MEMORY, 2009, 17 (6), 635–654

Psychology Press
Taylor & Francis Group

Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage

Nash Unsworth
University of Georgia, Athens, GA, USA

Thomas S. Redick
Georgia Institute of Technology, Atlanta, GA, USA

Richard P. Heitz
Vanderbilt University, Nashville, TN, USA

James M. Broadway and Randall W. Engle
Georgia Institute of Technology, Atlanta, GA, USA

Complex span tasks, assumed by many to measure an individual's working memory capacity, are predictive of several aspects of higher-order cognition. However, the underlying cause of the relationships between "processing-and-storage" tasks and cognitive abilities is still hotly debated nearly 30 years after the tasks were first introduced. The current study utilised latent constructs across verbal, numerical, and spatial content domains to examine a number of questions regarding the predictive power of complex span tasks. In particular, the relations among processing time, processing accuracy, and storage accuracy from the complex span tasks were examined, in combination with their respective relationships with fluid intelligence. The results point to a complicated pattern of unique and shared variance among the constructs. Implications for various theories of working memory are discussed.

Keywords: Working memory; Complex span; Fluid intelligence.

Working memory refers to a limited-capacity system responsible for active maintenance, manipulation, and retrieval of task-relevant information that is needed for on-going cognition. Indeed, in their seminal paper Baddeley and Hitch (1974) argued that working memory is used for temporarily storing and carrying out computational processes on mental representations necessary for successful task performance. As such, working memory is thought to be involved in many tasks in which we engage on a daily basis, including language and reading comprehension, novel reasoning, and problem solving. Given the overlap with a number of important areas of psychological study, many researchers have attempted to measure the capacity of working memory and examine its relation to other high-level cognitive processes.

Address correspondence to: Nash Unsworth, Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA. E-mail: nunswor@uga.edu

http://www.psypress.com/memory DOI:10.1080/09658210902998047

REVIEW OF COMPLEX SPAN TASK MEASURES OF WORKING MEMORY CAPACITY

In one of the first major attempts to measure the capacity of the working memory system, Daneman and Carpenter (1980) devised the reading span task. In this task participants were required to read a series of sentences and attempt to recall the last word of each sentence. Daneman and Carpenter argued that this task provides an effective measure of the capacity of working memory because it measures the ability to simultaneously process and store information, which is at the core of the working memory concept. Consistent with this, they found that the reading span was more strongly correlated with a measure of reading comprehension than a task that measured only storage alone (the word span). Daneman and Carpenter concluded that the ability to concurrently store and process information was the key limit to the working memory system and that individuals differed in their ability to simultaneously carry out storage and processing.

Building on these notions, a number of other complex span tasks were subsequently developed to measure the capacity of working memory, all of which have the same basic requirements of storage and processing. For instance, nearly all of these tasks have in common the requirement that the to-be remembered (TBR) items are interspersed with some type of cognitive activity nominally unrelated to the retention of the TBR items. Furthermore, at recall participants are typically required to recall all of the TBR items in the correct serial order (serial recall). Variations of these tasks typically depend on the nature of the processing activity and the nature of the TBR items. For instance, differences in the nature of the processing task include reading or listening to sentences, solving arithmetic problems, counting objects in different colours, deciding whether or not letters are mirror images or not, and judging whether spatial patterns are symmetrical. Differences in the TBR items include digits, letters, words, shapes, and spatial locations, all of which must be remembered in the correct order. Collectively, these types of tasks have become known as complex working memory span tasks, which can be differentiated from simple span tasks that only include the memory component.

Despite many variations in the type of processing task and type of TBR items, all of these tasks have been shown to have both good reliability and validity. Specifically, previous research has shown that these complex span tasks have moderate to high internal consistency estimates (e.g., Conway, Cowan, Bunting, Thierrault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Unsworth, Heitz, Schrock, & Engle, 2005b) and moderate to high test–retest reliabilities (Hitch, Towse, & Hutton, 2001; Klein & Fiss, 1999; Turley-Ames & Whitfield, 2003; Unsworth et al., 2005b). A number of studies have also shown that a variety of these tasks all load highly on the same broad working memory factor in both exploratory and confirmatory factor analyses (e.g., Ackerman, Beier, & Boyle, 2002; Conway et al., 2002; Engle et al., 1999; Kane et al., 2004; Oberauer, Süß, Wilhelm, & Wittmann, 2003). Furthermore, the validity of these tasks has repeatedly been demonstrated by the fact that they tend to correlate with both higher-order and lower-order cognitive processes which are thought to depend on working memory resources (for reviews see Conway, Jarrold, Kane, Miyake, & Towse, 2007; Engle & Kane, 2004; Unsworth & Engle, 2007a,b).

In addition, these tasks have been shown to be related to important phenomena such as early onset Alzheimer's (Rosen, Bergeson, Putnam, Harwell, & Sunderland, 2002), susceptibility to life-event stress (Klein & Boals, 2001), and susceptibility to stereotype threat (Schmader & Johns, 2003; see Unsworth, Heitz, & Engle, 2005a, for a review). Thus, complex working memory span tasks have been used not only to examine basic theoretical conceptions of the capacity of working memory, but also in more applied and clinical situations. Various neuropsychological disorders, including certain aphasias (Caspari, Parkinson, LaPointe, & Katz, 1998), Alzheimer's disease (Kempler, Almor, Tyler, Andersen, & MacDonald, 1998), schizophrenia (Stone, Gabrieli, Stebbins, & Sullivan, 1998), and Parkinson's disease (Gabrieli, Singh, Stebbins, & Goetz, 1996) have been linked to deficits in the capacity of working memory based on scores from various complex span tasks. Finally, complex span tasks have been used extensively in the developmental literature to examine age increases (e.g., Case, Kurland, & Goldberg, 1982; Towse & Hitch, 2007) and age decreases in working memory capacity (Hasher, Lustig, & Zacks, 2007; Salthouse & Babcock, 1991).

WHAT IS THE NATURE OF THE PROCESSING AND STORAGE RELATIONSHIP IN COMPLEX SPAN TASKS?

The fact that complex span tasks are sensitive to so many variations in behaviour points to their utility in a number of domains, and provides the beginnings of an assessment of their nomothetic span. However, in many of these studies the typical measure of interest is the recall score from the storage component, and little regard is given to the various aspects of processing (either speed or accuracy). Theoretically, the processing component should also provide some index of the capacity of working memory and thus should be related to the measures cited previously. That is, if the capacity of working memory is the capacity to simultaneously process and store information (as was the original intent of the complex span tasks) then measures of both processing and storage (or some combination of them) should be examined together in association with the criterion construct of interest. However, to date, much less research has investigated processing accuracy and/or speed relative to the number of studies that have examined the correlation between the storage component of these tasks with some other measure (although see Bayliss, Jarrold, Gunn, & Baddeley, 2003; Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005; Conway & Engle, 1996; Engle, Cantor, & Carullo, 1992; Friedman & Miyake, 2004; Turner & Engle, 1989; Unsworth et al., 2005b; Waters & Caplan, 1996).

Despite the fact that relatively few studies have examined the relationship between processing and storage, a number of theories of complex span performance suggest that specific relations should exist between the processing and storage components. Specifically, some theories suggest that processing time and accuracy should be negatively correlated with storage performance, and processing time and accuracy should mediate the correlation between the span scores and measures of higher-order cognition (e.g., Daneman & Carpenter, 1980; Daneman & Tardif, 1987). One explanation of this proposed relationship is that individuals who are more efficient at processing should be able to quickly arrive at the correct solution. Solving the processing task quickly leaves additional time for rehearsal, which should increase their recall scores on the storage component (e.g., Towse, Hitch, & Hutton, 1998). Participants who are less efficient at processing will have less time to rehearse and subsequently lower span scores. Thus, these theories are quite explicit in predicting a negative relation between processing time and storage accuracy and suggest that this is partially (Towse et al., 1998) or entirely (Daneman & Tardif, 1987) responsible for the relationship between complex span tasks and measures of higher-order cognition. At the same time, however, other theories suggest that processing time and accuracy should not mediate the relation between span scores and measures of higher-order cognition (e.g., Conway & Engle, 1996; Engle et al., 1992). In these views executive abilities, and not processing–storage trade-offs, are the primary cause for the relationship between the complex spans and higher-order cognition. Thus a number of plausible explanations currently exist in terms of the relation between processing and storage within the complex span tasks and their ability to predict higher-order cognition.

