

The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning

By: [Michael J. Kane](#), Stephen W. Tuholski, David Z. Hambrick, Oliver Wilhelm, Tabitha W. Payne, and Randall W. Engle

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 visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189-217

Made available courtesy of The American Psychological Association: <http://www.apa.org/>

***** Note: Figures may be missing from this format of the document**

Abstract:

A latent-variable study examined whether verbal and visuospatial working memory (WM) capacity measures reflect a primarily domain-general construct by testing 236 participants in 3 span tests each of verbal WM, visuospatial WM, verbal short-term memory (STM), and visuospatial STM, as well as in tests of verbal and spatial reasoning and general fluid intelligence (Gf). Confirmatory factor analyses and structural equation models indicated that the WM tasks largely reflected a domain-general factor, whereas STM tasks, based on the same stimuli as the WM tasks, were much more domain specific. The WM construct was a strong predictor of Gf and a weaker predictor of domain-specific reasoning, and the reverse was true for the STM construct. The findings support a domain-general view of WM capacity, in which executive-attention processes drive the broad predictive utility of WM span measures, and domain-specific storage and rehearsal processes relate more strongly to domain-specific aspects of complex cognition.

Article:

The generality versus specificity of cognitive abilities, mechanisms, and structures has triggered lively debate throughout psychology's history, for example, it surrounds questions of general versus multiple intelligences (e.g., Guilford, 1967; Jensen, 1998; Spearman, 1927; Thurstone, 1938), single versus multiple pools of attentional resources (e.g., Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980), and process versus systems approaches to long-term memory (LTM; e.g., Toth & Hunt, 1999; Tulving, 1985). Likewise, cognitive psychology's two most prominent models of immediate memory differ most markedly in their proposals about the unitary versus fractionated nature of the construct. Atkinson and Shiffrin (1968) proposed a unitary short-term store (STS) that was specialized for holding information in a speech-based code, and although they allowed that further research might elucidate additional codes, all of these codes were thought to operate within the monolithic STS. In contrast, Baddeley (1986, 2000) proposed a partitioned working memory (WM) system with two domain-specific storage structures: a phonological loop that is specialized for maintaining verbal-linguistic information, much like the STS, and a visuospatial sketchpad that is specialized for maintaining visual and spatial information. Evidence for the distinction between verbal and visuospatial storage comes from numerous empirical dissociations in dual-task, neuropsychological, and neuroimaging studies (see Henson, 2001; Jonides et al., 1996; Logie, 1995).

However, some domain generality is proposed within Baddeley's (1986, 2000) multicomponent model: A central executive acts as an attention-control structure and coordinator for the two storage components and their interaction with LTM. This general executive provides the WM model with its functional emphasis as a system that maintains information in the service of ongoing cognitive activity such as language comprehension, imagery, and reasoning. That is, the dynamic interaction of domain-specific storage with the domain-general central executive gives the WM model the power to address the role of active memory in real-world cognition. One might expect, then, that individual-differences research focusing on the relation between WM and aspects of complex cognitive ability would assume a mixture of domain-specific and domain-general WM contributions. However, this has not always been the case. Daneman and Carpenter (1980, 1983) invented the

first viable measure of WM capacity (WMC) and argued that it selectively engaged the central executive component of WM in a domain-specific fashion. Their reading span task presented short series of sentences for comprehension, followed by a memory test for all the sentence-final words in the series. Reading span is therefore a dual task, incorporating traditional memory span demands with a secondary processing task that putatively engages the central executive. What created widespread interest in reading span (and related WM span tasks), was Daneman and Carpenter's (1980) finding—now extensively replicated—that reading span strongly predicted comprehension abilities in ways that simple short-term memory (STM) storage tasks did not (for reviews, see Carpenter, Miyake, & Just, 1995; Daneman & Merikle, 1996).

