The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading

Cem Alptekin and Gülcan Erçetin

Department of Foreign Language Education, Faculty of Education, Boğaziçi University, Turkey

Reading span as a measure of L1 or L2 working memory capacity is shown to be related to L2 reading comprehension. Albeit limited, there is research on the relationship between both L1 and L2 reading spans and their roles in L2 reading comprehension, yet these studies take reading as a global construct rather than delving into its multi-level representational architecture. This study differentiates itself from previous research in that it investigates the relationship of both L1 and L2 reading spans to L2 reading comprehension, while taking into account two reading dimensions, namely, literal understanding and inferential comprehension. Findings reveal no significant difference between L1 and L2 storage capacities, although task processing in L1 is more accurate compared to that in L2. Generally, L1 and L2 processing tasks correlate positively as do L1 and L2 storage tasks. Finally, only L2 reading span has a meaningful relationship with L2 inferential comprehension.

In a broad sense, working memory (WM) is needed in a wide variety of cognitive operations that require controlled processing. With its dual functions of processing and storage, it is characterised by executive attention procedures, and is a limited-capacity system that is necessary for conducting numerous linguistic and conceptual tasks (Just & Carpenter, 1992). These include context-based vocabulary learning (Daneman & Green, 1986), note taking (Kiewra & Benton, 1988), writing (Benton, Kraft, Glover & Plake, 1984), reasoning (Barrouillet, 1996; Kyllonen & Christal, 1990), generating hypotheses (Dougherty & Hunter, 2003) and both listening and reading comprehension (Daneman & Carpenter, 1983; Daneman & Merikle, 1996).

Despite some differences of opinion regarding the operationalisation of WM capacity, span tasks (counting span, operation span and reading span) seem to be the most widely acknowledged measures for assessing WM, as they are found to have satisfactory validity and reliability (Conway et al., 2005). Span tasks involve a dual-task paradigm, which combines a memory span measure with a concurrent processing task. Based on the notion of a trade-off between processing and storage as interdependent components, they operate on the assumption that an increase in the amount of processing demands leads to a decrease in the number of storage items and vice versa.¹

In the case of reading, what is commonly used to measure WM capacity is the sentence-based reading span test (RST), although other span tests (e.g. operation span
tasks) are also known to be used. Developed by Daneman and Carpenter (1980), this test generally consists of unrelated simple sentences in the active voice, each sentence ending with a different word. The sentences are divided into sets. The number of sentences in sets gradually increases from two to six. As they view the sentences, participants indicate whether each sentence is accurate from the viewpoint of syntax, semantics or veracity. After they finish viewing the sentences at one level, they are asked to indicate the sentence-final words that they are able to recall. The sentence judgement procedure measures the processing function of WM capacity, whereas the word recall measures the storage function. The reading span is taken to be a composite score involving each participant’s sentence judgement score and total word recall (Waters & Caplan, 1996).

**Relationship between L1 and L2 WM capacities**

Research has shown that first language (L1) and second language (L2) WM capacities share substantial amounts of variance and that this significant relationship is not language-specific (Juffs, 2005; Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993). That is, high-span individuals have more attentional resources to draw on than low-span individuals, independent of the language (native or foreign) involved. For example, Osaka and Osaka (1992) found a strong relationship between reading spans for L1 (Japanese) and L2 (English) with highly proficient L2 users. This relationship was confirmed by Osaka et al. (1993) for proficient bilinguals of L1 (German) and L2 (French). In both of these studies, correlations between L1 and L2 WM measures exceeded .70. A number of other studies, however, found only a moderate relationship between L1 and L2 WM capacities, correlations ranging between .39 and .68 (Harrington & Sawyer, 1992; Juffs, 2004, 2005; Miyake & Friedman, 1998). These findings suggest that L1 and L2 may, by and large, share similar WM resources.

