Psychological Bulletin
2007, Vol. 133, No. 6, 1038–1066

0033-2909/07/$12.00 DOI: 10.1037/0033-2909.133.6.1038

On the Division of Short-Term and Working Memory: An Examination of Simple and Complex Span and Their Relation to Higher Order Abilities

Nash Unsworth
University of Georgia

Randall W. Engle
Georgia Institute of Technology

Research has suggested that short-term memory and working memory (as measured by simple and complex span tasks, respectively) are separate constructs that are differentially related to higher order cognitive abilities. This claim is critically evaluated by reviewing research that has compared simple and complex span tasks in both experimental and correlational studies. In addition, a meta-analysis and re-analyses of key data sets were conducted. The review and analyses suggest that simple and complex span tasks largely measure the same basic subcomponent processes (e.g., rehearsal, maintenance, updating, controlled search) but differ in the extent to which these processes operate in a particular task. These differences largely depend on the extent to which phonological processes are maximized and variability from long list lengths is present. Potential methodological, psychometric, and assessment implications are discussed and a theoretical account of the data is proposed.

Keywords: working memory, short-term memory, complex span tasks, simple span tasks

Individual differences in memory abilities have long interested psychologists and have played an integral role in psychometric batteries of intelligence. Indeed, memory span tasks have been part of intelligence batteries since their inception (e.g., Terman, 1916; see Dempster, 1981, for a review). In these short-term memory (or simple) span tasks, participants are given a list of to-be-remembered (TBR) items including letters, digits, words, or shapes and are then asked to recall the list in the correct serial order immediately after presentation of the last item. For example, in the letter span task, participants who receive the list “R, S, L, Q, T” must correctly recall the letters in their correct serial position. Any deviation (e.g., recalling “S” as the first letter) is counted as an error. Additionally, list length is typically varied such that participants are required to sometimes recall short lists (e.g., two items) and other times recall longer lists (e.g., seven items).

In working memory (or complex) span tasks, such as the simple span ones, participants recall a set of items in their correct serial order. The tasks differ in that complex span requires that participants engage in some processing activity unrelated to the memory task. This activity is interleaved between presentation of the individual TBR items. The processing component can include reading sentences, solving arithmetic problems, or assessing the symmetry of visual objects. For instance, in the operation span task, participants solve math problems while trying to remember unrelated items. A trial in this task may look like:

IS (8/2) − 1 = 1?R

IS (6*1) + 2 = 8?L

IS (10*2) − 5 = 15?S

IS (12/6) + 4 = 10?Q

IS (2*3) − 3 = 3?T

Here, participants are instructed to solve the math problems and remember the letters. At the recall signal (???), they must recall the letters in the correct serial order. Note that the list of TBR items is exactly the same as in the simple span task.

Beginning with research by Daneman and Carpenter in 1980, several studies have shown that complex span tends to correlate higher with measures of higher order cognition than does simple span (e.g., Cantor, Engle, & Hamilton, 1991; Conway & Engle, 1996; Conway et al., 2002; Daneman & Merikle, 1996; Dixon, LeFevre, & Twilley, 1988; Engle, Tuholski, Laughlin, & Conway, 1999; Kail & Hall, 2001; Masson & Miller, 1983; Turner & Engle, 1989). These findings have led some researchers to conclude that working memory (WM) is more important for higher order cognition than is short-term memory (STM). For example:

> The high correlation between reading span and the various comprehension measures is striking; the reading span task succeeds where previous short-term memory measures have failed. (Daneman & Carpenter, 1980, p. 463)
>
> WMC is more closely related with Gf and Reasoning than is short-term memory (STM). (Kane, Hambrick, & Conway, 2005, p. 66)
>
> Thus, the existing evidence, though scanty, is consistent with the hypotheses that WM and STM are distinct but related and that WM plays a greater role than STM in higher-order cognitive processes. (Kail & Hall, 2001, p. 2)

At the same time, other studies have found that simple span correlates nearly as well as complex span with measures of higher order cognition (Bayliss, Jarrold, Baddeley, & Gunn, 2005; Colom, Rebollo, Abad, & Shih, 2006; Kane et al., 2004; La Pointe & Engle, 1990; Mukunda & Hall, 1992; Shah & Miyake, 1996; Unsworth & Engle, 2006a). This suggests that simple and complex

---

Nash Unsworth, Department of Psychology, University of Georgia, and Randall W. Engle, School of Psychology, Georgia Institute of Technology.

We thank Rich Heitz and Tom Redick for helpful comments in all phases of this research project.

Correspondence concerning this article should be addressed to Nash Unsworth, Department of Psychology, University of Georgia, Athens, GA 30602-3013. E-mail: nunswor@uga.edu

span largely measure the same fundamental cognitive processes and whatever is common to the tasks predicts higher order cognitive processes. For example:

> More importantly, the simple word span task also significantly predicted comprehension and, in some cases, did so as well as did the complex span task. (La Pointe & Engle, 1990, p. 1129)
>
> Still another important theoretical implication is that complex span measures (of WM) must not be clearly distinguished from simple span measures (of STM). Both measures share something in common that could produce their *association* with cognitive ability measures. (Colom, Rebollo, Abad, & Shih, 2006, p. 167)
>
> At the very least, the results suggest that complex and simple span tasks should not be dichotomized to simply reflect working memory and short-term memory, but rather all immediate memory tasks require a number of processes which may be important for higher-order cognition. (Unsworth & Engle, 2006a, p. 77)

Comparing these discrepant findings and conclusions, however, is complicated by the fact that studies differ in the nature of their samples (younger adults, children, older adults) and in their presentation and scoring procedures. Furthermore, simple span tasks have been studied extensively with regard to particular experimental manipulations (e.g., word length, phonological similarity), whereas these manipulations have been used less often with complex span tasks. Thus, we suggest that a detailed examination of both experimental and correlational evidence may provide important insights regarding the extent to which simple and complex span assess the same or different constructs.

The examination of potential similarities and differences in simple and complex span (and, therefore, STM and WM) is important from a purely cognitive perspective and is also important for many other research areas that rely on these tasks. As we noted previously, these tasks have long been part of basic psychometric intelligence batteries, as well as neuropsychological assessments. Additionally, these tasks have played an integral role in memory research examining correlations with other abilities, including research in the areas of cognitive aging, developmental psychology, and psychopathology. Finally, these tasks have been used in research domains to make inferences about basic cognitive processes, including social psychology and personality theory. Thus, determining the extent to which memory span tasks measure the same or different constructs is an important endeavor for several areas of research.

In the current article, we argue that simple and complex span largely measure the same basic processes (e.g., rehearsal, maintenance, updating, controlled search) but differ in the extent to which these processes operate in a particular task. When we talk about the extent to which processes operate in a task, we refer to the notion that tasks are not process pure; rather, they reflect a combination of processes. For instance, two different tasks may involve the same set of processes; however, a particular process (or combination of processes) may affect performance more on one task than on the other. Furthermore, we argue that a number of factors, including the nature of the sample, the scoring method, and the administration procedures, can influence the extent to which simple and complex span predict measures of higher order cognition. In particular, we argue that simple and complex span predict higher level cognition similarly when performance from long list lengths is measured (especially in simple span tasks) and the role of phonological rehearsal is reduced. We begin with a brief review of the relevant literature. We then review similarities and differences between simple and complex span in experimental designs. This is followed by a section describing similarities and differences in how well simple and complex span predict measures of higher order cognition with an emphasis on the role of different scoring procedures.

The Division of Short-Term Memory and Working Memory

In order to address similarities and differences between STM and WM, we must first place these concepts in a theoretical context (see Engle & Oransky, 1999, for a review). Both concepts are cast in frameworks distinguishing information that is utilized over the short term from information that is utilized over the long term. Initially, STM was conceptualized as a somewhat passive repository of information before transfer to long-term memory (LTM). In modal models of memory, STM was limited in capacity, and information was maintained through rehearsal (primarily verbal). If not rehearsed, information was rapidly lost. Specific models of STM associated it with many important control processes including rehearsal, coding, organization, and retrieval strategies (Atkinson & Shiffrin, 1968, 1971). Atkinson and Shiffrin (1971) suggested that these control processes were important for coordinating the many subcomponent processes needed to process new information and to retrieve relevant old information. This conceptualization placed STM at the forefront of explaining complex cognitive activities.

Despite a wealth of data supporting a division between STM and LTM and the development of explicit models arguing for the importance of STM in cognition, it soon became clear that STM, as initially conceptualized, was overly simplistic with respect to its role in higher order cognitive functions such as reading and reasoning. With this limitation clearly in mind, Baddeley and Hitch (1974) reconceptualized STM as a more dynamic system that was important for storing information over the short term and for flexible cognitive tasks in which storage and manipulation are both required. In a series of experiments, Baddeley and Hitch demonstrated that participants could briefly store some information while processing or manipulating aspects of other information. Consequently, they argued for a WM system that was responsible for storage and many other cognitive operations simultaneously. Thus, WM included stores dedicated to briefly retaining verbal or spatial information and a general purpose processor involved in coordinating and manipulating information and in monitoring ongoing processing.

While Baddeley and Hitch (1974) were arguing for a dynamic WM memory system in the cognitive literature, neo-Piagetians were making similar arguments in the developmental literature (see, e.g., Case, Kurland, & Goldberg, 1982; Pascual-Leone, 1970; Pascual-Leone & Johnson, 2004). Like Baddeley and Hitch (1974), Case, Kurland, and Goldberg (1982) argued for a distinction between storage (storage space) and processing abilities (operating space) in order to account for developmental increases in memory span performance. Thus, the original conception of STM was replaced by a more dynamic conception of memory that combined storage and processing/manipulation/integration. Following Baddeley and Hitch's (1974) lead (see also Atkinson & Shiffrin, 1968, 1971; Miller, Galanter, & Pribram, 1960), WM has been seen as distinct from STM ever since.

In addition to basic theoretical differences between STM and WM, researchers also became interested in distinctions between putative measures of the two constructs and the extent to which some measures predicted higher order cognitive functioning better than others. A number of researchers argued that measures of STM should correlate fairly well with measures of reasoning, reading comprehension, and other higher order abilities. However, several researchers found that this was not always the case. For example, simple span tasks typically did not correlate well with measures of reading ability (see, e.g., Perfetti & Lesgold, 1977). Noting this apparent paradox (i.e., STM was prominent in models of complex cognition but failed to correlate with a complex task such as reading), Daneman and Carpenter (1980) suggested that simple span does not correlate with reading ability because simple span primarily measures the storage aspect of STM and does not adequately measure the processing aspects that are emphasized in models of WM. In order to test this hypothesis, Daneman and Carpenter devised a task that required a trade-off between storage and processing and captured the dynamics of the whole WM system rather than just one component. In the initial reading span task, Daneman and Carpenter instructed participants to read a series of sentences and recall the last word of each at the recall prompt. This task was contrasted with a simple word span task in which participants remembered a series of unrelated words. Both tasks theoretically measured storage abilities, but only the complex reading span task adequately measured the processing component. Daneman and Carpenter found that reading span correlated more highly with several measures of reading comprehension (including Verbal Scholastic Aptitude Test [SAT] performance) than did simple word span, in line with the argument that processing and storage are more important in higher order cognition than in storage alone.

For the most part, subsequent research has supported these initial findings, suggesting that complex span generally correlates with higher order abilities better than simple span. However, as noted previously, some researchers have reported significant and sometimes sizable correlations between simple span and performance on other cognitive measures. Part of the problem, as we discuss later, may be due to unreliability in traditional measures of simple span (see also Dempster, 1981). Thus, although there is evidence suggesting a distinction between simple STM span tasks and complex WM span tasks in both experimental and differential literatures, there is also evidence suggesting that these two types of tasks measure very similar processes and abilities.

Experimental Effects in Simple and Complex Span

One approach to determining whether simple and complex span tasks measure the same or different theoretical constructs is to identify dissociations in the influence of experimental variables. Here, we review how several experimental variables affect performance on simple and complex span tasks.

Serial Position Functions

Serial position functions, in which proportion correct for each item is plotted as a function of the item's presentation position, have played an important role in theories of memory. These functions have three prominent features: good performance for early items (primacy effects), poor performance for mid-list items, and good performance for late items (recency effects). The size of primacy and recency effects and the relative dominance of one over the other vary from task to task. For instance, small primacy and large recency effects are found in immediate free recall (in which items can be recalled in any order). In contrast, strong primacy and weak recency effects are found in immediate serial recall (in which output and input position need to match from first to last item), although this varies depending on the presentation modality. Additionally, immediate backward serial recall (in which input and output position must correspond from last to first item) produces strong recency effects and primacy effects that are somewhat similar to those in immediate free recall. Thus, serial position functions change systematically depending on how recall is structured. These systematic effects have been instrumental in thinking about distinctions among memory structures and memory functions. Both simple and complex span require immediate serial recall; however, they may have different serial position functions. If so, this would support the claim that STM and WM are somewhat distinct constructs. Typical serial position functions for simple and complex span are shown in Figure 1. Note that items were presented visually and recall was in the forward direction.

Performance is higher on simple span than complex span, and the serial position curves are notably different. The recency effect is small in simple span (as is typical with visual presentation of items) but is pronounced in complex span. In fact, the recency effect in complex span is similar to that found in simple span when backward serial recall is required. This suggests that slightly different processes may operate in the two tasks.

List-Length Effects

The list-length effect refers to the finding that the proportion of items recalled in the correct serial position decreases as a function of list length (although the absolute number of words recalled tends to increase; Murdock, 1962; Roberts, 1972). Previous research has demonstrated list-length effects in a number of paradigms, including free recall, serial recall, and recognition. Most explanations of the effect suggest that proportion correct decreases as a function of list length because of competition among items sharing the same cue (i.e., cue-overload; Watkins, 1979). In a previous study, we (Unsworth & Engle, 2006a) examined list-length effects in simple and complex span. The results are shown in Figure 2. Both simple and complex span show a clear list-length effect, but it is more pronounced in complex than in simple span (see Unsworth & Engle, 2006a, for an interpretation of these results based on differential recall from primary and secondary memory in simple and complex span tasks).

Both tasks also demonstrate an increase in the mean number of items recalled as list length increases, as shown in Figure 3. Thus, both aspects of the list-length effect—decrease in proportion correct and initial increase in number of items recalled (see, for example, Beaman, 2006)—are found in simple and in complex span. This suggests that similar processes operate in the two tasks, although the extent to which they do so likely differs (see Unsworth & Engle, 2006a, 2006b).

Recall Errors

Although proportion correct is the main indicator of performance in many studies, the type and frequency of recall errors are

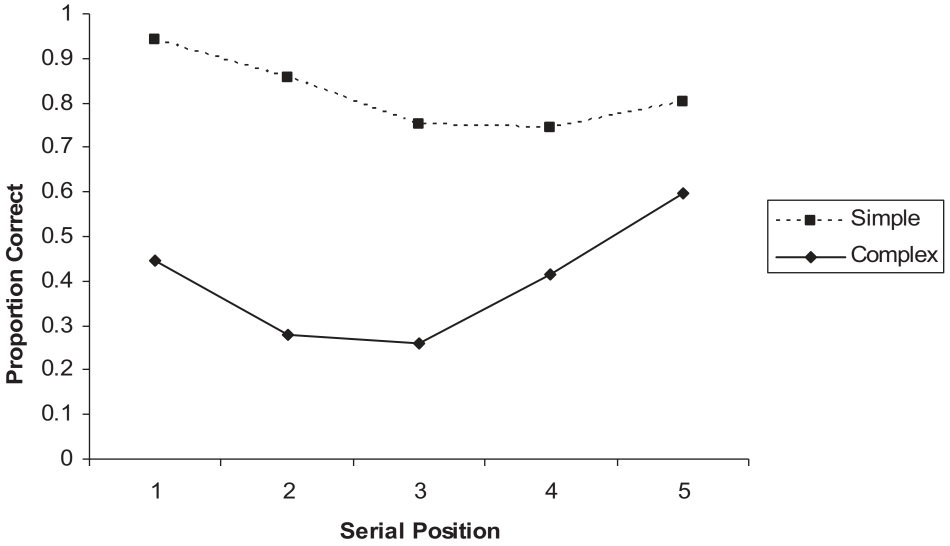

Figure 1. Proportion correct as a function of serial position and memory span task for a list length of five items. Simple span data are an average of word and letter span. Complex span data are an average of operation and reading span. The data are from Kane et al. (2004).

an additional source of evidence about the processes that influence performance (e.g., Bjork & Healy, 1974; Conrad, 1964; Estes, 1972; Henson, 1998). Errors in serial recall tasks are typically classified as item errors or order errors. Item errors include forgetting an item completely (omission) or recalling an item not presented on the current list (intrusion). Order errors occur when the correct item is recalled but in the incorrect serial position (transposition). Previous research has shown that the frequency of these error types changes as a function of serial position and individual differences (e.g., Maylor, Vousden, & Brown, 1999).

The error types for simple and complex span are shown in Figure 4. Transposition errors in simple span tend to be most frequent, followed by omissions, and then intrusions. In complex span, however, omissions are most frequent, followed by transpositions, and intrusions. Thus, order errors are most prevalent in simple span, but item errors (in particular omissions) are most prevalent in complex span. This suggests that simple and complex span differ somewhat in the processes that occur at recall.

