working memory participants outperforming their low–working memory counterparts on reading comprehension when background knowledge had been provided. This is despite the fact that the working memory scores did not capture additional variance in reading comprehension scores.

**TABLE 5**
Correlation Coefficients for the Background Condition

<table>
<thead>
<tr>
<th></th>
<th>Working memory</th>
<th>VST</th>
<th>C-Test</th>
<th>Reading comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working memory</td>
<td>–</td>
<td>.15</td>
<td>.32*</td>
<td>.30*</td>
</tr>
<tr>
<td>VST</td>
<td>–</td>
<td>–</td>
<td>.73**</td>
<td>.58**</td>
</tr>
<tr>
<td>C-Test</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>.63**</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note. VST = Vocabulary Size Test.
*p < .01, **p < .001.

**TABLE 6**
Results of the Regression Analysis for the Background Condition

<table>
<thead>
<tr>
<th>Step 1</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>p (β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.090</td>
<td>8.570</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>VST</td>
<td>0.003</td>
<td>0.001</td>
<td>.270</td>
<td>.040</td>
</tr>
<tr>
<td>C-Test</td>
<td>0.550</td>
<td>0.150</td>
<td>.470</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Step 2</td>
<td>B</td>
<td>SE B</td>
<td>β</td>
<td>p (β)</td>
</tr>
<tr>
<td>Constant</td>
<td>7.280</td>
<td>8.620</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>VST</td>
<td>0.003</td>
<td>0.001</td>
<td>.280</td>
<td>.030</td>
</tr>
<tr>
<td>C-Test</td>
<td>0.500</td>
<td>0.160</td>
<td>.430</td>
<td>.002</td>
</tr>
<tr>
<td>Working memory</td>
<td>3.130</td>
<td>2.780</td>
<td>.100</td>
<td>.280</td>
</tr>
</tbody>
</table>

*Note. VST = Vocabulary Size Test.

$R^2 = .69$, $p < .001$ for Step 1; $\Delta R^2 = .01$, $p = .28$ for Step 2.
in the background condition beyond the variance already captured by the C-test. This is further elaborated on below.

DISCUSSION

This study examined whether and how working memory and background knowledge combine to facilitate L2 reading. Our results aligned with Leeser (2007) in that they support the rich-get-richer model, which holds that the positive effect of prior knowledge on cognitive performance can be enhanced by high working memory capacity. This contradicts the findings reported in Alptekin and Erçetin (2011) and Payne and colleagues (2009), which favored the independent-influences model in which no interaction between working memory and background knowledge was found.

As previously mentioned, methodological differences likely contributed to the inconsistent findings across studies. First, the present study and Leeser (2007) used the same RST scoring system. To measure the storage component of working memory, the number of correct recalls of final words was recorded; to measure the processing component, we recorded not only the sentence judgment scores but also the reaction times, thus taking processing speed into account, which is rarely done (Waters & Caplan, 1996). Additionally, Leeser and the present study operationalized background knowledge as readers’ prior knowledge of the topic at hand and examined the effect of such topic-related knowledge on L2 reading whereas Alptekin and Erçetin (2011) and Payne and colleagues (2009) targeted cultural familiarity and domain knowledge, respectively.

Regarding the results of this study, low-working-memory readers did benefit from having background knowledge, but not as much as their high–working memory counterparts. This finding highlights the role of working memory in L2 reading in terms of using existing resources to one’s advantage, particularly because no significant difference on VST and C-test scores was found between the groups. The high–working memory group’s superior performance may be attributed to their larger storage and processing capacity. Some defining characteristics of good readers are efficient and automatic decoding, lexical access, parsing, inferencing, and integrating information. Assuming that L2 readers of comparable proficiency levels must devote similar amounts of working memory resources to lower-level processes, such as decoding, lexical access, and parsing, it follows that those with higher working memory capacity will have more of these resources left for higher level processes such as inferencing, integrating information, and strategy deployment, including using background knowledge.
Similarly, processing-oriented hypotheses support the notion that the differences between high- and low-working-memory readers lie in how efficiently they utilize working memory resources to access relevant background knowledge to “support the encoding and retrieval of the yet-to-be-recalled items” (Yeari, 2017, p. 2). This explains how our high-working memory participants were able to put their background knowledge to a better use than their counterparts, which resulted in a bigger boost to their reading comprehension scores. Given that using background knowledge was not part of the instructions during the task, the difference can be attributed to the participants’ ability to access and use resources available to them.

