Acta Psychologica 134 (2010) 16–28

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach

Nash Unsworth *

Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA

ARTICLE INFO

Article history:
Received 18 March 2009
Received in revised form 24 August 2009
Accepted 23 November 2009
Available online 21 December 2009

PsycINFO classification:
2343

Keywords:
Working memory
Long-term memory
Intelligence
Individual differences

ABSTRACT

The present study examined the extent to which working (WM) and long-term memory (LTM) reflect the same, related, or completely different constructs and how they relate to other cognitive ability constructs. Participants performed various WM, recall, recognition, general fluid (gF) and general crystallized intelligence (gC) measures. Confirmatory factor analyses suggested that the memory measures could be grouped into three separate yet correlated factors (WM, recall, and recognition) and that these factors were strongly related to gF, but were related less so with gC. Furthermore, it was found that the common variance from the three memory factors could be accounted for by a higher-order memory factor which was strongly related to gF, but less so with gC. Finally, structural equation modeling suggested that both the variance common to the WM tasks and the variance common to all the memory tasks accounted for a unique variance in gF. These results are interpreted within an embedded process model of memory and suggest that WM and LTM tasks measure both shared and unique processes, which are important for intelligence.

1. Introduction

The notion that there are separate memory systems for information over the short-term and the long-term is an old and enduring one (James, 1890). Many contemporary theories of memory suggest that a small subset of information can be actively maintained (usually through rehearsal) over the short-term via a working memory system, while the vast amount of information a person has at their disposal is usually stored in a long-term system (e.g., Healy & McNamara, 1996). At the same time, however, other theories suggest that there is no need to distinguish between working and long-term memory systems, instead arguing for a single unitary memory system that operates over both the short-term and the long-term (e.g., Nairne, 2002). Clearly then, the debate as to whether separate working memory (WM) and long-term memory (LTM) constructs are needed or whether a single memory construct is all that is needed is an ongoing one (see for example Ruchkin, Grafman, Cameron, & Berndt, 2003 and associated commentaries).

The question of the current study was "To what extent do WM and LTM represent the same or different constructs and how are these constructs related to higher-order cognition." Although a great deal of work has been done in the experimental, neuroscientific, and modeling literatures to examine similarities and differences between WM and LTM tasks, relatively less work has been done to examine the conceptual and construct validity of these tasks from an individual differences perspective (although see Carroll, 1993; Geiselman, Woodward, & Beatty, 1982; Herrmann et al., 2001; Kyllonen & Christal, 1990). Therefore, to examine the question mentioned above, the unique and shared variance across WM and LTM tasks was assessed and structural models were used to examine the underlying factor structure for the purported constructs. Below a brief overview of differences and similarities between WM and LTM is given, followed by the rationale for the present study.

1.1. Conceptual distinctions between working and long-term memory

Early theories of WM and LTM suggested that these two constructs represented qualitatively distinct and independent memory systems (see Baddeley, 2007; Healy & McNamara, 1996; Jonides et al., 2008 for excellent reviews). In these theories the WM system is responsible for maintaining and manipulating a small amount of information over a relatively short interval while LTM, on the other hand, is responsible for maintaining all the memories a person has acquired over their lifespan. Thus, it was postulated that these two systems represented functionally different aspects of memory and had different properties and limits in terms of capacity and duration. Furthermore, it was assumed that forgetting in the two systems was due to different mechanisms (decay vs. interference).

---

* Tel.: +1 706 542 8346.
E-mail address: nunswor@uga.edu

0001-6918/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2009.11.010

It should also be noted that within these views WM can be further subdivided into separate stores as in Baddeley (1986), Baddeley (2007) WM model and LTM can be further subdivided into separate systems (e.g., Schacter & Tulving, 1994).

As such, these views suggest fairly explicit differences between WM and LTM and suggest that because these two systems are guided by different properties, the two systems are independent of one another. This suggests that it should be possible to dissociate these two systems based on experimental, neuropsychological, and neuroimaging results. Indeed, a number of studies have suggested that certain manipulations affect WM while leaving LTM unchanged, and vice versa. For instance, Rose, Myerson, Roediger, and Hale (2008) have recently suggested that levels of processing manipulations influence performance in LTM, but not in WM. Additionally, early neuropsychological work suggested dissociations between WM and LTM, where some patients with medial temporal lobe damage typically had problems with LTM, but had preserved WM (e.g., Baddeley & Warrington, 1970; Scoville & Milner, 1957). Conversely, other patients (such as KF) tended to have deficits in WM, but preserved LTM (Shallice & Warrington, 1970). Finally, recent neuroimaging work has suggested a distinction between WM and LTM, with the medial temporal lobe being required exclusively on LTM aspects of a task (Talmi, Grady, Goshen-Gottstein, & Moscovitch, 2005; although see Nee & Jonides, 2008). Very much in line with these theories, this work suggests that WM and LTM represent qualitatively and functionally distinct systems with different properties and different neural substrates.

Conversely, other theories have roundly rejected the notion that there are separate WM and LTM systems. These theories suggest that there is a single unitary memory system that operates over both short- and long-time scales and thus, there is no need to postulate different memory systems (e.g., Crowder, 1982; Melton, 1963; Nairne, 2002; Surprenant & Neath, 2008). Important for these unitary memory models is the finding that similar effects are found in WM and LTM tasks, and thus suggest that WM and LTM seem to follow very similar rules. As such these theories argue that it is more parsimonious to conclude that a single memory system is responsible for remembering over both the short-term and the long-term. Important evidence for these notions comes from computationally explicit models that assume a single memory system that operates across many retention intervals. These models have successfully been able to handle a number of findings from both WM and LTM research within the same unitary framework. For instance, Nairne's feature model (1990) is a unitary memory model that has been successfully applied to a number of traditional WM effects. Similarly, the OSCAR model of Brown, Preece, and Hulme (2000) can usefully account for a number of serial order effects that have traditionally been taken as evidence in favor of WM. Likewise Brown, Neath, and Chater's SIMPLE model (2007) can account for a number of effects including serial position effects in free recall which have long been considered as evidence in favor of separate WM and LTM systems. Thus, the ability of these computationally explicit unitary memory models to account for WM and LTM effects within a single framework provides powerful evidence for a single memory system that operates over both the short-term and the long-term. Although it should also be noted that there are several computationally explicit models that assume separate WM and LTM (e.g., Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Raaijmakers & Shiffrin, 1980).

Finally, other theories argue for embedded processes for WM and LTM. In these theories it is assumed that WM is actually an activated subset of LTM (Atkinson & Shiffrin, 1971; Cowan, 1995; Hebb, 1949). In these views, it is usually assumed that a small subset of the currently activated representations are activated highly enough that they are in the capacity limited focus of attention (Cowan, 1995, 2001). Other representations have lower activation levels and therefore are not actively participating in on-going processing. This suggests that WM and LTM are not entirely separable and implies that there should be both similarities and differences between measures of WM and LTM. Like unitary memory models, this means that there should be similar experimental effects over both the short-term and the long-term to the extent that the activated LTM representations are similar across both retention intervals. At the same time there should also be situations where differences between WM and LTM tasks are observed. Likewise, these embedded process views suggest that there should be both unique and overlapping neural substrates for WM and LTM processes. Indeed, Ranganath and Blumenfeld (2005) have recently questioned the notion that there are distinct neural circuits for WM and LTM, arguing instead that a similar circuitry is involved in both based on the notion that WM represents the activated subset of LTM (see also Cabeza, Dolcos, Graham, & Nyberg, 2002). Thus, embedded process models propose that there are distinctions between WM and LTM, but unlike previous models that have argued for completely independent systems, embedded process models suggest that there are also many similarities between WM and LTM. In many ways these theories represent a hybrid view of the other two classes of theories in that it is suggested that there is a single overall memory system with similar properties, but within that system there are differences in the functional nature of WM and LTM. As such these views predict that WM and LTM should have both unique and overlapping properties.

