THEORETICAL REVIEW

Complex span and *n*-back measures of working memory: A meta-analysis

Thomas S. Redick · Dakota R. B. Lindsey

Abstract Working memory is a construct of primary relevance to many areas of psychology. Two types of tasks have been used to measure working memory, primarily in different research areas: Complex span tasks are commonly used in behavioral studies in the cognitive and individual-differences literature, whereas *n*-back tasks have been used more frequently in cognitive neuroscience studies investigating the neural underpinnings of working memory. Despite both categories of tasks being labeled as "working memory" measures, previous empirical studies have provided mixed evidence regarding the shared amount of overlapping processes between complex span and *n*-back tasks. The present meta-analysis showed that the complex span and *n*-back tasks are weakly correlated, although significant heterogeneity across studies was observed. A follow-up analysis of unpublished data indicated that the sample composition affects the relationship between the complex span and *n*-back tasks, following the law of diminishing returns. Finally, a separate meta-analysis indicated that the simple span and *n*-back tasks are correlated to the same extent as are the complex span and *n*-back tasks. The present findings indicate that the complex span and *n*-back tasks cannot be used interchangeably as working memory measures in research applications.

T. S. Redick (✉)
Purdue University, Psychological Sciences,
703 Third Street, West Lafayette, IN 47907, USA
e-mail: tredick@purdue.edu

D. R. B. Lindsey
Georgia Institute of Technology, Atlanta, GA, USA

> The studies that have directly investigated the relation between the *n*-back task and other WM measures always revealed weak correlations. (Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011, p. 148)
> Correlations between the *n*-back task and more complex span measures are variable, with some studies reporting relatively low correlations . . . and others reporting high correlations. (Li, Schmiedek, Huxhold, Röcke, Smith, & Lindenberger, 2008, p. 739)
> Theories of WM . . . all predict *n*-back and WM span tasks to measure largely the same thing, that is, to reflect primarily the same WM construct. Why don't they? (Kane, Conway, Miura & Colflesh, 2007, p. 621)

Working memory (WM) is a construct that has been studied extensively in the past 50 years, since it was first mentioned by Miller, Galanter, and Pribram (1960), and especially since the influential WM model proposed by Baddeley and Hitch (1974). The concept is a more dynamic version of the short-term memory construct that was present in initial information-processing models (e.g., Atkinson & Shiffrin, 1968). WM has been studied extensively not only in cognitive psychology, but also in other areas, including social, clinical, developmental, and personality research. WM is critical to activities involving the goal-directed use of immediate memory, the maintenance and manipulation of recently attended information, and switching and scheduling task priorities in multitasking situations. An important consideration for such research efforts is how to operationally define and measure WM. In the present research, we investigated the degree of overlap between two commonly used categories of WM measures: complex span and *n*-back tasks.

Complex span task measures of WM

Beginning with reading span (Daneman & Carpenter, 1980), "complex span" tasks became popular measures of WM, in contrast to existing "simple span" tasks such as digit span. Other complex span tasks followed: (a) counting span (Case, Kurland, & Goldberg, 1982); (b) operation span (Turner & Engle, 1989); (c) rotation span (Shah & Miyake, 1996); and (d) symmetry span (Kane et al., 2004), to name a few. Instead of being given a list of digits to serially recall, as in digit span, subjects taking an operation span task see a series of items such as the following: "IS (2 × 1) + 3 = 6 ? DOG". Because complex span tasks combine the recall of some items (e.g., words) while subjects also perform a secondary processing task (e.g., math operations), such tasks are also known as *storage-plus-processing tests*. Complex span tasks are consistently positively related to higher-order cognitive abilities, including reasoning, reading comprehension, and mathematics achievement (Daneman & Merikle, 1996; Unsworth & Engle, 2007).

N-back task measures of WM

The *n*-back task was reported first by Kirchner (1958). However, *n*-back tasks did not rise to prominence until the surge of cognitive neuroscience research, largely because the task administration is amenable to the methodological constraints (stimulus and response timing, response formats) for many neuroimaging techniques. In the *n*-back task, individuals are asked to report whether or not the item currently presented matches the item that had been presented *n* items back. Often in studies, the *n* varies across experimental blocks, in order to assess the effect of different memory demands upon behavioral performance and physiological correlates. In some studies, stimulus sequences are explicitly manipulated to include not only match and nonmatch trials, but also lure trials. Lure trials are those in which the current stimulus is the same as a recently presented stimulus, but is not in the correct serial position to be a match (e.g., in a three-back task, the second

letter *K* in the sequence *T–K–W–K* is a lure, because it matches the letter presented two back).

Similarities and discrepancies between complex span and n-back tasks

On the surface, both complex span and *n*-back tasks seem to tap related aspects of WM functioning. In both tasks, subjects need to maintain information from a set of possible stimuli (e.g., letters, words, digits, symbols, or locations). The to-be-remembered items are typically presented visually or aurally. Across trials, subjects must remember the currently relevant information and prevent interference from recently presented items. Items need to be accessible in memory for short periods of time (seconds), until the necessary retention interval has passed. Thus, complex span and *n*-back tasks both seem likely candidates for assessing the WM system. Indeed, performance on both kinds of tasks has provided converging evidence about similar research topics (Table 1).