Because the studies reviewed above were conducted with proficient users, it is not clear whether the relationship between L1 and L2 WM capacities would be affected by the degree of competence in the L2. An important study in this regard was conducted by Van den Noort, Bosch and Hugdahl (2006), who investigated the interaction between WM capacity and the level of language proficiency. The participants were native speakers of Dutch (L1) who were fluent in German (L2), and had started learning Norwegian (L3) during the 6 months before the study. The results indicated that the participants’ storage capacity and their speed of processing increased in relation to their proficiency level in the language (L1 > L2 > L3). Moreover, significant correlations were found among the scores on the reading span tasks in Dutch, German and Norwegian (all \( r > .75 \)), suggesting that the cognitive resources underlying WM capacity in the L1, L2 and L3 are closely related.

**WM capacity and L2 reading**

WM capacity is shown to correlate significantly with L2 reading comprehension (Harrington & Sawyer, 1992; Leeser, 2007; Walter, 2004). Harrington and Sawyer (1992), for instance, found a significant but moderate correlation between L2 reading span and L2 reading comprehension scores based on the reading section of the TOEFL \( (r = .57) \). Yet no significant correlation was found between the participants’ reading span
scores and their scores on a cloze test. Leeser (2007), in a study conducted with elementary learners of Spanish, assigned the participants to high- and low-WM capacity groups based on their L1 reading span scores. He found that high-WM capacity participants recalled a greater percentage of propositions from the L2 texts they read than those with a low-WM capacity.

An important issue in studying the relationship between WM capacity and L2 reading comprehension is whether WM capacity should be measured in the L1 or L2. If WM capacity is a predictor of reading comprehension, irrespective of whether individuals are using their first or second language, one would expect both L1 and L2 WM capacities to be associated with L2 reading performance due to the moderate-to-strong relationships found between the two constructs. However, previous studies investigating the links between WM capacity and L2 comprehension indicate that while the effects of L2 WM are direct, those of L1 WM remain indirect at best. For example, Chun and Payne (2004) found no correlations between L1 reading span and L2 reading comprehension. Nor did they find any correlations between L1 reading span and L2 text recall. Walter (2004), who examined the relationship between reading span and L2 reading comprehension for a group of upper- and lower-intermediate French learners of English, found a low correlation between L1 reading span and summary completion in the L2 (r = .33) but a strong relationship between L2 reading span and summary completion (r = .73), suggesting that L2 reading comprehension is associated more with L2 WM capacity than L1 WM capacity. Going beyond correlations, Miyake and Friedman (1998) examined the relationship between L1 and L2 listening span and L2 syntactic comprehension through a path analysis. They showed that L2 syntactic comprehension was directly linked to L2 WM capacity. Although there was a moderate relationship between WM capacity for L1 and that for L2, L1 WM capacity was found to be a mediator variable. That is, it was directly related to L2 WM capacity but indirectly related to L2 syntactic comprehension. In general, then, research findings suggest that WM capacity measured in the L2 plays an important role in L2 reading and listening comprehension.