The goal of the present investigation was to examine the relation between processing and storage in the complex span tasks, in an attempt to better understand why complex span tasks correlate so well with measures of higher-order cognition. Rather than adopt a particular theory of working memory or attempt to pit various theories against one another, we decided on a more empirical course whereby we addressed a number of questions to gain a better understanding of the important components in the complex span tasks and how these components are related to one another. This method has been beneficial in previous research on complex span tasks. For example, Engle et al. (1992) measured viewing time to each segment of the reading and operation processing tasks and the TBR words. Engle et al. observed that viewing time on the processing tasks did not mediate the relationship between storage and verbal SAT scores. More recently, Friedman and Miyake (2004) illustrated the distinction between experimenter- and participant-paced versions of reading span tasks in terms of their differential prediction of reading comprehension (see also St. Clair-Thompson, 2007b). In both studies empirical investigation of the processing component and the nature of the relationship with higher-order cognition provided important information about the sufficiency of various working memory theories. Thus, although we will not adopt a particular theoretical position, the results should have implications for nearly all theories of complex span task performance and individual differences in WMC.

QUESTIONS GUIDING THE CURRENT RESEARCH

The first question addresses the relationship between processing and storage. As mentioned above, some theories are quite explicit in terms of how processing and storage should be related, while others are less specific. Previous research has shown that indices of processing time and storage span scores measured during complex span performance are either not correlated (Barrouillet & Camos, 2001, Expts. 2 & 3; Engle et al., 1992; Towse, Hitch, & Hutton, 2000; Turner & Engle, 1989; Waters & Caplan, 1996) or are negatively correlated (Barrouillet & Camos, 2001, Expt. 1; Friedman & Miyake, 2004; Hitch et al., 2001; St. Clair-Thompson, 2007a,b; Towse & Hitch, 1995; Towse et al., 1998; Unsworth et al., 2005b). In addition, processing accuracy during complex span tasks is positively related to storage span scores across several studies (Daneman & Tardif, 1987; Engle et al., 1992, Expt. 2; Salthouse, Pink, & Tucker-Drob, 2008; Shah & Miyake, 1996, Expt. 1; Waters & Caplan, 1996), although other studies have observed no relationship between processing accuracy and storage performance (Engle et al., 1992, Expt. 1; Lépine, Barrouillet, & Camos, 2005; Shah & Miyake, 1996, Expt. 2; Towse et al., 2000; Turner & Engle, 1989). Note that the pattern of results is inconsistent with the hypothesis that individuals are engaging in a strategy whereby processing performance is sacrificed for better storage, or vice versa (Carpenter & Just, 1989; see Engle et al., 1992, for description). That is, if participants were purposely engaging in rehearsal during the processing component of the complex span tasks, one would expect to find that processing time is *positively* correlated with storage span scores, and that processing accuracy is *negatively* correlated with storage span scores. Clearly, previous research indicates this is not the typical strategy that participants use for to perform the span tasks.[1]

The second question concerns whether both processing performance measures (accuracy and time) reflect the same construct or whether they reflect different constructs. Early resource theories (Daneman & Tardif, 1987) suggested that processing time and processing accuracy reflect the same construct (i.e., processing efficiency), whereby processing time is traded for processing accuracy, and thus either one could be used as an index of processing. However, it is also possible that processing time and processing accuracy represent different underlying processes, and thus cannot be used interchangeably. Waters and Caplan (1996) observed that the time to read sentences and the number of errors committed on the processing task of reading span were not correlated with each other (for similar results with operation span, see Towse et al., 2000). Furthermore, if processing accuracy and processing time are not redundant measures of processing ability, then they could potentially show relationships of different magnitude and/or direction with higher-order cognition. In fact, Schweizer (2005) has suggested that the reason complex span tasks correlate so well with measures of intelligence is because the processing components of the complex span tasks (i.e., solving maths problems, reading sentences, and rotating spatial figures) are so similar to test items on the numerical, verbal, and spatial reasoning tasks used as criterion measures, and thus simply measure the exact same processes. However, whether this overlapping variance would manifest itself more with processing time, processing accuracy, or a combination of the two from the complex span tasks is an unanswered question.

The third question examines the extent to which processing performance (either accuracy or time) mediates the relationship between storage and higher-order cognition. As mentioned previously,

[1] Two recent studies (Friedman & Miyake, 2004; St. Clair-Thompson, 2007b) have shown that the relationships discussed between processing and storage must be qualified by whether or not the complex span task is participant- or experimenter-administered. When participants are allowed to control the timing of item presentation during complex span tasks, the processing time *positively* correlates with performance on the storage component, indicating that participants in these test situations are in fact altering their processing performance to engage in rehearsal and other mnemonic strategies (see Engle et al., 1992, for similar effect on word span). Importantly, participant-administered complex span scores and processing times do not correlate with higher-order cognition.

one of the key factors differentiating competing working memory theories is the extent to which the correlation between storage and higher-order cognition is mediated by the processing component (Engle et al., 1992; Friedman & Miyake, 2004). Considering the possible outcomes of a mediation analysis, processing could: (a) fully mediate the storage relationship with higher-order cognition, indicating that processing and storage require the same mental operations (e.g., Daneman & Tardif, 1987); (b) partially mediate the storage relationship with reasoning, indicating that processing and storage measure similar and distinct processes (e.g., Waters & Caplan, 1996); or (c) not affect the storage relationship with reasoning, indicating that processing and storage measure entirely distinct cognitive abilities (e.g., Engle et al., 1992). Note that these previous investigations examined the zero-order and partial correlations at the single-task level, instead of using multiple indicators to define a domain-general construct. Although previous studies are inconsistent in regard to this question, recent research (Friedman & Miyake, 2004; St. Clair-Thompson, 2007b, Unsworth et al., 2005b) leads us to expect that the storage relationship with higher-order cognition will remain significant when partialling out either processing time or accuracy.

The fourth question clarifies whether processing adds any predictive power over and above what is accounted for by storage. If processing time and/or accuracy does not fully mediate the storage–reasoning relationship, the possibility remains that processing and storage share considerable variance when predicting higher-order cognitive measures, but that there is also some unique variance accounted for by the processing component. Indeed, previous research has suggested that aspects of processing (time and/or accuracy) in the complex span tasks add to the predictive power when predicting various measures of higher-order cognition (e.g., Friedman & Miyake, 2004, Unsworth et al., 2005b; Waters & Caplan, 1996), although other research suggests that processing performance does not account for unique criterion-related variance (Engle et al., 1992, Expt. 2).

Our final question deals with the inter-relationship among all of the variables in the study, assuming that processing time and processing accuracy are not completely redundant. That is, how are processing time, processing accuracy, storage, and measures of higher-order cognition all related to one another? Despite the studies listed above that have examined the relationship between processing and storage performance, relatively few studies have examined processing accuracy, processing time, and storage accuracy inter-relationships obtained within complex span tasks in the same sample (Engle et al., 1992; Towse et al., 2000; Turner & Engle, 1989; Waters & Caplan, 1996). In addition, no studies have looked at each component's unique and shared contribution in predicting a multiply-determined fluid intelligence composite. This should provide a more fine-grained breakdown of the variance accounted for and suggest both commonalities and differences among the various task components.

DESIGN OF THE CURRENT STUDY

Although previous studies have considered some of the research questions we have outlined (e.g., Engle et al., 1992; Friedman & Miyake, 2004; Waters & Caplan, 1996), so far the results are somewhat mixed. These discrepant findings in the literature could be due to differences in the tasks used (whereby idiosyncratic task effects might bias the results), differences in the samples used (children, young adults, or older adults), as well as problems associated with examining correlation results with small sample sizes (i.e., unstable results). Given these issues, we designed the current study to address a number of limitations of the prior research. First, our interest is in gaining further understanding of the nature of individual differences in working memory capacity as measured by performance on complex span tasks. To this end we used a large sample of young adults to answer the research questions we have outlined. While developmental differences in working memory capacity are interesting in their own right, various studies using complex span tasks have found that manipulations that affect individuals of different ages do not necessarily translate to effects on individuals within the same age range (Towse et al., 1998, vs Towse et al., 2000; see also May, Hasher, & Kane, 1999). It is important to examine these issues in an adult sample to discover how they generalise to individual differences research. In addition, several previous studies have examined processing and storage relationships using extremely small sample sizes for correlational research (Barrouillet & Camos, 2001; Case et al., 1982; Daneman & Tardiff, 1987; Towse et al., 2000), especially those that have utilised factor-analytic techniques (Bayliss et al., 2003; although see Bayliss et al., 2005). Finally, our sample was chosen so as not to be restricted in ability range, a problem that has complicated previous working memory capacity research (see Kane et al., 2004, for further discussion).