However, other, more direct lines of experimental and correlational research have provided evidence suggesting a relatively weak relationship between complex span and *n*-back performance. First, training studies offer the opportunity to observe whether practice on one type of WM task affects performance on other WM measures. When subjects repeatedly practice *n*-back tasks (dual or single), they consistently show transfer to other versions of *n*-back tasks (Anguera et al., 2012; Jaeggi, Studer-Luethi et al., 2010; Li et al., 2008). Likewise, when subjects repeatedly practice complex span tasks, they show transfer to other, unpracticed versions of complex span tasks (Richmond, Morrison, Chein, & Olson, 2011). However, multiple WM-training studies using either single or dual *n*-back training did not demonstrate transfer to complex span measures (Chooi & Thompson, 2012; Jaeggi, Buschkuehl, Perrig, & Jonides, 2008; Jaeggi, Studer-Leuthi et al., 2010; Kundu, Sutterer, Emrich, & Postle, in press; Li et al., 2008; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Redick et al., 2013; Thompson et al., 2013; but see Anguera et al., 2012). Thus, many WM training studies suggest that

Table 1 Comparison of complex span and *n*-back tasks as measures of working memory (WM), with representative publications

Outcome/Effect	Complex Span	*N*-Back
Positive correlation with fluid intelligence tests (RAPM)	Kane et al. (2007)	Jaeggi, Buschkuehl et al. (2010)
Evidence for domain-general WM system	Kane et al. (2004)	Nystrom et al. (2000)
Meta-analysis evidence for aging deficit	Bopp and Verhaeghen (2005)	Spencer and Raz (1995)
Review of evidence for schizophrenia deficit	Barch et al. (2009)	Barch (2005)
Sensitivity to subsequent-task effects	Schmeichel (2007)	Scheibe and Blanchard-Fields (2009)

RAPM, Raven's Advanced Progressive Matrices

transfer is task-specific: Improving performance on one type of WM measure (*n*-back) does not lead to improved performance on another category of WM task (complex span). The lack of transfer suggests that improving the processes involved in *n*-back task performance via repeated practice does not result in changes to the processes involved in complex span task performance.

Second, and more directly, correlational research has indicated surprisingly weak relationships between complex span and *n*-back performance. For example, Kane et al. (2007) observed nonsignificant-to-weak correlations between operation span and two- and three-back letter tasks. The weak relationship in Kane et al. (2007) was not attributable to a lack of statistical power, given their large sample size ($N = 129$), nor was it due to low reliability for the operation span or *n*-back tasks. In fact, the nonsignificant-to-weak correlations observed by Kane et al. (2007) replicated those from previous studies (Oberauer, 2005; Roberts & Gibson, 2002). Notably, Kane et al. (2007) observed that the complex span and *n*-back tasks accounted for independent variance in Raven's Advanced Progressive Matrices, despite the weak association between the two WM measures. More recently, Jaeggi, Buschkuehl, Perrig, and Meier (2010) demonstrated similar nonsignificant complex span and *n*-back correlations.

Present research

The present research was designed to quantify the nature of the relationship between complex span and *n*-back tasks. In addition to Kane et al. (2007) and Jaeggi, Buschkuehl et al., (2010), the authors of several other recently published studies have administered complex span and *n*-back tasks. Interestingly, two studies using latent-variable analyses indicated substantial correlations between factors containing complex span tasks, on the one hand, and *n*-back tasks, on the other (Burgess, Gray, Conway, & Braver, 2011; Schmiedek, Hildebrandt, Lövdén, Wilhelm & Lindenberger, 2009). Therefore, we conducted a meta-analysis to determine the magnitude of the relationship between complex span and *n*-back tasks across studies. We also attempted to discover variables that might moderate the relationship, as will be seen below.

Because both Jaeggi, Buschkuehl et al., (2010) and Kane et al. (2007) cited studies suggesting that *n*-back tasks might be more strongly correlated with simple span measures than with complex span tasks, we also conducted a meta-analysis on the relationship between simple span and *n*-back tasks. Although complex and simple span tasks tend to correlate strongly and measure many similar processes (Unsworth & Engle, 2007), assessing the relationship between *n*-back and simple span tasks could provide information about what common WM processes the different tasks measure.

Method

Study selection

We identified studies by searching through the PsycINFO database using the keywords "working memory," "complex span," "simple span," and "*n*-back," along with the specific names of complex span tasks. Other studies were identified by searching through the publications and references of prominent WM researchers. Several studies were identified in which both complex span and *n*-back tasks were administered, but the correlations were not reported. In these cases, the authors were contacted and asked to provide the specific correlations, if possible. In some cases, the authors provided correlations for tasks or conditions that were administered but not reported separately in the published article. If the study contained separate analyses for young and older adults, only the young-adult data were used. If WM tasks were administered at both pre- and posttest, only the pretest data were used.

Design and analyses

Most studies have reported complex span task scores as performance on the storage aspect of the task (accuracy or number of items recalled), excluding performance on the processing part. This practice is relatively common in the WM literature and is based on examination of the psychometric properties of alternative scoring methods (Conway et al., 2005; Redick et al., 2012; Unsworth & Engle, 2007). However, there is no consensus on what is the dependent variable for the *n*-back task. Authors have reported performance in terms of mean accuracy and/or response times, and have included (a) overall accuracy; (b) accuracy and/or response times reported for specific trial types (targets, nontargets, lures); (c) hit and false alarm rates; and (d) signal detection measures of sensitivity and bias. Although the use of overall accuracy may not be ideal, we decided that it was the one measure of *n*-back performance that was both reported most often and the easiest for the authors to obtain upon request. In addition, multiple formulas can be used to derive signal detection measures, particularly in terms of the decision about how to correct for hit and false alarm rates at ceiling and at floor. Therefore, except where noted in Table S1, overall *n*-back accuracy was used as the dependent variable in the meta-analysis.[1]

---

[1] Because obtaining overall accuracy was not possible for all of the studies in the meta-analysis, we decided to include scoring method as a moderator of the complex span and *n*-back relationship. Our prediction was that the use of signal detection or corrected-accuracy (hits minus false alarms) as dependent variables would represent performance more accurately than would global accuracy measures. However, the moderator analysis (Table S2) indicated that in the present studies, no evidence was apparent that using overall *n*-back accuracy instead of signal detection or corrected-accuracy measures affected the observed correlations with complex span performance.