Operationalisation of reading comprehension

Research examining the relationship between WM capacity and reading in the L2 has always treated reading comprehension as a global construct, disregarding its multi-level representational architecture and the contribution of each level (the surface code, the text-base and the situation model) to comprehension (Kintsch, 1998). This is a problematic issue in that WM capacity may be differentially involved based on whether reading tasks are chiefly of a literal or inferential nature, especially in view of the different degrees of cognitive load (Sweller, 1994) associated with literal or inferential reading tasks and the different levels of cognitive activation associated with automatic or controlled processing. For example, Rupp, Ferne and Choi (2006) indicate that inference questions that cause readers to engage with the whole text are perceived as being very difficult. This is because generating inferences, particularly of an elaborative character, is essential to the construction of a situation model, that is, the mental representation of the text. As such, inferences of the type that produce novel knowledge derived from textual content, compared with those that simply add pre-existing information to the text, require more resource-demanding control processes in that they go beyond the text-base in order to integrate it with the reader’s prior topical knowledge. Thus, generating inferences
enables readers to move from the semantically ‘shallow’ text-base to the semantically ‘deep’ situation model (Perfetti, 1999). Literal understanding, however, which normally rests on lower-order conceptual and linguistic operations primarily at the text-base level, cannot be said to be as cognitively demanding as inferential comprehension, because it does not entail a high degree of controlled processing. Without the need for schema formation and automation, as is the case with inferential comprehension, WM is likely to deal with building a text-base with relative ease, at least for skilled readers. Normally, many of the text-base operations, in particular those at the microstructure level, involve the use of decoding and syntactic parsing. Clearly, this kind of sentence-based processing is quite different from comprehension as a complex multi-level process, which further includes but is not limited to text cohesion (e.g. anaphora resolution), proposition integration (e.g. argument overlap) and meaning construction (e.g. activation of relevant background knowledge from memory). Studies in both L1 (Yuill & Oakhill, 1991) and L2 (Walter, 2004, 2007) indicate that there is discontinuity between, for instance, decoding and comprehension in reading, corroborated by neurological evidence as well (Robertson et al., 2000). Walter, basing her views on Gernsbacher’s (1990) ‘structure building’ model of comprehension, states that L2 readers can understand the individual sentences in a text without adequately comprehending the text itself (2008, p. 456). Given the inadequacy of the microstructure to account for textual meaning to emerge, the act of comprehending further requires the reader to form a macrostructure that would result from the integration of propositional units through the use of coreferences, logical implications and simple cause-and-effect relationships. A deeper understanding of the text is not realised, however, unless the micro- and the macrostructural properties of the text-base are interactively integrated with the reader’s relevant prior knowledge so as to form a mental model of the situation the text describes (Kintsch & Rawson, 2007). Thus, in addition to textual features, the reader’s ‘situation model’ (Kintsch, 1998) comprises world knowledge, knowledge of text genre and the discourse model built up while processing the text.

Aims – hypotheses

Separating L2 reading comprehension into its literal and inferential dimensions, the present study explored the role of L1 and L2 WM capacities in L2 reading with proficient L2 users through reading span tests in L1 (Turkish) and L2 (English). The following research questions were investigated: (1) Are there performance differences in task-processing accuracy and storage capacity between L1 and L2 reading span tasks? (2) Is there a relationship between L1 and L2 reading spans? (3) Are individual differences in L1 and L2 reading spans related to literal and inferential comprehension tasks in L2 reading? In line with the findings of the few studies on the subject (e.g. Van den Noort et al., 2006), we hypothesised a larger storage capacity and more accurate task-processing in the L1 compared with the L2 (Hypothesis 1). Second, we expected a positive relationship between L1 and L2 WM capacities (Hypothesis 2), given the support in the literature on this point (e.g. Juffs, 2004, 2005; Osaka & Osaka, 1992; Osaka et al., 1993; Van den Noort et al., 2006). Finally, we assumed that a stronger relationship would be observed between L2 WM capacity and literal and inferential comprehension in the L2 than between L1 WM capacity and literal and inferential comprehension in the L2 (Hypothesis 3), in light of the findings by Miyake and Friedman (1998) and Walter (2004).2

© United Kingdom Literacy Association 2009
Method

Participants

The participants were Turkish university students enrolled in a Turkish university where the medium of instruction is English. They had been successful on the university’s English proficiency test whose minimum pass mark is accepted as the equivalent of 550 on the paper-and-pencil version of the TOEFL. They had also obtained high scores on the verbal sections of the national university entrance examination (ÖSYM), administered in Turkish and similar to the ‘Critical Reading Section’ of the SAT Reasoning Test. With a minimum competence level in the range of 550 on the TOEFL and a satisfactory reading proficiency in Turkish, they were considered to have the capacity to make use of higher-order reading comprehension processes, one of which would be to achieve deeper comprehension through the generation of inferences. The participants’ ages ranged from 20 to 23. Of the 43 participants 38 were female and 5 were male. They formed a homogeneous group in terms of their educational background in that they had all successfully completed a teacher training high school and were enrolled in university-level ELT (English Language Teaching) courses in order to become teachers of English.