Dehn (2008) asserts that an individual’s ability to efficiently process information depends on his or her inhibitory and attention control as well as updating ability governed by the central executive component of working memory. Thus, high-working-memory students likely approached the complex task of reading comprehension more strategically in the sense that they suppressed irrelevant information, including what they learned from the background texts, to keep relevant information active and ready for integration with textual information. Palladino and collaborators. (2001) claim that holding information in temporary storage is not enough for successful reading comprehension to occur. Rather, it is the flexible use of attentional control that allows for “coordination of information activation and suppression” (p. 346). In this respect, the high-working-memory readers’ selective attention, paired with their ability to go through the iterative process of updating information held in working memory, may have been key to taking advantage of their background knowledge and ultimately achieving better reading comprehension.

Our findings have implications for teaching because they reinforce the importance of not only providing background knowledge when it is not already present, as was the case with our participants, but also following up with explicit instructional support to help readers activate background knowledge. It comes as no surprise that providing background knowledge facilitated comprehension of L2 texts. Increasing familiarity with material is in fact considered highly beneficial in teaching contexts because it can ameliorate working memory loads (Gathercole & Alloway, 2008). However, our findings show that the picture is more complicated than that; not everyone benefits equally from background knowledge due to the constraints of working memory. In other words, raising familiarity with the topics was not enough to help low-working memory L2 readers fully overcome limitations imposed by working memory capacity.
Though measuring students’ working memory is not common practice in teaching contexts, a classroom is likely to include students with a wide range of working memory capacities as an individual-difference factor. In this regard, teachers should not assume that L2 students can utilize background knowledge to its full extent to aid understanding of a text just because they are familiar with the topic at hand. Further, providing background knowledge on an unfamiliar topic to facilitate reading comprehension may not be the end of a set of instructional activities, but a starting point, particularly for low–working-memory students who may not benefit from this intervention fully. Thus, additional instruction may still be necessary to help all students use what is available to them to achieve more successful comprehension.

As noted by an anonymous reviewer, the finding that short readings provided in the L1 to simply introduce the topic were able to make a difference in comprehension points to the importance of prereading activities; they provide new knowledge that helps orient the reader to the text or activate prior knowledge that is consistent with information in the L2 text to be read. Connecting text to background knowledge is in fact an empirically validated reading strategy (Grabe & Stoller, 2011). Providing relevant background knowledge on which readers can draw can be done by assigning a relatively easy and simple prereading task as the current study showed. Because reading itself can build knowledge, multiple readings on the same topic, dubbed topic-focused wide reading, as recommended by Neuman, Kaefer, and Pinkham (2014), can prepare students for reading informational or academic texts. Other ways that relevant background knowledge can be constructed include lectures, visual aids, and text previewing (Carrell, 1988). Text previewing, for instance, particularly helps students activate what they already know about the topic through recognizing the title, subheadings, and text structure (Alyousef, 2006). Taking this further, a brief discussion can be carried out to elicit relevant information prior to reading. Teachers can pose questions that will engage students in a discussion, which will also help students recognize what to look for while reading. In line with the findings of the current study, Carrell (1988) pointed out that even when readers are equipped with prior knowledge, problems can still arise if the knowledge is not used effectively or activated; prereading activities, therefore, should serve the goals of “building new background knowledge as well as activating existing background knowledge” (p. 248). These advantages in turn will be able to assist students particularly those with lower working memory who need explicit instructional support for accessing their prior knowledge and utilizing it to facilitate their comprehension.
A Note for Future Research on Using L2 Knowledge Measures