In addition to examining error frequency in span tasks, we can examine another important characteristic of errors. When a transposition error is made, the item is typically placed in a position adjacent to the correct one (e.g., Brown, Preece, & Hulme, 2000; Henson, 1998). Some researchers have argued that this is a ubiquitous finding in studies of serial order memory (Brown et al., 2000). If so, then simple and complex span should have similar transposition gradients. Transposition gradients for simple and

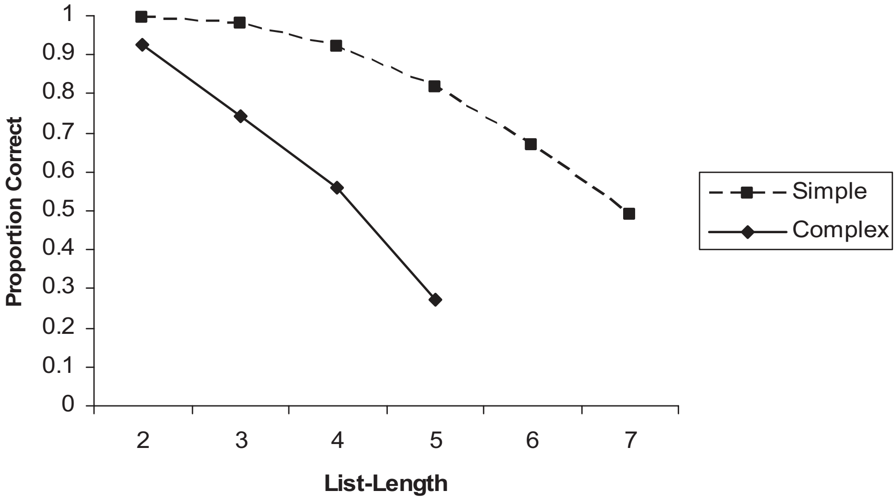

Figure 2. Proportion correct as a function of list length and memory span task. Simple span data are an average of word and letter span. Complex span data are an average of operation and reading span. The data are from Unsworth and Engle (2006a).

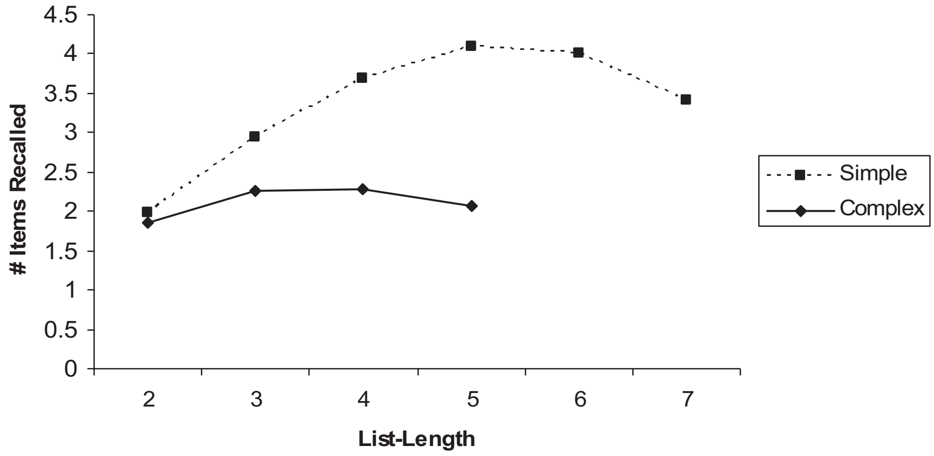

Figure 3. Number of items recalled as a function of list length and memory span task. Simple span data are an average of word and letter span. Complex span data are an average of operation and reading span. The data are from Unsworth and Engle (2006a).

complex span are shown in Figure 5. Both tasks have similar gradients, with most transpositions occurring in a position immediately adjacent to the correct one. Note that the gradient for simple span is based on a list length of six items, whereas the gradient for complex span is based on a list length of five items. These results suggest that similar processes are likely to operate in simple and complex span at the time of recall.

Word-Frequency Effects

Word-frequency effects refer to the finding that recall is better for lists composed of high-frequency than of low-frequency words in both free and serial recall (e.g., Baddeley & Scott, 1971). This finding has led some researchers to conclude that LTM processes influence performance on simple span tasks and, thus, simple span is not a pure measure of STM capacity (e.g., Hulme et al., 1997; Watkins, 1977). Thus, these effects have recently played a role in constraining various theories of short-term recall and especially simple span tasks.

Numerous studies have examined word-frequency effects in simple span, but the effect has been examined less so in complex span. To our knowledge, only one study (Engle, Nations, & Cantor, 1990) has specifically examined word frequency effects in both simple and complex span tasks. The results from Engle et al. (1990) for high- and low-frequency words in both simple and complex span are shown in Figure 6. Word frequency effects were present in both tasks. Recall was better for low-frequency than for high-frequency word lists. Additionally, word frequency effects were more pronounced in simple span than in complex span.

Thus, once again, experimental effects are similar in simple and complex span tasks, but the effects are more pronounced in one task compared with the other. Considerable debate has arisen as to how these effects should be interpreted in simple span, with some

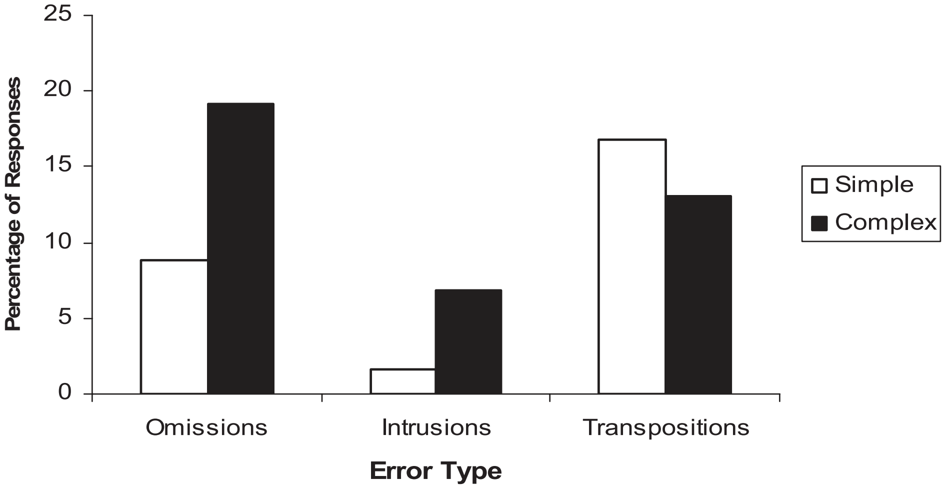

Figure 4. Mean percentage of errors as a function of task and error type. Simple span = letter span visual–young condition. Complex span = combination of operation and reading span. Simple span data are from Maylor et al. (1999). Complex span data are from Unsworth and Engle, 2006b).

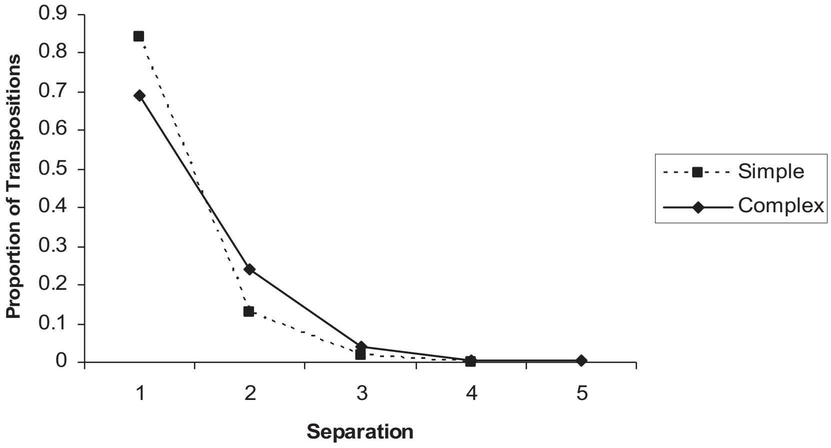

Figure 5. Mean number of transposition errors as proportion of total transposition errors as a function of task and separation. Simple span data are from young adults responding to a list length of six in Experiment 1 of McCormack et al. (2000). Complex span data with a list length of five are from Unsworth and Engle (2006b).

researchers arguing for a reconstruction process operating at retrieval (e.g., Hulme et al., 1997) and other researchers arguing for a rehearsal explanation (e.g., Wright, 1979). Although it is not yet clear which interpretation is accurate, these results suggest that both views must account for the presence of word-frequency effects in complex span tasks and for the fact that the effect seems to be smaller in complex span than in simple span.

Word-Length Effects

One prominent view of performance on simple span tasks is that individuals either overtly or covertly rehearse the presented items verbally and recall is determined by the extent to which the items can be rehearsed (Baddeley, 1986). This view suggests that any variable that reduces rehearsal should impair performance. With this notion in mind, Baddeley, Thomson, and Buchanan (1975) found that lists composed of long words were recalled worse than lists composed of short ones. This word-length effect has been used as the primary argument for the notion that individuals rely on rehearsal processes to recall items in the correct serial order in simple span tasks. At the same time, however, other researchers have argued for different interpretations of word-length effects, explanations that do not rely on rehearsal processes (see, e.g., Hulme et al., 1997). Thus, several researchers have wondered to what extent these effects are present in complex span tasks and whether they are of the same magnitude (e.g., La Pointe & Engle, 1990; Tehan, Hendry, & Kocinski, 2001).

The results from a study conducted by La Pointe and Engle (1990) examining word-length effects in simple and complex span tasks appear in Figure 7. Word-length effects are present in both simple and complex span but, like word-frequency effects, were more pronounced in the simple span than in the complex span tasks. Additionally, performance in the long-word-length condition was fairly similar for both simple and complex span. That is, performance was better in simple span than complex span (as is seen with lists composed of short words), but performance was nearly identical for simple and complex span when lists were composed of long words. Thus, whatever processes are hindered

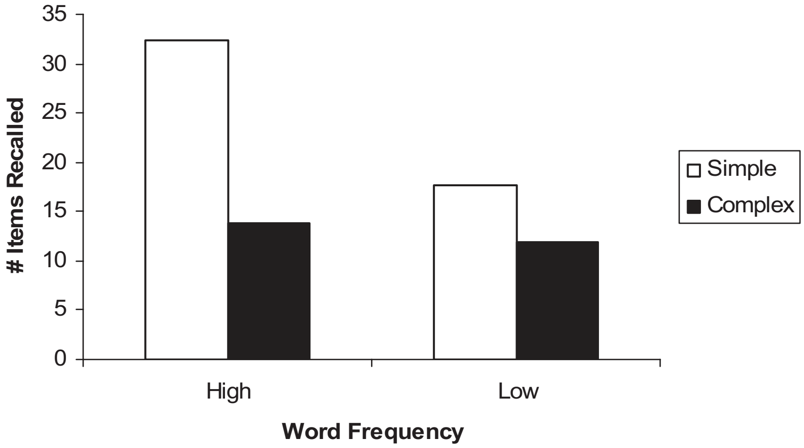

Figure 6. Mean number of items recalled as a function of word frequency and span task. Simple span data are from word span; complex span data are from operation span. Data are from Engle, Nations, and Cantor (1990).

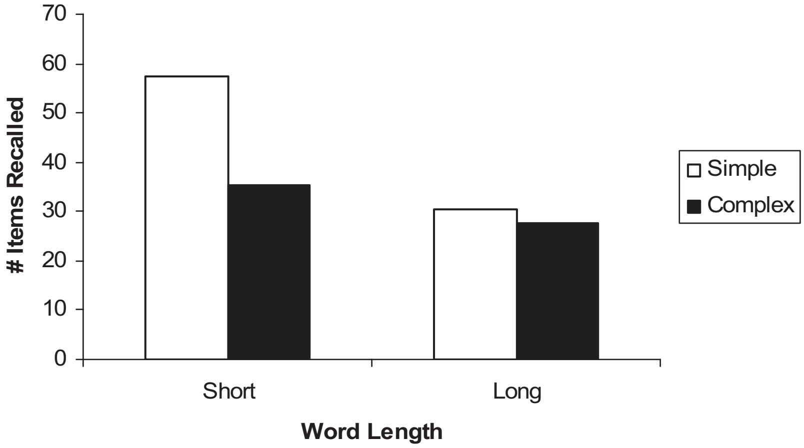

Figure 7. Mean number of items recalled as a function of word length and span task. Data are an average from Experiments 1 (reading span) and 2 (operation span) of La Pointe and Engle (1990).

by the inclusion of long words seem to increase the similarity of performance on simple and complex span tasks.

Similar word-length effects in simple and complex span (see also Tehan, Hendry, & Kocinski, 2001) suggest that similar processes operate in both tasks, although as mentioned previously, the extent to which these processes are involved in each task likely differs. Indeed, La Pointe and Engle (1990) commented, “This calls into question any theory that the two tasks measure completely different aspects of temporary memory such as storage and processing” (p. 1122).

Articulatory Suppression Effects

Another finding that has been used to argue for the role of rehearsal processes in simple span tasks is articulatory suppression. This effect refers to the finding that a repetitive articulation task (such as repeating “the”) typically worsens performance as compared with a control condition. According to a basic rehearsal explanation, performance declines because participants are unable to subvocally rehearse the items; thus, the items are forgotten (likely because of decay). However, other researchers have argued that articulatory suppression may affect recall performance as a result of a mechanism other than rehearsal (see, for example, Neath & Surprenant, 2003). In any case, articulatory suppression effects have been demonstrated numerous times in simple span tasks; thus, models of serial order recall must explain them. If performance on simple and complex span tasks is driven by fundamentally different processes, then articulatory suppression may affect performance in simple, but not in complex, span. However, if the tasks largely measure the same basic processes, then performance should be similarly affected by articulatory suppression manipulations. Relevant data from La Pointe and Engle (1990) are shown in Figure 8. As with word length effects, articulatory suppression reduced performance in both simple and complex span, but the effect was more pronounced in the simple than in the complex span task.

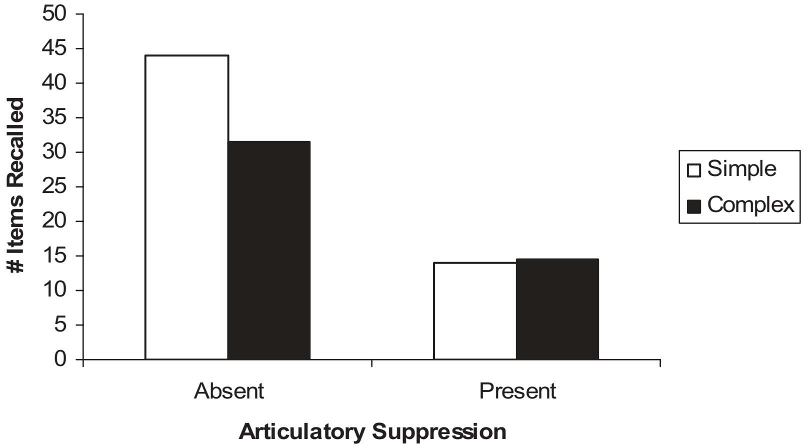

Figure 8. Mean number of items recalled as a function of presence versus absence of articulatory suppression and span task. Absent data are an average of long and short words from Experiments 1 (reading span) and 2 (operation span), and present data are an average of long and short words from Experiment 3 of La Pointe and Engle (1990).

Additionally, the recall advantage typically seen in the simple span task was absent when articulatory suppression was present and recall was equivalent for both simple and complex span. This suggests, in conjunction with word-length effects, that whatever process is affected by these manipulations (e.g., rehearsal) leads to similar performance on simple and complex span tasks. We discuss this point further below.

Phonological Similarity Effects

Several studies have found that lists composed of phonologically similar items (e.g., BDECG) are harder to recall than lists composed of phonologically dissimilar items (e.g., WXFHR) (Baddeley, 1966; Conrad, 1964). This finding has been used to argue that STM relies on an acoustic code (Baddeley, 1966; Conrad, 1964) and that participants actively rehearse items to ensure accurate recall (see, e.g., Baddeley, 1986). Phonological similarity effects have been extensively studied in simple span but not in complex span. To our knowledge, only three studies have examined phonological similarity effects in complex span (Copeland & Radvansky, 2001; Lobley, Baddeley, & Gathercole, 2005; Tehan et al., 2001). Although differing slightly in methods and overall results, all studies showed a phonological similarity effect on complex span tasks. The results from Copeland and Radvansky (2001) are shown in Figure 9. The simple span results are an average of word data from their Experiments 1 and 2, and the complex span results are operation span data from their Experiment 3.

Phonologically dissimilar items were recalled better than phonologically similar items, and the effect was largely the same in both simple and complex span tasks. Note that Copeland and Radvansky (2001) found a phonological similarity facilitation effect for the reading span task in which sentences composed of phonologically similar items led to better retention than sentences composed of phonologically dissimilar items. Copeland and Radvansky argued that participants used the whole sentence to recall sentence-final words (see also Cowan et al., 2003); thus, they were aided by similarity when sentences were composed of phonologically similar words. The operation span task, however, consisted of math operations and phonologically similar or dissimilar words. Thus, it was more like a simple span task. This is why only the operation span data from their Experiment 3 is examined in the current article. Regardless, the results indicate that phonological similarity effects appear in both simple and complex span tasks, reinforcing the notion that similar processes operate in both tasks.

Summary and Conclusion of Experimental Effects

Examining the influence of experimental variables on simple and complex span produced a number of interesting findings. First, both tasks are similarly affected by a number of experimental variables including articulatory suppression, word length, and list length, although performance is typically better in simple than in complex span. Furthermore, effects tend to be more pronounced in simple than in complex span. That is, experimental variables like phonological similarity and word length disrupt performance on simple span tasks more than on complex span tasks. This disruption serves to equate overall levels of performance. An examination of these effects suggests that most of them probably disrupt phonological processes (coding, rehearsal) that operate to a greater extent in simple than in complex span.

