Available online at www.sciencedirect.com
ScienceDirect
System 37 (2009) 627–639
SYSTEM
www.elsevier.com/locate/system

Assessing the relationship of working memory to L2 reading: Does the nature of comprehension process and reading span task make a difference?

Cem Alptekin [1], Gülcan Erçetin *

Bogaziçi University, Faculty of Education, Department of Foreign Language Education, 34342 Bebek, Istanbul, Turkey

Received 21 January 2009; received in revised form 11 June 2009; accepted 15 June 2009

Abstract

Although an important role has been ascribed to working-memory capacity in reading comprehension, little consensus exists on its conceptualization, operationalization, and measurement except for its recognition as a limited-capacity processing and storage system. One specific problem in the measurement of working memory comes from researchers' use of storage scores in reading span tests (RST) as an index of working-memory capacity, as if processing were inconsequential. The inadequacy of the focus on storage is compounded by the use of different testing procedures to measure it. Another issue stems from the use of L2 reading as a global construct, thus failing to account for L2 readers' text-boundedness, which often enhances their literal understanding at the expense of their inferential comprehension. Thus, the purpose of this study was first to compare the performance of L2 readers on two L2 RSTs that differed in task type used to measure storage. Second, the study aimed at investigating the relationships between L2 reading, compartmentalized into its literal and inferential dimensions of understanding, and composite RST scores representing both storage and processing performance. The findings indicate that, unlike recall tasks, recognition tasks fail to detect individual differences in working-memory storage. They further indicate that composite scores of storage and processing correlate with inferential rather than literal understanding in L2 reading when recall-based rather than recognition-based RSTs are used to measure storage.

Keywords: Working-memory capacity; Inferential comprehension; Literal comprehension; Reading span tests; L2 reading

1. Introduction

Although an important role has been ascribed to working-memory capacity in reading comprehension, little consensus exists on its conceptualization, operationalization, measurement, and assessment (Friedman and Miyake, 2004; Juffs, 2004; Koda, 2005; MacDonald and Christiansen, 2002; Traxler, 2006; Waters and Caplan, 1996) except for its recognition as a limited-capacity processing and storage system for conducting

* Corresponding author. Tel.: +90 212 359 4588; fax: +90 212 257 5036.
E-mail addresses: alptekin.cem@gmail.com (C. Alptekin), gulcaner@boun.edu.tr (G. Erçetin).
[1] Tel.: +90 212 359 4588; fax: +90 212 257 5036.

0346-251X/$ - see front matter
doi:10.1016/j.system.2009.09.007

a variety of cognitive tasks that require controlled attention. Baddeley's (1986, 2000) theoretical model posits working-memory capacity as involving a multi-component system which consists of a supervisory attentional mechanism (the 'central executive') and three auxiliary domain-specific subsystems, each of which is responsible for processing phonological ('phonological loop'), visual ('visuospatial sketch pad'), and long-term memory-related data ('episodic buffer'). There seems to be a general consensus that Daneman and Carpenter's (1980) concept of working memory, operationalized through a complex span task that requires the simultaneous processing and storage of items, maps onto Baddeley's construct of the central executive (Baddeley, 2003; Hitch, 2005; Just and Carpenter, 1992; Turner and Engle, 1989). Yet in Baddeley's model, 'the processing function of the central executive is supported by the independent storage functions of the phonological loop and visuospatial sketchpad', as indicated by Andrade (2006, p. 16).

Another difficulty stems from Daneman and Carpenter's claim that, as components of a unitary pool of attentional resources, processing and storage act in a trade-off relationship due to capacity constraints. Towse and colleagues (Towse et al., 2002; Towse and Hitch, 1995) argue that, rather than a trade-off between processing and storage, the critical factor of span is task switching, determined by the time spent processing versus recalling. 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 instead of being parts of a unitary pool. Alternatively, it has been suggested that the basic determinant of span may be the attention-switch itself rather than the time spent processing or recalling, given that working memory is characterized by a general, domain-free system of attention-control processes whose functions are to inhibit conditions of interference and distraction with a view to maintaining relevant information in long-term memory in an accessible state (Conway and Engle, 1994, 1996; Conway et al., 2005; Engle, 2002; Hambrick and Engle, 2003; Kane et al., 2007).

Furthermore, while it is generally acknowledged that working-memory capacity as a domain-general ability of controlled attention is not language-specific (Osaka and Osaka, 1992; Osaka et al., 1993), some evidence points to its interaction with language proficiency by showing how significant performance differences between first-language (L1) and second-language (L2) span tasks decrease as an individual's level of L2 knowledge increases (van den Noort et al., 2006; Service et al., 2002; Walter, 2004).

Finally, there is no consensus on how to score the span task, one of the common measures of working-memory capacity (Conway et al., 2005; Friedman and Miyake, 2005). Nor is there consensus on whether to measure the processing and storage components through tasks in which modality effect (e.g., tasks subserved by numerical, verbal or visuospatial properties) is a critical factor (Brünken and Seufert, 2007; Maehara and Saito, 2007) or else through tasks that simply necessitate effortful controlled processing regardless of the modality they characterize (Conway and Engle, 1996). Last but not least, different amounts of working-memory resources required by different cognitive tasks are not normally taken into consideration in the research. For example, in the case of reading comprehension, the likelihood of substantial differences occurring in the demands placed on working memory when readers answer comprehension questions with the text available or unavailable has not been accounted for, even though text availability leads to cognitive operations that cater to the search for the correct answer; yet its unavailability requires a significant dependence on the reader's memory skills (Andreassen and Bråten, 2007).