Materials and procedures

Materials for the study comprised an RST in L1 (Turkish), an RST in L2 (English) and an L2 reading comprehension test with multiple-choice items based on a short story in English.

RST. The RSTs used in the study were modified versions of the original test designed by Daneman and Carpenter (1980). They made use of the procedures described in Harrington and Sawyer (1992). That is, both versions consisted of 70 unrelated simple sentences in the active voice, each 11–13 words in length. Every sentence ended with a different word. The test involved four levels starting at two and extending up to five, with each level containing five trials. A grammaticality judgement test was incorporated into the RST to ensure that participants processed every sentence for syntax and did not focus simply on the final words. There were 35 grammatical (e.g. The only thing left in the kitchen cupboard was a broken cup) and 35 ungrammatical sentences (e.g. I dreamed that I was in with my field a sheep), arranged randomly. Each sentence appeared only once. Participants were tested on the same sets of sentences. After they finished reading all five sets at one level, they moved on to set 1 of the next level. Participants took the RST in Turkish and the one in English separately, with a time interval of 2 weeks between the two administrations. Using Cronbach’s α, the internal consistency reliability coefficients for the processing and storage tasks on the Turkish version were found to be .743 and .875 while those on the English version were .695 and .857, respectively.

The tests, administered in a computer lab, were delivered on-line by displaying one sentence after another at 7-second intervals until all the sentences in a set had been viewed. While processing the sentences, the participants pressed one of two computer keys to indicate whether a given sentence was grammatical or ungrammatical. After all the sentences in a set had been viewed, a field box appeared on the screen for the participants to enter the sentence-final words that they were able to recall. While the participants’ judgements concerning the grammaticality of the sentences represented the processing measure of their reading span, the total number of reliably recalled sentence-final words was taken as the measure of storage.
Scoring the tests involved obtaining composite scores by converting word recall and sentence judgement scores to z-scores and taking their average, as suggested by Waters and Caplan (1996).

**Reading text.** The text used for L2 reading comprehension was an American short story of 2,270 words by Delmore Schwartz (1978). The story, ‘In dreams begin responsibilities’, is autobiographical in nature and takes place in New York City in the early 1900s, when immigrants were struggling to find their way in the New World. The two conflicting themes are success in business and worldly accomplishment on one hand, and social problems caused by quick financial gains in a new culture on the other. The plot deals with the thoughts and deeds of a young man with nouveau riche tendencies, who is about to marry a woman from a well-established yet not particularly wealthy family.

**Reading comprehension test.** Given the bias stemming from the memory effect associated with commonly used recall tasks (Alptekin, 2006; Chang, 2006), the participants’ comprehension of the text was measured with a multiple-choice test. Based on Pearson and Johnson’s (1978) taxonomy of reading questions, the test contained a total of 20 questions, half of which were textually explicit, while the other half were textually or scriptally implicit. Questions were not grouped according to type; they simply followed the course of the story. Briefly, textually explicit questions measured readers’ literal understanding in that their answers could be derived directly from the text. The reader was supposed to understand the author’s propositional message, that is, form a coherent text-base, by extracting meaning from sentences, associating propositions between and across sentences and supplementing local inferences to make the text cohere. Textually or scriptally implicit questions, on the other hand, measured inferential comprehension which required extraction of the deeper meanings of the text by going beyond the text-base into constructing a situation model. Inferential questions tapped mainly two types of inferences: connective and elaborative. The former involved combining different pieces of textual information with prior knowledge of the topic in order to identify the author’s intended meaning, while the latter required going beyond the text-base to construct a mental model of what the text as a whole was about. Examples of test items for each category appear in Appendix A.