Greater reliance on phonological processes in simple than in complex span may also explain differences in serial position functions and frequency of errors. Many of the variables that presumably affect phonological processes (articulatory suppression, phonological similarity) can sometimes affect order errors more than item errors, suggesting that phonological representations may be important for recall of order information (e.g., Nairne & Kelley, 2004; Tehan, Hendry, & Kocinski, 2001). Thus, if phonological information is more important in simple than in complex span, one would expect more order errors in the former than the latter, which is precisely the case. Participants recall many of the TBR items in simple span, but not always in the correct serial position. Additionally, if complex span tasks rely less than simple span tasks on phonological information for recall, then problems in recall may be due to loss of item information, resulting

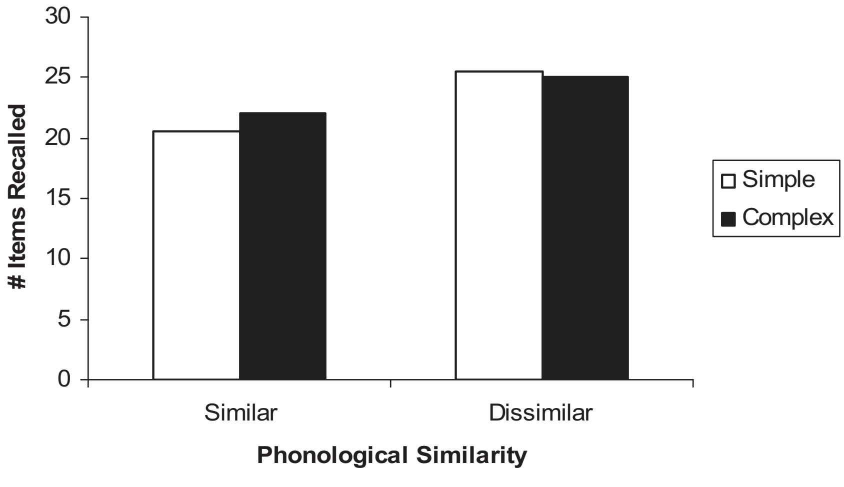

Figure 9. Mean number of items recalled as a function of phonological similarity and span task. Simple span data are an average of word span results from Experiments 1 and 2, and complex span data are operation span results from Experiment 3 of Copeland and Radvansky (2001).

in more item errors (omissions and intrusions) in complex than simple span, which is again the case. This differential involvement of order and item information in simple and complex span may be one reason why they demonstrate slightly different serial position functions and load on separate factors in factor analytic studies. Furthermore, simple span begins to resemble complex span in terms of overall performance when phonological processes are disrupted. This suggests that simple and complex span tasks likely measure the same processes but differ in the extent to which these processes operate in a particular task. Specifically, phonological processes may be more prominent in simple than in complex span, and other processes (e.g., temporal–contextual search) may be more prominent in complex than in simple span. However, all processes likely operate in both tasks. Thus, we have reached a conclusion very similar to that of Tehan, Hendry, and Kocinski (2001), who noted, "When it comes to performance on simple and complex span tasks, the data clearly indicate that both tasks are supported by a common storage system" (p. 346). As we discuss later, the common storage system likely involves a number of subcomponent processes (e.g., rehearsal, maintenance, updating, controlled search) that are needed in both simple and complex span tasks.

Simple and Complex Span as Predictors of Higher Order Cognitive Abilities

The review provided above suggests that simple and complex span are similarly affected by experimental manipulations (such as list length and articulatory suppression) and likely involve similar processes. Indeed, this claim is bolstered by the fact that simple span is almost always moderately or strongly correlated with complex span (Colom, Shih, Flores-Mendoza, & Quiroga, 2006; Conway et al., 2002; Engle et al., 1999; Kane et al., 2004). Thus, it is parsimonious to conclude that the two tasks largely measure the same construct (e.g., Colom, Shih, Flores-Mendoza, & Quiroga, 2006). However, the real debate is not necessarily about whether or not the two tasks measure common processes, but rather whether one task (complex span) measures processes over and above those measured in the other task (simple span). If so, is this why complex span predicts performance on many measures of higher order cognition better than simple span? As noted previously, many researchers (including us) have answered this question in the affirmative. Numerous studies have suggested that complex span is better than simple span at predicting higher order cognition (e.g., Ackerman, Beier, & Boyle, 2005; Daneman & Merikle, 1996); thus, researchers have concluded that WM is more important in complex cognitive activities than is STM. This conclusion, however, seems to fly in the face of decades of research suggesting that STM does contribute to higher order cognitive processes. As mentioned previously, simple span tasks have long been a part of IQ batteries, and previous research has shown that they are at least moderately correlated with measures of higher order cognitive functioning (see, e.g., Beier & Ackerman, 2004; Carroll, 1993).

Given these inconsistencies, we ask why this is the case. One explanation comes from a study that we recently conducted examining list-length effects in both simple and complex span and the extent to which memory span performance correlates with measures of higher order cognition (fluid abilities) as a function of list length (Unsworth & Engle, 2006a). Specifically, we examined list length functions for two complex span tasks (operation span and reading span) and two simple span tasks (word span and letter span) with list lengths ranging from 2 to 5 in complex span and ranging from 2 to 7 in simple span. The accuracy results were shown previously. The correlations between memory performance and fluid abilities as a function of span task and list length are shown in Figure 10.

Interestingly, the data indicate that the correlation between complex span and fluid abilities does not change as a function of list length but the correlation between simple span and fluid abilities does change. Specifically, the correlation between complex span and fluid abilities is approximately .40 across all lengths, but the correlation between simple span and fluid abilities increases with list length and eventually reaches a similar magnitude. Thus, all list lengths in complex span correlate moderately well with measures of higher order cognition, but only supraspan lengths do so in simple span. Additionally, partial correlation analyses suggested that all lengths in complex span and supraspan lengths in simple span share similar processes, although each

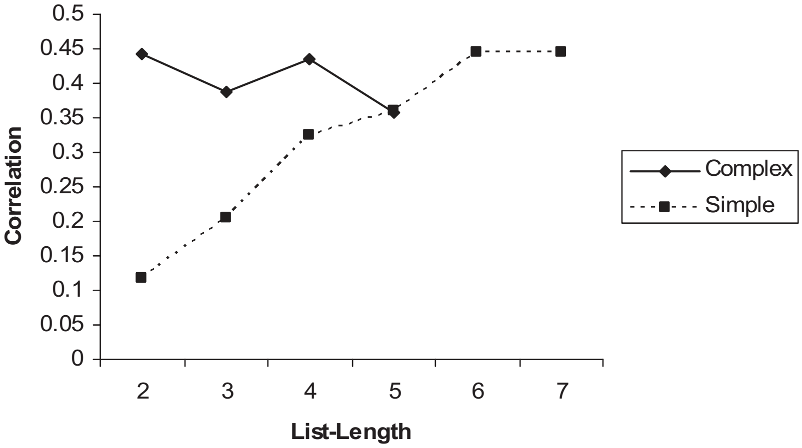

Figure 10. Zero-order correlations between memory span and fluid abilities as a function of span task and list length. The data are from Unsworth and Engle (2006a).

accounted for a small amount of unique variance in fluid abilities. Previous studies examining the differential predictive power of simple and complex span tasks may not have measured supraspan list lengths, leading to differences in the relation between simple and complex span and higher order cognition.

If this is, indeed, one of the primary reasons that simple span tasks typically have lower and less consistent correlations with measures of higher order cognition than complex span tasks, then why have supraspan list lengths been excluded from the data? Several reasons include the nature of the sample, the types of tasks used to measure simple and complex span, the administration procedures, and the scoring methods. Here, we explore how scoring methods may affect the predictive utility of span tasks.

Although correct responses in simple and complex span tasks can be scored in several different ways (see, e.g., Conway et al., 2005; Friedman & Miyake, 2005), we focus on two of the most popular methods. The first method is the "absolute" scoring (ABS) procedure in which a participant's score is the sum of all perfectly recalled list lengths. For example, if an individual recalls 2 items in a list of 2, 3 items in a list of 3, and 3 items in a list of 4, then his or her absolute span score is 5 (2 + 3 + 0). We have used this method extensively in both quasi-experimental and large-scale correlational designs (e.g., Engle et al., 1999; Kane, Bleckley, Conway, & Engle, 2001). The second scoring procedure is proportion correct scoring (PCS). This score is the proportion of items that a participant recalls in the correct serial position. This method is used predominantly when examining immediate serial recall of supraspan lists. Recently, this procedure has been used by a number of researchers (including us) to examine individual differences in complex span (Friedman & Miyake, 2004; Kane et al., 2004; Unsworth, Heitz, Schrock, & Engle, 2005).

One reason that we switched to the PCS method is because it demonstrates better psychometric properties than does the ABS method (for detailed discussions, see Conway et al., 2005; Friedman & Miyake, 2005). Another reason to prefer the PCS method is that it results in slightly higher correlations with criterion measures than does the ABS method. For instance, in examining different scoring procedures for the reading span task, Friedman and Miyake (2005) found that the PCS method, which they called "proportion words," correlated slightly higher with Verbal SAT scores than did the ABS method, which they called "correct set words," (*r*s = .44 and .40, respectively). Friedman and Miyake suggested that one reason for the discrepant correlations was that outliers might be more influential in one scoring method than in the other. Therefore, Friedman and Miyake reanalyzed the data with the outliers removed but still found a difference in the correlations; in fact, the difference was slightly larger than before (*r*s = .46 and .35, respectively). On the basis of these findings, Friedman and Miyake (2005; see also Conway et al., 2005) suggested that the PCS method may be more sensitive to subtle individual differences than the ABS method; thus, it correlates with the criterion measure more highly. With this we completely agree, albeit for a slightly different reason. We argue that the PCS method correlates higher with criterion measures than does the ABS method because of a part–whole relation (e.g., Carroll, 1988; J. Cohen & Cohen, 1983). That is, the PCS method contains the same information as the ABS method plus additional information from items on lists that were not perfectly recalled. This additional information may be an important source of variance that is not included when the ABS method is used. It may include individual differences in the ability to effectively retrieve items from supraspan lists or individual differences in the types of errors that individuals make on the span tasks.

Additional limitations of the ABS method are apparent when considering some of the experimental effects discussed previously. Specifically, one cannot examine either serial position functions or error responses using the ABS method because only trials in which all items are recalled in the correct position are included. Any information provided by differences in serial position functions or recall errors is effectively lost when using the ABS procedure. Thus, the binary nature (correct vs. incorrect) of the method reduces the overall level of variability because variability from the longest lists is excluded. The continuous nature of the PCS method includes variability from all list lengths; thus, the PCS is more likely than the ABS method to reveal subtle individual differences.[1]

Meta-Analysis of the Predictive Power of Simple and Complex Span

Much of the research demonstrating a distinction between simple and complex span in terms of predicting measures of higher order cognition has generally relied on something similar to the ABS procedure. For instance, Engle et al. (1999) and Conway et al. (2002) used the ABS procedure, and both found that a latent variable composed of variance common to complex span tasks predicted fluid abilities better than a latent variable composed of variance common to simple span tasks. In contrast, studies suggesting similarities between simple and complex span have tended to rely on something similar to the PCS method (e.g., Unsworth & Engle, 2006a). In order to examine precisely how simple and complex span relate to measures of higher order cognition, we conducted a meta-analysis. Specifically, we examined correlations between span tasks and measures of higher order cognition (verbal and spatial abilities) for both the ABS and the PCS scoring methods.

Literature Search

The studies included in our meta-analysis were a subset of those used in previous meta-analyses plus several recently published studies. A total of 22 studies met our criteria for inclusion (see below); 13 were used in previous meta-analyses (see, e.g., Ackerman, Beier, & Boyle, 2005; Daneman & Merikle, 1996; Mukunda & Hall, 1992) and 8 were recently published studies (Colom, Abad, Rebollo, & Shih, 2005; Hambrick, 2003; Kane et al., 2004; Lustig, Hasher, & May, 2001, young adults only; Mackintosh & Bennett, 2003; Süβ et al., 2002; Unsworth & Engle, 2005; and Unsworth et al., 2005). Additionally, we included one study (Crawford & Stankov, 1983) that had not appeared in previous meta-analyses but met our inclusion criteria. (All of the studies that were included in our meta-analysis are designated as such in the References section.)

Criteria for inclusion. Several excellent meta-analyses of relations among WM, STM, and higher order cognition have already been published (Ackerman et al., 2005; Daneman & Merikle,

---

[1] We thank Oliver Wilhelm for initially drawing our attention to the merits of the PCS method.

1996; Mukunda & Hall, 1992). Thus, we restricted our focus to studies that specifically examined simple and/or complex span tasks in young adults. We did not examine other putative WM or STM tasks, such as running memory span, memory updating, and mental counters. This was done because the specific question in this article is whether complex and simple span (as measures of WM and STM, respectively) represent similar or different constructs. Additionally, only studies with young adults (typically ages 17 to 35 years) were included because of differences that may arise when examining memory span performance in children or old adults. Additionally, as mentioned previously, studies were included only if it was clear that either the ABS or the PCS method was used to score the data. Using these criteria, several prominent WM studies were excluded, among these, Daneman and Carpenter (1980) and Turner and Engle (1989). Finally, reading comprehension and spatial/matrix reasoning have been the two most studied measures of higher order cognition associated with complex and simple span tasks; thus, we included only studies that used measures of verbal/reading comprehension or spatial/matrix reasoning. Thus, our inclusion criteria were the following:

1. The study had to clearly assess either simple or complex span;
2. The study had to have data from only young adults (17 to 35 years of age);
3. The study had to use either the ABS or the PCS scoring method; and
4. The criterion measure in the study had to be either verbal comprehension or spatial/matrix reasoning. For verbal comprehension, the measures included the Nelson–Denny Reading Comprehension Test, the Verbal SAT, and other tasks variously labeled as reading comprehension. The Verbal SAT measures more than reading comprehension; thus, we labeled this as verbal comprehension. Studies that measured vocabulary alone were not used. For spatial/matrix reasoning, the tasks included the Raven Progressive Matrices (Raven, Raven, & Court, 1998; both standard and advanced), Cattell's Culture Fair Test (Cattell, 1973), the Wechsler Abbreviated Scale of Intelligence (WASI), Matrix Reasoning subtest (Wechsler, 1999), and the Beta Examination (3rd ed.; Beta III), Matrix Reasoning subtest (Kellogg & Morton, 1999).

Clearly these criteria drastically reduced the number of usable studies; however, much of the previous research arguing for a distinction between simple and complex span has made this distinction on the basis of only simple and complex span tasks, primarily in young adults, and with verbal comprehension or spatial/matrix reasoning abilities as the criterion measure.

Aggregation of within-sample effect sizes. Several studies reported multiple correlations between the span tasks and the ability measures. Therefore, correlations from these studies were aggregated, and the mean correlation was used. Mean correlations from these studies were obtained with Fisher's r-to-z transformation.

Correlational analysis. The procedure described by Hedges and Olkin (1985) was used to compute meta-analytic estimates of effect size. Specifically, effect size for each memory/ability correlation was computed as the weighted (by sample size) average of all correlations entering into the computation. The estimated population correlation (ρ) was obtained with a z-to-r transformation of the estimated effect size.

Confidence intervals. Ninety-five percent confidence intervals (CIs) were calculated with formulas described by Hedges and Olkin (1985).

Results and Conclusions

The results of the different meta-analyses are shown in Table 1. As can be seen, the analyses yielded several interesting findings. First, and perhaps most important, simple and complex span demonstrated remarkably similar correlations with composites of higher order cognitive abilities. Complex span tasks had numerically larger correlations with all criterion indicators; however, the differences between the correlations were actually quite small, and in all cases the 95% CIs overlapped. Contrary to much previous research, this suggests that simple span tasks are as good as complex span tasks in predicting measures of higher order cognition. This is especially true when examining only those studies including simple and complex span tasks and those that did not involve developmental differences.

The results also suggest that simple and complex span predict performance on verbal and spatial tasks to a similar extent. Finally, and somewhat interestingly, we found no differences as a function of scoring method. The PCS method resulted in slightly larger correlations than did the ABS method; however, all of the 95% CIs overlapped. Thus, contrary to our hypothesis about possible dif-

Table 1
Estimated Population Correlation Between Memory Span Measures and Ability Measures With 95% Confidence Intervals (CIs)

Task scoring	Criterion	ρ	No. of samples	*N*	95% CI
Complex PCS	Average ability	.39	11	1,460	.35 to .43
Simple PCS	Average ability	.35	5	539	.27 to .41
Complex ABS	Average ability	.34	11	1,116	.28 to .39
Simple ABS	Average ability	.29	11	1,087	.24 to .36
Complex PCS	Verbal	.36	7	874	.30 to .42
Simple PCS	Verbal	.31	4	411	.22 to .40
Complex ABS	Verbal	.38	8	665	.31 to .45
Simple ABS	Verbal	.35	8	676	.29 to .42
Complex PCS	Matrix	.41	5	821	.35 to .46
Simple PCS	Matrix	.40	2	363	.31 to .49
Complex ABS	Matrix	.28	5	719	.21 to .35
Simple ABS	Matrix	.21	5	679	.13 to .28

Note. ABS = absolute scoring method; PCS = proportion correct scoring method; no. of samples = number of independent samples; *N* = aggregate sample size.

ferences between the ABS and PCS methods, the meta-analysis suggested only minor differences. In fact, the only difference that we found was that the PCS method had a higher magnitude correlation with spatial/matrix reasoning than did the ABS method.