2. Working-memory capacity and conceptualization of reading

In addition to theoretical and operational issues concerning working-memory capacity, researchers typically treat reading comprehension itself as a global construct, paying little attention to its multilevel representational architecture and the role played by each level in comprehension. Clearly, reading tasks involve in varying degrees the contribution of the surface code, the propositional textbase, and the situation model of comprehension (Kintsch, 1998). Depending on the reader's interaction with a given level, it is likely that working-memory capacity is differentially involved, as reading tasks characterizing each level require not only different degrees but also different types of cognitive activation associated with their complexity. As a case in point, inferential reading, which underlies the situation model, is found to be more difficult than literal reading (Rupp et al., 2006) because of the heavier demands it places on working-memory capacity in terms of the 'intrinsic' cognitive load its tasks involve (Sweller, 1994). Generating inferences, in fact, is known to entail effortful cognitive control processes that are not normally found in the relatively simpler operations of literal understanding, since these operations are generally characterized by automatic or less demanding controlled processing.

In general, literal understanding involves lower-order linguistic operations of lexical decoding and syntactic parsing to set the surface code as well as integrating micro- and macro-level propositions to construct a syntactically and semantically cohesive textbase. As such, it provides a shallow understanding of the text since it does not go beyond the representation of textual meaning as expressed by the author. A deeper understanding of the text, that is, one which gives rise to its coherent interpretation, comes with the integration of the surface code, the textbase, and the reader's relevant knowledge (i.e., world knowledge, knowledge of text genre, the discourse model built up during text processing) such that a mental model of the situation is formed in relation to what the author means.

3. Working-memory capacity and L2 reading

A number of studies exploring the relationship between working-memory capacity and L2 reading seem to have basically adopted Daneman and Carpenter's (1980) original measure of working memory, with span tasks (reading span, counting span, visuospatial span) used to assess input retention and processing as part of a dual-task paradigm (Chun and Payne, 2004; Harrington and Sawyer, 1992; Leeser, 2007; Walter, 2004). In general, span tasks have been found to be valid (Conway et al., 2005) and reliable (Friedman and Miyake, 2004; Whitney et al., 2001) instruments of working-memory assessment in a variety of fields. As indicated earlier, the underlying notion in a dual-task paradigm is the trade-off expected to emerge between the processing and storage functions of limited working-memory resources. However, span tasks are satisfactory with the one caveat that they focus on storage rather than processing (Waters and Caplan, 1996). Thus, to evaluate both storage and processing, the need arises to make use of composite scores. Unfortunately, the use of composite scores seems to have been overlooked in most L2 research on the subject (e.g., Harrington and Sawyer, 1992; Miyake and Friedman, 1998; Osaka and Osaka, 1992; Osaka et al., 1993), as the emphasis has been placed exclusively on the participant's recall score as an index for what the span task really measures.

Taking recall as the putative index of working-memory capacity, one encounters a fundamental discrepancy in its measurement in L2 research. An overview of the literature shows that differences exist in the type of testing procedure used. While most studies have employed a test where individuals are required to report the sentence-final words that they remember (e.g., Harrington and Sawyer, 1992; Leeser, 2007; van den Noort et al., 2006; Walter, 2004), recognition tasks where test participants select sentence-final words from a set of options have also been used (e.g., Chun and Payne, 2004). Thus, one wonders whether the type of testing procedure affects the conclusions drawn about individuals' storage capacity and, in turn, those drawn about their working-memory capacity in general.

What further necessitates a closer examination of potential links between working-memory capacity and reading comprehension is the generally acknowledged tendency of L2 readers to behave like inefficient L1 readers (Chun and Payne, 2004; Diao and Sweller, 2007; Horiba, 1996; Jonz, 1989; Taillefer, 1996) in that they attend to surface-level and textbase-level features of the text, often at the expense of coming to grips with higher-level conceptual processes of reading. It follows that this over-reliance on the text, which often stems from language proficiency problems, enhances literal understanding while impeding inferential comprehension. The propensity for text-based processing prevents the rapid retrieval of previously acquired or recently formed domain-specific schemas from 'long-term working memory' (Ericsson and Kintsch, 1995), not to mention its inhibitory effect on their cognitively efficient application in working memory.[2] L2 readers' excessive focus on text-bound literal understanding combined with the different nature and degree of the cognitive operations subserving literal versus inferential reading processes make it imperative for L2 reading to be compartmentalized into its principal dimensions of literal and inferential comprehension prior to being used as a dependent variable in working-memory research.

[2] According to Ericsson and Kintsch (1995), working memory comprises two functional components. The first is the 'short-term working memory', which is responsible for updating and manipulating representations, switching and dividing attention between tasks, and selecting relevant information while inhibiting irrelevant data. The second is the 'long-term working memory', which maintains access to relevant representations from long-term memory for retrieval purposes.