### Results

**Performance on L1 and L2 reading span tests**

Descriptive statistics for the processing and storage tasks of the L1 and L2 reading span tests are provided in Table 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 processing</td>
<td>61.56</td>
<td>4.98</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td>L1 storage</td>
<td>49.70</td>
<td>9.26</td>
<td>32</td>
<td>70</td>
</tr>
<tr>
<td>L2 processing</td>
<td>58.72</td>
<td>5.03</td>
<td>44</td>
<td>67</td>
</tr>
<tr>
<td>L2 storage</td>
<td>48.23</td>
<td>8.79</td>
<td>28</td>
<td>65</td>
</tr>
</tbody>
</table>
The participants demonstrated better performance on both storage and processing tasks in the L1. Nevertheless, the mean difference between the L1 and L2 storage tasks is negligible, with wider variability being observed between them compared with that between the processing tasks. Another finding lies in the average performance on processing tasks in both L1 and L2 being higher than that on the storage tasks in both languages. A one-way repeated measures ANOVA revealed significant mean differences, $F(3, 126) = 34.15, p < .001$, $\eta^2 = .45$, with a large effect size. Post hoc comparisons showed that participants perform better on the processing tasks compared with storage tasks. Yet task-processing in the L1 was still significantly more accurate. Finally, storage capacity in the L1 and L2 did not differ significantly (L1 processing $>$ L2 processing $>$ L1 storage $= L2$ storage).

**Relationships between L1 and L2 reading spans**

Correlations between the reading span tasks in both languages and correlations between the composite WM scores are shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>L1 storage</th>
<th>L2 processing</th>
<th>L2 storage</th>
<th>Composite L1 WMC</th>
<th>Composite L2 WMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 processing</td>
<td>$-.541^{**}$</td>
<td>$.529^{**}$</td>
<td>$-.424^{**}$</td>
<td>$.479^{**}$</td>
<td>$.088$</td>
</tr>
<tr>
<td>L1 storage</td>
<td>$-$</td>
<td>$-.339^{*}$</td>
<td>$.629^{**}$</td>
<td>$.479^{**}$</td>
<td>$.243$</td>
</tr>
<tr>
<td>L2 processing</td>
<td>$-$</td>
<td>$-$</td>
<td>$.287</td>
<td>$.199</td>
<td>$.597^{**}$</td>
</tr>
<tr>
<td>L2 storage</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$.214</td>
<td>$.597^{**}$</td>
</tr>
<tr>
<td>Composite L1 WMC</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$.346^{*}$</td>
</tr>
</tbody>
</table>

*Note: WMC, working memory capacity.

$p < .05; \; ^{*}p < .01.$

Correlations of L1 and L2 WM capacities with L2 reading comprehension

L1 processing, L1 storage and L2 storage scores correlated neither with literal nor with inferential comprehension in L2 reading. The only significant correlations were between inferential comprehension and L2 processing ($r = .327, p < .05$) and inferential comprehension and composite L2 WM scores ($r = .454, p < .01$). Considering the significant correlations between L1 and L2 processing and composite scores, partial correlations were obtained by controlling for these relationships. The correlation between inferential comprehension and L2 processing was slightly reduced after controlling for the relationship between L1 and L2 processing ($r = .305, p = .049$). Similarly, there was a slight decrease in the correlation between inferential comprehension and composite L2
WM scores after controlling for the relationship between L1 and L2 WM scores ($r = .396, p < .01$). These findings suggest that WM capacity plays an important role in inferential reading comprehension but not in literal reading comprehension. Moreover, L2 inferential comprehension seems to be directly related to L2 WM capacity rather than to L1 WM capacity.

**Discussion**

The findings related to the first research question, whether task-processing accuracy and storage capacity of WM differ between L1 and L2, indicate that proficient L2 users demonstrate more accurate task-processing in the L1, yet they have similar storage capacities in both L1 and L2. As such, the results contradict those of Van den Noort et al. (2006) concerning storage being larger in the L1 compared with the L2. In general, however, they support the view that WM capacity interacts with language proficiency in that differences between L1 and L2 reading spans are less significant in the case of proficient L2 users (Service, Simola, Metsaenheimo & Maury, 2002; Van den Noort et al., 2006; Walter, 2004).