As indicated previously, another advantage of the current study is the simultaneous examination of processing time, processing accuracy, and storage performance within the same sample. Investigations of processing-and-storage relationships have almost exclusively focused on how processing time affects recall of the TBR items (but see Conway & Engle, 1996; Daneman & Tardif, 1987; Engle et al., 1992; Waters & Caplan, 1996). These studies and others imply that processing accuracy is an often-overlooked component of complex span measures, probably because relatively few errors are committed on the various processing tasks. Despite relatively high overall accuracy, significant correlations between processing accuracy and both storage accuracy and higher-order cognition have been obtained (Engle et al., 1992; Daneman & Tardif, 1987; Waters & Caplan, 1996), suggesting that there is sufficient variability in performance of the processing tasks that is important to further explore.

Several previous studies have attempted to use single-task (processing-only) performance to mediate the relationship between storage performance and higher-order cognition (Bayliss et al., 2003, 2005; Case et al., 1982; Daneman & Tardif, 1987; Engle et al., 1992; Hitch & McAuley, 1991; Salthouse & Babcock, 1991). While examination of single-task processing performance may be interesting in testing particular hypotheses, especially considering Daneman and Tardif's argument that processing tasks alone are sufficient for measuring working memory capacity adequately, these approaches seem somewhat ill-suited to learn about how processing performance *during* the complex span tasks affects the ability to recall the TBR items. As noted by Towse and Hitch (1995), measuring processing performance during complex span tasks "has the advantage of measuring the rate at which [count] operations were executed rather than at the maximum rate at which operations could have occurred" (p. 116). This seems similar to an issue encountered in the human performance literature: Does performance on a single task performed alone (so called pure blocks) inform us about what occurs when the task is performed in conjunction with another task (Schumacher et al., 2001; Tombu & Jolicæur, P, 2004)? The cynical reader may note that this appears to be the approach we have taken in previous investigations of working memory capacity, by partialling out single-task storage performance (simple span tasks) from the complex span and fluid intelligence relationship (e.g., Engle et al., 1999). However, note that in this situation we have measures of storage performance in both the single- and dual-task conditions, whereas the majority of the studies listed above only measured processing performance independently of the complex span task. Because our primary interest is gaining further understanding of how the complex span task components combine to account for higher-order cognition, our processing indices are obtained during the complex span tasks themselves.

In a similar vein, recently various modifications of the processing component in complex span tasks have been investigated to advance specific theories of working memory capacity (Barrouillet, Bernadin, & Camos, 2004; Barrouillet, Bernadin, Portrat, Vergauwe, & Camos, 2007; Barrouillet & Camos, 2001; Hitch et al., 2001; Maehara & Saito, 2004; Saito & Miyake, 2004; Towse et al., 1998, 2000). While these approaches may be informative regarding these precise theories, they do not do as much to further our knowledge about why individual differences in complex span scores predict higher-order cognition as they are *typically* measured. For instance, how do these theories account for spatial processing and memoranda predicting criterion-related ability tests? What about the recent view we have advocated that indicates processing may not be necessary to obtain high correlations between span tasks and higher-order cognition (Unsworth & Engle, 2007a)? We wanted to measure processing and storage performance on commonly used complex span tasks to answer the research questions we have outlined. The complex span tasks used here are commonly used throughout the literature and have been administered to thousands of participants, both in our own research labs and in others (Salthouse et al., 2008; M. J. Kane, personal communication, 7 October 2008).

Again, while we acknowledge that previous research has examined processing and storage relationships, these studies have tended to focus on particular tasks such as reading span and have not examined these issues at the latent-variable level (Engle et al., 1992; Friedman & Miyake, 2004; Waters & Caplan, 1996; but see Bayliss et al., 2005). Even the few studies (Bayliss et al., 2003, 2005; St. Clair-Thompson, 2007b) that have examined multiple complex span tasks and criterion-ability measures have not investigated the full combination of numerical, verbal, and spatial domains, as is advocated in both intelligence (Marshalek, Lohman, & Snow, 1983) and working memory capacity (Oberauer, Schultze, Wilhlem, & Süß, 2005) research. In the present study we utilised a latent-variable approach to examine the relations among processing, storage, and higher-order cognition. This was done because previous results may be due to the fact that only a single task was used and thus may not provide the best evidence for more general principles across content domains. In order to derive latent variables for the constructs of interest, we used multiple indicators of each construct. The different components from the complex span tasks were derived from the operation span (Turner & Engle, 1989), reading span (Daneman & Carpenter, 1980), and symmetry span tasks (Kane et al., 2004). Thus, the processing component from each task could be numerical, verbal, or spatial, and in each task we examined processing time, processing accuracy, and storage accuracy. Instead of using a single complex span task, we extracted the common variance from three complex span tasks for each component. To examine how these components were related to higher-order cognition, we had participants perform a number of general fluid (gF) abilities tests representing different content domains (i.e., spatial, numerical, and verbal). As noted previously, a great deal of research has suggested that the storage component of complex working memory span measures shares a substantial amount of variance with measures of gF (e.g., Ackerman, Beier, & Boyle, 2005; Conway et al., 2002; Engle et al., 1999; Kane et al., 2004). Thus, the present investigation sought to simultaneously examine the relationships among various storage and processing components within complex span tasks and their relation to fluid abilities. Given this design, the current study provides a fairly unique dataset to examine the predictive power of complex span tasks.

METHOD

Participants

A total of 138 participants were recruited from the subject-pool at Georgia Institute of Technology and from the Atlanta, Georgia community through newspaper advertisements. Participants were between the ages of 18 and 35 and received either course credit or monetary compensation for their participation. Each participant was tested individually in two laboratory sessions lasting approximately 1 hour each.

Materials and procedure

After signing informed consent, all participants completed automated versions of the operation span (Ospan) task, the reading span (Rspan) task, the symmetry span (Symspan) task, and a brief computerised version of the Raven progressive matrices (Raven; Raven, Raven, & Court, 1998) in Session 1. In Session 2 all participants completed the Rspan task and the Symspan task again to obtain estimates of test–retest reliability as well as several paper-pencil reasoning measures. These included Inferences, Number Series, Surface Development, Verbal Analogies, and Necessary Arithmetic Operations. Raven and Surface Development test represented Spatial Reasoning, while Inferences and Verbal Analogies represented Verbal Reasoning, and Number Series and Necessary Arithmetic Operations represented Numerical Reasoning. All tasks were administered in the order listed above.

Tasks

Ospan. Participants solved a series of maths operations while trying to remember a set of unrelated letters (F, H, J, K, L, N, P, Q, R, S, T, Y). Before beginning the real trials, participants performed three practice sections. The first practice was simple letter span. A letter appeared on the screen and participants were required to recall the letters in the same order as they were presented. In all experimental conditions, letters remained on-screen for 1000 ms. At recall, participants saw a 4×3 matrix of letters. Recall consisted of clicking the box next to the appropriate letters (no verbal response was required) in the correct order. Participants had as much time as needed to recall the letters. After recall, the computer provided feedback about the number of letters correctly recalled in current set. Next, participants performed the maths portion of the task alone. Participants first saw a math operation [e.g. $(1*2) + 1 = ?$]. Participants were instructed to solve the operation as quickly as possible and then click the mouse to advance to the next screen. On the next screen a digit (e.g., "3") was presented and the participant was required to click either a "True" or "False" box to indicate the answer. After each operation participants were given accuracy feedback. The math practice served to familiarise participants with the maths portion of the task as well as to calculate how long it would take that person to solve the maths operations. Thus, the maths practice attempted to account for individual differences in the time required to solve maths operations without an additional storage requirement. After the maths alone section, the program calculated each individual's mean time required to solve the equations. This time (plus 2.5 standard deviations) was then used as the maximum time allowed for the maths portion of the dual-task section for that individual.