4. Research questions

Amidst the controversies surrounding the relationship between working-memory capacity and reading comprehension, the current study aimed to shed light on two specific issues associated with this relationship. The first dealt with the type of reading comprehension process being assessed, i.e., whether the reading task taps literal or inferential understanding. The second involved whether the storage function of the span task is likely to affect construct validity if it is measured through recall or recognition questions, as the use of these typologically different questions may induce response outcomes that shape the construct. It is interesting to note, for example, that Chun and Payne (2004), who used multiple-choice items to measure storage, were unable to find any significant relationship between span scores and reading comprehension in the L2, although positive relationships between the two constructs have been found in recall-based research (e.g., Harrington and Sawyer, 1992; Leeser, 2007; Walter, 2004), possibly due to researchers' use of different task requirements (recognition versus recall) and scoring procedures (storage scores versus composite scores). Based on these considerations, we sought to investigate the following:

1. Is there a relationship between processing and storage tasks of recognition- and recall-based reading span tests?
2. Is there a relationship between recall-based reading span and literal comprehension accuracy?
3. Is there a relationship between recall-based reading span and inferential comprehension accuracy?
4. Is there a relationship between recognition-based reading span and literal comprehension accuracy?
5. Is there a relationship between recognition-based reading span and inferential comprehension accuracy?

Given that, irrespective of their type (reading span, counting span, visuospatial span), span tasks account for the same variance in comprehension (Turner and Engle, 1989), it was decided to make use of a reading span test (RST) in the present study. It was hypothesized that the processing and storage tasks in the recognition-based RST would correlate with the processing task in the recall-based RST but not with the recall-based storage task, due to the identical nature of the cognitive processes underlying recognition (Hypothesis 1). This hypothesis was based on the understanding that both types of RST make use of the same grammaticality judgment test requiring participants to correctly discriminate syntactically accurate sentences from those that are inaccurate, as being representative of the processing function. Similarly, the recognition-based RST also requires participants to correctly discriminate the correct option from the distractors in the given multiple-choice questions. Although this procedure has been taken to tap the RST's storage function (e.g., Chun and Payne, 2004), what it in fact measures is the participant's ability to identify the correct option in relation to alternatives that are, cognitively speaking, 'externally presented retrieval cues' indicative of other recognition tasks rather than participants' internally generated cues for free recall of stored items (Unsworth and Engle, 2007, p. 112). It was further hypothesized that, given the general recall-based RST findings that point to a positive relationship between working-memory capacity and reading performance (e.g., Harrington and Sawyer, 1992; Leeser, 2007; Miyake and Friedman, 1998; Walter, 2004), there would be, in the case of the recall-based RST, a positive relationship between reading span and literal comprehension accuracy on one hand (Hypothesis 2), and inferential comprehension accuracy on the other (Hypothesis 3). Finally, no meaningful relationship was expected between reading span and either literal understanding (Hypothesis 4) or inferential comprehension (Hypothesis 5) in the case of the recognition-based RST, as research using recognition tasks as working-memory measures has failed to show significant correlations with L2 reading comprehension (e.g., Chun and Payne, 2004).

5. Methodology

5.1. Participants

The participants in the present study were Turkish undergraduate students enrolled in an English-medium university in Turkey. 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. Their ages ranged from 20 to 23, with an average of 21.52 years. Of the 30 students who participated in the study, 24 were female and 6 were male. They formed a homogeneous group in terms of their educational background in that they had all completed a teacher training high school and were enrolled in university-level ELT (English Language Teaching) courses in order to become teachers of English.

5.2. Materials and procedures

Materials for the study consisted of two versions of an RST in English (the participants' L2), one with recall-based and the other with recognition-based items designed to measure storage. In addition, a reading comprehension test with multiple-choice items based on a narrative text in English was used.

5.2.1. Reading span test (RST)

The RST in English was a modified version of Daneman and Carpenter's (1980) model, using the procedures described in Harrington and Sawyer (1992) and Chun and Payne (2004). Although the RST used in Chun and Payne was in the participants' native language, the current study made use of an RST in the L2, modeled after Harrington and Sawyer. This was thought to be appropriate in view of research findings that point to working memory in the L2 having a stronger relationship with L2 syntactic comprehension (Miyake and Friedman, 1998) and L2 reading comprehension (Harrington and Sawyer, 1992; Walter, 2004) than with L1 working memory.

A computerized version of the RST was used in the present study. It consisted of 70 unrelated simple sentences in the active voice, each 11–13 words in length. Every sentence ended with a different word. The sentences were divided into 20 sets. The test involved four levels, starting at two and extending up to five sentences, with each level containing five trials. A grammaticality judgment test was incorporated into the RST to ensure that participants processed every sentence syntactically and did not simply focus on the final words. There were 35 grammatical and 35 ungrammatical sentences, arranged randomly. Each sentence appeared only once. All participants were tested on the same sets of sentences. After the participants finished reading all five trials at one level, they proceeded to trial 1 of the next level. The total number of words recalled (or recognized) across all trials was recorded as the measure of the participant's storage capacity, given the concern for the reliability and criterion validity of RST scoring methods (Friedman and Miyake, 2005). The total number of sentences accurately judged, on the other hand, represented the measure of the participant's processing capacity. The maximum possible score for both processing and storage tasks was 70. Finally, storage and processing scores were converted to z-scores and their average was taken in order to obtain composite scores, as suggested by Waters and Caplan (1996).

The RST had two versions which differed in terms of the type of task used for the storage component. One version involved a recall task while the other involved a recognition task. In the recall task, the participants were required to remember the sentence-final words and enter them in any order into a textbox on a computer screen. The recognition task, on the other hand, required the participants to choose the sentence-final words from a list of given options.