The main analyses of interest were meant to specify the magnitude of the correlations between (a) complex span and *n*-back, and (b) simple span and *n*-back tasks. For studies with multiple measures of complex span or *n*-back, the correlations were averaged together to deal with sample-dependence issues. After calculating the averaged correlations for each study, mean-weighted correlation coefficients ($r+$) were calculated, along with the 95 % confidence interval for $r+$ (Hedges & Olkin, 1985). Finally, as a measure of the heterogeneity of the correlations derived from the different studies, Q_{total} was calculated and tested for significance in relation to a χ^2 distribution (degrees of freedom = number of studies – 1).

Moderator variables

In an effort to examine sources of heterogeneity in the literature, we conducted moderator analyses using the following variables:

N-back load Where possible, our analyses focused separately on correlations of the complex and simple span tasks with $n = 2$ and 3. Too few studies had *n* at other levels for us to include them.

Automated versus traditional complex span For complex span tasks only, we examined separately the correlations with *n*-back for traditional versus automated complex span tasks (Unsworth, Heitz, Schrock, & Engle, 2005). Automated complex span tasks are completely mouse-driven and require subjects to identify the to-be-remembered items by clicking among a matrix of possible stimuli at test. In contrast, traditional complex span tasks are presented in either paper-and-pencil or a combination of computerized and paper-and-pencil response formats, in which the subject must generate the to-be-remembered items at test. Although the researchers in previous work have concluded that the automated and traditional complex span tasks largely measure the same underlying WM construct (Unsworth et al., 2005), we included it as a possible moderator here.

Complex and simple span content Complex and simple span tasks with verbal stimuli (numbers, letters, or words) as to-be-remembered items were coded separately from complex and simple span tasks with nonverbal memoranda (arrows, locations, patterns, shapes, or symbols).

N-back content *N*-back tasks with verbal stimuli (numbers, letters, or words) were coded separately from *n*-back tasks with nonverbal stimuli (locations or shapes).

Recall order For digit span only, studies were coded separately if the authors had administered either the forward or the backward version of the test. Although

backward digit span is typically assumed to be a WM measure because item order must be manipulated, other studies have provided evidence that in young adults, forward and backward digit span measure largely common cognitive processes (Rosen & Engle, 1997; St. Clair-Thompson, 2010).

Results and discussion

Descriptive information about all of the studies included in the meta-analysis is provided in Table S1. As can be seen in that table, verbal, numerical, and visuospatial domains were represented across the studies. Also apparent in Table S1 is that in numerous studies the researchers administered multiple measures of complex span tasks but only one *n*-back task to subjects. Finally, the samples consisted mostly of young-adult subjects, as indicated by the mean age of the subjects, the age range of the subjects, or the fact that subjects were high school or undergraduate students.

Table 2 provides the aggregated correlations from the studies contributing to the meta-analysis. The meta-analysis results are presented in Table 3. Examining the confidence intervals reveals that all of the mean-weighted correlations were statistically greater than zero. The complex span and *n*-back correlation was $r+ = .20$ (95 % CI = .16 to .24), and the simple span and *n*-back correlation was $r+ = .25$ (95 % CI = .21 to .30). The overlap of the confidence intervals indicates that the difference between the two correlations was not significant. Although the test for heterogeneity was not significant for the simple span correlation, a significant Q_{total} value for the complex span correlation indicated heterogeneity among the correlations across the 20 samples.

As can be seen in Table 2, some studies have reported rather large zero-order correlations between complex span and *n*-back tasks. Several moderator variables were examined to try to account for this heterogeneity (Table 3). First, the complex span correlations did not differ as a function of *n*-back load of $n = 2$ or 3. Second, the administration method of the complex span tasks did not affect the correlation with *n*-back accuracy. However, moderator analyses determined that the verbal or nonverbal nature of the to-be-remembered items influenced the magnitude of the complex span and *n*-back correlation: *N*-back correlations were significantly higher with nonverbal ($r+ = .31$, 95 % CI = .26 to .36) than with verbal ($r+ = .18$, 95 % CI = .14 to .22) complex span tasks. However, complex span correlations were not statistically different for nonverbal ($r+ = .23$, 95 % CI = .17 to .30) versus verbal ($r+ = .16$, 95 % CI = .11 to .21) *n*-back tasks. Finally, the studies were further divided according to the content of both the complex span and *n*-back tasks. The correlation between nonverbal complex span and nonverbal *n*-back tasks was

Table 2 Correlations with the *n*-back task from individual studies in the meta-analysis

Study	N	Complex Span		Simple Span	
		j	*r*	*j*	*r*
Anguera et al. (2012, Study 2)	67	2	.22		
Burgess et al. (2011)	112	4	.41		
Colom et al. (2008, Study 2)	111	3	.18	3	.13
Colom et al. (2008, Study 3)	289	2	.24	2	.23
Gevins and Smith (2000)	48			1	.36
Greenstein and Kassel (2009)	33	4	.50		
Haatveit et al. (2010)	318			1	.37
Jaeggi, Buschkuehl et al. (2010, Study 1)	116	4	−.07		
Jaeggi, Buschkuehl et al. (2010, Study 2A)	70	2	.10	4	.18
Jaeggi, Buschkuehl et al. (2010, Study 2B)	70	2	.14	4	.16
Jaeggi, Studer-Luethi et al. (2010, Study 1)	104	9	.29	6	.13
Jaeggi, Studer-Luethi et al. (2010, Study 2)	83	3	.11		
Kane et al. (2007)	129	4	.14		
Klatzky et al. (2008, Study 2)	12			4	.55
Krumm et al. (2009)	199	3	.36	3	.26
Li et al. (2008)	46	8	.12		
McAuley and White (2011)	38			4	.33
Oberauer (2005, Study 2)	119	1	.17	1	.19
Roberts (1998, Study 2)	60	2	.21	3	.31
Roberts and Gibson (2002)	26	4	.01	3	.28
Samuels et al. (2007)	20			1	.23
Schmiedek et al. (2009)	96	6	.28		
Schmiedek et al. (2010)	145	5	.25		
Unsworth (2010)	165	3	.08		
Unsworth et al. (2009)	138	1	.08		
Walkenhorst and Crowe (2009)	60			4	.27