The final practice session had participants perform both the letter recall and maths portions together, just as they would do in the real block of trials. As before, participants first saw the maths operation and then clicked to advance to the comparison (True/False) screen. After they clicked the mouse button indicating that the response, the TBR letter was shown. If a participant took more time to solve the operations than their average time plus 2.5 *SD*, the program automatically moved on and counted that trial as an error. Participants completed three practice trials each of set-size two. After participants completed all of the practice sessions, the program progressed to the real trials. The real trials consisted of three trials of each set-size, with the set-sizes ranging from three to seven. This made for a total of 75 letters and 75 maths problems. Note that the order of set-sizes was random for each participant. The score was the number of correct items recalled in the correct position.

Rspan. Participants were required to read sentences while trying to remember the same set of unrelated letters as Ospan. As with the Ospan, participants completed three practice sessions. The letter practice was identical to the Ospan task. In the processing-alone session participants were required to read a sentence and determine whether the sentence made sense (e.g. "The prosecutor's dish was lost because it was not based on fact.?"). As with the Ospan, the time to read the sentence and determine whether it made sense was recorded and used as an overall time limit on the real trials. The final practice session combined the letter span task with the sentence task just like the real trials. In the real trials participants were required to read the sentence and to indicate whether it made sense or not. Half of the sentences made sense while the other half did not. Nonsense sentences were made by simply changing one word (e.g. "dish" from "case") from an otherwise normal sentence. There were 10–15 words in each sentence. After participants gave their response they were presented with a letter for 1000 ms. At recall, letters from the current set were recalled in the correct order by clicking on the appropriate letters. There were three trials of each set-size with list length ranging from three to seven. The same scoring procedure as Ospan was used.

Symspan. In this task participants were required to recall sequences of red squares within a matrix while performing a symmetry-judgement task. In the storage alone practice session participants saw sequences of red squares appearing in the matrix and at recall were required to click the correct locations in the matrix in the correct order. In the symmetry-judgement task alone session participants were shown an 8×8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical approximately half of the time. The same timing parameters used in the Ospan and Rspan were used. The final practice session combined the matrix recall with the symmetry-judgement task. Here participants decided whether the current matrix was symmetrical and then were immediately presented with a 4×4 matrix with one of the cells filled in red for 650 ms. At recall, participants recalled the sequence of red-square locations in the preceding displays, in the order they appeared by clicking on the cells of an empty matrix. There were three trials of each set-size with list length ranging from two to five. The same scoring procedure as Ospan and Rspan was used.

Raven Progressive Matrices. The Raven is a measure of abstract reasoning (Raven et al., 1998). This version of the Raven is a brief computer-administered version that consists of 12 items. Each item consisted of a matrix of geometric patterns with the bottom-right pattern missing. Participants were instructed to select from among either six or eight alternatives the one that correctly completed the overall series of patterns. Each matrix item appeared separately on screen along with the response alternatives. Using the mouse, the participant simply clicked on the response that they thought completed the pattern. The mouse click registered the response and moved the program onto the next problem. Participants were allotted 5 minutes to complete the task. Items were presented in ascending order of difficulty (i.e., the easiest item is presented first and the hardest item is presented last). A participant's score was the total number of correct solutions. Participants received two practice problems.

Surface development. In this task participants were presented with an illustration of a piece of paper that can be made into a three-dimensional object corresponding to a shape next to it when folded. Some of the edges on the unfolded paper were marked with letters and some of the edges on the folded shape were marked with numbers. Participants were required to match the lettered edges on the unfolded paper to the corresponding numbered edges on the folded shape. The test consisted of five unfolded and folded shapes, each with five numbered edges that required responses. After completing one practice problem, participants had six minutes to complete all 25 test items. These items represented Part 1 from the Educational Testing Service (ETS) version (Ekstrom, French, Harman, & Derman, 1976). A participant's score was the total number of items solved correctly.

Verbal inferences. In this task participants read a brief (one- to three-sentence) passage about a topic and were instructed to choose the conclusion (out of five presented) that could be inferred from the passage without assuming any additional information or knowledge. Participants received one practice item and ten real items. The items were selected from Part 1 of the ETS version (Ekstrom et al., 1976). Participants had 6 minutes to complete the task. A participant's score was the total number of items solved correctly.

Verbal analogies. In this task participants read an incomplete analogy and were required to select the one word out of five possible words that best completed the analogy. After one practice item, participants had 5 minutes to complete 18 test items. These items were originally selected from the Air Force Officer Qualifying Test (Berger, Gupta, Berger, & Skinner, 1990), and we used the same subset of items used in Kane et al. (2004). A participant's score was the total number of items solved correctly.

Number series. In this task participants saw a series of numbers and were required to determine what the next number in the series should be (Thurstone, 1962). That is, the series follows some unstated rule which participants are required to figure out in order to determine which the next number in the series should be. Participants selected their answer out of five possible numbers that were presented. Following five practice items, participants had 4.5 minutes to complete 15 test items. A participant's score was the total number of items solved correctly.

Necessary arithmetic operations. In this task participants were presented with maths story problem and, instead of answering the problem, they were required to indicate how it should be solved (i.e., using addition, subtraction, multiplication, etc.). Some problems required only one operation, while others required two operations. When answers required two operations, the operations were given in the order in which they should be performed. Following two practice problems, participants had 5 minutes to complete 15 test items. The items were selected from Part 1 of the ETS version (Ekstrom et al., 1976). A participant's score was the total number of items solved correctly.

RESULTS

Of the 138 participants, 75 were women and 63 were men with a mean age of 22.6 years ($SD =$ 4.44 years). Furthermore, 76% of these participants were currently enrolled college students at either Georgia Institute of Technology (71 participants), or at another Atlanta area university (34 participants). The other 33 participants were paid community volunteers who were not currently college students.

Descriptive statistics for the memory and reasoning tasks are shown in Table 1. For each of the three span tasks, three performance indices were measured. These were the total number of items recalled on the storage portion of the task (recall), proportion correct on the processing component of the task (processing accuracy), and mean of the median time to correctly complete the processing component of the task (processing time).[2] Scores for the reasoning tasks

[2] Given that some participants demonstrated either floor or ceiling performance for the recall scores, we removed these participants ($n = 8$) and reanalysed the data. All of the results were virtually identical to those reported in the paper.

TABLE 1
Descriptive statistics and reliability estimates for memory and reasoning measures

Measure	*M*	*SD*	*Range*	*Skew*	*Kurtosis*	α
OspanR	55.42	14.91	3–75	−1.22	1.41	.83
SymspanR	26.91	8.65	1–42	−.63	.22	.80
RspanR	51.60	16.75	2–75	−1.06	.76	.89
OspanA	.90	.09	.48–1.0	−2.79	9.70	.85
SymspanA	.93	.09	.50–1.0	−2.56	7.66	.84
RspanA	.93	.07	.56–1.0	−2.74	9.62	.83
OspanT	3451	1163	1389–7831	1.24	2.04	.98
SymspanT	2416	888	1191–7120	1.51	4.82	.99
RspanT	4036	1284	2245–10561	1.87	6.47	.98
Raven	8.19	2.07	3–12	−.48	−.45	.66
SurfDev	16.09	8.77	1–30	.15	−1.37	.94
Inference	6.09	2.43	0–10	−.30	−.72	.71
Analogy	10.40	3.99	1–18	−.50	−.36	.82
NumSer	8.30	3.10	0–14	−.19	−.61	.78
NAO	8.07	2.96	1–15	−.15	−.73	.77

Ospan = operation span; Symspan = symmetry span; Rspan = reading span; R = recall component; A = processing accuracy component; T = processing time component; Raven = Raven Progressive Matrices; SurfDev = Surface development; Inference = Verbal inferences; Analogy = verbal analogies; NumSer = number series; NAO = Necessary arithmetic operations.

were the total number of correct items. Cronbach's alpha was computed for each measure at the level of individual items. For the measures derived from the complex span task, item scores were the number recalled for a given trial (for the recall component) or the proportion of processing items solved correctly for a given trial (for the processing accuracy component). Therefore, for the Ospan and Rspan tasks there were 15 total trials, and for the Symspan task there were 12 trials. For the reasoning tasks item scores were binary (correct vs incorrect). As can be seen in Table 1, all measures had generally acceptable values of internal consistency.[3]