The test, administered in a computer lab, was delivered online by displaying one sentence after another in 7-s intervals until all the sentences in a set had been viewed. No reading aloud was done. While processing the sentences silently, the participants pressed one of two computer keys to indicate whether a given sentence was grammatical or ungrammatical. In the recognition version, after all the sentences in a set had been viewed, a list of words to choose from appeared on the screen. Following Chun and Payne (2004), three options were given (two distractors and the correct option) for each word to be recognized as the final word of each sentence. For sets containing five sentences, for instance, 15 words were presented. Distractors came either from nouns belonging to the same semantic category or from sentence-final nouns in previous sets. Participants selected the words which they thought were correct by clicking the checkbox next to the word. In the recall version, on the other hand, they entered the sentence-final words they remembered as soon as the message "Recall" appeared on the computer screen. In both versions, the participants' judgments concerning the grammaticality of the sentences and the number of correct words they remembered were recorded by the computer.

Internal consistency reliability coefficients for the recognition and recall tests were calculated on the storage and processing tasks separately. The Cronbach's alpha on the processing and storage tasks for the recognition test were found to be .634 and .691, respectively, while those for the recall test were .688 and .857.

The participants first took the recognition-based RST. Two months later they were given the recall version of the same test. The reason for the 2-month time interval between the two administrations was to ensure that they had forgotten all about the content of the sentences in the RST.

5.2.2. Reading text

The narrative text used for reading comprehension was an American short story by Delmore Schwartz (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 in the narrative 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.

Schwartz's story was selected for having a universally meaningful theme, taking place in a culturally unfamiliar social context. The plot deals with the thoughts and deeds of a young man with nouveau riche tendencies, who is about to marry a girl from a well-established yet not particularly wealthy family.

5.2.3. Reading comprehension test

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 an equal number of textually explicit questions on one hand, and textually and scriptally implicit questions on the other (10 of each type, a total of 20 points). These were presented in mixed order, following the course of the story. Textually explicit questions measured readers' literal understanding. That is, the answers could be derived directly from the text, based on the comprehension of an explicitly stated proposition or the identification of the relationships between explicit propositions. Textually implicit and scriptally implicit questions, on the other hand, were designed to measure readers' inferential comprehension, as the answers rested on the ability to make either connective or elaborative inferences. In the case of connective inferences, readers were required to integrate implicitly presented textual information and their schematic knowledge (textually implicit processing for local coherence); as for elaborative inferences, readers were required to draw inferences by moving beyond the text to construct a mental model of what it was about (scriptally implicit processing for global coherence). Examples of test items for each category appear in the appendix.

The participants had continuous access to the text on the computer screen and were instructed to read it silently. They were told to answer the comprehension questions that would appear one by one, alongside the text. They were informed that their comprehension accuracy scores would be based on the number of correct answers out of the total number of questions. During the 50 min they were given for the test as a whole, they were free in terms of the processing time for individual questions.

6. Results

6.1. Relationship between processing and storage tasks on the recognition- and recall-based RSTs

Descriptive statistics for the recognition and recall reading span tests, including the composite scores for each test, are provided in Table 1. The column labeled 'processing' shows the data for the sentence judgment task while the column labeled 'storage' shows the data for the total number of words recalled accurately.

The average performance on the processing task is similar across the recognition and recall tests. However, there is a noticeable difference between the two tests in terms of average performance on the storage task; the participants performed better when they were given the recognition-based RST. In terms of variability, the scores on the recognition-based test yielded less variability on both components, indicating that the participants performed similarly. Larger individual differences, on the other hand, are observed on the recall-based test.

Correlations among the processing, storage, and composite scores across the tests are provided in Table 2. Table 2 indicates that there is a moderate positive correlation between the processing and the storage tasks on the recognition-based RST. The positive relationship between the tasks and similar average performance on these tasks suggest that the two tasks may be tapping identical cognitive processes. However, no correlation between the processing and storage tasks is observed on the recall-based RST. The large mean difference

Table 1
Descriptive statistics for the reading span tests.

	Recognition			Recall		
	Processing	Storage	Composite	Processing	Storage	Composite
M	59.66	58.01	−.16	58.05	48.27	−.07
SD	3.68	5.10	.80	5.01	8.62	.57
Min.	53	45	−1.53	44	28	−1.50
Max.	67	67	1.28	67	67	1.54

Table 2
Correlations among the working-memory tasks across recognition- and recall-based tests.

	Recognition processing	Recognition storage	Recall processing	Recall storage	Recognition composite
Recognition storage	.60***				
Recall processing	.60***	.30*			
Recall storage	.01	.23	−.23		
Recognition composite	.80***	.84***	.45**	.14	
Recall composite	.51***	.44**	.64***	.60***	.48***

* $p < .10$
** $p < .05$.
*** $p < .01$.

between the processing and storage tasks (t (29) = 4.90, $p < .001$) and the lack of relationship between them on the recall-based test may point to an independent functioning of these components. In addition, the processing task on the recall-based RST has a significant positive correlation with the processing task on the recognition-based test, and it correlates with the storage task on the recognition-based test at the .10 level. The recall-based storage task, on the other hand, does not correlate with any other task. Finally, the composite scores on the recognition- and recall-based RSTs seem to correlate moderately with each other.