Overall, the results from the meta-analysis were fairly straightforward in suggesting that simple and complex span correlate similarly with higher order cognition when using fairly strict inclusion criteria involving only those studies with simple or complex span, ABS or PCS scoring methods, young adults, and criterion measures of verbal comprehension or spatial/matrix reasoning. However, as noted above, our hypothesis that differences in scoring method would influence the correlations was not supported. Why might this be the case? Two possibilities seem most relevant. First and most obvious, scoring procedure may have no influence on the magnitude of correlations between span and complex cognition. This would mean that Unsworth and Engle's (2006a) results suggesting differences as a function of list length were an anomaly. Second, perhaps differences in scoring procedures do matter, but these differences were somehow obscured by idiosyncrasies in the studies that were included in our meta-analysis. Although we attempted to select studies that used the same basic methodology, perhaps other factors (e.g., nature of the sample, specific tasks, group vs. individual administration) influenced some of the correlations. Indeed, one large difference among the studies examined here was that some used a procedure in which list lengths were presented in an ascending order (e.g., Lustig, Hasher, & May, 2001), whereas others used a procedure in which list lengths were presented randomly (e.g., Kane et al., 2004). Therefore, in order to test our scoring hypothesis more directly, we decided to rescore and reanalyze one data set that has previously been used to argue for differences between WM and STM and their relation to higher order cognition (i.e., Engle et al., 1999).

Rescoring and Reanalyses of Previous Data Sets

Rescoring and Reanalysis of Engle, Tuholski, Laughlin, and Conway (1999)

Engle et al. (1999) examined relations among WM, STM, and general fluid intelligence (gF) with a latent variable analysis of complex span, simple span, and pencil-and-paper measures of matrix reasoning. These included three complex span tasks, three simple span tasks, two measures of gF, several other putative measures of WM, and participants' SAT scores. Engle et al. found that latent variables consisting of complex and simple span tasks were strongly related (.68); however, complex span was more highly related to gF than was simple span. Moreover, the structural equation models (SEMs) suggested that the correlation between simple span and gF was mediated by performance on complex span. Accordingly, Engle et al. argued that WM is more important for higher order cognition than is STM, echoing previous claims (e.g., Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Turner & Engle, 1989). However, all of the span tasks in the Engle et al. (1999) study were scored with the ABS method. Thus, data from the longest list lengths on the simple span tasks were usually excluded. If, as argued previously, variability from these long lists is important in predicting performance on higher order cognitive tasks (see Unsworth & Engle, 2006a), then using the ABS method likely reduced the amount of variability from these list lengths and attenuated the correlations. If the PCS method had been used, an entirely different pattern of results may have emerged such that simple and complex span predicted higher order cognition similarly. In addition, and perhaps more importantly, results derived from the PCS method may indicate that complex span does not mediate the relation between simple span and higher order cognition (specifically, gF). That is, much has been made of the fact that complex span tends to predict variance in measures of higher order cognition over and above that predicted by simple span (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Conway et al., 2002; Engle et al., 1999), suggesting that additional important processes (executive functions) are tapped by complex span, but not simple span, tasks. In order to examine these issues more thoroughly, we performed several confirmatory factor analyses (CFAs) and examined several SEMs using the original data from Engle et al. (1999). Additionally, we tested the same models after rescoring all of the span tasks using the PCS method. This allowed us to directly compare the influence of scoring methods on the correlations within the same data set.

Confirmatory factor analyses and structural equation models. The first question in these analyses was whether scoring method affected correlations among the latent span variables and the latent gF and SAT variables. In order to examine this question, we conducted separate CFAs for simple and complex span and for each scoring procedure. Note that ABS correlations can be obtained from the original Engle et al. (1999) study; the PCS correlations are provided in the Appendix of the current article. The first CFA examined latent correlations among STM (composed of simple span tasks: Bspan = backward word span; Fspan = forward word span with phonologically similar words; Fspand = forward word span with phonologically dissimilar words), gF (composed of Raven and Cattell tests), and SAT (Verbal and Quantitative SAT scores) with the ABS scoring method. The resulting model is shown in Figure 11a. Fit statistics for all models are shown in Table 2.[2] As can be seen in the figure, STM is moderately correlated with gF and is strongly correlated with SAT scores, consistent with the original results of Engle et al. Additionally, as would be expected, gF and SAT are highly correlated.

Next, we examined the same model, but this time with the PCS scoring method. The model is shown in Figure 11b. Once again, STM is highly correlated with the SAT latent variable, and SAT and gF latent variables are also highly correlated. However, now the correlation between STM and gF is much higher than in the ABS model. Thus, as suggested previously, scoring the simple

---

[2] The chi-square statistic reflects whether there is a significant difference between the observed and reproduced covariance matrices. Therefore, nonsignifcant values are desirable. However, with large sample sizes even slight deviations can result in a significant value; therefore we also report the ratio of chi-square to the number of degrees of freedom. Ratios of two or less usually indicate acceptable fit. We also report the root-mean-square error of approximation (RMSEA) and the standardized root-mean-square residual (SRMR), both of which reflect the average squared deviation between the observed and reproduced covariances. In addition, we report the normed fit index (NFI), non-normed fit index (NNFI), and confirmatory fit index (CFI), all of which reflect the proportion of the observed covariance matrix explained by the model. NFI, NNFI, and CFI values greater than .90 and SRMR values less than .05 are indicative of acceptable fit (Kline, 1998). Additionally, Hu and Bentler (1999) suggested that the combination of fit indices such as CFI $>$ .95 and SRMR $<$ .05 are the best indicators of model fit.

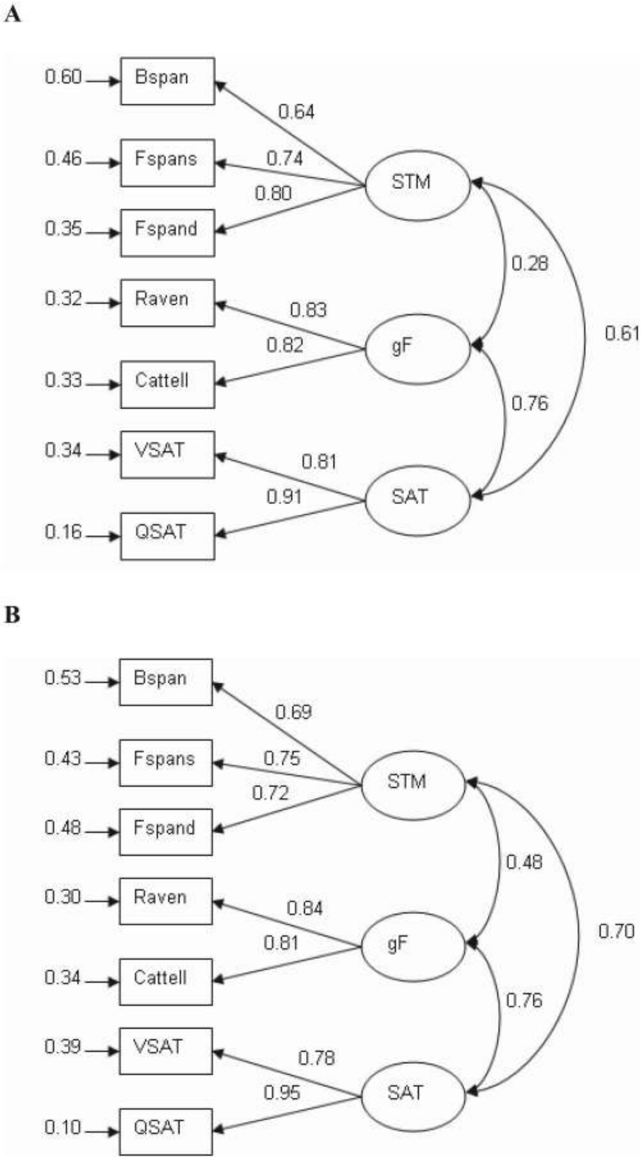

Figure 11. Confirmatory factor analysis of (a) short-term memory (STM), general fluid intelligence (gF), and Scholastic Aptitude Test (SAT) scores for the absolute scoring method (ABS) and of (b) STM, gF, and SAT for the proportion correct scoring method (PCS). Bspan = backward word span; Fspans = forward word span with phonologically similar words; Fspand = forward word span with phonologically dissimilar words; Raven = Raven Progressive Matrices; Cattell = Cattell Culture Fair Test; VSAT = Verbal SAT; QSAT = Quantitative SAT.

span tasks differently and allowing for variance from long list lengths increased the correlations among the STM and higher order constructs.

These same analyses were also performed for the complex span tasks (Ospan = operation span; Rspan = reading span; Cspan = counting span) and their relation to gF and SAT performance. The resulting model with the ABS scoring method is shown in Figure 12a, and the model from the rescored data with the PCS method is shown in Figure 12b. As can be seen in both figures, the WM latent variable correlated highly with both gF and SAT, and gF and SAT were highly correlated, consistent with previous research.

Additionally, unlike our analysis of the data from simple span tasks, changing the scoring procedure did not affect the latent correlations from the complex span tasks. Thus, it appears that scoring method affects simple span more than it affects complex span. Furthermore, note the substantial difference between correlations of STM and WM with gF when using the ABS method (.28 vs. .51) and the small difference between the correlations when using the PCS method (.48 vs. 52). This suggests that the correlation between STM higher order cognition (particularly gF) is similar to the correlation between WM and higher order cognition when variability from long lists in simple span is included. Additionally, note that the magnitudes of the latent correlations between STM and gF and between WM and gF are similar to those of meta-analytically derived latent correlations for these same constructs (Ackerman et al., 2005).

The models described above suggested that scoring procedure affects correlations among WM, STM, and higher order cognition; thus, we examined whether WM mediates the correlation between STM and higher order cognition, as has been found previously. We recreated the mediation model from Engle et al. (1999), shown in their Figure 3b. The resulting model appears in Figure 13a of this article. Note that the model is exactly the same as the original Engle et al. model, but the parameter values are slightly different as a result of the different modeling software used. As can be seen, WM and STM are highly correlated, but only WM uniquely predicts variance in gF; thus, the correlation between STM and gF is fully mediated by WM. Note also that the solid line from WM to gF is significant ($t > 2.58$), but the dotted line from STM to gF is not significant ($t < 1.0$).[3]

Next we examined the same model with the rescored data using the PCS method for both simple and complex span tasks. The resulting model is shown in Figure 13b. Here, WM and STM are highly correlated, as in the original model, but in this case neither WM nor STM uniquely predicts gF, as indicated by the dotted lines in Figure 13b. Specifically, the path from WM to gF was not significant and neither was the path from STM to gF. What are we to conclude from this? Are neither WM nor STM related to gF once the data are rescored according to the PCS method, or are both WM and STM related to gF, with neither uniquely related? The latter possibility seems correct given that the overall R^2 for the gF latent variable was .28. Thus, the variance common to both WM and STM accounted for approximately 28% of the variance in gF, with neither task contributing unique variance.

These results suggest that the variance common to both STM and WM is important in predicting higher order constructs, not that WM predicts variability over and above that accounted for by STM. Another way to examine whether WM predicts unique variance is to have a latent variable composed of variance common to all span tasks and another latent variable composed of only the residual variance common to the complex span tasks. Utilizing this model, Colom, Rebollo, Abad, and Shih (2006) reanalyzed data

---

[3] Note that in the *Rescoring and Reanalyses* sections, because of the large sample sizes and the large number of statistical tests used to compare correlations and path coefficients across models, we adopted a significance level of $p < .01$. Adopting a significance level of $p < .05$ led to virtually identical results.

Table 2
Fit Indices for the Confirmatory Analyses and Structural Equation Models

Model	χ^2	*df*	χ^2/df	RMSEA	NFI	NNFI	CFI	SRMR
Short-term memory								
ABS CFA	24.94	11	2.67	.10	.96	.95	.95	.05
PCS CFA	9.56	11	0.87	.00	.98	1.0	1.0	.03
Working memory								
ABS CFA	18.47	11	0.68	.07	.97	.97	.99	.04
PCS CFA	18.69	11	1.70	.08	.97	.98	.99	.04
Engle et al. (1999)								
ABS SEM gF	21.01	17	1.24	.04	.96	.99	.99	.04
PCS SEM gF	15.16	17	0.89	.00	.98	1.0	1.0	.03
Colom et al. (2006)								
ABS SEM gF	15.63	15	1.04	.02	.96	1.0	1.0	.04
PCS SEM gF	10.81	15	0.72	.00	.98	1.0	1.0	.03
ABS SEM SAT	22.97	15	1.53	.06	.97	.98	.98	.04
PCS SEM SAT	14.57	15	0.97	.00	.98	1.0	1.0	.03

Note. ABS = absolute scoring method; PCS = proportion correct scoring method; CFA = confirmatory factor analysis; SEM = structural equation model; gF = general fluid intelligence; RMSEA = root-mean-square error of approximation; NFI = normed fit index; NNFI = nonormed fit index; CFI = confirmatory fit index; SRMR = standardized root-mean-square residual.

from a number of studies examining relations among WM, STM, and higher order cognition, including the Engle et al. (1999) study. Colom, Rebollo, et al. (2006) found that in all cases the variance common to the span tasks significantly predicted variance in latent variables composed of higher order cognitive measures. Additionally, Colom, Rebollo, et al. (2006) found that, in some cases, the residual variance common to only the complex span tasks did not predict unique variance in the higher order factors. Specifically, of the six models that explicitly examined simple and complex span (as defined here), three models suggested that WM residual variance significantly predicted higher order cognition and the other three models suggested that the WM residual variance was not related to higher order cognition. Interestingly, for the current discussion, the three models that supported the predictive utility of WM residual variance all relied on data utilizing the ABS method (data from Conway et al., 2002, and Engle et al., 1999), whereas the three models that found no support for the predictive utility of WM residual variance relied on data utilizing the PCS method (data from Kane et al., 2004). Thus, as with the results described above, WM predicts variance in higher order cognition over and above that accounted for by STM, but only when the ABS method is used. When the PCS method is used to score the data, the variance common to both WM and STM predicts higher order cognition, and neither WM nor STM is a unique predictor. This is largely due to the fact that, as shown in the CFAs, STM correlations with higher order cognition change with scoring procedure because variance from the longest list lengths is included.

In order to examine this issue further, we modeled the Engle et al. (1999) data using both the ABS (original data) and the PCS scoring methods with the model structure from Colom, Rebollo, et al. (2006). The model utilized by Colom, Rebollo, et al. is recreated in Figure 14a from their Figure 2 and is applied to the Engle et al. (1999) data. Fit statistics are presented in Table 2.

As Colom, Rebollo, et al. (2006) found, the resulting model suggested that variance common to WM and STM predicts gF and that the residual WM variance also uniquely predicts gF at the $p <$ .01 level. However, once the data were rescored with the PCS method, the WM residual no longer predicted gF. Specifically, as shown in Figure 14b, the variance common to both WM and STM significantly predicted gF, but the residual WM variance was not a unique predictor. Additionally, using the PCS method to score the data increased the overall R^2 for the gF latent variable from .24 with the ABS method to .28 with the PCS method. This suggests that the PCS method accounted for variance over and above that accounted for by the ABS method and that the PCS method renders simple and complex span more similar in terms of their correlations with higher order cognition.

These same basic conclusions also hold when SAT performance is used as the criterion variable, as shown in Figures 15a and 15b. Figure 15a shows a recreation of Colom, Rebollo, et al.'s (2006) model in their Figure 3 applied to the Engle et al. (1999) data with the ABS method. Here, both the common and residual variances significantly predict SAT performance. With the PCS method, however, the common variance predicts SAT performance, but the residual WM is not a reliable predictor. Additionally, as with the gF model, the PCS method accounts for slightly more variability ($R^2 = .57$) than does the ABS method ($R^2 = .52$).