Most of the measures were approximately normally distributed with values of skewness and kurtosis under the generally accepted values (i.e., skewness < 2 and kurtosis < 4; see Kline, 1998), except for the processing accuracy measures. These measures had high skewness and kurtosis values. Therefore, in order to obtain more normally distributed values, we transformed the three processing accuracy measures using an arcsin transformation suggested by Stevens (2002). This led to more generally acceptable values of skewness and kurtosis. These transformed measures were used in all subsequent analyses. These transformed values were used in all subsequent analyses. We also examined multivariate kurtosis with Mardia's coefficient. Although Mardia's index was high (10.79), removing two participants reduced it down to an acceptable value (3.30). Removing these two participants led to exactly the same results as the full sample, therefore all reported analyses are based on the full sample. Correlations, shown in Table 2, were moderate to large in magnitude, irrespective of the particular content domain.[4]

Confirmatory factor analyses

In order to examine the main questions of interest, several confirmatory factor analyses (CFAs) were conducted to examine the structure of the data. Specifically, we examined whether the

[3] In addition to measuring internal consistency we also measured test–retest reliability for the recall components for the Rspan and Symspan tasks to compare with a previous estimate of test–retest reliability for the Ospan task (Unsworth et al., 2005b). The correlation from Time 1 to Time 2 for Rspan was .82, and for Symspan was .77 (*M* time between testing = 49.76 days, *Med* time between testing = 6 days). These values compare well with the test–retest reliability for Ospan (.83) from Unsworth et al. (2005b). Note that all remaining analyses utilised Symspan and Rspan performance at Time 1 only.

[4] Because the three complex span tasks were mouse-driven tasks, we also administered a task to assess mouse skill. In this task participants saw a square appear randomly at one of four locations onscreen. Participants were required to click on the square as quickly as possible. Response time and errors (i.e., not clicking directly on the square) were recorded. All analyses were rerun after partialling out potential differences in mouse skill. All results were exactly the same as the reported results.

TABLE 2
Correlations for memory and reasoning measures

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1. OspanR	—														
2. SymspanR	.66	—													
3. RspanR	.77	.70	—												
4. OspanA	.47	.35	.40	—											
5. SymspanA	.24	.42	.28	.24	—										
6. RspanA	.30	.34	.45	.45	.51	—									
7. OspanT	−.37	−.26	−.24	−.17	−.22	−.35	—								
8. SymspanT	−.41	−.43	−.29	−.20	−.22	−.22	.55	—							
9. RspanT	−.32	−.31	−.34	−.40	−.13	−.34	.50	.58	—						
10. Raven	.49	.51	.52	.30	.37	.46	−.41	−.41	−.32	—					
11. SurfDeve	.43	.53	.48	.40	.30	.44	−.38	−.47	−.39	.53	—				
12. Infer	.41	.47	.48	.35	.38	.50	−.41	−.57	−.48	.52	.57	—			
13. Analogy	.49	.52	.57	.45	.45	.61	−.38	−.50	−.52	.59	.69	.65	—		
14. NumSer	.42	.45	.40	.43	.22	.45	−.51	−.46	−.51	.48	.65	.45	.59	—	
15. NAO	.53	.53	.47	.44	.32	.48	−.51	−.53	−.52	.56	.67	.62	.66	.70	—

Ospan = operation span; Symspan = symmetry span; Rspan = reading span; R = recall component; A = processing accuracy component; T = processing time component; Raven = Raven Progressive Matrices; SurfDev = surface development; Inference = verbal inferences; Analogy = verbal analogies; NumSer = number series; NAO = necessary arithmetic operations. Correlations greater than .22 are significantly different from zero at the $p < .01$ level, and correlations greater than .17 are significantly different from zero at $p < .05$ level.

data were better represented by a one-factor WM model (with all WM measures loading on a single factor), a two-factor model differentiating processing and storage (with all the recall measures loading on one factor and all the processing time and processing accuracy measures loading on another factor), or a three-factor model differentiating the three WM components (recall, processing time, and processing accuracy). Additionally, for each model a gF latent variable was formed based on the six separate gF measures, and the correlations between the WM components and the gF factor were examined. Latent variable analyses were conducted with Lisrel 8.80 (Jöreskog & Sörbom, 2006). Model fits were assessed via the combination of several fit statistics. These include chi-square, root mean square error of approximation, standardised root mean square residual, normed fit index, the non-normed fit index, and the comparative fit index. The chi-square statistic reflects whether there is a significant difference between the observed and reproduced covariance matrices, and thus non-signifcant values are desirable. However, with large sample sizes even slight deviations can result in a significant value. Therefore we also report the root mean square error of approximation (RMSEA) and the standardised 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 comparative fit index (CFI), all of which compare the fit of the specified model to a baseline null 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) suggest that the combination of fit indices such as CFI > .95 and SRMR < .05 are the best indicators of model fit.

Shown in Table 3 are the fit statistics for the resulting models. As can be seen, the three-factor model fitted significantly better than either the one-factor model, $\Delta\chi^2(5) = 154.92$, $p < .01$, or the two-factor model, $\Delta\chi^2(3) = 58.03$, $p < .01$. We also examined an alternate three-factor model in which all the components from a given complex span task loaded on separate factors (i.e., all of the components from the Ospan on one factor, all the components from the Symspan on another factor, and all of the components from the Rspan on another factor). This was done to examine the extent to which the data simply reflected the different tasks used. As shown in Table 3, the fit of this model was poor, and thus the original three-factor model was favoured. Finally, we examined a CFA based on the content of all of the tasks with one factor representing all of the verbal measures, one factor representing all of the numerical measures, and one factor representing all of the spatial measures for both the WM and gF

TABLE 3
Fit indices for the confirmatory analyses and structural equation models

Model	χ^2	*df*	*RMSEA*	*NFI*	*NNFI*	*CFI*	*SRMR*
One-factor CFA	340.11	89	.14	.89	.90	.92	.08
Two-factor CFA	243.22	87	.11	.92	.93	.95	.07
Three-factor CFA	185.19	84	.09	.94	.95	.96	.06
Alt Three-factor CFA	282.96	84	.13	.90	.91	.93	.08
Content CFA	315.43	87	.14	.85	.90	.92	.08
Recall-PTime SEMa	124.14	52	.10	.95	.96	.97	.07
Recall-PTime SEMb	96.26	51	.08	.96	.97	.98	.05
Recall-PAcc SEMa	111.26	52	.09	.95	.96	.97	.06
Recall-PAcc SEMb	107.40	51	.09	.95	.97	.97	.05

CFA = confirmatory factor analysis: SEM = structural equation model; RMSEA = root mean square error of approximation; NFI = normed fit index; NNFI = nonormed fit index; CFI = comparative fit index; SRMR = standardised root mean square residual.

measures. This was done to see if the structure of the data simply reflected the content of the tasks rather than potential differences attributed to processing and storage components. As shown in Table 3 the fit of this model was poor, and the fit was significantly worse than the fit of the endorsed three-factor model, $\Delta\chi^2(3) = 130.24$, $p < .01$. The endorsed three-factor model is shown in Figure 1.[5] As can be seen in the model, Recall and Processing Time were moderately negatively correlated. Also as shown in Figure 1, Recall and Processing Accuracy were moderately positively correlated, and Processing Time and Processing Accuracy were moderately negatively correlated. Additionally, the Recall latent variable, the Processing Time latent variable, and the Processing Accuracy latent variable were all strongly correlated with the gF latent variable. In fact, each component from the complex span task was more highly correlated with gF than they were with each other. Indeed, Processing Accuracy and Processing Time both correlated with gF at .80, but they were only correlated with each other at –.49. Thus, it would seem that Processing Accuracy and Processing Time are not redundant and do not provide the same index of processing efficiency. Rather, Processing Accuracy and Processing Time seem to be measuring slightly different processes, which are associated with fluid abilities.