6.2. Correlations with reading comprehension

In order to determine which type of RST was a better measure of working-memory capacity, literal and inferential comprehension accuracy scores were used as criterion measures and their correlations with working-memory scores were obtained.

The participants performed better in understanding literal questions ($M = 7.67$, $SD = 1.34$, $Min. = 5$, $Max. = 10$) than inferential questions ($M = 4.27$, $SD = 1.28$, $Min. = 2$, $Max. = 6$). The large difference between the means of literal and inferential comprehension scores (t (29) = 8.96, $p < .001$) and the low correlation between them ($r = -.246$, $p > .05$) point to different dimensions of reading comprehension. Therefore, correlations between working-memory scores and comprehension accuracy were calculated separately for each dimension of reading.

To examine whether reading span was related to comprehension accuracy, Pearson product–moment correlations were calculated not only between the composite scores but also the storage scores for the recognition-based and recall-based RSTs with literal and inferential comprehension (see Table 3).

Table 3 shows a moderate correlation between composite scores on the recall-based RST and inferential comprehension. Other correlations with composite scores were low or negligible as were those with storage

Table 3
Correlations between working-memory scores and reading comprehension dimensions.

	Recognition		Recall	
	Storage	Composite	Storage	Composite
Literal comprehension	.17	.19	.04	.04
Inferential comprehension	.07	.19	.20	.40*

* $p < .05$.

scores. These findings provide support for Waters and Caplan's (1996) thesis that the measurement of working-memory capacity based simply on performance on the storage component of a complex span task is questionable in terms of the reliability and stability of the measure.

A simple regression analysis was conducted to determine whether composite recall scores predicted any variability in inferential comprehension. The analysis revealed that recall-based composite span scores significantly predicted inferential comprehension ($r^2 = .16$; $F\ (1, 28) = 5.21$, $p < .05$).

These findings confirm our first hypothesis, which predicted that the processing and storage tasks in the recognition-based RST would correlate with the processing task in the recall-based RST but not with the recall-based storage task, due to the identical nature of the cognitive processes underlying recognition. It is likely that recognizing the grammaticality of given sentences and the last word of each sentence from a set of lexical options taps similar or perhaps identical processes of discrimination, namely responding to externally presented retrieval cues. That is, the availability of either the list of sentences or the multiple-choice questions provides the participants with the opportunity to retrieve relevant information from their long-term working memory by means of recognition procedures. On the other hand, remembering the last word of sentences from memory is a case of free recall that involves internally generated retrieval cues whose mission is to focus the search on relevant representations in long-term working memory. It follows that there is a qualitative difference between recall and recognition tasks in terms of the cognitive operations they require. Moreover, free recall tasks may be more demanding than recognition tasks in that item retrieval through the use of internally generated cues is based on the ability to effectively delimit the search process to only relevant information. Conversely, the use of externally presented retrieval cues is likely to facilitate the recovery of information related to the target items in that it delimits the search set naturally.

Our second hypothesis, which posited a positive relationship between recall-based working-memory scores and literal comprehension, was not confirmed. However, our third hypothesis, which predicted a positive relationship between recall-based working-memory scores and inferential comprehension, was confirmed. Finally, we found no relationship between working-memory scores from the recognition-based RST and either dimension of reading comprehension, as expected.

7. Discussion

The present findings suggest that working memory is a significant factor affecting the dependent variable of reading, but only in the case of inferential comprehension and with storage load measured through a recall-based procedure. However, no meaningful relationship emerges when a recognition-based procedure is utilized to measure the storage load, as was also the case with Chun and Payne's (2004) research findings. These results can best be explained in light of the interaction between the qualitatively and functionally distinct components of working memory. It should be noted that short-term working memory, with its limited capacity, actively maintains a distinct number of separate representations for ongoing processing by means of the continued allocation of attention. However, with the switching of attention due to trying to tackle a distracting secondary task or the presence of more incoming information ('cue overload'), item representations can be displaced from short-term working memory. In this case, items must be retrieved from long-term working memory through the use of internally generated cue-dependent search processes that aim at accessing relevant information by means of a variety of retrieval cues (e.g., contextual, categorical, temporal). It follows that both short-term working memory and long-term working memory contribute to performance in free recall, accounting for the meaningful relationship found between the recall-based measurement of the storage load in working-memory capacity and inferential reading comprehension.

However, as suggested by Unsworth and Engle (2007), performance in recognition tasks is driven by two separate mechanisms.[3] One is a fast-acting and fairly automatic process of familiarity, which is said to operate independent of retrieval processes associated with working-memory capacity. The other is the slower and

[3] Our views generally follow Unsworth and Engle's (2007) operational model of working memory, which accounts for the roles played by primary and secondary memories in the performance of recall and recognition tasks. However, we make use of Ericsson and Kintsch's (1995) operational constructs, which comprise the two functional components of working-memory capacity, namely, short-term working memory and long-term working memory.

more controlled recollection process that aims at recovering information about the target item from long-term working memory. In this sense, only in cases where it becomes necessary for a controlled search that uses externally presented cues to identify the target item can recognition procedures play a role in retrieving pertinent information from long-term working memory. Otherwise, it is likely for automatic processes to lead the individual to the target item on account of its familiarity. Hence, the use of recognition-based procedures in RSTs with a view to measuring storage can be inadequate in that the participants' performance could be partially contingent on a strategically controlled search process of long-term working memory or on the automatic retrieval of information from long-term working memory. Whereas short-term working memory is activated in the former procedure, it plays no role in the latter.