Reanalysis and Rescoring of Kane et al. (2004)

The results obtained from our reanalysis of the Engle et al. (1999) data set were straightforward in suggesting that the variance shared by simple and complex span, but not the WM residual, predicts higher order constructs when variability is present from the supraspan lists in simple span tasks (due to rescoring the data using the PCS method instead of the ABS method). In order to confirm this result, we decided to rescore and reanalyze another prominent data set suggesting that shared variance in simple and complex span underlies their predictive power (i.e., Kane et al., 2004). Kane et al. (2004) assessed performance on a number of complex span tasks, simple span tasks, and measures of fluid reasoning to examine differences between verbal and spatial aspects of WM and STM. Kane et al. found that shared variance in simple and complex span tasks (both verbal and spatial versions) strongly predicted variance in a broad-based gF construct. Colom, Rebollo, et al. (2006) reanalyzed the Kane et al. data set and found that shared variance in simple and complex span tasks strongly predicted variance in the reasoning construct, but in no case was

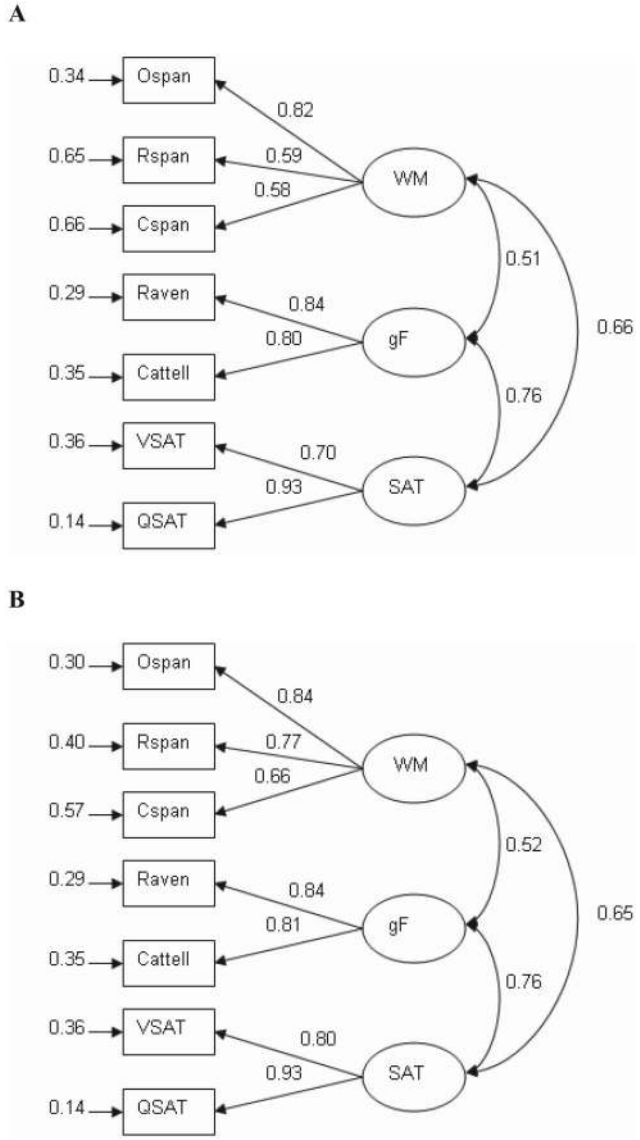

Figure 12. Confirmatory factor analysis of (a) working memory (WM), general fluid intelligence (gF), and Scholastic Aptitude Test (SAT) scores for the absolute scoring method (ABS) and of (b) WM, gF, and SAT for the proportion correct scoring method (PCS). Ospan = operation span; Rspan = reading span; Cspan = counting span; Raven = Raven Progressive Matrices; Cattell = Cattell Culture Fair Test; VSAT = Verbal SAT; QSAT = Quantitative SAT.

the residual WM variance a unique predictor. As noted previously, the Kane et al. data set was scored with the PCS method; thus, the finding that residual variance did not account for unique variance in gF is consistent with our reanalysis of the Engle et al. (1999) data. Nonetheless, an analysis of the Kane et al. data set with the ABS method would provide important converging evidence that scoring method influences the magnitude of the correlation between simple span and higher order cognition. To be consistent with our reanalysis of the Engle et al. (1999) data set, we initially examined three complex verbal span tasks (operation span, reading span, counting span), three simple verbal span tasks (word span, letter span, digit span), and three general matrix reasoning tasks (WASI, Raven Advanced Progressive Matrices, and the Beta III).

CFAs examining relations among WM, STM, and gF for both the ABS procedure and the PCS procedure appear in Figures 16a and 16b. Fit indices are shown in Table 3. As can be seen, scoring method did not affect the results. Additionally, note that the correlation between WM and STM (with both scoring procedures) is very similar to the correlation between WM and STM found in the Engle et al. (1999) reanalyses with the PCS version. Thus, these results suggest that WM and STM (as measured by simple and complex spans) demonstrate similar correlations with gF regardless of scoring procedure, a finding that is inconsistent with our hypothesis that the ABS method excludes important variance from long list lengths in simple span.

To examine these results further, we modeled the Kane et al. (2004) data using the structure from Colom, Rebollo, et al. (2006) for both PCS and ABS methods. Recall that these analyses examine the extent to which shared variance in simple and complex span and variance common to complex span (the residual variance) differentially predict higher order constructs. The resulting models are shown in Figures 17a and 17b. As with the CFAs, these analyses suggest little difference between the ABS and PCS methods: shared variance in simple and complex span, but not WM residual variance, related to gF. Thus, like the CFA analyses reported above, both scoring procedures provide results that are very similar to the PCS analyses from the Engle et al. (1999) data set.

Similar results were found when we included spatial complex and simple span tasks in the model using a structure similar to that provided in Colom, Rebollo, et al.'s (2006) Figure 7. The complex span tasks included navigation span, rotation span, and symmetry span; the simple span tasks included ball span, arrow span, and matrix span (see Kane et al., 2004, for task details). The resulting models for the ABS and PCS methods when both verbal and spatial memory span tasks were included appear in Figures 18a and 18b. The results are very similar to those including only the verbal tasks from Kane et al. and are consistent with the PCS analyses from Engle et al. (1999). The variance common to simple and complex span predicts substantial variance in gF, but the residual WM variance accounts for nothing further.

Comparing and Contrasting Engle et al. (1999) and Kane et al. (2004)

Clearly, the results obtained from rescoring and reanalyzing Engle et al.'s (1999) data do not match the results obtained from rescoring and reanalyzing Kane et al.'s (2004) data. In particular, scoring method had a large influence on the data from Engle et al., but not on the data from Kane et al. The Engle et al. results suggested that residual WM variance predicted gF when using the ABS, but not the PCS, method. The Kane et al. results suggested that shared variance in simple and complex span tasks accounted for variance in gF, but the residual WM variance accounted for no additional variance even when the ABS method was used. Furthermore, this did not change when both verbal and spatial memory span tasks were examined.

What are we to make of this? Is the Engle et al. data set anomalous, or do the Engle et al. and Kane et al. studies differ in ways that may account for the discrepancy? We think that the latter may be the case. We have argued that long list lengths in simple

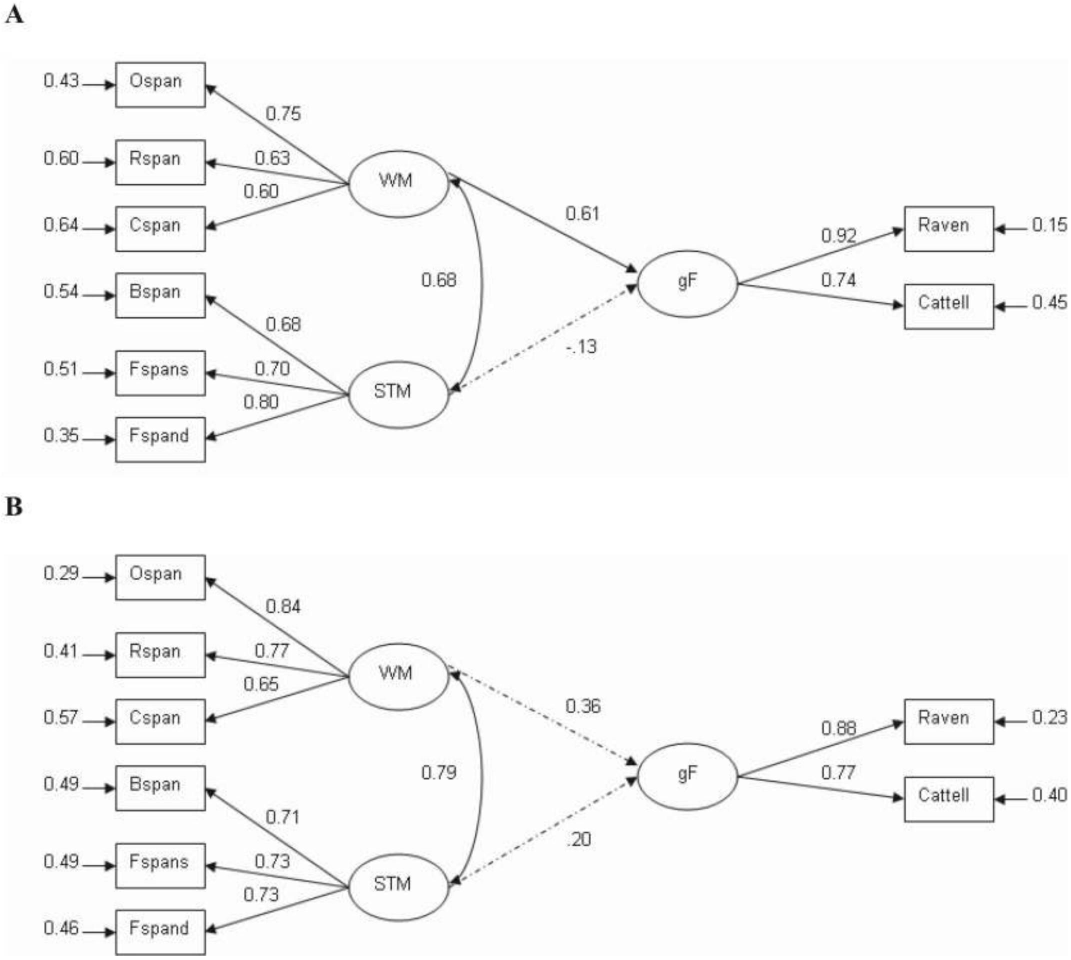

Figure 13. Structural equation model of (a) working memory, short-term memory (STM), and general fluid intelligence (gF) for the absolute scoring method (ABS) and of (b) WM, STM, and gF for the proportion correct scoring method (PCS). Ospan = operation span; Rspan = reading span; Cspan = counting span; Bspan = backward word span; Fspans = forward word span with phonologically similar words; Fspand = forward word span with phonologically dissimilar words; Raven = Raven Progressive Matrices; Cattell = Cattell Culture Fair Test.

span add variability that is related to higher order cognition. The ABS method can reduce variance from long lists in simple span, but we also believe that it is not the only factor that can have this effect. The two studies may have differed such that the Kane et al. data set may have included more variability from long lists in simple span than did the Engle et al. data set, even when the ABS method was used.

In order to test this possibility, we examined the number of individuals who successfully recalled at least one complete trial at the two longest lengths, lists of six and seven items. In the Engle et al. data set, 46 out of 133 participants had at least one completely correct trial at a list length of six, whereas only 16 out of 133 participants had at least one completely correct trial at a list length of seven. Thus, roughly 12% of the sample contributed to variance at the longest list lengths. In the Kane et al. data set, 190 out of 235 participants had at least one completely correct trial at a list length of six, whereas 110 participants out of 235 had at least one completely correct trial at a list length of seven. Thus, roughly 47% of the sample contributed to the variance at the longest list length. Clearly, many more individuals contributed to the variance at the long list lengths with the ABS procedure in the Kane et al. data set than in the Engle et al. data set.

Furthermore, we observed greater variability among individuals in the Kane et al. (2004) study than in the Engle et al. study. Specifically, the standard deviation and range for lists of seven items determined with the ABS method in Engle et al. were 0.52 and 3.89, respectively. In Kane et al., however, the standard deviation and range for a list length of seven conducted with this method were 1.34 and 7.00. Thus, we found large differences in the amount of variability at the longest simple span lists in the two studies, and these differences likely led to the differences in our patterns of results. Scoring did not matter much in the Kane et al. data set because the data contained adequate variability from long list lengths even when the ABS method was used. However, scoring had a large effect in the Engle et al. data set because variability from long list lengths was only present when the PCS method was used.

Next we examined how these differences in variability would affect the correlations between each simple span list length (verbal

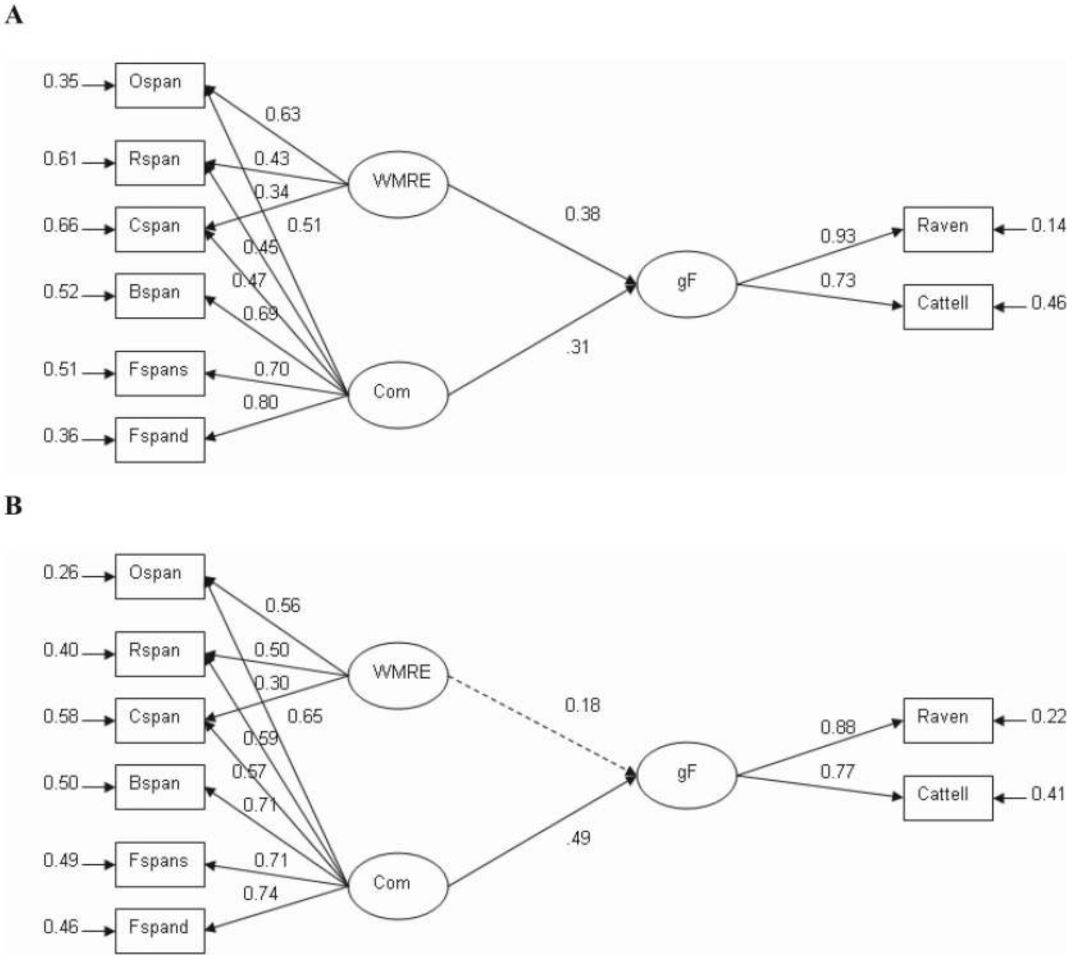

Figure 14. Structural equation model of (a) common variance (Com), working memory residual variance (WMRE), and general fluid intelligence (gF) for the absolute scoring method (ABS) and of (b) Com, WMRE, and gF for the proportion correct scoring method (PCS). Ospan = operation span; Rspan = reading span; Cspan = counting span; Bspan = backward word span; Fspans = forward word span with phonologically similar words; Fspand = forward word span with phonologically dissimilar words; Raven = Raven Progressive Matrices; Cattell = Cattell Culture Fair Test.

span only) with the gF composite as a function of scoring method (ABS or PCS) for the two data sets. The correlations are shown in Figures 19a and 19b. As can be seen, scoring method had a large influence on correlations in the Engle et al. data set but little influence in the Kane et al. data set. Consistent with previous research (Unsworth & Engle, 2006a; see Figure 10 of the current article), the correlations increase as a function of list length with the PCS version in both studies.[4] The correlations also rise using the ABS method in the Kane et al. data set, but less so in the Engle et al. data set. No rise in the correlations is seen in lists longer than four items. Thus, the Kane et al. data set, but not the Engle et al. one, appears to include adequate variability at the long list lengths.

An important question is why the Kane et al. (2004) data set included more variability from long lists in simple span than did the Engle et al. data set? We believe that three factors may be responsible for the difference: the nature of the sample, the specific simple span tasks, and the procedures used to administer the tasks. Engle et al. evaluated all college students from one comprehensive state university, whereas Kane et al. studied a combination of students (from three universities) and community volunteers. Thus, the Kane et al. sample was more heterogeneous than the Engle et al. sample.

The Engle et al. (1999) and Kane et al. (2004) studies also differed in the specific span tasks that were used and in the administration of those tasks. Specifically, Engle et al. used three variations of a word span task, whereas Kane et al. used a word span task, a letter span task, and a digit span task. Previous research (Crannell & Parrish, 1957) has found that list-length effects are steeper for words than for either letters or digits. Thus, the use of diverse tasks in the Kane et al. data set could have

---

[4] Note that we also examined the correlation between the verbal complex span tasks and gF as a function of list length and found that, consistent with previous research (Unsworth & Engle, 2006a; see Figure 10 of the current article), the correlations remained fairly stable across all list lengths. Specifically, the correlations between performance on the verbal complex spans and gF for list lengths 2–6 with the PCS method were .41, .38, .42, .40, and .29.