1.2. Measurement and prediction of working and long-term memory

Given the preceding discussion and similarities and differences between WM and LTM, one question that naturally arises is what constitutes a WM task and what constitutes a LTM task. Clearly, in order to differentiate these two constructs there must be a set of putative measures of WM and set of putative measures of LTM. Traditionally, two task characteristics have differentiated WM and LTM: number of to-be-remembered (TBR) items and retention interval (Cowan, 2008). Specifically, WM tasks usually consist of a set of TBR items that are within theoretical capacity limits (i.e., 4 + 1, Cowan, 2001; 7 + 2, Miller, 1956), whereas LTM tasks usually consist of a set of TBR items that exceed these capacity limits. Additionally, WM tasks are usually associated with either no retention interval (i.e., immediate recall) or with a very brief retention interval of only a few seconds (e.g., Cowan, 2008; Jonides et al., 2008; Ranganath, Johnson, & D'Esposito, 2003), whereas in LTM tasks the retention interval is usually much longer. Thus, in many studies WM tasks have been operationalized based on a small number of items that have to be retained for a few seconds, whereas LTM tasks are based on a larger number of items which have to be retained for much longer.

Based on this, a number of putative WM tasks have been developed to measure WM and assess the extent to which WM predicts higher-order cognitive performance. Specifically, beginning with Daneman and Carpenter (1980) a number of complex WM span tasks have been developed in which a processing task (reading sentences, solving math operations, judging symmetry, etc.) is combined with TBR items (letters, digits, words, spatial locations, etc.). Both these complex WM span tasks and simple span tasks (i.e., tasks without the processing component) have been shown to be powerful predictors of a number of important higher-order cognitive constructs including reading comprehension, reasoning, and intelligence (Ackerman, Beier, & Boyle, 2002; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Christal, 1990; Unsworth & Engle, 2007b). Furthermore, other putative measures of WM have also been shown to predict higher-order cognitive performance (Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) similar to more traditional complex WM span tasks. Thus, there is quite a bit of evidence to suggest that multiple different types of WM tasks seem to measure similar processes and these processes are important for higher-order cognitive performance.

LTM tasks can usefully be classified into two main types of tasks: recall and recognition tasks. In recall tasks (including free recall, serial recall, and cued recall) participants are presented with a set of TBR items and after a brief delay are required to recall the TBR items. In recognition tasks (item recognition and source recognition) participants are presented with a number of TBR items and at test they are usually presented with a single item and they must decide if the item is new or old (and if old what is the source in source recognition tasks). Thus, in recall tasks one must generate candidate responses, whereas in recognition tasks the item is given and one must make an assessment about whether the item was presented previously or not. These two types of tasks have a long history in memory research and debate still continues as to whether these tasks measure the same basic processes or whether they measure qualitatively distinct processes. Indeed, a recent study by Quamme, Yonelinas, Widaman, Kroll, and Sauvé (2004) found that recall and recognition measures largely measured the same basic processes, but that there was also some differentiation between the two. Thus, LTM tasks can be grouped into either recall or recognition tasks and current results seem to suggest that these two types of tasks measure both overlapping and unique processes.

Furthermore, there is a long history of examining the link between LTM measures and measures of intelligence (e.g., Christal, 1959; Ingham, 1952; Kelley, 1964). For instance, Beier and Ackerman (2004) recently reanalyzed two large datasets (Christal, 1959; Kelley, 1964) with many LTM measures and several intelligence measures. Beier and Ackerman (2004) found that these memory tasks correlated quite substantially with the intelligence measures (*r*'s > .70). Additionally, Carroll (1993) reanalyzed many datasets and found that measures of associative LTM also tended to have moderate to substantial relations with higher-order ability factors. Thus, these results suggest that LTM tasks predict higher-order cognition in a similar manner as WM tasks.

1.3. The present investigation

The goal of the present investigation was to examine the extent to which WM and LTM reflect the same, related, or completely differently constructs. If one assumes that WM and episodic LTM reflect two completely different systems then we should find that WM and episodic LTM tasks are unrelated. Likewise, if various tests of episodic LTM (e.g., recall and recognition) rely on different processes then these tests should also not be related. However, if WM measures and various episodic LTM measures rely on a similar set of processes then they should be strongly related and this overlapping variance should be related to other cognitive abilities.

Furthermore, a latent variable approach was used to examine the relations among WM, LTM, and higher-order cognition. This was done because the 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 constructs (although see Park et al., 1996). In order to derive latent variables for the constructs of interest, multiple indicators of each construct were used. These included five putative measures of WM and nine putative measures of episodic LTM that either required recall or recognition.

The recall and recognition measures were used for two reasons. First, both classes of tasks are considered to be traditional long-term memory tasks given that the number of presented items tend to be larger than the hypothesized capacity of WM and the test of the items tends to occur after a significant delay unlike most WM tasks where there is usually little or no delay after the last item has been presented. Second, these tasks were used because previous work (e.g., Unsworth & Engle, 2007a) has explicitly stated that individual differences in WM are partially due to differences in context-based retrieval processes which operate in nearly all episodic memory tasks. Given the importance of context-based retrieval processes in both recall and recognition tasks, this would suggest that all three types of tasks should be related to one another regardless of the fact that they are WM or LTM tasks.

Using these putative measures of WM and episodic LTM, several latent variable models can be constructed to test the extent to which the data is represented by various constructs (see also Nyberg, 1994; Nyberg et al., 2003 for similar analyses concerning the distinction between episodic and semantic LTM). For instance, Model A shown in Fig 1a represents a unitary memory latent variable that could occur if, for instance, there is a single memory system that cuts across all tasks with little or no specific variance associated with a class of tasks. Model B represents separate WM and episodic LTM constructs that are either unrelated (Model B1) or related (Model B2). In the non-correlated model (Model B1) this would occur if WM and LTM tasks measure fundamentally different processes whereby one set of processes governs memory over the short-term and another set of processes governs memory over the long-term. Finally, Model C represents separate latent variables for classes of tasks based on WM, recall, or recognition and these constructs can be completely unrelated (Model C1), related only for the two episodic LTM constructs (Model C2), or all constructs can be related (Model C3). Models C1 and C2 reflect not only differences between WM and LTM, but also differences within LTM. This would suggest that recall measures represent a fundamentally different set of processes than recognition measures, both are different from WM. Model C3 suggests that there are in fact differences between the different constructs, based on slightly different sets of processes, but overall the constructs share a number of overlapping processes and, hence, variance. An alternative (and statistically equivalent) to Model C3 would be to assume that the three interrelated constructs could be accounted for by a higher-order memory construct as depicted in Model D. This model is similar to hierarchical models of intelligence that suggest lower-order specific factors as well as broad higher-order factors (e.g., Carroll, 1993). Each model represents a different possible theoretical configuration of the data. Importantly, although some of these configurations may seem intuitively more plausible than others, these different models have not yet been tested against one another with a large sample of participants and tasks.

Additionally, Unsworth and Engle (2007a) argued that tasks that draw heavily on WM control processes will be strongly related to measures of fluid abilities and weakly related to measures of crystallized abilities. This is because measures of fluid abilities presumably require many of the same control processes as those indexed by tasks like complex WM span tasks leading to a large amount of overlapping variance. For instance, WM processes are likely needed on gF tasks to the extent that these tasks require information (i.e., partial solutions) to be actively maintained while testing other possible solutions and relying on controlled retrieval processes to bring relevant information into an active state in WM. Tasks that tap primarily crystallized abilities (vocabulary and general information tests), however, rely more on associative/automatic processes and thus should be weakly related to the memory measures because of the small amount of overlapping variance. Furthermore, the fluid and crystallized abilities measures were included to examine the notion that WM measures better predict higher-order cognitive abilities than long-term memory (LTM) measures. For instance, Baddeley (2007) has recently claimed that “working memory span also predicts cognitive functioning much more effectively than measures of either simple word span or episodic LTM” (p. 146; see also Engle et al., 1999). Conversely, Mogle, Lovett, Stawski, and Sliwinski (2008) have