Given then the differences in the cognitive processes associated with recognition and recall, the participants' recall-based storage tasks not correlating with any of the other measures in the present study is not surprising. Nor should one find it unusual that there are meaningful correlations between constructs that seemingly appear to be at odds with each other: the participants' storage performance in recognition-based tasks correlates not only with their performance in the processing tasks on the same RST, but also with their performance in processing on the recall-based RST. In fact, these results point to the likelihood that the judgment task on the syntactic plausibility of the RST sentences and the recognition of sentence-final words tap similar, if not identical, computational procedures, as indicated earlier.

The results further suggest that reading comprehension itself, as a dependent variable, should be operationalized and measured in terms of its principal components of literal and inferential reading rather than being treated as a holistic entity. Literal reading requires readers to expend their cognitive resources on less controlled or, at times, automatic processing of textually explicit data, which is relatively less demanding in terms of 'intrinsic' cognitive load (Sweller, 1994), particularly for L2 readers with advanced proficiency. In contrast, inferential reading involves tackling a series of complex tasks associated with textually and scriptally implicit information. It follows that, unlike literal understanding, inferential comprehension necessitates controlled and effortful processing even for proficient L2 learners.

Another interesting issue in the present study involves the lack of a significant relationship between literal comprehension and recall-based L2 working memory. At first, this may appear surprising since a number of researchers, such as Harrington and Sawyer (1992) and Leeser (2007), found meaningful relationships between working-memory capacity and overall reading comprehension. This discrepancy should be expected, however, in view of the important differences in conceptualization and methodological design between the present research and the aforementioned studies. First, unlike the present study, Leeser examined the relationship between working-memory capacity measured in the L1 and text recall in the L2. His use of free recall as a measure of reading comprehension may explain his findings showing a significant relationship between working-memory capacity and text recall, as free recall, with its strong reliance on memory (Koda, 2005, p. 237), is a common denominator of performance on both working-memory tasks and recall-based reading tasks.

In contrast, like the present study, Harrington and Sawyer investigated the relationship between L2 working-memory capacity and L2 text comprehension. Nevertheless, like Leeser, they treated reading comprehension ability as a global construct, bypassing an in-depth understanding of its underlying constituents that seem to be associated with different types and degrees of cognitive activation. Thus, the possible confounding effects of literal understanding in relation to inferential comprehension or, for that matter, inferential comprehension to literal understanding were simply ignored in the assessment of the link between reading ability and working-memory capacity.

Secondly, Harrington and Sawyer, in tune with most of the related research in the L2 field (e.g., Chun and Payne, 2004; Osaka and Osaka, 1992), operationalized working-memory capacity as depending exclusively on storage scores. Leeser, on the other hand, following Waters and Caplan's (1996) suggestions, made use of composite scores representing both the processing and storage components of the RST. Aside from such major variations in working-memory assessment that could shape the construct, none of the L2 studies examining the relationship between working-memory capacity and reading comprehension took account of the differences in processing that characterize recognition versus recall tasks, the majority of the researchers making use of recall-type working-memory tasks while others using recognition-based tasks, as if the types of processing underlying such tasks were one and the same.

8. Pedagogical implications

The findings of the present study have direct implications for the assessment of reading comprehension. The fact that the participants performed significantly better on literal comprehension than inferential comprehension and the lack of relationship between these two dimensions of reading suggest that global measures of comprehension may not be the best indicators of reading performance. Cognizance of these two dimensions of reading could open the way to a more appropriate weighting of test items in that weights could be assigned to questions according to the type and depth of comprehension they require. In this sense, the intrinsic cognitive load of the reading task could be the determining factor, with inferential questions normally receiving more weight than literal ones.

The findings from the recall- versus recognition-based RSTs also have indirect implications for the type of assessment used to measure reading comprehension. Recall and multiple-choice questions are the two most commonly used types of assessment. The current findings indicate that there will be a larger variability in performance due to differences brought about by working-memory capacity in the case of recall rather than recognition tasks. In other words, a task like free recall relies on memory to a greater extent than a recognition task (Koda, 2005, p. 237). Unsworth and Engle (2007) explain why there is greater variability in working-memory performance in the case of recall tasks. They argue that individual differences in working-memory capacity result from the ability to delimit the search set in retrieving information. In performing recall tasks, individuals have to use internally generated cues to delimit the search set, and the search set of low working-memory individuals will include more irrelevant information compared to those with high working-memory capacity. In the case of recognition tasks, the search set will already be delimited by externally given cues. Therefore, individual differences stemming from efforts to delimit the search set will not arise. These findings from working-memory research can be transferred to designing proper reading comprehension measures in that using recognition-based assessment tools may reduce the cognitive load on working-memory capacity imposed by task requirements. In other words, when the inherent difficulty of a complex task (e.g., inferential elaboration) is coupled with the extraneous load brought about by the design of the assessment tool (e.g., free recall), there may be an unnecessary cognitive overload on working-memory capacity (Sweller, 1994). Thus, the use of recognition tasks in reading comprehension assessment can, at least, reduce the detrimental effects of construct- irrelevant factors.