Daneman, Carpenter, and their colleagues (e.g., Carpenter & Just, 1988; Daneman & Carpenter, 1980) originally hypothesized that the resource trade-off between processing and storage in the reading-span task is specific to language comprehension (although they subsequently advocated more agnostic, if not more domain-general, views of WMC within symbolic domains; see, e.g., Daneman & Merikle, 1996; Just & Carpenter, 1992). By the domain-specific view, reading-span performance is primarily a function of reading efficiency. With more efficient reading skills, good readers consequently have greater WMC remaining to store the products of that processing. Accordingly, span measures of WMC (i.e., the number of products effectively stored) are necessarily tied to a specific processing task, such as reading: The relatively large WMC of good readers is peculiar to reading-comprehension tasks. In task domains in which good readers are less skilled, their residual storage capacity during processing should be relatively limited.

The strongest evidence for such domain specificity in WMC has come from studies contrasting the utility of span tasks consisting of verbal versus spatial materials for predicting complex verbal versus spatial abilities. If WM span tasks measure domain-specific capacities, then verbal span tasks should have limited value in predicting spatial ability, and vice versa. Indeed, Daneman and Tardif (1987) first reported that whereas span tasks using verbal and numerical materials correlated significantly with verbal ability measures ($r_s = .61$ and $.51$, respectively), a spatial span task did not ($r = -.09$). Using different tasks, Morrell and Park (1993) similarly found that verbal and numerical span tasks, but not spatial span tasks, predicted standardized measures of text comprehension, and only spatial span predicted object assembly performance from diagrammatic, visuospatial instructions.

Shah and Miyake (1996) reported the most influential and compelling demonstration of the verbal–spatial dissociation in WMC. In one study, reading span predicted verbal Scholastic Aptitude Test (SAT) scores more strongly than did a spatial task requiring mental rotation of letters and memory for their orientations ($r_s = .45$ and $.07$, respectively). Conversely, spatial span predicted a composite of standardized visuospatial tests better than did reading span ($r_s = .66$ and $.12$). In an exploratory factor analysis the spatial span and spatial ability tests yielded one factor, and reading span and verbal SAT yielded another. Similar findings with these same span tasks have been reported by other researchers, with reading span predicting only verbal performance and rotation span predicting only spatial performance (Friedman & Miyake, 2000; Handley, Capon, Copp, & Harper, 2002).

In a second study, Shah and Miyake (1996) crossed the domains of the processing and storage components to create two span tests each with matching and mismatching domains. For the matching tasks, one task had both verbal processing and storage stimuli and another had both spatial processing and storage stimuli (reading–word span, rotation–arrow span). The other two tasks had domain-mismatching processing and storage stimuli (reading–arrow span and rotation–word span). Here, domain of the storage items (words vs. arrows), rather than the processing items, most strongly influenced the correlations with verbal and spatial ability measures. For example, spatial-ability scores were best predicted by rotation–arrow and reading–arrow tasks ($r_s = .68$ and $.65$, respectively) compared with rotation–word and reading–word tasks ($r_s = .02$ and $.17$). Although partial-correlation analyses indicated that the processing domain of the tasks had a small effect on the correlations, domain specificity in WMC tasks seemed to be driven primarily by the memory items, indicative of domain-specific storage mechanisms at work, leaving open the possibility that the executive contributions to these span tasks are more domain general in nature.

Despite findings that seem to indicate the domain specificity of WMC, we have proposed a more domain-general view, which is also more consistent with the Baddeley model's (1986, 2000) proposal of a domain-general executive interacting with domain-specific storage structures or processes. Our view is that WM span tasks involve joint contributions of a general executive-attention capability and domain-specific rehearsal, coding, and storage processes. Of importance, the shared variance among measures of WM span and complex cognition reflects primarily the contribution of domain-general attention control, rather than domain-specific storage or rehearsal (e.g., Engle & Kane, 2004; Engle, Kane, & Tuholski, 1999). We argue that the critical, executive-attention capability is one by which memory representations—for action plans, goal states, or environmental stimuli—are maintained in a highly active and easily accessible state. WM span tasks elicit such active maintenance by providing proactive interference that accumulates over trials, making LTM retrieval difficult and slow (e.g., Lustig, May, & Hasher, 2001). Simultaneously, WM span tasks challenge that active maintenance by imposing shifts of attentional focus, to the unrelated secondary task, between the presentations of each to-be-remembered item (e.g., Towse, Hitch, & Hutton, 2000). Therefore, when we use the term *WMC* we really mean the domain-general executive component of the WM system.