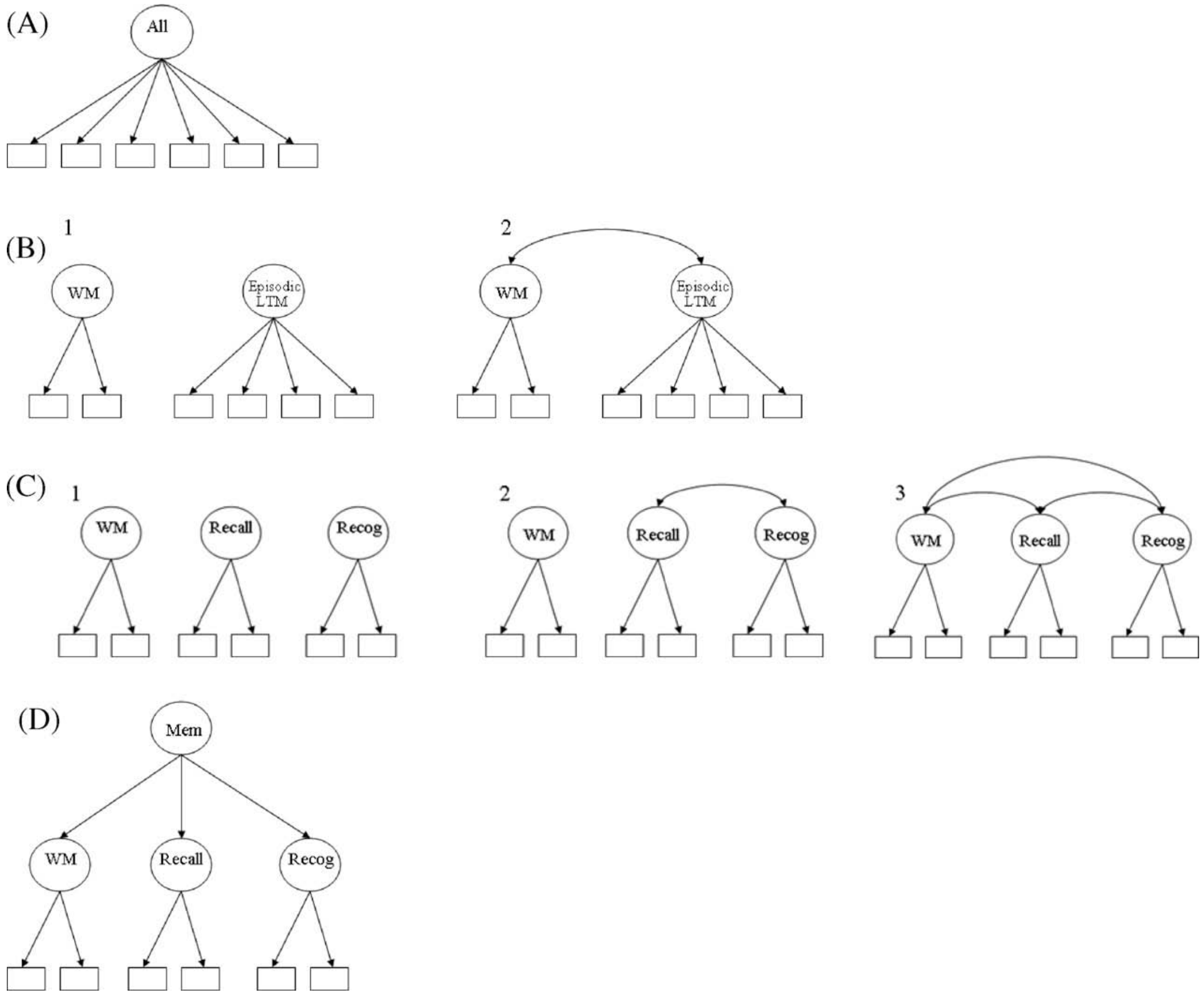

Fig. 1. Seven alternative structures for the measures of memory. See text for more details.

suggested that LTM measures predict higher-order cognition better than WM measures. Although in Mogle et al.'s (2008) study they used only one measure of fluid intelligence and did not include any measures of crystallized intelligence. Thus, it remains an open question as to whether WM is a better predictor than episodic LTM or whether both predict higher-order cognition. As argued throughout, if WM and episodic LTM tasks largely measure the processes within an embedded process model, then both WM and episodic LTM measures should be related to higher-order cognitive functioning. Thus, it is expected that both the WM and the episodic LTM measures will be strongly related to fluid abilities (due to the use of overlapping control processes in each), but will be weakly related to measures of crystallized abilities (which rely on primarily automatic processes). These hypotheses were examined in the context of several latent variable analyses with each construct represented by multiple indicators.

2. Method

2.1. Participants

A total of 165 participants (61% females) were recruited from the subject-pool at the University of Georgia. Participants were between the ages of 18 and 35 (M = 19.20, SD = 1.71) and received course credit for their participation. Each participant was tested individually in two laboratory sessions lasting approximately for 2 h each.

2.2. Materials and procedure

After signing informed consent, all participants completed the Operation span (Ospan) task, the Symmetry span (Symspan) task, the Reading span (Rspan) task, a brief computerized version of the Raven progressive matrices (Raven, Raven, & Court, 1998), a brief paper pencil verbal analogies test, a version of Thurstone, 1962) Number Series test, and delayed free recall with category switches in Session 1. In Session 2, all participants completed a delayed free recall task, a cued-recall directed forgetting task (Tolan & Tehan, 1999), a version of the list-before-last task (Ward & Tan, 2004), a list-discrimination task (Underwood, Boruch, & Malmi,1978), a paired associates cued recall task, the gender item and source recognition tasks, the picture item and source recognition tasks (Cansino, Maquet, Dolan, & Rugg, 2002), a brief paper pencil synonym vocabulary test, a brief paper pencil general knowledge test, a brief paper pencil antonym vocabulary test, and a 3-back version of the N-back (Gray, Chabris, & Braver, 2003). All tasks were administered in the order listed above.

2.3. Tasks

2.3.1. Working memory (WM) tasks

2.3.1.1. Ospan. Participants solved a series of math 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 correct order. The recall phase was untimed. After recall, the computer provided feedback about the number of letters correctly recalled in the current set. Next, participants performed the math 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 depending on their answer. After each operation participants were given accuracy feedback. The math practice served to familiarize participants with the math portion of the task as well as to calculate how long it would take that person to solve the math operations. Thus, the math practice attempted to account for individual differences in the time required to solve math operations without an additional storage requirement. After the math 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 a time limit for the math portion of the main session for that individual. Participants completed 15 math operations in this session.

The final practice session had participants perform both the letter recall and math portions together, just as they would do in the real block of trials. Here participants first saw the math operation and after they clicked the mouse button indicating that they had solved it, they saw the letter to be recalled. 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 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 3 to 7. This made for a total of 75 letters and 75 math problems. Note that the order of set-sizes was random for each participant (see Unsworth, Heitz, Schrock, and Engle (2005) for more task details). The score was the number of correct items recalled in the correct position.

2.3.1.2. Symspan. In this task participants were required to recall sequences of red squares within a matrix while performing a symmetry-judgment 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-judgment 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. Participants performed 15 trials of the symmetry-judgment task alone. The same timing parameters used in the Ospan were used. The final practice session combined the matrix recall with the symmetry-judgment 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 2 to 5. The same scoring procedure as Ospan was used.

2.3.1.3. 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. ?"). Participants were given 15 sentences, roughly half of which made sense. 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 3 to 7. The same scoring procedure as Ospan was used.

2.3.1.4. Cued-recall directed forgetting. In this task participants were presented with a block of four serially presented words and were required to recall the word from the most recently presented list that matched a specific cue (Tolan & Tehan, 1999). Each word was presented on screen for 1 s. Immediately after the presentation of the last word, and before the cue, participants performed a distracting task for 8 s. In the distracting task participants saw a three-digit number and were required to arrange the digits in the descending order. Each three-digit number was onscreen for 2 s. In 8 of the trials, participants saw only one block (one-block trials) of words before performing the distracting task and then were cued to recall one of the words. Immediately following the distracting task participants were presented with a cue and were instructed to recall the one word from the most recent list that matched the cue. Participants had 5 s for recall.

In 8 of the trials, participants saw two blocks (two-block trials) of words before performing the distracting task and then were cued to recall one of the words. In the two-block trials after the first block of words was presented participants were signaled via an exclamation point (!) that another block of words was about to appear and were instructed to only remember words from the most recent list (i.e., directed forgetting). In these trials the cue only matched one of the words from the most recent list that was presented. In another 8 trials, participants were presented with two blocks of words like the two-block trials, however here, the cue matched a word from both the most recent list and a word from the immediately preceding list (e.g., lure trials). All three trial types were presented in a fixed random order. The number of correctly recalled words was the primary measure of interest.