Mediation analyses

Next we examined whether performance on the processing component would mediate the relation between storage span scores and measures of higher-order cognition. For the mediation analyses we relied on structural equation modelling to test

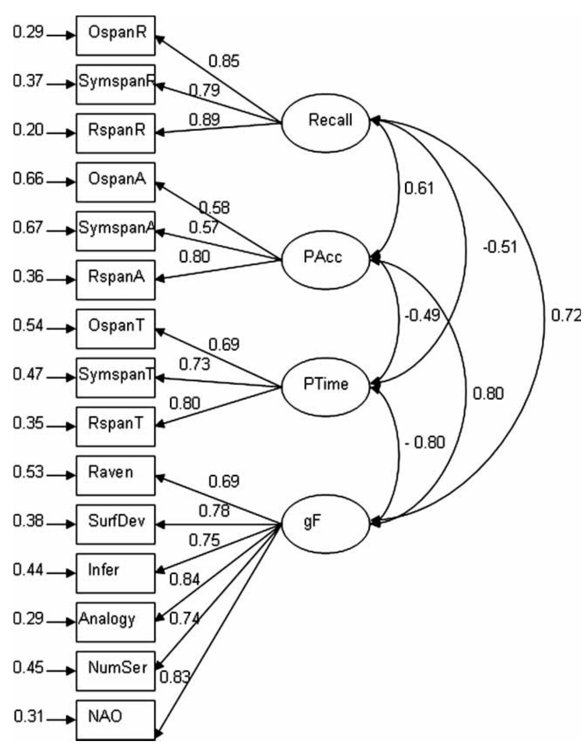

Figure 1. Confirmatory factor analysis for Recall, Processing Accuracy (PAcc), Processing Time (PTime), and general fluid intelligence (gF). Paths connecting latent variables (circles) to each other represent the correlations between the constructs, the numbers from the latent variables to the manifest variables (squares) represent the loadings of each task onto the latent variable, and numbers appearing next to each manifest variable represent error variance associated with each task.

[5] Note that the model fits could have been improved if the error variances for each component from the span tasks were allowed to correlate (i.e., all of the errors associated with the Ospan). Doing so did not change any of the parameter values or the relative fit of the models. Thus, the simpler, non-correlated error models were used throughout.

for full and partial mediation as suggested by James, Mulaik, and Brett (2006). In order to examine whether processing time would mediate the relation between the storage component and gF, two structural equation models (SEMs) were specified. In the first model, the Recall latent variable was allowed to have a direct effect on the Processing Time latent variable, but no direct effect on gF. However, the Processing Time latent variable was allowed to have a direct effect on gF. Thus, the first model tested whether Processing Time fully mediates the relation between Recall and gF. The resulting model is shown in Figure 2a and fit statistics are presented in Table 3. As can be seen there is a strong effect of Recall to Processing Time and a strong effect of Processing Time to gF. Next, we examined a partial mediation model where the same model was used, but now the direct path from Recall to gF was freed. The resulting model is shown in Figure 2b and fit statistics are presented in Table 3. As can be seen, in the second model both Processing Time and Recall significantly predicted gF and Recall significantly predicted Processing Time, all evidence in favour of partial mediation. In fact, the difference in chi-square for the two models was significant, $\Delta\chi^2(1) = 27.88$, $p < .05$, suggesting that the partial mediation model fitted the data significantly better than the full mediation model. This indicates that variation in Processing Time does not fully account for the relation between Recall and gF.

We also examined the extent to which Processing Accuracy would either fully or partially mediate the relationship between Recall and gF. The full mediation model is shown in Figure 3a and fit statistics are presented in Table 3. As shown in Figure 3a there is a strong effect of Recall to Processing Accuracy and strong effect from Processing Accuracy to gF. The fit of the model was acceptable. Next the partial mediation model where the direct path from Recall to gF was freed was examined. The partial mediation model is shown in Figure 3b and fit statistics are presented in Table 3. Consistent with partial mediation, both Processing Accuracy and Recall significantly predicted gF and Recall significantly predicted Processing Accuracy. Furthermore, and consistent with the Processing Time analyses, the partial mediation model fit the data significantly better than the full mediation model, $\Delta\chi^2(1) = 3.86$, $p < .05$. These results indicate that storage span scores and indices of processing performance account for both shared and unique variance in gF.

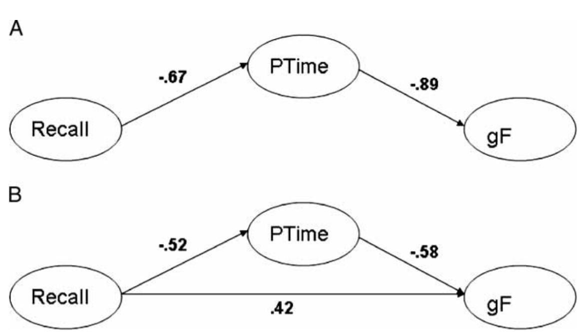

Figure 2. Structural equation model analysis for Recall, Processing Time (PTime), and general fluid intelligence (gF). (A) Recall–gF effect fully mediated by PTime. (B) Recall–gF effect partially mediated by PTime.

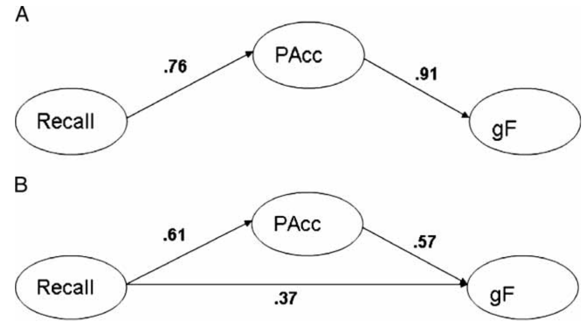

Figure 3. Structural equation model analysis for Recall, Processing Accuracy (PAcc), and general fluid intelligence (gF). (A) Recall–gF effect fully mediated by PAcc. (B) Recall–gF effect partially mediated by PAcc.

Variance partitioning analyses

To explore the shared and unique contribution of each latent component with gF further, we utilised variance partitioning methods that have been used previously (e.g., Chuah & Maybery, 1999; Cowan et al., 2005; Unsworth & Engle, 2006). Variance partitioning, or communality analysis (Pedhazur, 1997), attempts to allocate the overall R^2 of a particular criterion variable (here gF) into portions that are shared and unique to a set of predictor variables (here Recall, Processing Accuracy, and Processing Time). A series of regression analyses was carried out to obtain R^2 values from different combinations of the predictor variables (see Table 4) in order to partition the variance. For each variable entering into the regression, a factor composite was computed for all the measures making up that factor. For instance, for the Recall composite recall scores, each of the three complex span tasks was entered into a factor analysis and factor scores were

TABLE 4
R^2 values for regression analyses predicting gF for various predictor variables

Predictor variables	R^2	*F*
1. Recall, PTime, PAcc	.69	98.61
2. Recall, PTime	.61	104.63
3. Recall, PAcc	.55	82.64
4. PTime, PAcc	.61	106.26
5. Recall	.42	97.27
6. PTime	.45	111.58
7. PAcc	.37	80.84

All R^2 values are significant at $p < .01$. PTime = Processing Time component; PAcc = Processing Accuracy component.

computed for each participant. These factor scores were then used in all regression analyses.[6]

As shown in Figure 4, the results suggest that 69% of the variance in gF is accounted for by the three components derived from the complex span tasks. Furthermore, the results suggest a complicated breakdown of the variance. The largest chunk of gF variance was accounted for by variance shared among all three components, but substantial gF variance was also accounted for by variance shared between only two of the components as well as by variance unique to each component. Similar to the previous analysis, although processing accuracy and processing time contain some overlapping variance in predicting gF, each component also contains unique variance to contribute to the prediction gF, suggesting that they are not identical indices of processing efficiency. Importantly, consistent with previous research, the recall/storage scores accounted for 42% of the variance in gF, but partialling out either processing accuracy or processing time led to a substantial reduction in the predictive power of the recall scores. This suggests that part of the reason that the recall correlates so with higher-order cognition is because of the shared variance between recall and aspects of processing. Thus, although storage and processing measures have been shown to be related to gF before, the current results suggest that these measures can be further broken down into various components, each of which are important in predicting gF.

[6] Using z-score composites led to nearly identical results as using the factor composites.

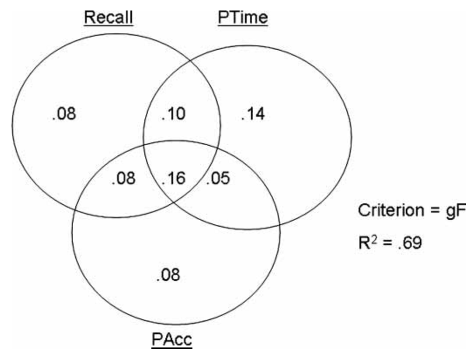

Figure 4. Venn diagrams indicating the amount of variance accounted for in gF by the Recall component, the Processing Time Component (PTime), and the Processing Accuracy component (PAcc). Numbers are based on regressions from Table 4.