The second research question, focusing on the relationship between L1 and L2 WM capacities, was answered affirmatively. The results demonstrate that there is a positive correlation between L1 and L2 processing tasks, on the one hand, and between L1 and L2 storage tasks on the other. Yet the relationship between L1 and L2 WM scores decreases when the composite scores are correlated. These findings corroborate the results of other studies (Juffs, 2005; Miyake & Friedman, 1998; Osaka & Osaka, 1992; Osaka et al., 1993) which indicate that the relationship between L1 WM capacity and L2 WM capacity is consistent across languages. The significant correlations between L1 and L2 processing on one hand and between L1 and L2 storage on the other suggest that the cognitive resources underlying WM capacity in the L1 are analogous to those in the L2. The negative correlation found between processing and storage in the L1 is indicative of a trade-off between the processing and storage components of WM, as postulated by Daneman and Carpenter (1980). This finding casts doubt upon the argument that processing and recall are positively correlated in that ‘subjects who recall the most target items also perform most accurately on the processing task’ (Conway et al., 2005, p. 774).

As for the relationship between processing and storage in the L2, the nonsignificant correlation observed is rather unexpected but not necessarily out of line given that trade-off ‘may take place only when working memory is not fully occupied by processing’ (Towse & Hitch, 1995, p. 123). Participants taking the RST in their L2 normally expend their cognitive resources much more than in the L1, as L2 processing is less automatised than that in the L1 (Stowe & Sabourin, 2005). In fact, the results of the present study show the mean for the processing accuracy in the L2 to be significantly lower than that in the L1. It may be argued that L2 processing places a heavier demand on WM, causing participants to switch their attention from processing to storage task demands (Towse & Hitch, 1995). Thus, the lack of a significant correlation between processing and storage in the L2 indicates that processing and storage components act in a functionally independent manner (Duff & Logie, 2001; Towse & Hitch, 1995) when the tasks are performed in the second language. It is possible, as suggested by Towse, Hitch and Hutton (1998), that the processing component of the task allows time for stored information to be forgotten, due to interference-based information loss. Nonetheless, processing still determines the
amount of time that information needs to be stored and forgotten in memory (see Friedman & Miyake, 2004 for a detailed treatment of the relationship between processing and storage).

The third research question, concerning the relationship between individual reading span differences in the L1 and L2 and literal and inferential reading comprehension in the L2, yielded interesting results. For one thing, literal understanding did not correlate with any WM capacity measure. This suggests that literal understanding, as a data-driven process, is essentially dependent on the level of language proficiency and surface readability features (e.g. decoding, syntactic parsing) and, consequently, does not impose on WM capacity a heavy ‘intrinsic’ cognitive load, that is, the load determined by the inherent complexity of the reading task itself (Sweller, 1994; Sweller, van Merriënboer & Paas, 1998). The highly proficient L2 user is able to make use of a high level of automatic processing in dealing with microstructural and, to a reasonable degree, macrostructural operations. Put simply, with proficient L2 readers, processing the linguistic and propositional properties of a text does not present a serious challenge to their WM capacity. In the absence of the task demand for an inordinate amount of controlled processing, the executive attention procedures of WM capacity are not called for in a significant way (Conway et al., 2005).

For another, inferential comprehension correlated with L2 processing, yet not with L2 storage. This finding is in tune with Waters and Caplan’s (1996) observation that ‘most of the shared variance between sentence span tasks and reading comprehension tasks is accounted for by the processing component of the sentence span task, but that there is a small but significant contribution of the recall measure’ (p. 26). It is possible, as Koda (2005) suggests, that ‘tasks used to measure working-memory capacity sometimes tap similar, if not identical, abilities underlying reading’ (p. 199). Hence, according to Koda, when Daneman (1991) assumes that the computational processes used in reading the sentences on the RST would consume less of the WM resources of the better readers, leaving more residual capacity to store the sentence-final words, one sees that the processing dimensions of WM capacity and reading ability are one and the same (pp. 199–200).