2.3.1.5. N-back. Participants were presented with a fixation point onscreen for 800 ms followed by a word for 2500 ms. The participants' task was to determine if the current word was the same word that was presented three trials back as quickly and accurately as possible. Participants pressed the F key for yes responses and the J key for no responses. Trials could be either non-targets (the word had not be presented previously), targets (the words matched the word presented three trials back), or lures (the presented word matched the word presented two or four trials back). There were 40 non-target trials, 20 target trials, and 20 lure trials. The dependent variable was total accuracy across all trial types.

2.3.2. Recall tasks

2.3.2.1. Delayed free recall. In this task participants were given 6 lists of 10 words each. All words were common nouns that were presented for 1 s each. After list presentation, participants engaged in a 16 s distractor task before recall: Participants saw 8 three-digit numbers appear for 2 s each, and were required to write the digits in the ascending order. After the distractor task participants saw ???, which indicated that they should type as many words as they could remember from the current list in any order they wished. Participants had 45 s for recall. A participant's score was the total number of items recalled correctly.

2.3.2.2. List-before-last recall. This task was a variant of the list-before-last task developed by Shiffrin (1970) and modified by Ward and Tan (2004). On each trial in this task participants were presented with two lists of 10 words each. There were 6 trials (12 lists total). All words were common nouns that were presented for 1 s each. Each list was labeled as either List 1 or 2 and the list labels preceded each list for 3 s. Participants were told to remember both lists and at recall they would be cued to either recall List 1 or to recall List 2. During recall participants had 45 s to type as many words from the cued list as possible. A participant's score was the total number of items recalled correctly from the cued list.

2.3.2.3. Delayed free recall with category switches. Participants received 6 lists of 10 words each broken down into two blocks (three lists per block). All words in each block came from the same semantic category (e.g., professions and fruits; Kane & Engle, 2000). The first three lists allowed for proactive interference (PI) to accrue and the first list in the next block allowed for a "release from PI." Following the last word in a list participants were required to count backwards by three's as quickly and accurately as possible from a three-digit number onscreen for 15 s and to write the numbers down as they go. After the distractor task participants saw ???, which indicated that they should type as many words as they could remember from the current list in any order they wished. Participants had 45 s for recall. A participant's score was the total number of items recalled correctly.

2.3.2.4. Cued recall. In this task participants were given 3 lists of 10 words pairs each. All words were common nouns and the word pairs were presented vertically for 2 s each. Participants were told that the cue would always be the word on the top and the target would be at the bottom. After the presentation of the last word participants saw the cue word and ??? in place of the target word. Participants were instructed to type in the target word from the current list that matched cue and then to press ENTER to indicate their response. The cues were randomly mixed so that the corresponding target words were not recalled in the same order as they were presented. Participants had 5 s to type in the corresponding word. The same procedure was done for all three lists. A participant's score was proportion of items recalled correctly.

2.3.3. Recognition tasks

2.3.3.1. Gender item and source recognition. In this task participants heard words (40 total words) in either a male or a female voice. Participants were explicitly instructed to pay attention to both the word (item) and the voice the word was spoken in (source). At test participants were presented with 40 old and 40 new words one at a time in the center of the screen. On 50% of test trials, participants were required to indicate if the word was new or old and, if old, what voice it was spoken in (source trials). Specifically, participants pressed the "1" key if the word was presented in a male voice, the "2" key if the word was presented in a female voice, or the "4" key if the word was new. On the other 50% of trials, participants simply judged if the item was new or old (item trials). Specifically, participants pressed the "4" key if the item was new or the "5" key if the item was old. Source and item trials were randomly mixed. For all test trials, participants had 5 s to press the appropriate key to enter their response. A participant's item recognition score was the proportion of correct responses on item trials and a participant's score on source recognition trials was the proportion of correct responses on source trials.

2.3.3.2. Picture item and source recognition. This task was adapted from Cansino et al. (2002). During the encoding phase, participants were presented with a picture (40 total pictures) in one of four different quadrants onscreen for 1 s. Participants were explicitly instructed to pay attention to both the picture (item) and the quadrant it was located in (source). At test participants were presented with 40 old and 40 new pictures one at a time in the center of the screen. On 50% of test trials participants were required to indicate if the picture was new or old and, if old, what quadrant it was presented in (source trials). Specifically, participants pressed the "1" key if the item was presented in the top left quadrant, the "2" key if the item was presented in the top right quadrant, the "3" key if the item was presented in the bottom left quadrant, the "4" key if the item was presented in the bottom right quadrant, and the "5" key if the item was new. On the other 50% of trials, participants simply judged if the item was new or old (item trials). Specifically, participants pressed the "5" key if the item was new or the "6" key if the item was old. Source and item trials were randomly mixed. For all test trials, participants had 5 s to press the appropriate key to enter their response. A participant's item recognition score was the proportion of correct responses on item trials and a participant's score on source recognition trials was the proportion of correct responses on source trials.

2.3.3.3. List discrimination. In this task participants were presented with three lists of 10 words each. All words were common nouns that were presented for 1 s each. Each list was labeled as List 1, List 2, or List 3 and the list labels preceded each list for 3 s. At test participants were presented with one of the words onscreen for 5 s and were required to indicate which list the word belonged to. Specifically, if the word was from List 1 participants pressed the "1" key, if the word was from List 2 participants pressed the "2" key, and if the word was from List 3 participants pressed the "3" key. A participant's score was the proportion of correct responses.

2.3.4. General fluid intelligence (gF) tasks

2.3.4.1. Raven advanced progressive matrices. The Raven is a measure of abstract reasoning (Raven et al., 1998). The test consists of 36 items presented in the ascending order of difficulty (i.e., easiest–hardest). Each item consists of a display of 3 × 3 matrices of geometric patterns with the bottom right pattern missing. The task for the participant is to select among eight alternatives, the one that correctly completes the overall series of patterns. Participants had 10 min to complete the 18 odd-numbered items. A participant's score was the total number of correct solutions. Participants received two practice problems.

2.3.4.2. Verbal analogies. In this task participants read an incomplete analogy and were required to select the one word of five possible words that best completed the analogy. After one practice item, participants had 5 min to complete 18 test items. These items were originally selected from the Air Force Officer Qualifying Test (AFOQT; 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.

2.3.4.3. 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 what the next number in the series should be. Participants selected their answer of five possible numbers that were presented. Following five practice items, participants had 4.5 min to complete 15 test items. A participant's score was the total number of items solved correctly.

2.3.5. General crystallized intelligence (gC) tasks

2.3.5.1. Synonym vocabulary. In this task participants were given 10 vocabulary words and were required to select the best synonym (of five possible choices) that best matched the target vocabulary word (Hambrick, Salthouse, & Meinz, 1999). Participants were given 2 min to complete the 10 items. A participant's score was the total number of items solved correctly.

2.3.5.2. Antonym vocabulary. In this task participants were given 10 vocabulary words and were required to select the best antonym (of five possible choices) that best matched the target vocabulary word (Hambrick et al., 1999). Participants were given 2 min to complete the 10 items. A participant's score was the total number of items solved correctly.

2.3.5.3. General knowledge. In this task participants were given 24 general information questions and were required to select the best answer (of four possible choices) to the question (Hambrick et al., 1999). Topics included American politics, sports, music, literature, history, art, and economics. Participants were given 5 min to complete the 24 items. A participant's score was the total number of items solved correctly.

3. Results

Descriptive statistics for the memory and intelligence measures are shown in Table 1. As can be seen in Table 1, all measures had generally acceptable values of internal consistency and 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). Correlations, shown in Table 2, were moderate and generally positive.