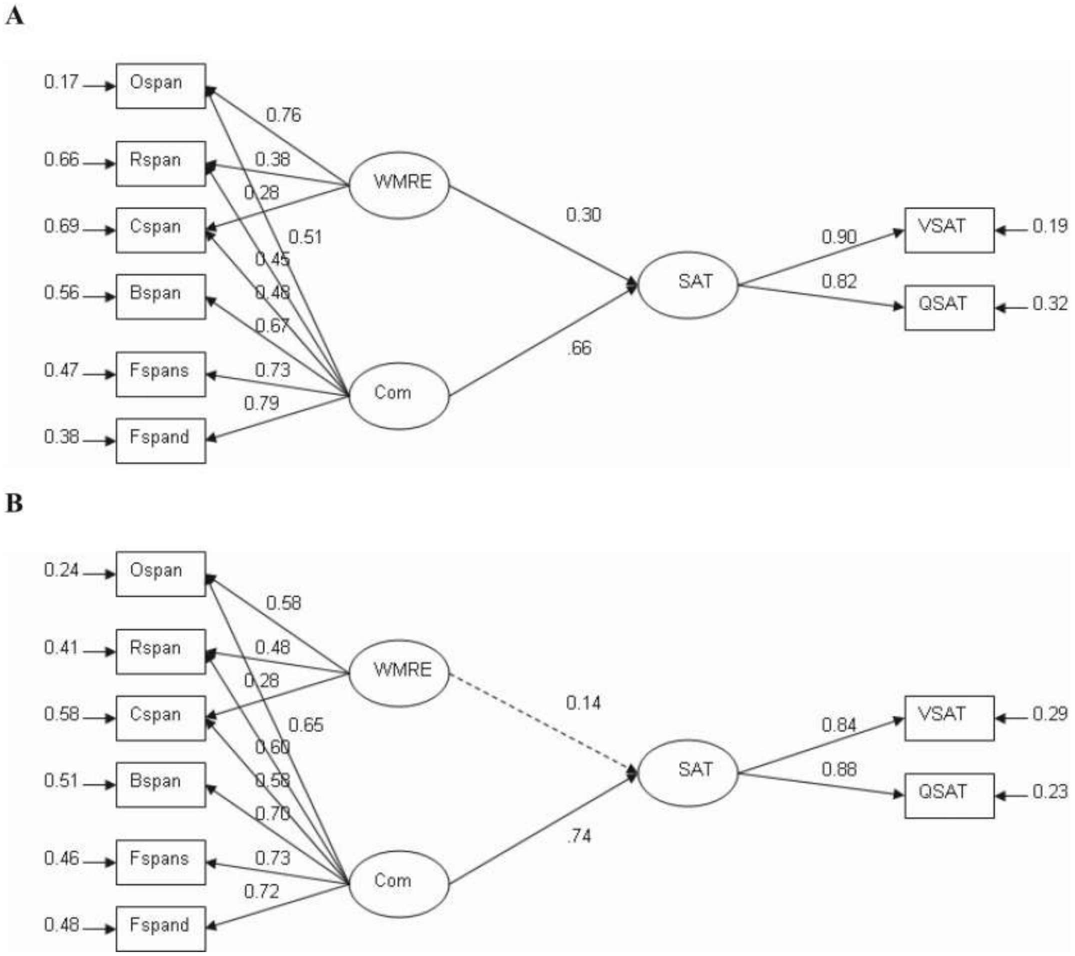

Figure 15. Structural equation model of (a) common variance (Com), working memory residual variance (WMRE), and Scholastic Aptitude Test (SAT) scores for the absolute scoring method (ABS) and of (b) Com, WMRE, and SAT for the proportion correct scoring method (PCS). Ospan = operation span; Rspan = reading span; Cspan = counting span; Bspan = backward word span; Fspans = forward word span with phonologically similar words; Fspand = forward word span with phonologically dissimilar words; VSAT = Verbal SAT; QSAT = Quantitative SAT.

increased the amount of variability at the longest list lengths. Furthermore, the administration procedure for each task was different in the two studies. Specifically, Engle et al. administered list lengths in ascending order beginning with a length of two and ending with a length of seven. In Kane et al., however, list lengths were presented in a random order. Participants in the Engle et al. study may have been more likely to give up on long list lengths than participants in the Kane et al. study. This might have occurred for two reasons. First, participants may have become fatigued over the course of the study. This would not be a problem in Kane et al. because list lengths were dispersed randomly over the session, whereas the long list lengths were clustered at the end of the session in Engle et al. Second, proactive interference may have built up over the course of the study (e.g., Lustig, Hasher, & May, 2001). This is a greater problem in Engle et al. than in Kane et al. because long list lengths appeared only at the end of the session in the Engle et al. study.

The differences in samples, tasks, and procedures may have resulted in the different findings obtained by Engle et al. (1999) and Kane et al. (2004). Despite these differences, the overall story is the same. Specifically, long lists matter in simple span. When variability is present at supraspan lengths, simple and complex span are correlated similarly with measures of higher order cognition. Thus, as we have shown here, scoring procedure is just one factor than can influence the magnitude of the correlations; other factors may include sample, task, and administration procedures.[5]

Summary and Conclusion of Predictive Power of Simple and Complex Span

The analyses described above suggest that simple and complex span generally predict higher order cognitive abilities to the same

---

[5] In further support of this claim, all analyses in this section were redone using only the longest list lengths (list lengths 5–7) from the simple span tasks. The results were identical to those presented in the current article, suggesting that excluding the shortest list lengths does not change the predictive power of the simple span tasks.

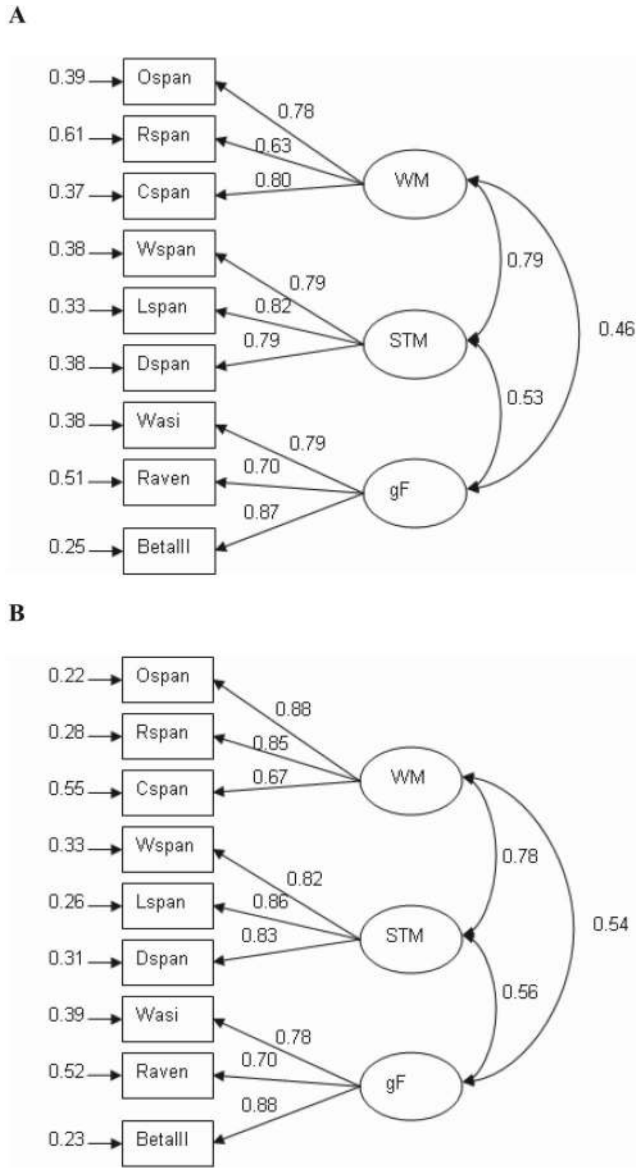

Figure 16. Confirmatory factor analysis of (a) working memory (WM), short-term memory (STM), and general fluid intelligence (gF) for the absolute scoring method (ABS) and of (b) WM, STM, and gF for the proportion correct scoring method (PCS). Ospan = operation span; Rspan = reading span; Cspan = counting span; Wspan = word span; Lspan = letter span; Dspan = digit span; Wasi = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning subtest; Raven = Raven Advanced Progressive Matrices; Beta III = Beta Examination (3rd ed., revised), Matrix Reasoning subtest.

extent when steps are taken to ensure that variability from long list lengths is present in both tasks. Long list lengths are particularly important for the simple span tasks.

Our re-analyses of the Engle et al. data set suggest that scoring procedure is one factor that can influence variability at long list lengths. Complex span predicted variability in gF and SAT performance better than simple span when the ABS method was used; however, both tasks predicted higher order cognition when the PCS procedure was used. The reason for this, we suggest, is that the ABS method typically taps performance from short list lengths in simple span, thereby reducing overall variability.

Our re-analysis of the Kane et al. (2004) data set revealed that factors other than scoring procedure may affect the amount of variability that is included from long list lengths in simple span. This data set contained substantial variability at supraspan list lengths even when the data were scored with the ABS method. We suggested that this additional variability may have been due to a heterogeneous sample, the inclusion of letter and digit span tasks, and a procedure that interleaved long list lengths among shorter ones. Taken together, these results suggest that, similar to the experimental effects reviewed previously, simple and complex span largely measure the same basic processes and that these processes are important for higher order cognitive abilities. The similarity in their relation to higher order cognition will only be apparent when information is obtained regarding performance on both short and long lists. When this occurs, the shared variance in both simple and complex span, but not the WM residual, predicts higher order cognition.

General Discussion

The current study examined whether simple and complex span measure the same or different constructs. We reviewed the extent to which simple and complex span are affected by common experimental variables and reviewed the extent to which simple and complex span predicted performance on measures of higher order cognitive abilities similarly. We draw two main conclusions on the basis of these reviews and re-analyses. First, simple and complex span are affected by many experimental variables similarly, although simple span tends to be somewhat more affected than complex span when an experimental variable disrupts phonological processing. This disruption results in similar performance on simple and complex span. Second, simple and complex span typically have correlations with higher order cognitive abilities that are similar in magnitude, especially when controlling for other variables (such as scoring). In particular, simple and complex span tasks demonstrate similar predictive utility when variance from long lists in simple span is included in the data. Across both experimental and differential perspectives, the evidence reviewed here suggests that simple and complex span largely measure the same basic processes; thus, the notion that STM and WM are largely different constructs is not warranted (see also Colom, Rebollo, et al., 2006). Rather, our review suggests that the variance common to simple and complex span is responsible for their predictive power (see also Kane et al., 2004).

This conclusion is contrary to previous research (including our own) suggesting that WM and STM are different constructs and that complex span predicts higher order abilities better than simple span (Cantor, Engle, & Hamilton, 1991; Conway & Engle, 1996; Conway et al., 2002; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Dixon, LeFevre, & Twilley, 1988; Engle et al., 1999; Kail & Hall, 2001; Masson & Miller, 1983; Turner & Engle, 1989). In the present article, we have suggested that one likely reason that researchers reached this conclusion was that their studies relied on the ABS scoring method. This method reduced the amount of variability from long list lengths in simple span tasks. We have argued that this variability is important in predicting complex cognition. Thus, previous research arguing for a distinction between STM and WM (on the basis of analyses of

Table 3
Fit Indices for the Confirmatory Analyses and Structural Equation Models

Model	χ^2	*df*	χ^2/df	RMSEA	NFI	NNFI	CFI	SRMR
ABS CFA	19.21	24	0.80	.00	.99	1.0	1.0	.02
PCS CFA	26.50	24	1.10	.02	.99	1.0	1.0	.03
Colom et al. (2006)								
ABS V SEM gF	13.46	22	0.61	.00	.99	1.0	1.0	.02
PCS V SEM gF	24.97	22	1.14	.02	.99	1.0	1.0	.02
ABS VS SEM gF	170.83	78	2.19	.07	.96	.97	.98	.06
PCS VS SEM gF	168.73	78	2.16	.07	.97	.98	.98	.06

Note. ABS = absolute scoring method; PCS = proportion correct scoring method; CFA = confirmatory factor analysis; V = verbal memory span tasks; S = spatial memory span tasks; SEM = structural equation model; gF = general fluid intelligence; RMSEA = root-mean-square error of approximation; NFI = normed fit index; NNFI = nonormed fit index; CFI = confirmatory fit index; SRMR = standardized root-mean-square residual.

simple and complex span) may have exaggerated differences between STM and WM.

Understanding the conditions under which simple and complex span predict higher order cognition similarly and the conditions under which they demonstrate differential predictive properties is important for understanding the processes that underlie performance on these tasks and for understanding their shared variability with other tasks. Two conditions that we have discussed through-

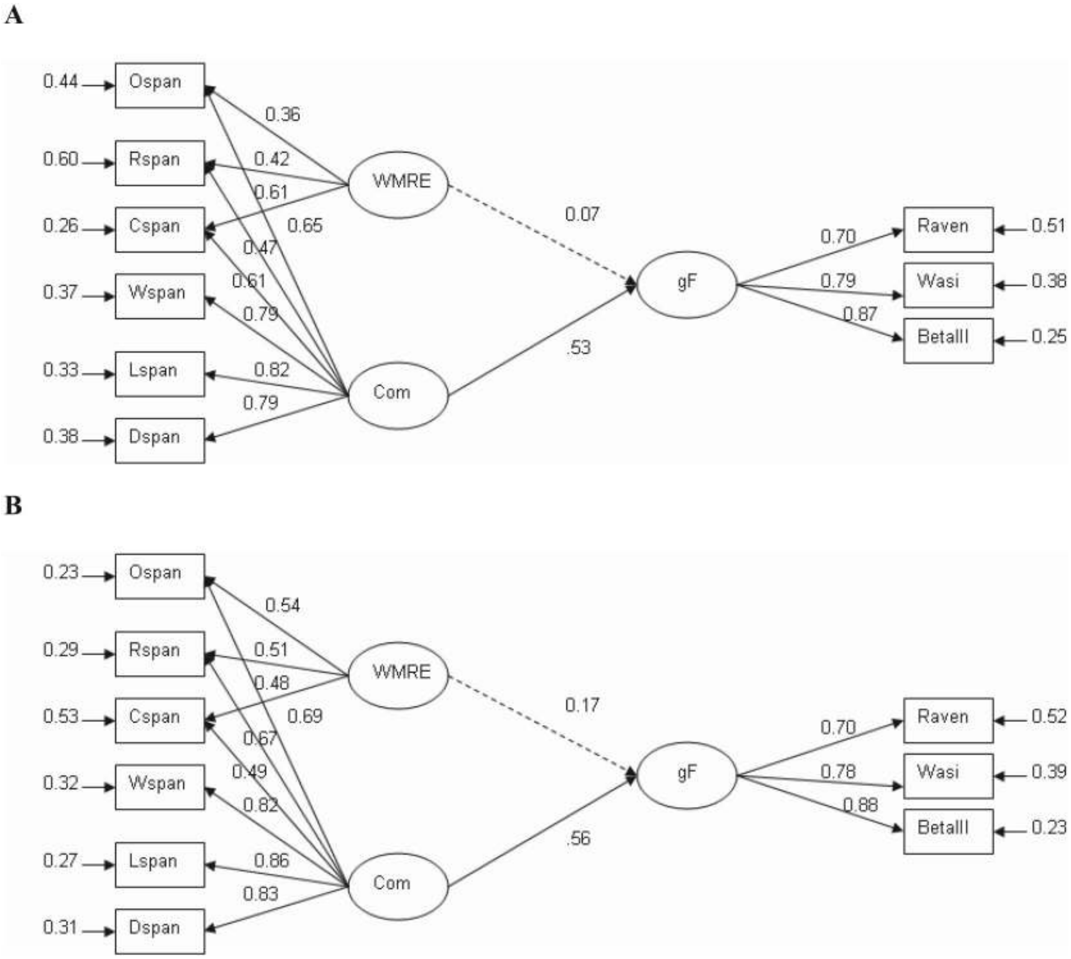

Figure 17. Structural equation model of (a) common variance (Com) variance, working memory residual variance (WMRE), and general fluid intelligence (gF) for the absolute scoring method (ABS) and of (b) Com, WMRE, and gF for the proportion correct scoring method (PCS). Ospan= operation span; Rspan = reading span; Cspan = counting span; Wspan = word span; Lspan = letter span; Dspan = digit span. Wasi = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning subtest; Raven = Raven Advanced Progressive Matrices; Beta III = Beta Examination (3rd ed., revised), Matrix Reasoning subtest.