GENERAL DISCUSSION

The goal of the present study was to examine the relationship between processing and storage components in the complex working memory span tasks to better understand why complex span tasks are effective in predicting higher-order cognitive abilities. To this end we examined the relationships among processing time, processing accuracy, and the number of items recalled from the storage component of multiple complex span tasks, along with a multiply-determined fluid abilities composite. Five primary questions were asked in order to guide our analyses. Each question will be addressed in turn.

(1) How are processing and storage related?

The results suggested that processing time and storage were negatively related, consistent with a number of the studies mentioned previously (e.g., Bayliss et al., 2005; Friedman & Miyake, 2004; Unsworth et al., 2005b). This suggests that participants who worked quickly on the processing components tended to remember more items than participants who worked more slowly on the processing components. This result is very much in line with several theories of working memory that suggest that processing and storage compete for a limited resource (Daneman & Carpenter, 1980), or suggest that the more time that is spent on the processing component the greater the opportunity for items to be forgotten and less time for rehearsal/refreshing processes (Towse et al., 1998). Furthermore, although not typically examined, it was found that processing accuracy was positively related to storage/recall. Thus, participants who were both fast and accurate on the processing component of the complex span tasks also tended to recall more TBR items. The pattern of results observed here, whereby those participants who performed the processing tasks more quickly and more accurately also remembered the most items, is wholly inconsistent with the strategic allocation hypothesis discussed earlier. Note that this notion of processing–storage trade-offs is also inherent in complex span-scoring procedures that exclude data for participants below a criterion level of processing accuracy (e.g., 85%; Conway et al., 2005) or exclude items in which the processing component was not answered correctly (e.g., Hambrick & Oswald, 2005). The positive relationship between processing accuracy and storage performance suggests that these scoring procedures are unnecessary.

(2) Are both processing components (accuracy and time) the same?

The second question that guided our analyses was whether both processing components reflect the same construct or whether they reflect different constructs. As noted previously, the early resource theories suggest that both processing time and processing accuracy provide some index of processing efficiency, and thus either processing accuracy or processing time can be used (Daneman & Tardif, 1987). The results, however, indicate that processing accuracy and processing time are not redundant, but rather index different constructs. In particular, although both components show similar correlations with the storage component and with gF, the two were only moderately related (i.e., $r = -.49$). In fact, as pointed out previously, the two processing components are more highly correlated with gF than they are with each other. As an illustrative example, accuracy on the maths operations in Ospan was more strongly correlated with the number of items correctly solved on the verbal analogies test ($r = .45$) than with the time to solve the maths operation ($r = -.17$). Furthermore, the results from the variance partitioning analysis suggested that both components accounted for significant variance in gF independent of the other component. Thus, the results suggest that processing time and processing accuracy do not reflect the same underlying construct (processing efficiency), but rather index two slightly different constructs. We thus agree with Waters and Caplan (1996) that inspection of both processing accuracy and time can be informative when analysing the results of complex span performance, but disagree with them in their approach of combining processing performance into one composite score, given that the two indices do not approach unity in their relationship.

(3) Does processing (either accuracy or time) mediate the relation between storage and higher-order cognition?

We examined whether processing would mediate the correlation between storage and higher-order cognition. In many ways this is the key question because several theories predict that the correlation between storage and higher-order cognition should be mediated by both of the processing components, while other theories suggest that the correlation between the storage component and higher-order cognition should not be mediated by either of the processing components (see Engle et al., 1992, and Friedman & Miyake, 2004, for reviews). In order to answer this question we tested mediation models using SEM for both the processing time and processing accuracy components. In both cases a partial-mediation model fit the data better than a full-mediation model, meaning that the processing components do not fully mediate the relation between storage and fluid intelligence. This result is consistent with previous work suggesting that the storage component is related to various measures of higher-order cognition even after controlling for processing performance (e.g., Conway & Engle, 1996; Engle et al., 1992; Friedman & Miyake, 2004; Unsworth et al., 2005b).

(4) Does processing (either accuracy or time) add any predictive power over and above what is accounted for by storage/recall?

Because the processing components do not share all of the same variability as the storage component in predicting higher-order cognition, as indicated above, then one or both processing measures should add predictive power over and above what is accounted for by performance on the storage component. As shown in the SEM analyses, both processing time and processing accuracy predicted significant variance in gF over and above what was accounted for by the storage component. Thus, both processing components add incremental validity when predicting higher-order cognition. Indeed, the variance partitioning analyses suggested that each processing components added substantial unique variance when predicting gF above that accounted for by the storage component and the other processing component. This is particularly interesting given that processing accuracy is usually assumed to be very high and thus, have little systematic variability associated with it (e.g., Conway et al., 2005, p. 774). However, in the current study it was found that processing accuracy had a substantial relation with gF. Thus, processing accuracy provides unique information over and above storage and processing time. However, we reiterate that the additional predictive utility of the processing components of the complex span tasks does not fully account for the storage–gF relationship.

(5) What is the relationship among all of the variables?

The final issue we addressed was to determine the relationship among all of the performance indices derived from the complex span task in predicting gF. That is, given that both processing components provide unique information and are not redundant, we wished to examine the relationship among all of the variables in the current study in order to provide a more fine-grained breakdown of the variance accounted for by each component. We examined this issue via CFA and variance partitioning for the two processing measures, the storage component, and gF. Somewhat surprisingly, the results suggested that all three of the components derived from the complex span tasks were more highly related to gF than they were to each other. Furthermore, the variance partitioning analyses suggested that there was a substantial amount of both shared and unique variance accounted for in gF by the three components. In particular, of the 69% of the total variance accounted for in gF by the three processing components, only 16% was shared by all three components, while the rest was fairly evenly accounted for by a combination of two components or by one component alone. It is clear that neither of the two processing components fully mediated the relation between storage/recall and gF, but at the same time there is a large amount of overlapping variance that is accounted for when the processing components are taken into account.

This suggests that the complex span tasks are multifaceted tasks that rely on many components, all of which are related to fluid intelligence (e.g., Unsworth & Engle, 2007a). Specifically, examining the breakdown of shared and unique variance associated with the components suggests that the shared variance between all of the components likely reflects overall resource-sharing/executive processes which are needed on all aspects of the tasks. The shared variance between processing time and recall likely reflects processing and storage trade-offs; faster processors have more time to rehearse or refresh memory traces, leading to overall better span scores. The shared variance between processing time and processing accuracy likely reflects general processing abilities on the tasks, whereby higher-ability participants are both more accurate and faster on the processing tasks than low-ability participants. The shared variance between processing accuracy and recall likely reflects processing efficiency specifically related to the accuracy of responses (e.g., Daneman & Tardif, 1987). The unique variance shared between processing time and gF likely reflects more basic differences in speed of processing abilities that are independent of the other abilities (e.g., Salthouse, 1996). The unique variance shared between processing accuracy and gF likely reflects overlapping task-specific variance that is shared between the tasks (e.g., Schweizer, 2005). Finally, the unique variance shared between recall and gF likely reflects more basic memory and attention abilities that are needed on the complex span tasks. Note that simple span and immediate free recall tasks, which do not have an explicit processing component, seem to account for the same variance in gF as the complex span tasks (Unsworth & Engle, 2007a,b).

Implications for working memory capacity theories

Collectively the above results paint a fairly ‘‘complex’’ picture of performance in complex working memory span tasks and their relation to measures of higher-order cognition, and point to the multifaceted nature of the complex span tasks. In particular, the results suggest that although both processing components are related to each other and to storage, each of the components seem to account for a good deal of unique variance in gF. Consistent with previous research (e.g., Bayliss et al., 2005; Engle et al., 1992; Friedman & Miyake, 2004; Unsworth et al., 2005b), the current results suggest that measures of processing did not fully account for the predictive power of complex span tasks, but rather seem to add predictive power over and above that accounted for by recall. These results, in conjunction with recent work suggesting that simple span tasks do as good a job as complex span tasks in predicting higher-order cognition (e.g., Colom, Shih, Flores-Mendoza, & Quiroga, 2006; Kane et al., 2004; Unsworth & Engle, 2007a), imply that theories which indicate that processing and storage are needed to account for traditional complex span correlations are incorrect, but at the same time suggest that all three components are important for higher-order abilities. Thus, as we (Engle et al., 1992; Unsworth et al., 2005b) and others (Cowan et al., 2003; Friedman & Miyake, 2004) have argued, if the goal of the study is to understand why the storage component of complex span tasks predicts higher-order cognition, then an examination of the processing components does not offer a full account. However, if the goal of the study is to account for as much variance as possible in higher-order cognitive measures, then examining the processing components will aid in this regard.