9. Conclusion

The findings of the current study demonstrate that working-memory capacity can be considered a predictor in L2 reading so long as RSTs assess storage through recall-based tasks and measurement makes use of composite scores that account for both storage and processing. In addition, even if working-memory capacity is generally accepted not to be language-specific, the findings point to a possible relationship between the L2 learner's proficiency level and the contribution of working memory to reading comprehension in the L2. In this light, L2 literal reading could be associated with working-memory capacity only for readers with low L2 proficiency levels because of these readers' propensity for text-based processing. Otherwise, the specific construct that needs to be correlated with working-memory capacity can be said to be inferential reading in that, unlike literal reading, it shows a consistently meaningful relationship with working memory.

Limitations of the current study as well as suggestions for future research are worth mentioning. To begin with, this study focused exclusively on the 'central executive' component of Baddeley's (1986, 2000) model of working memory. It should be realized that working-memory capacity comprises but is not limited to the central executive. Thus, the results should be interpreted with the caveat that they are not applicable to all aspects of working memory. Secondly, caution should be exercised in generalizing the results to other learners with different degrees of L2 proficiency. The participants in this study were future teachers of English—adults with a good command of the L2. The fact that their recall-based L2 working-memory scores correlated with L2 inferential reading scores may indicate that high-level questions tapping deep level comprehension and span tasks cannot be distinguished. That is, reasoning beyond the text or integrating several relevant pieces of information across paragraphs requires similar cognitive abilities as those characteristic of span tasks. In this sense, Koda's (2005, pp. 199–200) argument for identical processes underlying both working-memory capacity and reading ability may be valid so far as reading entails chiefly inferencing.

In the same vein, it can be conjectured that the significant effects found between the recall-based working-memory tasks and inferential comprehension on the L2 reading text could be related to the fact that the working-memory tasks were conducted in the participants' L2. It would thus be interesting to replicate the study with L2 learners with different proficiency degrees and age levels, with RSTs administered in both the L1 and L2. It would perhaps be even more interesting if a longitudinal study were undertaken to probe the interactions among L1 and L2 working-memory capacities, the L2 learner's evolving interlanguage system at different stages, and literal and inferential comprehension processes in L2 reading. Such studies could shed additional light on whether working-memory capacity interacts with the degree of language proficiency or is independent of language altogether. They could further show whether literal understanding and inferential comprehension in L2 reading have a relationship of interface or noninterface, depending on the learner's proficiency level in the second language and/or individual differences in relation to the relative contribution of working memory to the reading process. In the final analysis, working-memory capacity seems to be a significant factor distinguishing good and poor readers only if certain conditions are met. Thus, the findings of the present study should be considered exploratory, being suggestive of the need for new directions and areas to be investigated regarding the role of working-memory capacity in reading comprehension in the L2.

Appendix A. Examples of test items

A.1. Sample textually explicit literal question

They are going to Coney Island this afternoon, although my mother really considers such pleasures inferior. She has made up her mind to indulge only in a walk on the boardwalk and to enjoy a pleasant dinner, avoiding the riotous amusements as being beneath the dignity of so dignified a couple.

According to the narrator's mother, what kind of place is Coney Island?

(a) A place where one witnesses riotous amusements.
(b) A place where dignified couples take a walk.
(c) A place where inferior pleasures are avoided.
(d) A place where people enjoy a pleasant dinner on the boardwalk.

A.2. Sample textually implicit inferential question

My father tells my mother how much money he has made in the week just past, exaggerating an amount which need not have been exaggerated. But my father has always felt that actualities somehow fall short, no matter how fine they are. Suddenly, I begin to weep.

Why does the narrator begin to weep?

(a) He is aware of his father's exaggerations to his mother on all matters.
(b) He dislikes his father's talking to his mother on how much money he has made.
(c) He feels that his father should tell his mother about his real earnings.
(d) He knows that his father prefers to run away from actualities even if they are fine.

A.3. Sample scriptally implicit inferential question

What seems to mentally preoccupy the narrator's father with regard to the future?

(a) His popularity
(b) His business
(c) His marriage
(d) His family

References

Andrade, J., 2006. An introduction to working memory. In: Andrade, J. (Ed.), Working Memory in Perspective. Psychology Press, Hove, pp. 3–30.

Andreassen, R., Bråten, I., 2007. Working memory influences on reading comprehension assessed with different measures. In: Paper presented at the 12th EARLI Biennial Conference, Budapest, Hungary.

Baddeley, A.D., 1986. Working Memory. Clarendon Press, Oxford.

Baddeley, A.D., 2000. The episodic buffer: a new component of working memory? Trends in Cognitive Science 4, 417–423.

Baddeley, A.D., 2003. Working memory and language: an overview. Journal of Communication Disorders 36, 189–208.

Brünken, R., Seufert, T., 2007. Presentation modality and working memory capacity. In: Paper presented at the 12th EARLI Biennial Conference, Budapest, Hungary.

Chun, D.C., Payne, J.S., 2004. What makes students click: working memory and look-up behavior. System 32, 481–504.

Conway, A.R.A., Engle, R.W., 1994. Working memory and retrieval: a resource-dependent inhibition model. Journal of Experimental Psychology: General 123, 354–373.

Conway, A.R.A., Engle, R.W., 1996. Individual differences in working memory capacity: more evidence for a general capacity theory. Memory 4, 577–590.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., Engle, R.W., 2005. Working memory span tasks: a methodological review and user's guide. Psychonomic Bulletin and Review 12, 769–786.

Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior 19, 450–466.