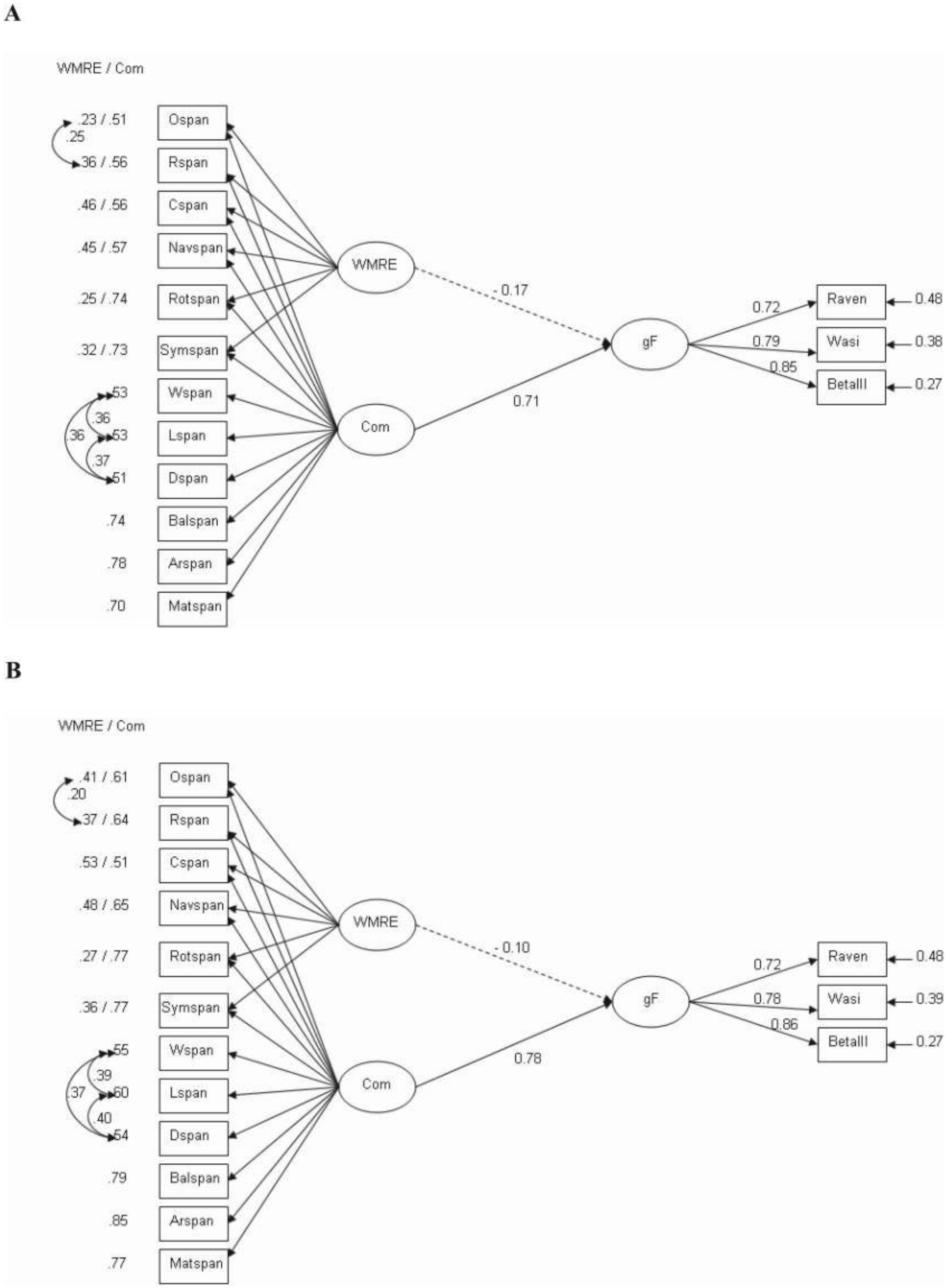

Figure 18. Structural equation model of (a) common variance (Com), working memory residual variance (WMRE), and general fluid intelligence (gF) for the absolute scoring method (ABS) and of (b) Com, WMRE, and gF for the proportion correct scoring method (PCS). The numbers in the WMRE column represent the factor loadings for each memory task on that factor; the numbers in the Com column represent the factor loadings for each memory task on that factor. Ospan = operation span; Rspan = reading span; Cspan = counting span; Navspan = navigation span; Rotspan = rotation span; Sympspan = symmetry span; Wspan = word span; Lspan = letter span; Dspan = digit span; Balspan = ball span; Arspan = arrow span; Matspan = matrix span. Wasi = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning subtest; Raven = Raven Advanced Progressive Matrices; Beta III = Beta Examination (3rd ed., revised), Matrix Reasoning subtest.

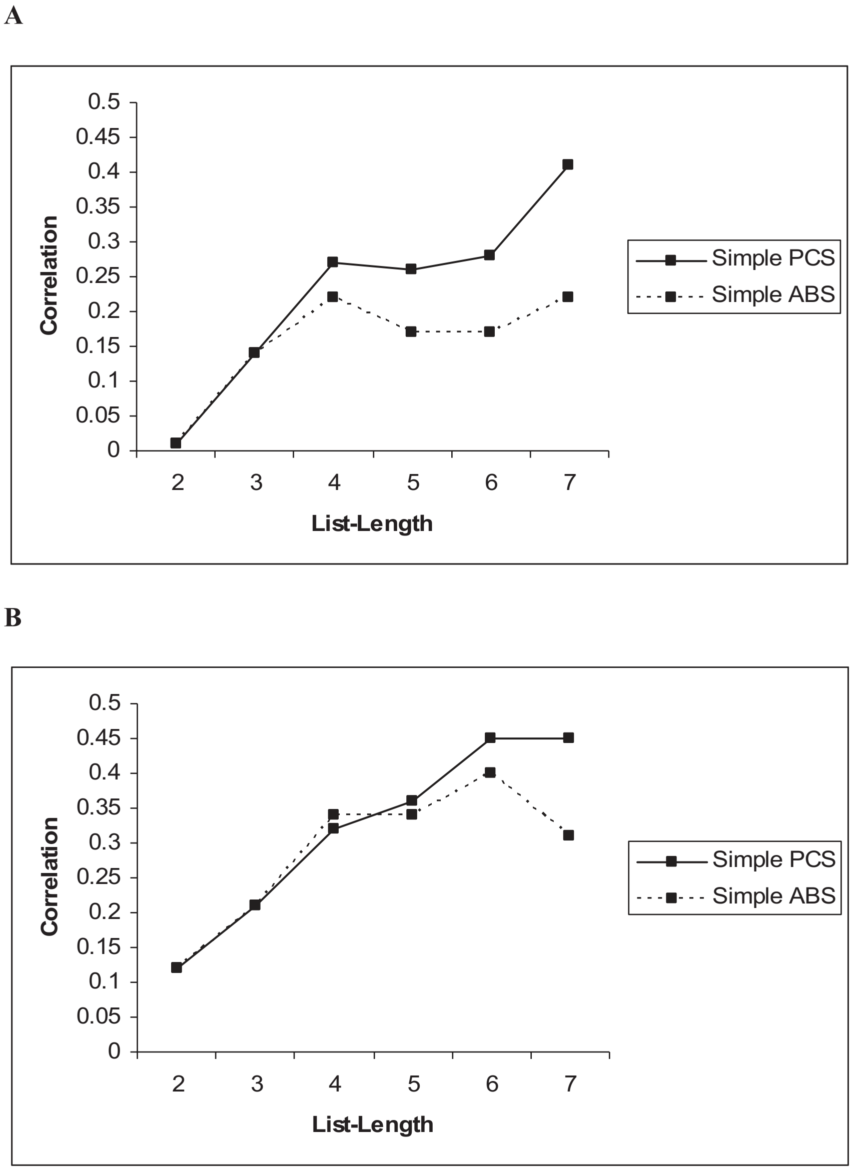

Figure 19. Zero-order correlations between verbal simple memory span and general fluid intelligence (gF) as a function of scoring method (absolute [ABS] vs. proportion correct [PCS]) for (a) the Engle et al. (1999) and (b) the Kane et al. (2004) data sets.

out this article that hint as to the nature of these processes are list length and phonological processing. When list length is increased or when phonological processes are disrupted, simple and complex span tend to behave similarly. At short list lengths or when phonological processes are maximized, simple and complex span tend to diverge in their relation to measures of higher order cognition. Thus, future research should be directed toward manipulating aspects of simple and complex span to fully understand how their similarities and differences relate to complex cognition.

Potential Methodological, Psychometric, Diagnostic, and Neuropsychological Implications

Our review suggests that methodological considerations are important in interpreting the relations among simple span, complex span, and higher order cognition. These considerations include scoring procedures, presentation procedures (e.g., auditory vs. visual), ordering of list lengths (e.g., ascending, descending, random; Lustig, Hasher, & May, 2001), output procedures (e.g., written vs. oral), the type of span task (e.g., word span, letter span, digit span), and the nature of the participant sample. We recommend using a variety of tasks to measure the construct of interest (i.e., a broad range of simple span tasks, including word span, letter span, and digit span), a broad range of list lengths for all participants, a scoring procedure that includes variability from all list lengths (i.e., the PCS method), and a large, broad sample of participants.

Previous discrepancies in the literature concerning the nature of STM and WM as measured by simple and complex span are likely due to correlations among STM, WM, and complex cognition that were obscured or reduced by methodological issues. As noted previously, these methodological considerations tend to influence simple span tasks more than the complex span tasks. This points to potential psychometric problems with these tasks such that they may be reliable in a given sample of participants (i.e., have acceptable internal consistency and test–retest reliability) but may be unreliable across different samples as a result of slight differences in the administration of the tasks. Furthermore, these methodological differences may hamper the validity of simple span tasks because the processes that affect performance can change from one situation to another. Clearly, this is a major problem with these tasks, but one that is, hopefully, rectifiable by considering some of the methodological recommendations noted above.

Finally, our review has implications for how span tasks are used for diagnostic and neuropsychological assessment purposes. Simple span tasks in intelligence and neuropsychological batteries have been used to assess STM functioning in mild cognitive impairment (e.g., Xu, Meyer, Thornby, Chowdhury, & Quach, 2002), generalized brain damage (e.g., Black, 1986), frontal lobe damage (e.g., Baddeley & Wilson, 1988), amnesia (e.g., Baddeley & Warrington, 1970; Della Sala, Logie, Cubelli, & Marchetti, 1998), Korsakoff's syndrome (e.g., Janowsky, Shimamura, Kritchevsky, & Squire, 1989), Alzheimer's (e.g., Cherry, Buckwalter, & Henderson, 1996), schizophrenia (e.g., Aleman, Hijman, de Haan, & Kahn, 1999), normal aging (e.g., Bopp & Verhaeghen, 2005), and other disorders (see Lezak, Howieson, & Loring, 2004; Wilde, Strauss, & Tulsky, 2004, for reviews). However, the administration and scoring of simple span tasks in these batteries are usually done such that participants start with a list length of two and progress to longer lists until they can no longer perfectly recall the sequence of items. This is essentially a stringent form of the ABS method; thus, this method is unlikely to capture much, if any, variance from the longest list lengths. As we argued throughout the article, this variance is critical in relating STM to other cognitive functions; thus, this scoring and administration method reduces the ability of theses tasks to adequately measure processes of interest and lowers their discrimination power (e.g., Chapman & Chapman, 1973, 1978). For instance, patients with frontal lobe lesions perform more poorly than control participants on simple span tasks in some studies (e.g., Janowsky, Shimamura, Kritchevsky, & Squire, 1989), whereas no group differences have been found in other studies (e.g., D'Esposito & Postle, 1999). Similar discrepancies can be found in the literature on schizophrenia. Schizophrenic patients demonstrate poorer performance on simple span tasks than control participants in some studies (e.g., Aleman, Hijman, de Haan, & Kahn, 1999), whereas no differences have been found in other studies (e.g., J. D. Cohen, Barch, Carter, & Servan-Schreiber, 1999). These discrepancies, like those between simple and complex span tasks, may be due to differences in the use of simple span tasks across situations and samples; thus, administering and scoring them such that variability from the longest list lengths is obtained may increase the power of these tasks to discriminate among groups and may reduce or eliminate discrepancies in the research areas that rely on these tasks. Thus, we believe that the issues raised in the current article are important in basic cognitive research and in other research domains that rely on these tasks for assessment purposes. Future consideration of these issues may shed light on the underlying processes that are measured in these tasks and lead to greater understanding of memory impairments in a number of domains.

Possible Theoretical Account

Although no clear theoretical proposal can handle all of the evidence as yet, here we offer our own current theoretical viewpoint. Recently, we (Unsworth & Engle, 2006a, 2007) have suggested that performance on simple and complex span tasks can be interpreted in terms of a dual-component framework that combines an active maintenance component (primary memory: PM) with a controlled cue-dependent search and retrieval process of information that cannot be maintained (secondary memory: SM). This framework, like similar dual-component models (e.g., Atkinson & Shiffrin, 1968; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Raaijmakers & Shiffrin, 1981), suggests that items are initially maintained in PM but are displaced to SM by other incoming items or by distracting information. Items that have been displaced must be retrieved via controlled search of SM at recall. Items that have not been displaced from PM are simply unloaded during recall; thus, recall is nearly perfect. Using this framework, Unsworth and Engle (2006a) suggested that the processing component in complex span tasks will function to displace items from PM; thus, these items must be recalled from SM at recall. Simple span tasks, on the other hand, require both unloading from PM and retrieval from SM for accurate performance (Craik, 1971; Watkins, 1977). This is because no intervening activity displaces items from PM; thus, items are displaced only by new incoming items. The extent to which SM is required in simple span, however, will depend on how many items are presented (list length). We assumed that PM is limited to approximately four items based on research by Cowan (2001) and Broadbent (1975); thus, performance on short list lengths is mainly determined by unloading from PM. As list length increases beyond the capacity of PM, however, performance will also be determined by retrieval from SM. As list length increases, more items are displaced from PM and must be retrieved from SM.

The scenario described above is the reason why we believe that the correlation between simple spans and gF increases as a function of list length. Short lists primarily measure the capacity of PM and and individuals' ability to maintain items in PM, whereas long lists also measure the ability to retrieve information from SM. Complex span shows consistent correlations across list lengths because some items are displaced from PM even at the smallest list lengths and have to be retrieved from SM. Thus, the extent to which a memory task will be correlated with measures of higher order abilities is determined, in part, by the extent to which it measures both maintenance in PM and retrieval from SM.

This framework can be used to explain why the ABS and PCS methods can produce different results. The ABS method typically measures variability from only the shortest list lengths in simple span, whereas the PCS method measures variability from all list lengths. Thus, the ABS method measures PM abilities only, whereas the PCS method measures PM and SM abilities, similar to complex span. Initial evidence for the notion that simple span tasks measure both PM and SM abilities came from a study by Craik (1971). Craik examined the correlation between simple span (word span) and estimated components of PM and SM from immediate free recall (see Watkins, 1974, for a review). Craik found that both the SM and the PM components correlated with word span performance (.72 and .49, respectively). This suggests that PM and SM processes are important for performance on simple span tasks. However, these correlations must be viewed with caution considering that they are based on only 18 participants.

In order to provide converging evidence about how PM and SM contribute to complex and simple span, we once again turn to the data from Engle et al. (1999). As mentioned previously, in this study, participants were asked to perform measures of simple span, complex span, fluid abilities, and a number of other memory tasks. Relevant to the current discussion is the fact that Engle et al. obtained estimates of PM and SM from an immediate free recall task. Correlating these estimates with composite measures of WM and STM (both utilizing the PCS scoring method) suggests results very similar to those of Craik (1971). Specifically, PM correlated with both the simple span/STM composite ($r = .41$) and the complex span/WM composite ($r = .26$). Additionally, the SM component correlated with both the simple span/STM composite ($r = .39$) and the complex span/WM composite ($r = .53$). Two interesting aspects of these results should be noted. First, as expected by the dual-component framework, simple span tasks seem to measure PM to a greater extent than complex span tasks, and complex span measures SM to a greater extent than does simple span. Thus, both abilities are indexed in both span tasks; however, the relative contribution of each changes as a function of the span task. Second, these abilities seem largely independent of one another, as indicated by the fact that the PM and SM components are not correlated ($r = -.02$). Thus, the variability indexed by PM and SM in both simple and complex span represents unique variance.

We also examined relations among SM, PM, simple span, complex span, and higher order abilities in CFA and SEM analyses using the results from the PCS procedure. Specifically, we investigated a CFA in which a PM factor was formed by allowing all of the span tasks (both simple and complex) to load on it along with the PM estimate from immediate free recall. An SM factor was formed by allowing all span tasks to load on it along with the SM estimate from immediate free recall. Additionally, these two factors were not allowed to correlate. The fit of the resulting CFA was good, $\chi^2(14) = 22.38$, $p > .07$, root-mean-square error of approximation (RMSEA) = 0.07, standardized root-mean-square residual (SRMR) = .04, normed fit index (NFI) = .98, nonnormed fit index (NNFI) = .97, confirmatory fit index (CFI) = .99, suggesting that the model provided an acceptable account of the data. Next, we examined the relation between the PM and SM factors with gF and SAT. The resulting model is shown in Figure 20. The fit of this model was also good, $\chi^2(44) = 80.48$, $p < .05$, RMSEA = 0.08, SRMR = .06, NFI = .94, NNFI = .96, CFI = .97.[6]

The model is notable in several respects. First, the complex span tasks primarily load onto the SM factor, whereas simple span tasks load strongly on the PM factor. Note that all factor loadings were significant except for the Rspan task on the PM factor. Thus, like the zero-order correlations presented earlier, this suggests that complex span measures SM more than PM and simple span measures PM more than SM. Second, both the PM and the SM factors account for significant variance in gF and SAT. Furthermore, the variance accounted for by each memory factor is completely unique because the correlation between the two factors was set to zero. Specifically, 40% of the roughly 61% of the variance accounted for in SAT was uniquely predicted by SM abilities, and 21% was uniquely predicted by PM abilities. Additionally, 25% of the roughly 35% of the variance accounted for in gF was uniquely predicted by SM abilities and 10% was uniquely predicted by PM abilities. These results suggest that performance in simple and complex span arises from two separate processes on which individuals differ: the ability to actively maintain information in PM and the ability to retrieve information from SM. Both of these abilities constrain performance in a number of tasks, including basic memory tasks and measures of higher order cognitive abilities.

Our dual-component framework may also be useful in interpreting experimental similarities and differences between simple and complex span. First, differences in serial position functions may arise because of differential unloading from PM and retrieval from SM in the two tasks. That is, several items are retrieved from PM in simple span tasks and, depending on list length, fewer items are retrieved from SM. The opposite is true for complex span, in which many items are retrieved from SM and few (perhaps only one) items are unloaded from PM. Thus, this differential involvement of unloading and retrieval may explain differences in the serial position curves as well as the differences in frequency of errors across serial position. Indeed, in a previous study, we applied a temporal–contextual retrieval account of SM to errors in complex span and argued that much of the error data could be explained by assuming that individuals use temporal contextual cues to search for items from SM. This same notion has also been applied to errors and serial position functions in simple span tasks (e.g., Brown, Vousden, McCormack, & Hulme, 1999; Maylor et al., 1999). Additionally, this notion of differential involvement of PM and SM has been applied to differences in list length functions for simple and complex span (i.e., Unsworth & Engle, 2006a). Thus, although differences between simple and complex span on some experimental variables are slight, these differences can be explained by assuming that both tasks measure the same basic processes (e.g., rehearsal, maintenance, updating, controlled search), but the extent to which these processes operate in a given task likely differs.