Admittedly, multiple-choice questions probe the reader’s ability in problem-solving and general reasoning with regard to a text which is present. As such, they are conducive to inference generation. Those individuals who are able to control their selective attention to suppress irrelevant items and focus on relevant ones, in relation to both existing textual content and immediate access to available prior knowledge in long-term memory, appear to be not only better readers but also to have high WM capacity. On the other hand, individuals whose reading performance is measured through recall tasks cannot be said to have had their reading ability assessed adequately. Recall tasks, in fact, fail to measure inferential comprehension properly, for they are built on a ‘memory bias’ (Chang, 2006) that equates remembering with comprehending, creating uncertainty as to the degree of recall based on elements extracted from the text and the amount of information retrieved from prior knowledge sources (Koda, 2005, p. 237). Moreover, their focus on the salient points of the text, often at the expense of the less obvious items, hinders the generation of deep inferences, as the integration of pertinent textual data and relevant prior knowledge fails to materialise in the absence of sufficiently rich incidental information (Alptekin, 2006).

Yet another finding has to do with composite L2 WM capacity scores correlating positively with inferential comprehension. This result is in line with Waters and Caplan’s (1996) suggestion that when recall scores are included as part of composite scores, WM
capacity becomes a better predictor of reading comprehension. This may be the outcome of recall being instrumental in lower-level cognitive processes of reading such as syntactic parsing and lexical decoding. For example, recall is involved in finding the referents of pronouns and other anaphoric elements in addition to the reader’s computational procedures underlying textual understanding. More importantly, this finding shows how inferential comprehension, which requires reasoning beyond the text while representing a satisfactory understanding of what the text says, is considerably more cognitively demanding. As such, inferential comprehension places a higher demand on the WM control of attention function, which is responsible for maintaining relevant information from the text in an active and accessible state while inhibiting irrelevant data. When inferential requirements exceed the upper bounds of capacity limitations, it is likely for effortful attentional control to fail and comprehension impairment to follow. Without adequate attentional control, an overloaded WM is unable to retrieve relevant information from available ‘long-term working memory’ (Ericsson & Kintsch, 1995), which normally contributes to a deeper understanding of the text.

One final finding of the study is that L2 WM capacity scores and L2 processing scores show slightly reduced correlations with L2 inferential comprehension when their relationships with L1 WM capacity scores and L1 processing scores are controlled. Considering the lack of a significant correlation between L1 WM capacity and inferential comprehension, this finding may confirm Miyake and Friedman’s (1998) argument that L1 WM capacity does not have a direct impact on L2 reading performance except to mediate the relationship between L2 WM capacity and L2 reading. All in all, however, as indicated before, the two constructs seem to tap the same or, at least, similar underlying cognitive resources that are utilised in inferential rather than literal reading processes.

**Conclusion**

WM capacity is a major individual-differences variable in various contexts of understanding human behaviour. The present study illustrated one such context, namely, reading comprehension in the L2. Using the reading span test, which is by now one of the most widely used complex span measures, the study demonstrated the significant relationship between L1 WM and L2 WM capacities as well as the significant relationship between L2 WM capacity and inferential reading in the L2. In general, the findings suggest that, in addition to the commonly held view about the contribution of L2 proficiency level to L2 reading performance (e.g. Bernhardt & Kamil, 1995; Lee & Schallert, 1997; Taillefer, 1996; Usó-Juan, 2006), WM capacity plays an important role in reading comprehension, especially in those cognitively demanding reading tasks characterising inferential reading.