Overall, the results are consistent with a number of accounts of complex span performance and WM more broadly (for reviews see Conway et al., 2007; Miyake & Shah, 1999). In particular, as suggested by resource-sharing and task-switching models (e.g., Daneman & Tardif, 1987; Towse et al., 1998), processing time and storage were negatively related, and processing time accounted for some of the variance between recall and higher-order cognition. Similarly, consistent with Daneman and Tardif (1987) and Schweizer (2005), processing accuracy accounted for some of the variance between recall and higher-order cognition, suggesting that part of the shared variance is due to overlapping processes in the processing accuracy and reasoning tasks. Furthermore, and consistent with previous research (e.g., Conway & Engle, 1996; Engle et al., 1992), processing time and accuracy did not fully mediate the relationship between storage and higher-order cognition. Thus there was evidence consistent with several accounts of complex span performance. This suggests that debates centering on what accounts for complex span are unlikely to be resolved by a single explanatory mechanism such as resource sharing, processing efficiency, or task switching. Rather, each of these mechanisms likely plays a role in complex span performance and accounts for the substantial variance between complex span and measures of higher-order cognition.

At the same time the results suggest that nearly all of these accounts fall short in explaining performance on complex span tasks as well as explaining their predictive power. That is, none of the current theories of complex span performance accounts for the data in the current study. Rather, each account explains a small piece of the puzzle, but each also leaves out a critical piece. The results from the current study (see also Bayliss et al., 2005) suggest that each account explains some aspect of complex span performance and the relation with measures of higher-order cognition, but none of these theories can fully explain the "complex" pattern of shared and unique variance associated with each of the three components and higher-order cognition. This suggests that although a great deal of work has been done successfully explaining aspects of complex span performance, more work is needed to offer a complete account of these influential tasks. It seems clear from the current results that working memory is not an undifferentiated pool of resources that can be allocated to different tasks; rather working memory represents a set of unique processes, each of which is important for higher-order cognitive processes. What is needed is a theory that encompasses all of these components (recall, processing accuracy, processing time), as well as their own subcomponents, into a single model and acknowledges that each component influences the other components while at the same time each component accounts for unique variance. More work is needed to examine the multifaceted nature of working memory and to provide a finer-grained breakdown of the different working memory processes that influence not only performance on the complex span tasks, but also influence performance on all those tasks that have been found to be dependent on working memory. Only when all components of working memory are successfully integrated into the same model will we have a fuller understanding of working memory functioning.

Conclusion

The current study examined the relation between components of processing and storage in complex working memory span tasks and their relation to higher-order cognition. It was found that processing and storage components are strongly related to one another and to fluid abilities. However, contrary to several theories of processing and storage (e.g., Daneman & Tardif, 1987) in complex span tasks, neither processing time nor processing accuracy fully mediated the relation between recall and gF. Thus, something other than a processing–storage trade-off accounts for the shared variability between storage span scores and measures of gF. Furthermore, consistent with previous research, adding information from the processing components increased the predictive power of the complex span tasks. Finally, examining the four-way relationship between the two processing measures, the storage component, and gF suggests a complex pattern of unique and shared variance between all of the variables. These results are consistent with the notion that complex span tasks are multifaceted and rely on multiple processes that are important for higher-order cognitive activities.

Manuscript received 20 October 2008
Manuscript accepted 24 April 2009
First published online 17 June 2009

REFERENCES

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.

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.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570–585.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, *45*, 1–20.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, D. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, *41*, 579–597.

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.

Berger, F.R., Gupta, W.B., Berger, R.M., & Skinner, J. (1990). *Air Force Officer Qualifying Test (AFOQT) form P: Test manual* (AFHRL–TR-89-56). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition*, *37*, 205–223.

Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology*, *73*, 7–44.

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.

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., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (Eds.). (2007). *Variation in working memory*. New York: Oxford University Press.

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.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (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., et al. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*, 113–132.

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., & Tardif, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Services.

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, Cognition*, *18*, 972–992.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York: Elsevier.

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.

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.

Gabrieli, J. D. E., Singh, J., Stebbins, G. T., & Goetz, C. G. (1996). Reduced working memory span in Parkinson's disease: Evidence of the role of a frontostriatal system in working and strategic memory. *Neuropsychology*, *10*, 322–332.

Hambrick, D. Z., & Oswald, F. L. (2005). Does domain knowledge moderate involvement of working memory capacity in higher-level cognition? A test of three models. *Journal of Memory and Language*, *52*, 377–397.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. R. A Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*. New York: Oxford University Press.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, *82*, 375–386.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*, 184–198.

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.

James, L. R., Mulaik, S. A., & Brett, J. M. (2006). A tale of two methods. *Organizational Research Methods*, *9*, 233–244.

Jöreskog, K., & Sörbom, D. (2006). *Lisrel 8.80*. Lincolnwood, IL: Scientific Software International Inc.

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.

Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. C. (1998). Sentence comprehension deficits in Alzheimer's disease: A comparison of off-line and on-line processing. *Brain and Language*, *64*, 297–316.

Klein, K., & Boals, A. (2001). The relationship of life event stress and working memory capacity. *Applied Cognitive Psychology*, *15*, 565–579.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavioral Research Methods, Instruments, Computers*, *31*, 429–432.

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

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, *12*, 165–170.

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

Marshalek, B., Lohman, D. F., & Snow, R. E. (1982). The complexity continuum in the radix and hierarchical models of intelligence. *Intelligence*, *7*, 107–127.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, *27*, 759–767.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Oberauer, K., Schulze, R., Wilhlem, O., & Süß, H-M. (2005). Working memory and intelligence – their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61–65.

Oberauer, K., Süß, H-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory – storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. New York: Harcourt Brace College Publishers.

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

Rosen, V. M., Bergeson, J. L., Putnam, K., Harwell, A., & Sunderland, T. (2002). Working memory and apolipoprotein E: What's the connection? *Neuropsychologia*, *40*, 2226–2233.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*, 425–443.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–777.

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*, 464–486.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., et al. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science*, *12*, 101–108.

Schweizer, K. (2005). An overview of research into the cognitive basis of intelligence. *Journal of Individual Differences*, *26*, 43–51.

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.

St. Clair-Thompson, H. L. (2007a). The effects of cognitive demand upon relationships between working memory and cognitive skills. *The Quarterly Journal of Experimental Psychology*, *60*, 1378–1388.

St. Clair-Thompson, H. L. (2007b). The influence of strategies upon relationships between working memory and cognitive skills. *Memory*, *15*, 353–365.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Stone, M., Gabrieli, J. D. E., Stebbins, G. T., & Sullivan, E. V. (1998). Working and strategic memory deficits in schizophrenia. *Neuropsychology*, *12*, 278–288.

Thurstone, T. G. (1962). *Primary mental abilities*. Chicago: Science Research Associates.

Tombu, M., & Jolicæur, P. (2004). Virtually no evidence for virtually perfect time sharing. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 795–810.

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

Towse, J. N., & Hitch, G. J. (2007). Variation in working memory due to normal development. In A. R. A Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*. New York: Oxford University Press.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory & Language*, *39*, 195–217.

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, *28*, 341–348.

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.

Unsworth, N., & Engle, R. W. (2006). 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. (2007a). On the division of short-term and working memory: An examination of simple and complex spans and their relation to higher-order abilities. *Psychological Bulletin*, *133*, 1038–1066.

Unsworth, N., & Engle, R. W. (2007b). 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., & Engle, R. W. (2005a). Working memory capacity in hot and cold cognition. In R. W. Engle, G. Sedek, U. Hecker, & D. N. McIntosh (Eds.), *Cognitive limitations in aging and psychopathology: Attention, working memory, and executive functions* (pp. 19–43). New York: Oxford University Press.

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

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