Any discrepancies between the *N*s reported here and those in the articles cited reflect the values provided by the corresponding author for the correlations of interest

highest ($r+ = .32$, 95 % CI = .23 to .40) and was significantly greater than the correlation between verbal complex span and verbal *n*-back tasks ($r+ = .14$, 95 % CI = .09 to .19).

For the simple span correlations, none of the moderator analyses based on the content (verbal/nonverbal) of the simple span or the *n*-back tasks produced a significant difference. In addition, the simple span correlations did not differ as a function of *n*-back load ($n = 2$ or 3). However, when we analyzed the order of administration for the digit span, the correlation with *n*-back was significantly greater for the backward ($r+ = .31$, 95 % CI = .24 to .37) than for the forward ($r+ = .16$, 95 % CI = .10 to .23) condition.

Discussion

The meta-analysis results can be summarized as follows. First, the correlations of the complex span measures with *n*-back

tasks were significantly greater than zero, but still weaker than would be expected for measures of the same underlying WM construct. In this respect, the meta-analysis results confirm those of Kane et al. (2007) and Jaeggi, Buschkuehl et al., (2010), individual studies that were specifically designed to address the magnitude of the relationship between complex span and *n*-back. Of course, the meta-analytic results extend the findings of Kane et al. (2007) and Jaeggi, Buschkuehl, et al. by measuring the relationship across multiple samples, different complex span tasks, and variations in the *n*-back tasks administered.

As we mentioned in the introduction, some studies have reported rather large zero-order correlations among complex span and *n*-back tasks, indicating more overlap in the processes involved in the successful completion of both tasks. The meta-analysis results indicated significant heterogeneity among the complex span and *n*-back correlations in the literature. Moderator analyses determined that the verbal or

Table 3 Overall meta-analysis results and results for moderators of complex and simple span correlations with the *n*-back task

Task/Moderator	k	N	r+	LL/UL	Q	p
Complex Span	20	2,178	.20	.16/.24	34.67	.02
Load						
N = 2	10	1,051	.15	.09/.21	17.29	.04
N = 3	14	1,374	.20	.15/.25	28.52	<.01
Automated/Traditional Administration						
Automated	6	590	.16	.08/.24	8.67	.12
Traditional	14	1,588	.21	.17/.26	24.77	.02
Complex Span Content						
Nonverbal	10	1,300	.31	.26/.36	22.13	<.01
Verbal	20	2,178	.18	.14/.22	34.06	.02
N-Back Content						
Nonverbal	8	781	.23	.17/.30	24.97	<.01
Verbal	13	1,447	.16	.11/.21	16.12	.19
Complex Span/*N*-Back Content						
Nonverbal/nonverbal	4	445	.32	.23/.40	3.59	.31
Nonverbal/verbal	6	789	.25	.18/.31	10.66	.06
Verbal/nonverbal	8	781	.19	.12/.26	19.57	<.01
Verbal/verbal	13	1,658	.14	.09/.19	15.40	.22
Simple Span	15	1,544	.25	.21/.30	12.44	.41
Load						
N = 2	9	1,060	.26	.20/.31	12.49	.13
N = 3	6	336	.23	.13/.33	1.54	.91
Simple Span Content						
Nonverbal	5	778	.26	.20/.33	2.81	.59
Verbal	14	1,425	.24	.19/.29	16.82	.21
N-Back Content						
Nonverbal	7	531	.25	.17/.33	4.45	.62
Verbal	10	1,111	.26	.21/.32	8.78	.46
Order						
Digit span forward	9	828	.16	.10/.23	5.03	.75
Digit span backward	9	718	.31	.24/.37	8.75	.16

LL/UL, Lower limit/upper limit of the 95 % confidence interval of *r*+. *Q* provides a measure of the homogeneity of the correlations across studies, with *k* − 1 degrees of freedom

visuospatial nature of the to-be-remembered items influenced the magnitude of the complex span and *n*-back correlation. One prediction was that higher correlations would be obtained when the different WM tasks used stimuli from the same domain, instead of from across domains. For example, Redick et al. (2012) observed, in a sample of over 6,000 subjects, that the correlation between verbal complex span tasks (*r* = .68) was higher than the correlation between verbal and visuospatial complex span (*r* = .52–.53) tasks. The moderator analyses both partially support and partially undermine this prediction. The highest correlation was obtained when both the complex span and *n*-back tasks used visuospatial

content. However, the lowest correlation was obtained when both the complex span and *n*-back tasks used verbal to-be-remembered information. Thus, as can be seen in Table 3, a more accurate description of the complex span and *n*-back results would be that the correlation between the types of tasks tended to be higher if one or both included visuospatial content as to-be-remembered memory items.

For the simple span results, the simple span and *n*-back correlation did not differ from the complex span and *n*-back correlation. Interestingly, when examining digit span specifically, the digit span backward correlation with *n*-back was greater than the digit span forward correlation with *n*-back.