An additional factor that likely influences differences between simple and complex span is the extent to which performance is based on phonological rehearsal processes, as noted previously. One reason that simple and complex span differ is that simple span tasks are more amenable to rehearsal processes than are complex span tasks (e.g., Cowan, 2005; Engle, Cantor, & Carullo, 1992). Support for this notion comes from an examination of those experimental variables that presumably disrupt rehearsal processes. These variables have a larger effect on simple span than on complex span. Manipulations that disrupt rehearsal processes (e.g., articulatory suppression, word length) make performance on simple and complex span tasks similar and lead to similarities in the magnitude of their correlations with measures of higher order cognition. Indeed, previous research has suggested that rehearsal processes likely inflate performance but at the same time attenuate the correlation between memory span and higher order abilities (e.g., Turley-Ames & Whitfield, 2003). Thus, rehearsal acts to augment other processes (e.g., maintenance and retrieval), thereby

---

[6] Note that the fit of the model could have been increased if QSAT were allowed to cross-load on both the SAT and gF factors. However, because this was not motivated by our framework, we decided not to free the path from gF to QSAT and instead to only report results for our initial model.

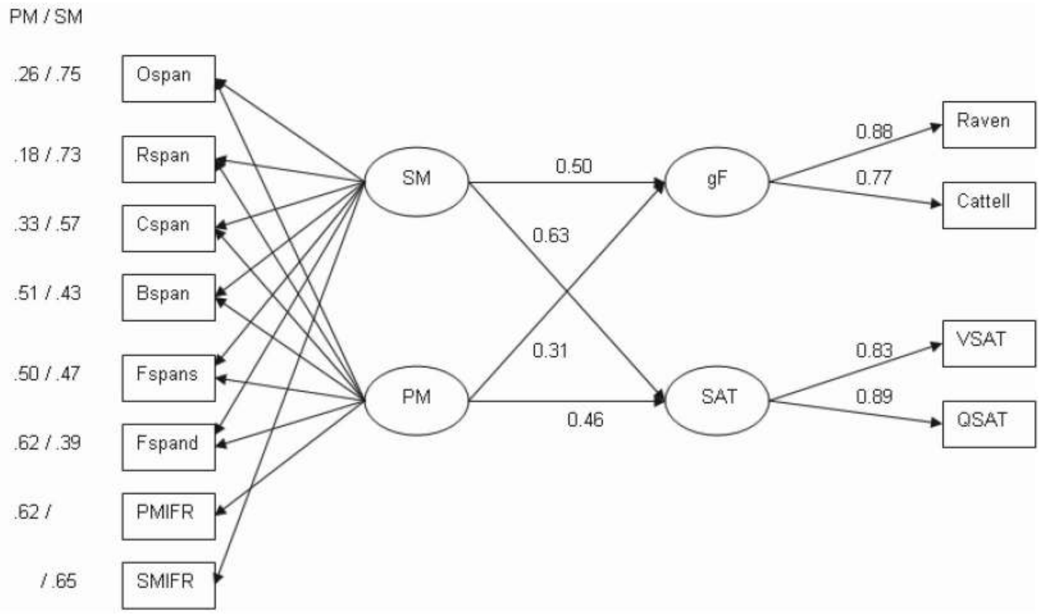

Figure 20. Structural equation model for primary memory (PM) and secondaary memory (SM) factors and their relation to general fluid intelligence (gF) and Scholastic Aptitude Test (SAT) scores. All paths are significant at the $p < .05$ level. The numbers in the PM column represent the factor loadings for each memory task onto the PM factor; the numbers in the SM column represent the factor loadings for each memory task onto the SM factor. PMIFR = PM immediate free recall estimate; SMIFR = SM immediate free recall estimate. PMIFR loads only onto the PM factor, and SMIFR loads only onto the SM factor. Ospan = operation span; Rspan = reading span; Cspan = counting span; Bspan = backward word span; Fspans = forward word span with phonologically similar words; Fspand = forward word span with phonologically dissimilar words; Raven = Raven Progressive Matrices; Cattell = Cattell Culture Fair Test; VSAT = Verbal SAT; QSAT = Quantitative SAT.

boosting performance and reducing the predictive power of some memory tasks.

Many of the suggestions that we describe above are speculative, and other perspectives may also explain the current results. For instance, Baddeley's working memory model (Baddeley, 1986) specifically addresses effects related to phonological coding and suggests that rehearsal processes are the main determinant of performance. Additionally, Nairne's (1990) feature model explains many of these effects but suggests that they are due to the ways in which cues interact with stored representations and the manner in which interference can affect some features but not others. Although, Nairne's framework does not address individual differences, they could be incorporated into the model. Finally, Cowan's (1995) embedded process model may address many of the experimental effects that we discussed as well as the differences in the correlational results. Specifically, Cowan et al. (2005) have suggested that differences in the capacity of attention (which is very similar to our concept of PM) can account for performance in memory span tasks and their relation to performance on cognitive abilities measures. These are just three of the many alternative frameworks that could potentially explain the current results. The key for any framework, however, is its ability to successfully integrate both the experimental and correlational findings.

Conclusion

In the current article, we examined the extent to which STM and WM, as measured by simple and complex span, represent the same or different constructs and the extent to which they are similarly affected by experimental variables and demonstrate similar correlations with measures of higher order cognitive abilities. Most of our analyses suggested that simple and complex span are remarkably similar in terms of performance indicators, susceptibility to experimental variables, and magnitude of correlation with higher order abilities. Thus, we conclude, as a matter of parsimony, that simple and complex span largely measure the same basic processes, and we reject the notion that STM and WM are different constructs. Rather, we suggest that all immediate memory tasks measure the same basic processes, accounting for their predictive power across a wide range of tasks. However, the extent to which a particular task measures all of these abilities is determined, in part, by the scoring procedure and the presence or absence of other processes (e.g., rehearsal) that may affect performance. Furthermore, we have tried to advocate for the benefits of a combined experimental and correlational program of research (e.g., R. L. Cohen, 1994; Cronbach, 1957; Underwood, 1975). A joint examination of these effects can provide a more complete understanding of the processes that are important for performance on these tasks and their ability to predict performance on other tasks.

References

*References marked with an asterisk were included in the meta-analysis.

*Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131,* 567–589.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60.

Aleman, A., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (1999). Memory impairments in schizophrenia: A meta-analysis. *American Journal of Psychiatry, 156,* 1358–1366.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225,* 82–90.

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18,* 302–309.

Baddeley, A. D. (1986). *Working memory.* London: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

*Baddeley, A. D., Logie, R. H., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluid reading. *Journal of Memory & Language, 24,* 119–131.

Baddeley, A. D., & Scott, D. (1971). Word frequency and the unit-sequence interference hypothesis in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 35–40.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 176–189.

Baddeley, A. D., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7,* 212–230.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., & Gunn, D. M. (2005). The relationship between short-term memory and working memory: Complex span made simple? *Memory, 13,* 414–421.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132,* 71–92.

Beaman, C. P. (2006). The relationship between absolute and proportion scores of serial order memory: Simulation predictions and empirical data. *Psychonomic Bulletin & Review, 13,* 92–98.

Beier, M. E., & Ackerman, P. L. (2004). A reappraisal of the relationship between span memory and intelligence via "best evidence synthesis." *Intelligence, 32,* 607–619.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13,* 80–97.

Black, F. W. (1986). Digit repetition in brain-damaged adults: Clinical and theoretical implications. *Journal of Clinical Psychology, 42,* 770–782.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology: Series B, 60B,* 223–233.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). New York: Wiley.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

Brown, G. D. A., Vousden, J. I., McCormack, T., & Hulme, C. (1999). The development of memory for serial order: A temporal–contextual distinctiveness model. *International Journal of Psychology, 34,* 389–402.

*Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15,* 229–246.

Carroll, J. B. (1988). Individual differences in cognitive functioning. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (pp. 813–862). New York: Wiley.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386–404.

Cattell, R. B. (1973). *Measuring intelligence with culture fair tests.* Champaign, IL: Institute for Personality and Ability Testing.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79,* 380–385.

Chapman, L. J., & Chapman, J. P. (1978). The measurement of differential deficit. *Journal of Psychiatric Research, 14,* 303–311.

Cherry, B. J., Buckwalter, J. G., & Henderson, V. W. (1996). Memory span procedures in Alzheimer's disease. *Neuropsychology, 10,* 286–293.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context processing deficits in schizophrenia: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology, 108,* 120–133.

Cohen, R. L. (1994). Some thoughts on individual differences and theory construction. *Intelligence, 18,* 3–13.

*Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent variable approach. *Intelligence, 33,* 623–624.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34,* 158–171.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. A. (2006). The real relationship between short-term memory and working memory. *Memory, 14,* 804–813.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84.

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

*Conway, A. R. 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 & Review, 12,* 769–786.

Copeland, D. W., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition, 29,* 774–776.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford, England: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 97–185.

Cowan, N. (2005). *Working memory capacity.* Hove, England: Psychology Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42–100.

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliot, E. M., Lacey, J. F., Moreno, M. V., & Hitch, G. J. (2003). Children's working memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132,* 113–132.

Craik, F. I. M. (1971). Primary memory. *British Medical Bulletin, 27,* 232–236.

Crannell, C. W., & Parrish, J. N. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology, 44,* 319–327.

*Crawford, J., & Stankov, L. (1983). Fluid and crystallized intelligence and primacy/recency measures of short-term memory. *Intelligence, 7,* 227–252.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

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

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

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112,* 3–42.

Della Sala, S., Logie, R. H., Cubelli, R., & Marchetti, C. (1998). Dissociation between recency and span: Neuropsychological and experimental evidence. *Neuropsychology, 12,* 533–545.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89,* 63–100.

D'Esposito, M., & Postle, B. R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia, 37,* 1303–1315.

*Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80,* 465–472.

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

*Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology, 82,* 799–804.

Engle, R. W., & Oransky, N. (1999). The evolution from short-term to working memory: Multi-store to dynamic models of temporary storage. In R. Sternberg (Ed.), *The nature of cognition* (pp. 514–555). Cambridge, MA: MIT Press.

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

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, DC: V. H. Winston.

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, Instruments, & Computers, 37,* 581–590.

*Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? A longitudinal study of real-world knowledge acquisition. *Memory & Cognition, 31,* 902–917.

Hedges, L. V., & Olkin, I. (1985). *Statisical methods for meta-analysis.* Orlando, FL: Academic Press.

Henson, R. N. A. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology, 36,* 73–137.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1217–1232.

*Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108,* 151–181.

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103,* 548–560.

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition, 29,* 1–9.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled attention view of working memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 66–71.

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

Kellogg, C. E., & Morton, N. W. (1999). *Revised Beta Examination* (3rd ed.). San Antonio, TX: Psychological Corporation.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

*La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Lobley, K. J., Baddeley, A. D., & Gathercole, S. E. (2005). Phonological similarity effects in verbal complex span. *Quarterly Journal of Experimental Psychology, 58A,* 1462–1478.

*Lustig, C., Hasher, L., & May, C. P. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130,* 199–207.

*Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence, 31,* 519–531.

*Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory for text. *Journal of Educational Psychology, 75,* 314–318.

Maylor, E. A., Vousden, J. I., & Brown, G. D. A. (1999). Adult age differences in short-term memory for serial order: Data and a model. *Psychology and Aging, 14,* 572–594.

McCormack, T., Brown, G. D. A., Vousden, J. I., & Henson, R. N. A. (2000). Children's serial recall errors: Implications for theories of short-term memory development. *Journal of Experimental Child Psychology, 76,* 222–252.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt.

Mukunda, K. V., & Hall, V. C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence, 16,* 81–97.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Verbal Learning and Verbal Behavior, 64,* 482–488.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Nairne, J. S., & Kelley, M. R. (2004). Seperating item and order information through process dissociation. *Journal of Memory and Language, 50,* 113–133.

Neath, I., & Surprenant, A. M. (2003). *Human memory: An introduction to research, data, and theory* (2nd ed.). Belmont, CA: Wadsworth.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 63,* 301–345.

Pascual-Leone, J., & Johnson, J. (2004). A dialectical constructivist view of developmental intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Understanding and measuring intelligence* (pp. 177–201). New York: Sage.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141–183). Hillsdale, NJ: Erlbaum.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive matrices.* Oxford, England: Oxford Psychologists Press.

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology, 92,* 365–372.

*Rosen, V. M., & Engle, R. W. (1997). Forward and backward serial recall. *Intelligence, 25,* 37–47.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27.

*Süβ, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261–288.

Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9,* 333–348.

Terman, L. M. (1916). *The measurement of intelligence.* Boston: Houghton Mifflin.

*Tirre, W. C., & Pena, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology, 84,* 462–472.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49,* 446–468.

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

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30,* 128–134.

*Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and Raven. *Intelligence, 33,* 67–81.

Unsworth, N., & Engle, R. W. (2006a). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54,* 68–80.

Unsworth, N., & Engle, R. W. (2006b). A temporal–contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language, 54,* 346–362.

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, 114,* 104–132.

*Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505.

Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin, 81,* 695–711.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529–534.

Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cueing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 173–195). New York: Academic Press.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Psychological Corporation.

Wilde, N. J., Strauss, E., & Tulsky, D. S. (2004). Memory span on the Wechsler scales. *Journal of Clinical and Experimental Neuropsychology, 26,* 539–549.

Wright, C. C. (1979). Duration differences between rare and common words and their implication for the interpretation of word frequency effects. *Memory & Cognition, 7,* 411–419.

Xu, G., Meyer, J. S., Thornby, J., Chowdhury, M., & Quach, M. (2002). Screening for mild cognitive impairment (MCI) utilizing combined Mini-Mental-Cognitive Capacity Examinations for identifying dementia prodromes. *International Journal of Geriatric Psychiatry, 17,* 1027–1033.

(*Appendix follows*)

Appendix

Correlations for Confirmatory Factor Analyses and Structural Equation Modeling Analyses

Table A1

Correlations for All Measures in the Engle et al. (1999) Proportion Correct Scoring-Method Analyses

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Ospan	—											
2. Rspan	.52	—										
3. Cspan	.45	.31	—									
4. Bspan	.35	.35	.42	—								
5. Fspans	.38	.29	.32	.43	—							
6. Fspand	.39	.36	.32	.55	.58	—						
7. PMIFR	.29	.22	.27	.20	.23	.30	—					
8. SMIFR	.46	.36	.39	.30	.38	.42	−.02	—				
9. Raven	.34	.28	.34	.27	.22	.19	.18	.25	—			
10. Cattell	.27	.24	.28	.19	.21	.05	.07	.22	.68	—		
11. VSAT	.49	.36	.39	.32	.50	.46	.21	.38	.46	.45	—	
12. QSAT	.47	.26	.44	.34	.45	.37	.29	.39	.61	.59	.74	—

Note. $N = 133$. Ospan = operation span; Rspan = reading span, Cspan = counting span; Bspan = backweard span; Fspans = forward span with phonologically similar words; Fspand = forward span with phonologically dissimilar words; PMIFR = primary memory, immediate free recall; SMIFR = secondary memory, immediate free recall; VSAT = Verbal Scholastic Aptitude Test; QSAT = Quantitative Scholastic Aptitude Test.

Table A2

Correlations for Measures in the Kane et al. (2004) Absolute Scoring-Method Analyses

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Ospan	—														
2. Rspan	.62	—													
3. Cspan	.46	.54	—												
4. Navspan	.42	.47	.51	—											
5. Symspan	.38	.51	.50	.54	—										
6. Rotspan	.38	.48	.55	.60	.64	—									
7. Wspan	.54	.47	.34	.20	.36	.31	—								
8. Lspan	.54	.50	.40	.26	.35	.32	.64	—							
9. Dspan	.49	.49	.39	.23	.39	.39	.63	.64	—						
10. Balspan	.32	.35	.43	.44	.55	.56	.38	.37	.33	—					
11. Arspan	.38	.41	.42	.44	.58	.57	.41	.40	.33	.61	—				
12. Matspan	.31	.36	.35	.41	.62	.52	.38	.33	.36	.50	.53	—			
13. Raven	.24	.26	.21	.24	.28	.38	.22	.33	.24	.48	.44	.31	—		
14. Wasi	.27	.29	.28	.29	.34	.37	.32	.35	.34	.39	.45	.39	.56	—	
15. Beta III	.30	.30	.28	.29	.35	.40	.37	.41	.36	.42	.50	.38	.61	.68	—

Note. $N = 235$. Ospan = operation span; Rspan = reading span, Cspan = counting span; Navspan = navigation span; Symspan = symmetry span; Rotspan = rotation span; Wspan = word span; Lspan = letters span; Dspan = digit span; Balspan = ball span; Arspan = arrow span; Matspan = matrix span; Raven = Raven Advanced Progressive Matrices; Wasi = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning subtest; Beta III = Revised Beta Examination (3rd ed.), Matrix Reasoning subtest.

Received December 4, 2006
Revision received June 21, 2007
Accepted July 5, 2007 ■