3.1. CFAs and SEMs

Confirmatory factor analysis (CFA) was used to examine the primary measurement question, "Are these memory constructs distinguishable or do they represent the same basic construct?" CFA was used because it allows one to test various models against one another to determine which model is most consistent with the observed pattern of correlations. For instance, CFA can be used to determine whether a one-factor memory model fits as well as a two-factor memory model. Additionally, structural equation modeling (SEM) was used to examine how the separate memory constructs differentially relate to intelligence. Thus, not only is the underlying structure of the data taken into account, but also models can be tested to examine how the different constructs are related to one another and account for separate and unique sources of variance in another construct like intelligence. If WM and LTM reflect different constructs, then a two-factor memory model should fit better than a single factor memory model, and the two memory factors should account for variance in intelligence (specifically gF).

Table 1
Descriptive statistics and reliability estimates for memory and intelligence measures.

Measure	M	SD	Range	Skew	Kurtosis	α
Ospan	60.38	11.00	26–75	−1.35	1.87	.79
Symspan	29.32	7.73	9–42	−.2	−.6	.77
Rspan	57.79	12.10	18–75	−.97	.70	.78
CRDF	21.93	2.45	12–24	−1.81	3.36	.77
N-back	.64	.16	.12–.95	−1.58	2.32	.75
DFR	32.51	6.63	15–51	.23	.28	.70
LBL	4.25	1.20	1–8.67	.42	.69	.77
DFRC	34.49	5.26	21–48	−.07	−.29	.63
CR	.45	.21	.07–1.0	.39	−.43	.72
GenRec	.57	.14	.13–.82	−.69	.55	.77
GenSour	.46	.15	.15–.90.	.20	−.23	77
PicRec	.76	.14	.23–.93	−1.45	2.29	.90
PicSour	.72	.17	.20–.97	−.97	.49	.90
LD	.54	.17	13–.90	−.15	−.44	.73
Raven	10.32	2.48	4–16	−.32	−.17	.72
Analogy	11.48	3.01	4–25	.26	1.68	.66
NS	9.53	2.48	3–15	−.32	−.17	.70
Syn	4.02	1.91	1–9	.22	−.38	.61
Ant	4.30	1.83	1–9	.25	−.29	.61
GenK	14.67	3.50	7–23	−.06	−.25	.68

Note. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; CRDF = cued-recall directed forgetting; DFR = delayed free recall, LBL = list-before-last recall; DFRC = delayed free recall with category switches; CR = cued recall; GenRec = gender item recognition; GenSour = gender source recognition; PicRec = picture item recognition; PicSour = picture source recognition; LD = list discrimination; Raven = Raven Progressive Matrices; Analogy = verbal analogies; NS = number series; Syn = synonym vocabulary; Ant = antonym vocabulary; GenK = general knowledge test.

Model fits were assessed via the combination of several fit statistics. These include χ^2, root mean square error of approximation, standardized root mean square residual, the non-normed fit index, and the comparative fit index. The χ^2 statistic reflects whether there is a significant difference between the observed and reproduced covariance matrices. Therefore, non-significant values are desirable. However, with large sample sizes even slight deviations can result in a significant value, therefore the ratio of χ^2 to the number of degrees of freedom is also reported. Ratios of two or less usually indicate acceptable fit. Test between nested models are examined via a χ^2 difference test. Also reported are the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) which reflect the average squared deviation between the observed and reproduced covariances. In addition, the non-normed fit index (NNFI) and the comparative fit index (CFI) which compare the fit of the specified model to a baseline null model are reported. NNFI and CFI values greater than .90 and RMSEA and SRMR values less than .08 are indicative of acceptable fit (Kline, 1998). Finally, the Akaike information criterion (AIC) examines the relative fit between models in which the model with the smallest AIC is preferred.

3.2. Can we distinguish separate memory constructs?

To examine the structure of the data further several confirmatory factor analyses (CFA) were conducted. As shown in Fig 1, seven models were constructed to examine the factor structure of the data.[1] For all models the residuals for the gender item and source

---

[1] Note for all models the N-back was not included given that it correlated weakly with the other memory measures and had a very low communality ($h^2 = .07$) when examined with the other memory tasks. Including N-back in the models lead to a non-significant factor loading on the WM factor. Thus, this version of the N-back did not seem to be related to the other WM measures in the current study (see also Kane, Conway, Miura, and Colflesh, 2007).

Table 2
Correlations for memory and intelligence measures.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
WM																				
1. Ospan	–																			
2. SymS	.51	–																		
3. Rspan	.62	.43	–																	
4. CRDF	.40	.27	.54	–																
5. N-back	.04	.10	.11	.18	–															
Recall																				
6. DFR	.29	.21	.20	.38	.13	–														
7. LBL	.11	.20	.15	.29	.09	.46	–													
8. DFRC	.12	.02	.15	.28	.04	.39	.41	–												
9. CR	.07	.15	.05	.22	.16	.44	.49	.27	–											
Recognition																				
10. GenR	.09	.06	.07	.17	.13	.25	.26	.32	.23	–										
11. GenS	.23	.19	.18	.26	.10	.21	.33	.30	.24	.63	–									
12. PicR	.15	.11	.15	.23	.11	.16	.17	.18	.11	.34	.26	–								
13. PicS	.21	.26	.20	.43	.24	.25	.26	.28	.31	.44	.42	.57	–							
14. LD	.05	.11	.01	.27	.09	.29	.40	.20	.38	.28	.29	.12	.31	–						
gF																				
15. Raven	.30	.30	.37	.34	.15	.16	.07	.01	.18	.07	.17	.34	.36	.14	–					
16. Ang	.11	.11	.22	.28	.25	.06	.17	.24	.21	.10	.19	.25	.21	.13	.31	–				
17. Ns	.24	.29	.23	.20	.16	.23	.18	.16	.07	.08	.16	.18	.18	−.07	.24	.22	–			
gC																				
18. Syn	.12	−.02	.10	.15	.17	.06	.12	.14	.08	.03	.09	.22	.14	.10	.27	.35	.12	–		
19. Ant	.11	.04	.16	.18	.12	.09	.15	.17	−.04	.06	.09	.17	.18	.00	.18	.29	.15	.43	–	
20. GenK	.07	.03	.20	.21	.07	.08	.07	.07	.07	.10	.09	.22	.15	−.02	.17	.39	.20	.43	.37	–

Note. Ospan = operation span; SymS = symmetry span; Rspan = reading span; CRDF = cued-recall directed forgetting; DFR = delayed free recall, LBL = list-before-last recall; DFRC = delayed free recall with category switches; CR = cued recall; GenR = gender item recognition; GenS = gender source recognition; PicR = picture item recognition; PicS = picture source recognition; LD = list discrimination; Raven = Raven Progressive Matrices; Ang = verbal analogies; NS = number series; Syn = synonym vocabulary; Ant = antonym vocabulary; GenK = general knowledge test. Correlations > .15 are significant at the $p < .05$ level.

recognition tasks were allowed to correlate as were the residuals for the item and source picture location task. This was done because there is likely a great deal of method specific variance shared between these tasks that is independent of the broader factors. Not allowing these correlations lead to significant reductions in model fit, but qualitatively identical results as the models reported below.

The first model examined a one-factor model of the data in which all the memory tasks loaded on a single factor (CFA A). Next, two different two-factor models were tested differentiating WM (with the three complex span tasks and the cued-recall directed forgetting task loading on the WM factor) and episodic LTM (with both the recall and recognition tasks loading on the episodic LTM factor). In one of these models the WM and LTM factors were uncorrelated (CFA B1), while in the other model they were allowed to correlate (CFA B2). Three three-factor models were also tested differentiating the WM tasks, the Recall tasks, and the Recognition tasks. In the first of these models all three factors were uncorrelated (CFA C1). In the second of these models only the two episodic LTM factors (Recall and Recognition) were allowed to correlate (CFA C2), and in the third model all three factors were allowed to correlate (CFA C3). Finally, an alternative to the three-factor correlated model (CFA C3) was constructed in which a higher-order memory factor accounted for the shared variance among the three lower-order factors (CFA D1). The fit indices for all models are shown in Table 3. An inspection of Table 3 suggests that the best fitting model is one that assumes three correlated factors best represent the data (CFA C3). In fact, this model fits significantly better than all the other models (all $\Delta\chi^2$'s > 10, p's < .01) and had the smallest AIC value. This suggests that the different memory tasks can be grouped together based on the processes that they are thought to tap (WM, LTM, etc.) as well as more task specific processes (e.g., Recall and Recognition). At the same time the results suggest that the factors are correlated (WM–Recall = .36, WM–Recognition = .40, Recall–Recognition = .72), and thus share a good deal of common variance. In fact, as shown in Table 3, the higher-order memory model (CFA D1) fit exactly the same as the correlated three-factor model (CFA C3) suggesting that the shared variance among the lower-order factors could be accounted for

Table 3
Fit indices for all models.