The C-test used in this study, intended as a language proficiency control alongside the VST, appears to have captured much of the same pool of variance in reading comprehension scores as our working memory test, to the point of rendering the working memory test disposable as a unique predictor of reading comprehension. However, as evidenced by the ANCOVA model, despite the fact that the two constructs overlap (with approximately 10% of shared variance), it was not the C-test construct but the working memory construct that interacted with background when it came to reading comprehension. If anything, this finding underscores the inadequacy of the C-test as a proxy for proficiency in studies of working memory and reading. It is likely that many of the same working memory resources relevant for reading comprehension are also relevant for the successful completion of a C-test. The VST, however, does not suffer from the same problem, as evidenced by the absence of a correlation with working memory scores in the present study.

Future studies of working memory and reading would probably benefit from using more comprehensive indicators of L2 proficiency such as TOEFL, IELTS, or TOEIC scores or CEFR levels as controls, alongside a test of vocabulary size. If, on the other hand, the goal is to predict how a group of readers will perform on comprehension, a test of working memory is not necessary, and a combination of the VST and C-test will suffice and account for approximately 70% of the variability in reading comprehension scores. Under such circumstances, the deployment of a sophisticated and resource-intensive working memory test is not likely to be of any practical relevance.1

CONCLUSION AND FUTURE DIRECTIONS

This study set out to investigate whether and how working memory and background knowledge interact in L2 reading

---

1 As an anonymous reviewer pointed out, Juffs (2017) has recently put forth the possibility that phonological working memory (PWM) may be an epiphenomenal construct critically hinging on phonological proficiency and sharing variance with it. Although this study measured a different construct, working memory, through an L2 reading span task, it is intriguing that the task we used appeared to tap resources similar to the C-test. Juffs suggests that more research into the nature of PWM is needed to probe the ways in which it shares variance with phonological proficiency. In a similar vein, more research is needed into the relationship between the L2 RST and the C-test to establish the cognitive resources they do and do not share. Such work would be valuable both as a basic scientific endeavor and as a methodological one.
comprehension and simulated a plausible situation where L2 students encounter texts on familiar and unfamiliar topics. The results provide support for the rich-get-richer hypothesis, suggesting that high-working-memory students are equipped with cognitive faculties that allow for effective use of existing resources such as background knowledge to enhance comprehension. Low-working-memory students, on the other hand, were not able to utilize background knowledge as efficiently, which led to less success at L2 reading comprehension.

The findings of this study could be enriched by conducting further research to investigate which component of working memory (storage or processing) and which underlying cognitive skills (e.g., updating, inhibiting) are particularly helpful when using background knowledge to facilitate L2 reading comprehension. Including more tasks to directly examine readers’ updating and inhibiting ability by tapping executive functions in a more fine-grained manner should cast additional light on how working memory and background knowledge interact. As an anonymous reviewer pointed out, an RST administered in the L1 or a nonlinguistic task (e.g., an operation span task) might have been a purer measure of working memory. We would add that it would, consequently, have been likely to share less variance with the C-test. Our findings, therefore, should be interpreted in light of the fact that the construct of working memory measured in the current study involved storage and processing in the L2.

As another anonymous reviewer pointed out, a question not addressed by the present study is how two subconstructs of reading comprehension, literal comprehension and inferential comprehension, might interact with working memory differently. Future studies that set out to investigate this important issue should feature reading comprehension tests with sufficient numbers of rigorously designed and tested items aimed at differentiating between the two subconstructs reliably and with sufficient statistical power.

As outlined above, the present study used a C-test as a primary proficiency control, with TOEFL and TOEIC scores (as available) converted to CEFR levels used as a secondary proficiency control. An anonymous reviewer pointed out that TOEFL and TOEIC tests differ in important ways and that, therefore, it would be preferable not to combine results from the two. Although we were unable to collect scores from a single standardized test of English proficiency for our participants, researchers conducting similar studies in the future might consider collecting scores from a single test where feasible.