In addition, the digit span backward correlation with *n*-back ($r + = .31$, 95 % CI = .24 to .37) was greater than the verbal complex span correlation with *n*-back ($r + = .18$, 95 % CI = .14 to .22). A speculative interpretation is that the digit span backward task requires subjects to temporally reorder information after the digits have been encoded, thereby forcing them to update the relative serial position of the items in memory (e.g., 1–6–4–7 must be recoded as 7 in the first serial position, 4 in the second serial position, etc.). Similarly, across trials in the *n*-back task, subjects must change the serial position of items that have been previously encoded as new items are continuously presented (e.g., H goes from being item *n*, to item *n* − 1, to item *n* − 2, etc.). The similarity of these reordering processes may account for the higher *n*-back correlation with digit span backward than with either digit span forward or verbal complex span tasks (see Oberauer, 2005, and Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011, for more on the role of binding items to the appropriate temporal context within the *n*-back task).

Sample composition as a moderator variable

An often overlooked aspect of individual-differences studies is whether the sample included a wide variation in cognitive abilities. Redick et al. (2012) showed that complex span intracorrelations varied as a function of the sample type—correlations among complex span tasks were smaller for samples from more selective universities than for samples from more diverse universities and samples composed of community volunteers. The pattern of results in Redick et al. (2012) is consistent with Spearman's (1927) "law of diminishing returns," in which correlations among mental-ability tests are smaller in individuals with higher IQ. Although IQ estimates are not available for all subjects in the present meta-analysis, it is apparent that, across studies, the samples represented different points along the mental-ability continuum.

For example, Roberts and Gibson (2002) used a sample of students at Massachusetts Institute of Technology and obtained an average complex span and *n*-back correlation of $r = .01$. In two separate samples of University of Georgia students, Unsworth (2010) and Unsworth et al. (2009) obtained average complex span and *n*-back correlations of $r = .08$. In Study 1 of Jaeggi, Buschkuehl, et al. (2010), 95 % of the subjects had a college degree or higher level of education—and the average complex span and *n*-back correlation was $r = −.07$. In contrast, Burgess et al. (2011) used a combination of Washington University St. Louis students and community volunteers and obtained an average complex span and *n*-back correlation of $r = .41$. Greenstein and Kassel (2009) recruited Chicago area community members and obtained an average complex span and *n*-back

correlation of $r = .50$. Greenstein and Kassel's results are particularly interesting, given that they used the same two- and three-back task as Kane et al. (2007), who found low and nonsignificant correlations with operation span using an undergraduate-only sample. In fact, Greenstein and Kassel argued that their "sample was probably more diverse in general cognitive ability" (p. 87) than was the Kane et al. (2007) sample. Making some assumptions about the IQ range of the subjects in these example studies, the pattern of correlations is consistent with the law of diminishing returns.

In order to further investigate the role of sample composition in complex span and *n*-back correlations, we analyzed previously unpublished data.[2] One hundred fifty-five subjects from 18 to 35 years old completed, as part of a larger study, automated operation span, automated symmetry span, and a three-back task with letter stimuli. Importantly, we also had access to the self-reported college status of these subjects. Seventy-five of the subjects were enrolled as students at Georgia Tech (GT) or Georgia State University (GSU), whereas 80 subjects were not college students or had attended another area college (primarily technical- and associate-degree-granting schools). This division of the sample led to roughly equal sample sizes that were also large enough to afford meaningful conclusions.

The correlations are presented in Table 4. In the overall sample, the zero-order correlations were moderate and larger than the complex span and *n*-back correlation obtained in the meta-analysis, but similar in magnitude to those from studies using similar sampling methods (e.g., Burgess et al., 2011). When examining the subsamples, the patterns of correlations differed, despite the similar sample sizes. In the GT + GSU subsample, the complex span and *n*-back correlations were not significant, but in the None + Other subsample, the complex span and *n*-back correlations were significant. For symmetry span, the subsample results indicated 10.6 % ($.35^2$–$.13^2$) more shared variance

---

[2] The original sample contained 171 subjects, but 16 of them were outliers and were removed. The automated operation and symmetry span tasks were administered using the versions described in Redick et al. (2012). The subjects solved math operations while trying to remember letters interleaved with each trial, and made symmetry decisions while trying to remember square locations interleaved with each trial. The three-back task required subjects to make a target/nontarget decision to each visually presented letter. Each of the two blocks of 40 trials contained 25 % target, 50 % nontarget nonlure, and 25 % lure trials. Of the ten lure trials presented in each block, eight were two-back lure trials, and one each were one-back and four-back lure trials. Descriptive statistics for each task are provided in the supplemental materials.

The operation and symmetry span data were used in other publications in relation to conditional go/no-go performance (Redick, Calvo, Gay, & Engle, 2011) and change detection performance (Shipstead, Redick, Hicks, & Engle, 2012). The *n*-back data have not been reported previously.

Table 4 Complex span and *n*-back correlations as a function of college status

Sample	N	Operation	Symmetry	*z*WM Composite
Overall	155	.26*	.31*	.34*
GT + GSU	75	.17	.13	.18
None + Other	80	.22^	.35*	.33*

GT, Georgia Tech; GSU, Georgia State University; None, no college; Other, other college; WM, working memory. * *p* < .05, ^ *p* = .05

with the *n*-back in the None + Other subsample than in the GT + GSU subsample.