Model	χ^2	df	χ^2/df	RMSEA	NNFI	CFI	SRMR	AIC
CFA A	279.67	63	4.44	.15	.79	.83	.15	335.67
CFA B1	127.20	63	2.02	.08	.90	.92	.13	183.20
CFA B2	116.89	62	1.89	.07	.92	.93	.09	174.89
CFA C1	171.96	63	2.73	.10	.86	.89	.18	227.96
CFA C2	116.05	62	1.87	.07	.91	.93	.13	174.05
CFA C3	105.24	60	1.75	.07	.93	.95	.08	167.24
CFA D1	105.24	60	1.75	.07	.93	.95	.08	167.24
CFA Mem gF gC	210.60	140	1.50	.06	.93	.94	.07	310.60
SEM Mem gF gC	210.60	140	1.50	.06	.93	.94	.07	310.60
SEM Com Mem gF gC	209.13	141	1.48	.05	.94	.95	.07	307.13

Note. RMSEA = root mean square error of approximation; NNFI = non-normed fit index; CFI = comparative fit index; SRMR = standardized root mean square residual; AIC = Akaike information criterion.

by a single higher-order memory factor with each of the lower-order factors having strong loadings on the higher-order factor (WM loading = .44, Recall loading = .81, Recognition loading = .89). Note that the fit of these two models (CFAs C3 and D1) fit exactly the same because they are statistically equivalent. Collectively these results suggest that the data are best represented by three separate factors (WM, Recall, and Recognition) and that these three factors share a considerable amount of variance which can be accounted for by a single higher-order factor.[2]

3.3. How do the memory constructs relate to intelligence?

Next in order to examine how each of these three memory factors were related to higher-order cognitive abilities, fluid (gF) and crystallized (gC) factors were added to the three-factor correlated memory model. The fit of the model was acceptable as shown in Table 3 (labeled as CFA Mem gF gC). As shown in Table 4, each task significantly loaded on its corresponding factor and all the factors were significantly correlated with one another. In terms of the correlations between the memory factors and gF and gC, it is clear that all the memory factors were moderately correlated with gF, but were weakly related with gC. In fact, the correlations between the memory factors and gF were significantly stronger than the correlations with gC (all t's > 3.83, p's < .01). Thus, consistent with previous work (Unsworth & Engle, 2007a), the results suggested that the memory measures were more strongly related with gF than gC. Furthermore, in terms of the differential relation of the memory factors with gF both the WM and Recognition factors were more strongly correlated with gF than the Recall factor (both t's > 2.96, p's < .01) but the WM-gF and the Recognition-gF correlations were not significantly different, $t < .1.82$. Thus, it is not the case that WM is necessarily a better predictor of higher-order abilities than measures of episodic LTM.

In order to examine this further, a structural equation model (SEM) with each of the three memory factors predicting both gF and gC was specified. This allows for an examination of how each of the three memory factors uniquely predicts variance in the intelligence factors. As shown in Table 3, the fit of the model was acceptable. Note that given the large correlation between gF and gC their residual variances were allowed to correlate. The resulting model is shown in Fig 2. All three factors accounted for roughly 55% of the variance in gF, but only 10% of the variance in gC. Additionally as shown by the solid lines in Fig 2, only the WM factor accounted for unique variance in gF which is inconsistent with the claims of Mogle et al. (2008). Although it should be noted that the path coefficient from the Recognition factor to gF was quite substantial (i.e., .39) but was not significant due to an unusually large standard error (i.e., .22). Thus, the variance shared between Recall, Recognition, and gF was completely shared between the memory factors and WM accounted some unique variance in gF. None of the factors accounted for a unique variance in gC. This suggests that over half of the variance is gF is accounted for by the memory factors with the WM factor accounting for some unique variance and all three factors accounting for some shared variance.

Table 4
Confirmatory factor analysis for WM, Recall, Recognition, gF, and gC measures.

Measure	Latent factor				
	WM	Recall	Recognition	gF	gC
Ospan	.74*				
Symspan	.57*				
Rspan	.81*				
CRDF	.63*				
DFR		.66*			
LBL		.73*			
DFRC		.55*			
CR		.64*			
GenRec			.57*		
GenSour			.60*		
PicRec			.44*		
PicSour			.68*		
LD			.50*		
Raven				.57*	
Analogy				.55*	
NS				.40*	
Syn					.70*
Ant					.59*
GenK					.63*
Interfactor correlations					
WM	–				
Recall	.36*	–			
Recognition	.41*	.70*	–		
gF	.67*	.41*	.56*	–	
gC	.26*	.21*	.26*	.73*	–

* $p < .05$.

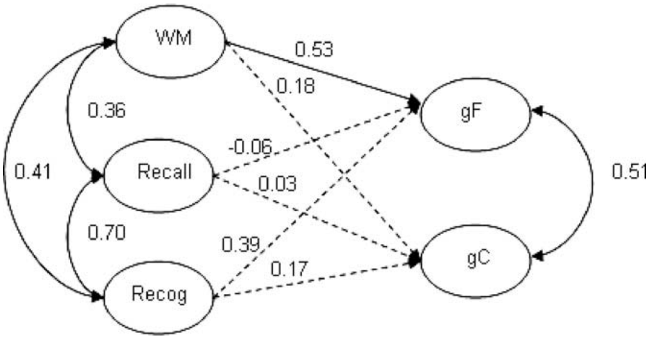

Fig. 2. Structural equation model predicting general fluid (gF) and general crystallized (gC) intelligence with working memory (WM), recall (Recall), and recognition (Recog). Single-headed arrows connecting latent variables (circles) to each other represent standardized path coefficients indicating the unique contribution of the latent variable. Double headed arrows connecting the memory factors represent the correlations among the factors. The double headed arrow connecting gF and gC represents the correlation between the residual variances for those two factors. Solid lines are significant at the $p < .05$ level and dotted lines are not significant at the $p < .05$ level.

The shared variance between the tasks likely reflects shared controlled retrieval processes that are required in each, while the unique variance likely reflects active maintenance processes that are needed specifically in the WM tasks (Unsworth & Engle, 2007a). In order to examine this notion more thoroughly, the next model tested the idea that WM tasks have two sources of variance, one of which is shared with all of the LTM tasks (controlled retrieval) and one which is unique to the WM tasks (active maintenance), both of which are important for higher-order cognitive abilities. To test this, two factors were specified for the memory tasks. The first factor allowed all the memory tasks to load on it and represents the shared variance across all of the memory tasks (Mem). The other factor only had loadings from the four WM tasks on it and represents the shared variance across the WM tasks that is independent of the variance shared with the LTM tasks (WMRE).

---

[2] Models were also examined to determine whether specific content factors (verbal and spatial) could be extracted and account for the data. Specifically, in one model separate verbal and spatial factors were specified and allowed to correlate. Although these two factors were highly correlated ($r = .94$) the fit of the model was quite poor, χ^2 (62) = 279.12, $p < .01$, RMSEA = .15, SRMR = .11, NNFI = .78, CFI= .83. Additionally, another model was attempted in which verbal and spatial factors were added to model CFA C3. In this model two separate sources of variance were extracted from each task with one source being attributable to specific memory processes and the other being attributable to the specific content of the stimuli. Unfortunately this model failed to converge on an acceptable solution after 1,000 iterations. Thus, it does not seem that the data could be accounted for by specific content related factors.