Diao, Y.L., Sweller, J., 2007. Redundancy in foreign language reading comprehension instruction: concurrent written and spoken presentations. Learning and Instruction 17, 78–88.

Engle, R.W., 2002. Working memory capacity as executive attention. Current Directions in Psychological Science 11, 19–23.

Ericsson, K.A., Kintsch, W., 1995. Long-term working memory. Psychological Review 102, 211–245.

Friedman, N.P., Miyake, A., 2004. The reading span test and its predictive power for reading comprehension ability. Journal of Memory and Language 51, 136–158.

Friedman, N.P., Miyake, A., 2005. Comparison of four scoring methods for the reading span test. Behavior Research Methods 37, 581–590.

Hambrick, D.Z., Engle, R.W., 2003. The role of working memory in problem solving. In: Davidson, J.E., Sternberg, R.J. (Eds.), The Psychology of Problem Solving. Cambridge University Press, Cambridge, pp. 176–206.

Harrington, M., Sawyer, M., 1992. L2 working memory capacity and L2 reading skill. Studies in Second Language Acquisition 14, 25–38.

Hitch, G.J., 2005. Working memory. In: Braisby, N., Gellatly, A. (Eds.), Cognitive Psychology. Oxford University Press, Oxford, pp. 307–341.

Horiba, Y., 1996. Comprehension processes in L2 reading. Studies in Second Language Acquisition 18, 433–473.

Jonz, J., 1989. Textual sequence and second language comprehension. Language Learning 39, 207–249.

Juffs, A., 2004. Representation, processing and working memory in a second language. Transactions of the Philological Society 102, 199–226.

Just, M.A., Carpenter, P.A., 1992. A capacity theory of comprehension: individual differences in working memory. Psychological Review 98, 122–149.

Kane, M.J., Conway, A.R.A., Hambrick, D.Z., Engle, R.W., 2007. Variation in working memory capacity as variation in executive attention and control. In: Conway, A.R.A., Jarrold, C., Kane, M.J., Miyake, A., Towse, J.N. (Eds.), Variation in Working Memory. Oxford University Press, New York, pp. 21–48.

Kintsch, W., 1998. Comprehension: A Paradigm for Cognition. Cambridge University Press, Cambridge.

Koda, K., 2005. Insights into Second Language Reading: A Cross-linguistic Approach. Cambridge University Press, Cambridge.

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

MacDonald, M.C., Christiansen, M.H., 2002. Reassessing working memory: a comment on Just & Carpenter (1992) and Waters & Caplan (1996). Psychological Review 109, 35–54.

Maehara, Y., Saito, S., 2007. The relationship between processing and storage in working memory span: not two sides of the same coin. Journal of Memory and Language 56, 212–228.

Miyake, A., Friedman, N.P., 1998. Individual differences in second language proficiency: working memory as language aptitude. In: Healy, A.F., Bourne, L.E. (Eds.), Foreign Language Learning: Psycholinguistic Studies on Training and Retention. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 339–364.

van den Noort, M., Bosch, P., Hugdahl, K., 2006. Foreign language proficiency and working memory capacity. European Psychologist 11, 289–296.

Osaka, M., Osaka, N., 1992. Language-independent working memory as measured by Japanese and English reading span tests. Bulletin of the Psychonomic Society 30, 287–289.

Osaka, M., Osaka, N., Groner, R., 1993. Language-independent working memory: evidence from German and French reading span tests. Bulletin of the Psychonomic Society 31, 117–118.

Pearson, P.D., Johnson, D.D., 1978. Teaching Reading Comprehension. Holt, Rinehart and Winston, New York.

Rupp, A.A., Ferne, T., Choi, H., 2006. How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. Language Testing 23, 441–474.

Schwartz, D., 1978. In dreams begin responsibilities. In: Atlas, J. (Ed.), In Dreams Begin Responsibilities and Other Stories. New Directions Publishing Corporation, New York, pp. 1–9.

Service, E., Simola, M., Metsaenheimo, O., Maury, S., 2002. Bilingual working memory span is affected by language skill. European Journal of Cognitive Psychology 14, 383–407.

Sweller, J., 1994. Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction 4, 295–312.

Taillefer, G., 1996. L2 reading ability: further insight into the short-circuit hypothesis. Modern Language Journal 80, 461–477.

Towse, J.N., Hitch, G.J., 1995. Is there a relationship between task demand and storage space in tests of working memory capacity? Quarterly Journal of Experimental Psychology 48A, 108–124.

Towse, J.N., Hitch, G.J., Hutton, U., 2002. On the nature of the relationship between processing activity and item retention in children. Journal of Experimental Child Psychology 82, 156–184.

Traxler, M.J., 2006. Commentary on Clahsen and Felser. Applied Psycholinguistics 27, 95–97.

Turner, M.L., Engle, R.W., 1989. Is working memory capacity task dependent? Journal of Memory and Language 28, 127–154.

Unsworth, N., Engle, R.W., 2007. The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. Psychological Review 14, 104–132.

Walter, C., 2004. Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. Applied Linguistics 25, 315–339.

Waters, G.S., Caplan, D., 1996. The measurement of verbal working memory capacity and its relation to reading comprehension. The Quarterly Journal of Experimental Psychology 49A, 51–79.

Whitney, P., Arnett, P.A., Driver, A., Budd, D., 2001. Measuring central executive functioning: what's in a reading span? Brain and Cognition 45, 1–14.