One possible reason that the samples with high-ability subjects have tended to produce smaller correlations is that there is insufficient variability on the tasks, as compared to more diverse samples. Restriction of range would certainly be a problem that could limit the correlation magnitude. Although we cannot speak for the studies in the meta-analysis, restriction of range did not appear to be a problem in the GT + GSU subsample here. As can be seen in Table S3, although the GT + GSU subsample had a higher mean than the None + Other subsample on operation span, symmetry span, and three-back accuracy (all *p*s < .01), the variability did not appear to be very different between the subsamples. Levene's tests for variance differences were only significant for operation span (*p* = .02), but not for symmetry span (*p* = .99) or a three-back task (*p* = .46).

Implications for WM research

The most important finding from the meta-analysis is that complex span and *n*-back tasks are weakly correlated, which is important, considering that the tasks are thought to both be measures of the same WM system. For comparison, in Daneman and Merikle's (1996) meta-analysis, the complex span correlations with global and specific language comprehension measures ranged from *r* = .30 to .52. The correlations in Daneman and Merikle are not assumed to be measures of the same underlying WM construct, yet they are higher than the meta-analysis correlation reported here between complex span and *n*-back. Because of positive manifold, in which reliable tests produce positive correlations with each other, regardless of the exact underlying construct that the measure is designed to assess, it is striking how small the correlation is between complex span and *n*-back tasks.

The meta-analysis results validate what Kane et al. (2007) discussed. Namely, WM researchers should be specific in interpreting the results of studies that use complex span versus *n*-back tasks as WM measures. The results of WM research using the different categories of tasks cannot simply be used interchangeably. Therefore, as Kane et al.

(2007) stated, the large body of cognitive neuroscience research using the *n*-back task is informative for understanding the neural substrates of WM, but it may shed little light on the nature of individual differences in WM as measured by complex span tasks. Likewise, for WM-training studies using *n*-back tasks, attempts to find "near transfer" to complex span tasks are somewhat misguided. In light of the small amount of shared variance between the types of WM tasks, practice on the *n*-back task need not affect performance on complex span measures of WM at all, and vice versa.

At first glance, our results indicating minimal processing overlap between complex span and *n*-back tasks seem completely inconsistent with the findings of Schmiedek et al. (2009). Those researchers created two WM latent variables, one with loadings from multiple updating tasks, including a three-back task, and a complex span factor composed only of reading, counting, and rotation span. The best-fitting model in Schmiedek et al. (2009) indicated that the correlation between these two WM latent variables was *r* = .96. However, closer inspection reveals that their results do fit with those of our meta-analysis. Reading span and counting span (both verbal tasks) had small, nonsignificant correlations with the nonverbal three-back task. In contrast, rotation span had sizeable, significant correlations with three-back; the content of both rotation span and three-back were nonverbal stimuli. The correlation between the two latent variables was so high at least partially because the complex span factor loading was much stronger for rotation span (.70) than for reading (.34) or counting span (.37).

Obviously, the reliability of the WM measures will place an upper limit on the maximum correlation that can be obtained between them. Given their frequent use in the individual-differences literature, the reliability of various complex span tasks has been assessed, and it is often quite high (Redick et al., 2012). However, less psychometric work has been carried out using the *n*-back task, and the findings have been largely inconsistent. For example, Jaeggi, Buschkuehl et al. (2010) examined the split-half reliability of verbal (auditory modality) and spatial (visual modality) one-, two-, and three-back tasks in three different samples. The reliabilities for the corrected-accuracy (hits minus false alarms) dependent variables were generally low (two-back, *r* = .09 to .85; three-back, *r* = .39 to .60). Indeed, Jaeggi, Buschkuehl et al. (2010) concluded that "the *N*-back task does not seem to be a useful measure of individual differences in WMC, due to its low reliability" (p. 409).

However, other studies have reported acceptable reliabilities for the *n*-back task. For example, the following studies included in the meta-analysis reported reliabilities greater than .70: Jaeggi, Studer-Luethi et al. (2010); Krumm et al. (2009); Oberauer (2005); Schmiedek et al. (2009); and Unsworth (2010). In addition, although the complex span

and *n*-back tasks might not be correlated strongly, the fact that *n*-back tasks are significantly correlated with other measures, such as Raven's Advanced Progressive Matrices (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Jaeggi, Buschkuehl et al., 2010; Jaeggi, Studer-Luethi et al., 2010; Kane et al., 2007), indicates that the *n*-back can capture meaningful individual-differences variation. In contrast to the Jaeggi, Buschkuehl, et al. quote above, Kane et al. (2007, p. 618) concluded that "*n*-back is a reliable individual-differences indicator of some construct(s)." This, then, is the puzzle—why would two measures of WM correlate less strongly than measures of putatively different constructs?

In contrast to viewing WM as a monolithic construct, we agree with other theories (e.g., Oberauer, Süß, Wilhelm, & Sander, 2007; Unsworth & Spillers, 2010b) that WM is a multifaceted system that relies on multiple processes (encoding, maintenance, recall, recognition, familiarity, updating, temporal ordering, binding, attention, and inhibition). If complex span and *n*-back tap largely separate components of the WM system, then one can account for why two WM measures might be only weakly correlated.

One notable distinction between complex span and *n*-back tasks is whether retrieval is based on recall or recognition. On complex span tasks, subjects must generate the sequence of stimuli presented on a given trial, whereas *n*-back tasks require subjects to recognize a current item as an item that was recently presented in the correct serial position. The importance of this distinction between recall versus recognition is evident in two separate lines of research.