The study has several implications for further research. First, research investigating the role of WM capacity in L2 reading should measure WM capacity in the L2 rather than L1. Despite the significant relationship between L1 and L2 reading spans, L2 reading span seems to be more directly related to L2 reading comprehension. Second, such research should focus on literal and inferential dimensions of reading separately rather than on treating reading comprehension as a global construct, as WM capacity seems to be differentially related to these two dimensions of reading ability. Finally, the lack of a relationship between WM capacity and literal comprehension with proficient users may explain why literal comprehension is easier for such individuals. It follows that it would
be interesting to see whether a meaningful relationship exists between L2 WM capacity and literal reading in the case of less proficient L2 users and, if it does, how it affects the association between L2 WM and inferential reading in the L2.

Acknowledgement

This research was supported by a grant (07D602) from the Boğaziçi University Research Fund.

Notes

1. Recently, Towse and colleagues (Towse & Hitch, 1995; Towse, Hitch & Hutton, 2002) argued that, rather than a trade-off between processing and storage, the critical determinant of span is task switching based on the time spent processing versus recalling. That is, the timing of the processing component (or the time spent away from storage) is what drives span performance. This view is indirectly supported by the notion that processing and storage components are functionally independent, with each component relying on different aspects of the cognitive system (Duff & Logie, 2001). Alternatively, Engle and colleagues (e.g. Conway & Engle, 1996; Engle, 2002; Kane, Conway, Hambrick & Engle, 2007) consider the fundamental determinant of span to be the attention switch itself rather than the time spent processing because, in their view, WM is characterised by a general system of attention-control processes whose functions are actively to maintain and retrieve relevant information in an accessible state, with a view to achieving the set goals amid conditions of interference or distraction.

2. The reader should be cautioned, however, that, unlike in the present study, the studies by Miyake and Friedman (1998) and Waller (2004) treated reading as a global construct, without any differentiation made between its literal and inferential dimensions.

3. A higher level of processing accuracy in relation to storage is to be expected given that processing accuracy is typically close to ceiling as ‘task instructions emphasize processing-task accuracy to ensure that subjects are attending to the secondary task’ (Conway et al., 2005, p. 774).

4. With reference to long-term memory, what is precisely referred to here is Ericsson and Kintsch’s (1995) construct of ‘long-term working memory’, which differentiates between short-term WM and long-term WM. Long-term working memory is said to provide an immediate and effortless access to long-term memory-based knowledge and skills congruent with the nature of the task that is processed.

References


**Appendix A**

**Examples of test questions**

Sample textually explicit question with corresponding story extract:

*They have in common a great interest in health, although my father is strong and husky, and my mother is frail. They are both full of theories about what is good to eat and not good to eat, and sometimes have heated discussions about it, the whole matter ending in my father’s announcement, made with a scornful bluster, that you have to die sooner or later anyway.*

What do the narrator’s parents discuss at times?

a. What is worth eating and what is not.
b. Dying sooner or later.
c. Their common interest in health problems.
d. Whether it is good to be strong and husky or frail.

Sample textually implicit question with corresponding story extract:

*Overhead the sun’s lightning strikes and strikes, but neither of them are at all aware of it. The boardwalk is full of people dressed in their Sunday clothes and casually strolling.*
My father and mother lean on the rail of the boardwalk and absently stare at the ocean. But I stare at the terrible sun which breaks up sight. I forget my parents. I stare fascinated, and finally, shocked by their indifference, I burst out weeping once more.

Why do you think the narrator’s parents are unaware of the ‘terrible sun’?

a. They are insensitive to heat.
b. They are oblivious to the future.
c. They are indifferent to each other.
d. They are interested in the ocean.

Sample scriptally implicit question:
What kind of relationship exists between the narrator’s father and his future father-in-law?

a. One of respect.
b. One of uneasiness.
c. One of doubt.
d. One of fear.

Received 13 December 2008; revised version received 22 July 2009.

Address for correspondence: G ülcan Erçetin, Department of Foreign Language Education, Faculty of Education, Boğaziçi University, 34342 Bebek, Istanbul, Turkey. E-mail: gulcaner@boun.edu.tr