Nonetheless, this study adds to the growing body of research that aims to shed light on working memory’s contribution to L2 reading.
comprehension and reveals a complex picture of the interplay between two individual-difference factors, working memory and background knowledge.

A STATISTICAL NOTE

Similar to comparable previous studies of L2 reading and working memory (Alptekin & Erçetin, 2011; Leeser, 2007), we divided participants into high– and low–working-memory groups, thereby rendering a continuous variable categorical. Under certain circumstances, dichotomizing has been linked to a loss of statistical power equivalent to removing a third of the sample (Cohen, 1983; Cohen, Cohen, West, & Aiken, 2003). There have also been justified concerns about possibly increasing the likelihood of spurious effects (MacCallum, Zhang, Preacher, & Rucker, 2002). However, in 2 × 2 factorial designs, the main effect of the dichotomized variable is not affected, and any adverse effects on the main effect of the orthogonal variable and the interaction of the variables are likely to be minimal (Iacobucci et al., 2015). Dichotomization simplifies statistical analyses and may reflect researchers’ and readers’ mental models more closely when group contrasts are the main focus (DeCoster, Iselin, & Gallucci, 2009; Iacobucci et al., 2015). Moreover, continuous and dichotomized variables usually lead to identical conclusions, albeit with reduced effect sizes (Cohen et al., 2003; Iacobucci et al., 2015). Crucially for the present study, in 2 × 2 factorial designs where the two orthogonal variables are not correlated, as is the case with the working memory and overall reading scores in this study, dichotomization has been shown to have no adverse effect on the reliability of the conclusions drawn from the ANOVA model (Iacobucci et al., 2015).

Still, this study’s results cannot be interpreted as having identified some specific break in working memory scores that would differentiate among learners for practical purposes. Our findings simply suggest that dedicating class time to activities related to background knowledge, even as a modest intervention, will benefit students, but that more time is probably better because it will result in greater reading comprehension gains for more students.

ACKNOWLEDGMENTS

We wish to thank Charlene Polio and three anonymous reviewers for their help and suggestions during the revision process. We also thank Jesse Conway for his helpful comments. Any remaining errors are our own.
THE AUTHORS

Jihye Shin is a PhD student in applied linguistics at Northern Arizona University. Her research interests lie in psycholinguistics, specifically in the link between working memory and reading comprehension. She is also interested in second language reading fluency and reading strategies assessment and instruction.

Vedran Dronjic is an assistant professor of applied linguistics at Northern Arizona University, where he directs the Language and Memory Lab. He holds a PhD from the University of Toronto. His research focuses on morphological processing, learning, and representation; language and memory; morphological awareness; vocabulary; and reading.

Boonjoo Park is an associate professor in the English Department at Daegu Catholic University, in South Korea. She has a PhD in second language acquisition and teaching from the University of Arizona. Her research interests are second language processing in production, syntactic and semantic priming, and implicit learning.

REFERENCES


APPENDIX
C-TEST (KLEIN-BRALEY, 1997)

Text 1
Within the last twenty to thirty years the blood groups of peoples in all parts of the world have been studied. The most interesting results of these studies have been that, with few exceptions, nearly every human group examined has been found to consist of a mixture of the same foetal blood groups; human races differ in the relative numbers of persons with whom each of the four 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 citizens asserted their rights, and complained their neighbours had broken his obligations, and these spoke of taking him to court. But in all except the most important matters between the king’s most powerful subjects, the court to which they referred was a local court, and the rules which they 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 my interest in the psychological aspects of communication is even older than automatons computers I can remember when those days before computers were like. When I tried to communicate with the pre-computer I can understand a summary statement more appropropriate than the 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
Many students of society—historians, political scientists, philosophers—find the study of works of literature useful and readily say so. They do not feel threatened by a different kind of discipline or term to overstress their own subject special mystique. The highest degree of imagination necessary for distinctive work between the human or social sciences ensures that minds with the power different not misunderstand the technical boundaries between academic discipline for divisions within human experience.