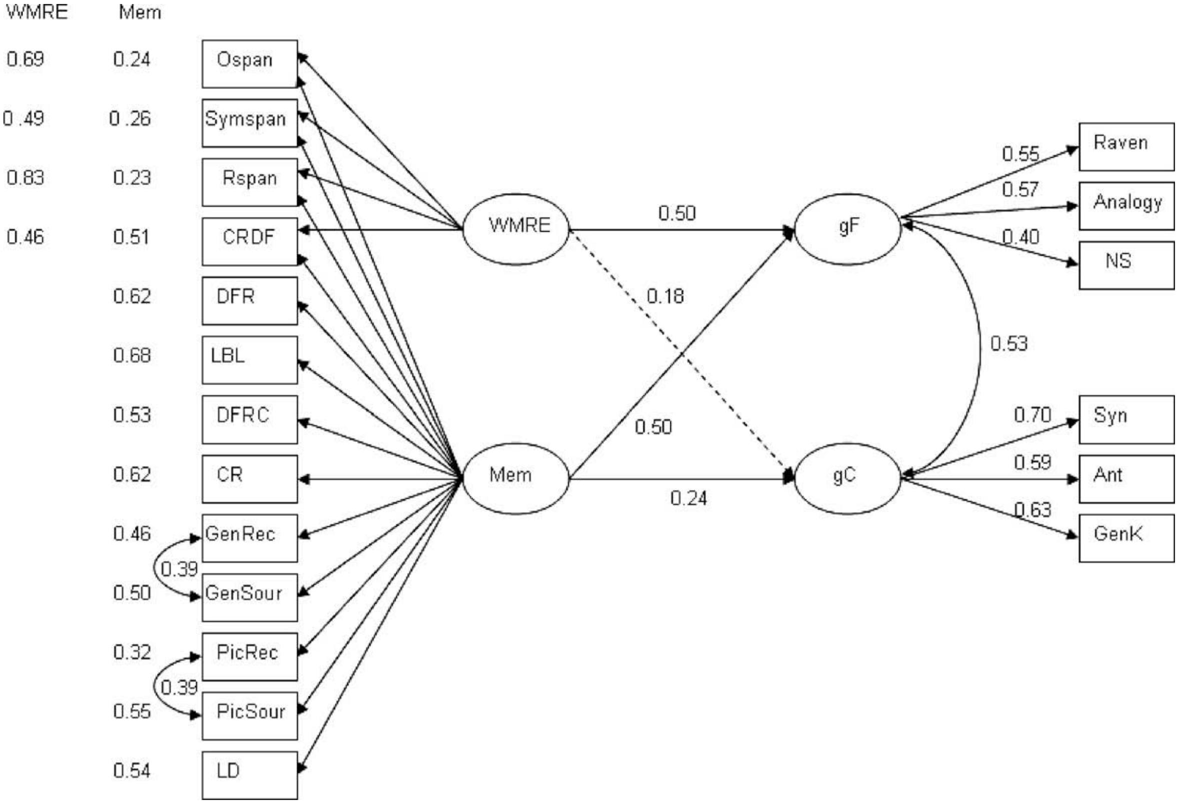

Fig. 3. Structural equation model for the common variance shared across all the memory tasks (Mem) and the residual variance common to only the working memory tasks (WMRE) predicting general fluid (gF) and general crystallized (gC) intelligence. All loadings are significant at the $p < .05$ level. Significant paths are indicted by solid lines and non-significant paths are indicated by dashed lines. The numbers in the WMRE column represent the factor loading for the four working memory tasks on the WMRE factor and the numbers in the Mem column represent the factor loadings for all the memory tasks onto the Mem factor.

Both factors were then set up to predict both gF and gC as in the previous SEM. If it is the case that WM tasks can be decomposed into two separate factors (active maintenance and controlled retrieval) we should see that the WM tasks cross-load on both factors and that they account for unique variance in intelligence.

The resulting model is shown in Fig 3and the fit statistics are shown in Table 3 (where the model is labeled SEM Com Mem gF gC). As can be seen the model fit was acceptable and all tasks loaded on their respective factors. Furthermore, the WM tasks loaded significantly on both the overall Mem factor and the WMRE factor. Overall both factors accounted for variance in gF ($R^2 = .50$), with both factors accounting for roughly 25% unique variance. Thus, half of the variance in gF was accounted for by the different memory measures and this could be further subdivided into a fourth of the variance being accounted for by specific WM variance and a fourth of the total variance being accounted for by the shared variance across all the memory measures. Additionally, the Mem factor was also uniquely related to gC (accounting for roughly 9% of the variance). Given that most of the memory measures were verbal in nature; this unique variance likely reflects the contribution of word knowledge to performance. Like the other SEM model, these results suggest that there are both shared and unique sources of variance for the WM and LTM tasks which are important for higher-order cognitive abilities which is inconsistent with recent claims made by Mogle et al. (2008).

4. General discussion

The goal of the current study was to address two questions: (1) Are WM and LTM distinct, related, or the same constructs? and (2) How do these memory constructs differentially relate to higher-order constructs like intelligence? In order to examine these questions a latent variable approach was used in which the common variance across multiple putative measures of WM and LTM was extracted and various models were tested to determine the relations between the constructs.

In regard to the first question, the results from the current study suggested that WM and LTM should be regarded as related yet distinct constructs. Confirmatory factor analyses suggested that the model that best accounted for the underlying structure of the data was one that assumed three separate, yet related constructs. Unitary models of memory and models that assume distinct WM and LTM constructs that are independent of one another were ruled out because these models fit the data significantly worse than the multifactor model. Thus, although WM and LTM seem to be somewhat distinct, they are not fully independent constructs. Rather it seems that WM and LTM share a good deal of common variance, but at the same time there is some unique variance associated with each.

This common variance likely reflects the fact that not all tasks are process pure and thus, performance on measures of WM (such as complex span tasks) reflect contributions of multiple constructs. For instance, recent work (Unsworth & Engle, 2007a) suggests that WM tasks measure both the need to actively maintain information in the face of distraction (e.g., Engle & Kane, 2004) and the need to strategically search LTM when representations cannot be maintained. These represent separate control processes that may act in concert to determine performance on a given task. Thus, the common variance shared between WM and LTM likely reflects the need to utilize similar processes in both types of tasks. Furthermore, the common variance shared between WM and LTM not only reflects shared control processes across tasks, but also reflects the fact that these two constructs (irrespective of the actual tasks used) are intimately linked because WM comprises activated LTM representations (e.g., Atkinson & Shiffrin, 1971). Thus, similar to the model initially proposed by Atkinson and Shiffrin (1971), WM represents both the activated portion of LTM and the set of control processes that act on those activated representations in order to bring them into a heightened state of activation and actively maintain them in the face of distraction.

In terms of the second question that asked how WM and LTM are related to higher-order constructs like intelligence, the results from the current study suggest that both WM and LTM are related to intelligence. Specifically, the results suggested that all three memory constructs were substantially related to gF, but were related less so with gC. Furthermore, the results from SEMs demonstrated that all three components accounted for variance in gF, but only WM accounted for unique variance. Thus, this suggests that WM tasks measure many of the same processes as LTM tasks, as well as some additional processes, both are related to intelligence.

The current results are consistent with the notion that memory tasks will vary in the extent with which these component processes are needed (e.g., Craik, 1983; Johnson, 2005), which will in turn impact the correlations among the memory tasks themselves and the correlations between the memory tasks and other cognitive constructs such as fluid and crystallized intelligence. That is, the predictive power of complex WM tasks is not localized solely to these tasks, but rather other episodic memory tasks can predict performance on higher-order cognitive tasks just as well (see Beier & Ackerman, 2004; Unsworth & Engle, 2007b). Thus, contrary to the belief that WM tasks are better predictors of higher-order cognition than LTM tasks (see Baddeley, 2007) or that LTM tasks are better predictors than WM tasks (Mogle et al., 2008), both types of tasks seem to be able to predict performance on higher-order fluid ability tasks. The differences in the results from the current study and those of Mogle et al. (2008) are likely due to the fact that Mogle et al. relied only on one measure of gF (Ravens) rather than on multiple measures. Indeed, an examination of their correlation matrix suggests that a single correlation (between word recognition and Ravens) likely accounts for their finding that LTM correlates more strongly with gF than WM does. Using a much larger and broader sample of WM, LTM, and intelligence tasks, the current results suggest that the common variance shared between WM and LTM accounts for a substantial portion of the variance in gF and WM accounts for variance in gF independently of the shared variance with LTM.