First, Shelton and colleagues (Shelton, Elliot, Hill, Calamia, & Gouvier, 2009; Shelton, Metzger, & Elliot, 2007) measured the relationship between complex span tasks and a variant of the *n*-back task called the *modified lag task*. In this task, subjects are presented with a series of stimuli of unknown length. During the presentation of the stimuli, no response is required, but after the sequence is presented, subjects are required to recall either the last item, the penultimate item, or the antepenultimate item. The item that the subjects will need to recall is also unknown until the stimulus presentation is complete. As can be seen, the method of the modified lag task is quite different from the typical *n*-back task, in which, for example, subjects know in advance that they are making target/nontarget decisions about every item presented and know that the decision is based on whether the current item matches the one presented *n* items ago. Germane to the present work, across three unique samples, Shelton and colleagues observed complex span and modified lag task correlations ranging from $r = .38$ to .51. Note that in the Shelton et al. studies, the samples were composed entirely of college students, so the stronger correlations do not

seem to reflect the use of a more diverse sample involving community nonstudents. Instead, the stronger correlations using the modified lag task may reflect increased overlap in the processes involved because of the use of explicit recall, without the opportunity to respond relying upon a familiarity signal. In fact, the modified lag task seems more similar to a running memory span task (Broadway & Engle, 2010; Cowan et al., 2005), in that subjects do not know exactly which part of the upcoming sequence they will need to recall, but do know that the last three items are the most important items to remember.

Second, neuroimaging studies seem to point to a common neural underpinning for performance on both complex span and *n*-back tasks—namely, involving prefrontal, parietal, and cingulate cortex regions (for an *n*-back fMRI meta-analysis, see Owen, McMillan, Laird, & Bullmore, 2005). In addition, although hundreds of fMRI studies have been conducted using *n*-back tasks, relatively few fMRI studies have been conducted using complex span tasks. One reason is that complex span tasks do not easily lend themselves to be administered in fMRI studies, because there is typically little control over the timing of events such as processing task decisions, and especially the retrieval process for each trial. As such, no meta-analysis of complex span fMRI activation has been conducted. However, two recent complex span fMRI studies (Chein, Moore, & Conway, 2011; Faraco et al., 2011) suggest an interesting difference in brain activity patterns versus the *n*-back fMRI results. Both studies showed significant involvement of the medial temporal lobe during performance of various complex span tasks. The medial temporal activity, which is not typically present during *n*-back tasks, suggests again that the retrieval processes involved in the explicit search and recall of items during complex span tasks differentiate the two types of WM tasks.

Limitations and future directions

The advantage of a meta-analysis is that the method aggregates across multiple studies in order to estimate the magnitude of the relationship on the basis of larger samples. However, a limitation of meta-analysis is that aggregation may obscure factors that affect whether or not a relationship is observed. Although we tested potential moderators, and also examined the role of sample composition, many of the moderators were coarse, given the limited number of available studies. The verbal/nonverbal moderator variable collapsed across different types of stimuli within each category. Other *n*-back task variables that varied across the studies included (a) response type (target-only or target/nontarget decision), (b) presentation rate, (c) presentation modality, and (d) the size of the pool of possible stimuli. Also, the frequency and inclusion

of lure trials may be an important consideration. As Szmalec et al. (2011) stated, "an *n*-back procedure with and one without lure trials are almost two different tasks in terms of what they measure" (p. 148). Many of the studies included here did not clearly indicate whether the *n*-back task included lures, and if so, on what proportion of trials.

In addition, the meta-analysis collapsed over variables that affect complex span correlations with other measures, including (a) the scoring procedure (Unsworth & Engle, 2007); (b) experimenter- versus subject-paced tasks (Friedman & Miyake, 2004); and (c) random versus ascending list-length presentation (St. Clair-Thompson, 2012). Research assessing whether these variables affect correlations with *n*-back performance in theoretically meaningful ways (e.g., a scoring procedure that minimizes contributions from secondary memory—Unsworth & Engle, 2007) may prove fruitful for understanding more about what complex span and *n*-back tasks do share in common.

In addition, there is strong evidence (e.g., Bailey, Dunlosky, & Kane, 2011; Unsworth & Spillers, 2010a) that individual differences in strategy use contribute to performance on complex span tasks. However, less is known about the explicit strategies that different subjects engage in during the performance of *n*-back tasks. Because strategy use during complex span task performance has been shown to mediate the relationship with other cognitive measures (Turley-Ames & Whitfield, 2003), more knowledge about the strategies underlying *n*-back performance would be informative. One possibility is that the presentation rate of stimuli during an *n*-back task may affect the particular strategies employed: A slower rate may allow covert rehearsal and active updating of the current memory set, whereas a faster rate may force subjects to rely more on familiarity matching. The *n*-back task is often assumed to measure updating, with subjects actively updating the current contents of a limited portion of temporary memory. However, as was illustrated by Szmalec et al. (2011), the underlying processes contributing to *n*-back performance are not simply the reflection of a mental counter that updates with the relevant information as necessary.

In general, many of the studies in the meta-analysis used multiple complex and/or simple span measures, but only one *n*-back task. Future latent-variable studies with multiple complex span and multiple *n*-back tasks would be helpful for determining the relationship at the construct level.

Conclusion

The results of the meta-analysis indicate that two categories of tasks used to measure WM, complex span and *n*-back tasks, are only weakly related. Although tasks using nonverbal stimuli have produced higher correlations, overall the findings demonstrate little shared variance among the two types of tasks. In fact, the digit span backward task was more strongly correlated with *n*-back performance than the verbal complex span tasks were. The present findings indicate that complex span and *n*-back tasks cannot be used interchangeably as WM measures in research applications. WM researchers should consider how the lack of this relationship affects interpretation of their own results and the work of others.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuehl, M., Benson, B. L., Jennett, S., . . . Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, *228*, 107–115.

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

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

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory and Cognition, 39*, 447–461. doi:10.3758/s13421-010-0034-3

Barch, D. M. (2005). The cognitive neuroscience of schizophrenia. *Annual Review of Clinical Psychology, 1*, 321–353.