The results of the present study have a number of important implications for theories of WM and episodic LTM. In particular, the results of the present study suggest that WM tasks are not inherently special tasks, because many of these tasks require many of the same component processes as other episodic memory tasks. The common variance across a number of tasks (that presumably reflects the extent to which these tasks require the same component processes) suggests that tasks should not be classified simply as working memory, short-term memory, or long-term memory, but rather many of these tasks require the same basic component processes. Thus, simply because a task is labeled as a WM or LTM task it does not mean that it represents a single process as indicated by the label. Rather, we must be aware that performance on a given task is likely determined by many processes and we must attempt to understand these processes from both experimental and differential perspectives. Indeed, Johnson (2005) has recently commented that "If experimental approaches stay alert to commonalities across tasks (and are not satisfied with local theories of very specific tasks), and individual differences approaches stay alert to components that may be represented in their latent variables (and are not satisfied with global explanatory constructs like episodic memory and executive function), these approaches should converge on a cumulative and cohesive picture of cognitive function" (p. 530). In many ways this quote nicely captures the overall message of the current study. It is unlikely that a given task reflects a specific construct like WM, rather performance on a given task arises from a number of component processes which individuals likely differ on. Models that focus exclusively on complex span tasks, on free recall tasks, or on recognition tasks must somehow account for the fact that these tasks largely measure the same processes.

Collectively, the results of the current study are most consistent with an embedded process model of WM and LTM. As noted previously, in this type of model WM comprises the activated subset of LTM along with various control processes that are needed to maintain activation in the presence of interference and distraction (Atkinson & Shiffrin, 1971; Engle et al., 1999). This type of model readily accounts for the fact that WM and LTM should share considerable variance given that WM and LTM will rely on many of the same processes and neural substrates. At the same time, this type of model suggests that there should also be differences between WM and LTM to the extent that some processes will be unique to WM, and some processes will be unique to LTM. Accordingly, there should be both similarities and differences between WM and LTM, which is consistent with the data.

It should also be noted that the data are consistent with Baddeley's (2007) model of WM, in that the episodic buffer is important for the interaction between WM and LTM and could potentially account for the overlapping variance between WM and LTM tasks, while the unique variance may be due to specific WM processes. Similarly, Logie's (2003) model that conceptualizes WM as a mental workspace that interacts with incoming perceptual information and stored knowledge would also likely be able to account for the results from the present study. Clearly more work is needed to better differentiate between an embedded process model and models with dedicated buffers necessary for interacting with episodic LTM or models that suggest WM is a mental workspace necessary for interacting with both perceptual information and stored knowledge.

5. Conclusion

The current study examined the distinction between WM and LTM from an individual differences perspective. Using a large sample of participants and tasks thought to measure both WM and LTM, it was found that WM and LTM represent distinct, but related constructs. In particular, the results suggested that there is a good deal of shared variance between WM and LTM measures that accounts for higher-order cognition. At the same time, there was also a unique variance shared across the WM tasks that was independent of the LTM measures and was related to the higher-order cognitive measures. These results were interpreted within an embedded process model of memory in which WM represents the subset of active LTM representations plus separable control processes of active maintenance and controlled retrieval (Atkinson & Shiffrin, 1971; Unsworth & Engle, 2007). A combination of individual differences studies that examine commonalities across tasks, experimental studies that examine the component processes in those tasks, and modeling and neuroscientific/neuropsychological studies that attempt to integrate the two should provide a greater understanding of WM and LTM relations.

Acknowledgements

The author thanks Aaron Biddle for data collection assistance and Gene Brewer, Greg Spillers, Mike Kane, and an anonymous reviewer for comments on an earlier version of the article.

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.

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

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

Baddeley, A. D. (2007). *Working memory, thought, and action*. NY: Oxford University Press.

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.

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.

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.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539–576.

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

Cabeza, R. E., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage, 16*, 317–330.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex, 12*, 1048–1056.

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

Christal, R. E. (1959). Factor analytic study of visual memory. *Psychological Monographs: General and Applied, 72*, 1–24.

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.

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. (2008). What are the differences between long-term, short-term, and working memory? In W. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *The essence of memory*. Amsterdam: Elsevier/Academic Press.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society, London, Series B, 302*, 341–359.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica, 50*, 291–323.

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

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.

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). NY: 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.

Geiselman, R. E., Woodward, J. A., & Beatty, J. (1982). Individual differences in verbal memory performance. A test of alternative information processing models. *Journal of Experimental Psychology: General, 111*, 109–134.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*, 316–322.

Hambrick, D. Z., Salthouse, T. A., & Meinz, E. J. (1999). Predictors of crossword puzzle proficiency and moderators of age-cognition relations. *Journal of Experimental Psychology: General, 128*, 131–164.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still work? *Annual Review of Psychology, 47*, 143–172.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Herrmann, D. J., Schooler, C., Caplan, L. J., Lipman, P. D., Grafman, J., Schoenbach, C., et al. (2001). The latent structure of memory: A confirmatory factor-analytic study of memory distinctions. *Multivariate Behavioral Research, 36*, 29–51.

Ingham, J. G. (1952). Memory and intelligence. *British Journal of Psychology, 43*, 20–32.

James, W. (1890). *Principles of psychology*. New York: Henry Holt.

Johnson, M. K. (2005). The relation between source memory, episodic memory: Comment on Siedlecki. *Psychology and Aging, 20*, 529–531.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193–224.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 615–622.

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference, and divided attention: Limits on long-term retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 333–358.

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.

Kelley, H. P. (1964). Memory abilities: A factor analysis. *Psychometric Monographs, 11*.

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

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*, 389–433.

Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. *The psychology of learning and motivation* (Vol. 42, pp. 37–38). Academic Press.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1–21.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationship among working memory, secondary memory, and fluid intelligence. *Psychological Science, 19*, 1071–1077.

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

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53–81.

Nee, D. E., & Jonides, J. (2008). Neural correlates of access to short-term memory. *Proceedings of the National Academy of Academy of Sciences, 105*, 14228–14233.

Nyberg, L. (1994). A structural equation modeling approach to the multiple memory systems question. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 485–491.

Nyberg, L., Maitland, S. B., Rönnlund, M., Bäckman, L., Dixon, R. A., Wahlin, A., et al. (2003). Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychology and Aging, 18*, 149–160.

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J., Frieske, D., Zwahr, M., et al. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*(4), 621–637.

Quamme, J. R., Yonelinas, A. P., Widaman, K. F., Kroll, N. E. A., & Sauvé, M. J. (2004). Recall and recognition in mild hypoxia: Using covariance structural modeling to test competing theories of explicit memory. *Neuropsychologia, 42*, 672–691.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14). New York: Academic Press.

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9*, 374–380.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia, 41*, 378–389.

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

Rose, N. S., Myerson, J., Roediger, H. L., III., & Hale, S. (2008). *Working memory, secondary memory, and long-term memory: Insights from the levels-of-processing span task.* Paper presented at the 49th annual meeting of the psychonomic society, Chicago, IL.

Schacter, D. L., & Tulving, E. (Eds.). (1994). *Memory systems 1994*. Cambridge, MA: MIT Press.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26*, 709–777.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*, 11–21.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22*, 261–273.

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science, 168*, 1601–1603.

Surprenant, A. M., & Neath, I. (2008). The 9 lives of short-term memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain*. Hove, UK: Psychology Press.

Süß, 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.

Talmi, D., Grady, C. L., Goshen-Gottstein, Y., & Moscovitch, M. (2005). Neuroimaging the serial position curve. *Psychological Science, 16*, 716–723.

Thurstone, T. G. (1962). *Primary Mental Abilities*. Chicago: Science Research Associates.

Tolan, G. A., & Tehan, G. (1999). Determinants of short-term forgetting: Decay, retroactive interference, or proactive interference? *International Journal of Psychology, 34*, 285–292.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107*, 393–419.

Unsworth, N., & Engle, R. W. (2007a). 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., & Engle, R. W. (2007b). 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., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.

Ward, G., & Tan, L. (2004). The effect of the length of to-be-remembered lists and intervening lists on free recall: A re-examination using overt rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1196–1210.