Barch, D. M., Berman, M. G., Engle, R. W., Jones, J. H., Jonides, J., MacDonald, A. W., . . . Sponheim, S. R. (2009). CNTRICS final task selection: Working memory. *Schizophrenia Bulletin*, *35*, 136–152.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology, 60B*, P223–P233.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*, 563–570. doi:10.3758/BRM.42.2.563

*Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*, 674–692.

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.

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage, 54*, 550–559.

Chooi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40*, 531–542.

*Colom, R., Abad, F. J., Quiroga, M.Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence, 36*, 584–606.

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. doi:10.3758/BF03196772

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. doi:10.1016/j.cogpsych.2004.12.001

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466. doi:10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433. doi:10.3758/BF03214546

Faraco, C. C., Unsworth, N., Langley, J., Terry, J. D., Li, K., Zhang, D., et al. (2011). Complex span tasks and hippocampal recruitment during working memory. *NeuroImage, 55*, 773–787.

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.

*Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10*, 829–839.

*Greenstein, J. E., & Kassel, J. D. (2009). The effects of smoking and smoking abstinence on verbal and visuospatial working memory capacity. *Experimental and Clinical Psychopharmacology, 17*, 78–90.

*Haatveit, B. C., Sundet, K., Hugdahl, K., Ueland, T., Melle, I., & Andreassen, O. A. (2010). The validity of d prime as a working memory index: Results from the "Bergen *n*-back task." *Journal of Clinical and Experimental Neuropsychology, 32*, 871–880.

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

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*, 6829–6833. doi:10.1073/pnas.0801268105

*Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the *n*-back task as a working memory measure. *Memory, 18*, 394–412.

*Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence, 38*, 625–635.

*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. doi:10.1037/0278-7393.33.3.615

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. doi:10.1037/0096-3445.133.2.189

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*, 352–358.

*Klatzky, R. L., Giudice, N. A., Marston, J. R., Tietz, J., Golledge, R. G., & Loomis, J. M. (2008). An *n*-back task using vibrotactile stimulation with comparison to an auditory analogue. *Behavior Research Methods, 40*, 367–372. doi:10.3758/BRM.40.1.367

*Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37*, 347–364.

Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (in press). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. The Journal of Neuroscience.

*Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging, 23*, 731–742.

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual *n*-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review, 20*, 135–141. doi:10.3758/s13423-012-0335-6

*McAuley, T., & White, D. A. (2011). A latent variables explanation of processing speed, response inhibition, and working memory typical development. *Journal of Experimental Child Psychology, 108*, 453–468.

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

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, F. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *NeuroImage, 11*, 424–446.

*Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134*, 368–387. doi:10.1037/0096-3445.134.3.368

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York: Oxford University Press.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). *N*-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46–59. doi:10.1002/hbm.20131

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., et al. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*, 164–171.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 308–324.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology. General, 142*, 359–379. doi:10.1037/a0029082

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging, 26*, 813–822. doi:10.1037/a0023631

*Roberts, R. M. (1998). *Pruning the right branch: Memory and understanding sentences*. Unpublished doctoral dissertation, California State University, Los Angeles.

*Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*, 573–98.

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

*Samuels, M. H., Schuff, K. G., Carlson, N. E., Carello, P., & Janowsky, J. S. (2007). Health status, psychological symptoms, mood, and cognition in l-thyroxine-treated subjects. *Thyroid, 17*, 249–259.

Scheibe, S., & Blanchard-Fields, F. (2009). Effects of regulating emotions on cognitive performance: What is costly for young adults is not so costly for older adults. *Psychology and Aging, 24*, 217–223.

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology. General, 136*, 241–255. doi:10.1037/0096-3445.136.2.241

*Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1089–1096. doi:10.1037/a0015730

*Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience, 2*:27, 1–10. doi:10.3389/fnagi.2010.00027

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. doi:10.1037/0096-3445.125.1.4

Shelton, J. T., Elliot, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical memory tests and their prediction of fluid intelligence. *Intelligence, 37*, 283–293.

Shelton, J. T., Metzger, R. L., & Elliot, E. M. (2007). A group-administered lag task as a measure of working memory. *Behavioral Research Methods, 39*, 482–493.

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory, 20*, 608–628.

Spearman, C. C. (1927). *The abilities of man*. London: Macmillan.

Spencer, W. D., & Raz, N. (1995). Differential age effects on memory for content and context: A meta-analysis. *Psychology and Aging, 10*, 527–539.

St. Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology, 22*, 286–296.

St. Clair-Thompson, H. L. (2012). Ascending versus randomised list lengths in working memory span tasks. *Journal of Cognitive Psychology, 24*, 335–341.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology. Human Perception and Performance, 37*, 137–151.

Thompson, T. W., Waskom, M. L., Garel, K. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. PLOS One, 8, 1–15.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*, 446–468. doi:10.1016/S0749-596X(03)00095-0

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

*Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134*, 16–28. doi:10.1016/j.actpsy.2009.11.010

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*, 1038–1066. doi:10.1037/0033-2909.133.6.1038

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. doi:10.3758/BF03192720

*Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences, 30*, 194–200.

Unsworth, N., & Spillers, G. J. (2010a). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual-retrieval. *Psychonomic Bulletin & Review, 17*, 200–205. doi:10.3758/PBR.17.2.200

Unsworth, N., & Spillers, G. J. (2010b). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392–406.

*Walkenhorst, E., & Crowe, S. F. (2009). The effect of state worry and trait anxiety on working memory processes in a normal sample. *Anxiety, Stress, and Coping, 22*, 167–187.