Our view is that correlations between WM span and complex cognition are jointly determined by general executive-attention and domain-specific storage but primarily by executive attention. Thus, a WMC measure should be quite general in predicting cognitive function. That is, the memory span test could be embedded in a secondary processing task that is unrelated to any particular skill or ability and still predict success in a higher level task. Evidence supporting this view comes from three sources: (a) manipulating the processing demands of verbal WM span tasks and noting their relations to comprehension, (b) examining the link between verbal WM span and measures of general fluid intelligence, and (c) examining the link between verbal WM span and low-level attention capabilities.

Here, we review these findings briefly. WM span tasks that require either reading comprehension or equation solution as the secondary task account for the same variance in measures of reading comprehension, indicating that reading per se does not drive the correlation between WM span and comprehension (e.g., Turner & Engle, 1989). Moreover, skill in the secondary processing task has no apparent impact on the correlation between WM span and comprehension: Statistically accounting for participants' strategic allocation of processing time to the secondary task, or matching the difficulty of the secondary task to each participant's skill level, does not diminish the correlation between WM span and verbal-ability measures (Conway & Engle, 1996; Engle, Cantor, & Carullo, 1992).

Recent work using more sophisticated latent-variable methods has demonstrated further a strong relation between verbal WMC measures and nonverbal tests of general fluid intelligence (Gf), such as Ravens Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 1998). Here, the variance common to WM span tests representing different verbal or symbolic domains was statistically extracted to yield a pure measure, or latent variable, of the WMC construct. That is, measurement error associated with the task-specific variance in reading comprehension (reading span), mathematical ability (operation span) and enumeration (counting span) was partialled out, leaving only the variance shared among all the tasks. This latent variable correlated substantially (approximately .60) with the variance shared among Gf tasks, whereas a latent variable composed of verbal STM tasks accounted for no unique variance (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999). Thus, as measured by tasks with a variety of symbolic materials, WMC demonstrates predictive power well beyond the verbal domain and represents an important aspect of intelligence.

Quasi-experimental research has shown further that individuals identified as having high or low WMC, on the basis of a verbal WM span task, differ significantly in their performance on attention tasks that bear little surface similarity to span tasks or each other. For example, in a dichotic-listening task, in which participants' names were spoken once in the distractor ear, high-span participants were much less likely to report hearing their names than were low-span participants, indicating that high-span participants more effectively blocked the processing of the irrelevant message (Conway, Cowan, & Bunting, 2001). In an antisaccade task, in which

participants had to quickly look away from an attention-capturing cue to identify a pattern-masked letter appearing in an opposite location, high-span participants were less error prone in blocking eye movements toward the cue, faster in making correct eye movements, and faster in correcting eye-movement errors than were low-span participants (Kane, Bleckley, Conway, & Engle, 2001). Finally, in a Stroop (1935) task in which participants had to name the color of conflicting color words, high-span participants were either faster or less error prone than were low-span participants, depending on task context (Kane & Engle, 2003; Long & Prat, 2002). Verbal measures of WMC thus predict performance on simple attention tasks, presenting either verbal or nonverbal stimuli, either visually or aurally. Of importance is that these attention tasks demand that participants block a prepotent response in favor of a novel goal-directed one, in our view by using executive attention to actively maintain the goal (see also De Jong, 2000, 2001; Duncan, 1995; Roberts & Pennington, 1996).

We see the diverse findings reviewed above as compelling evidence that WMC reflects primarily a domain-general, attentional construct that is important to a range of intellectual abilities. Our view thus conflicts with the findings discussed previously that suggest a strong dissociation between verbal and visuospatial WMC and reasoning (Daneman & Tardif, 1987; Friedman & Miyake, 2000; Handley et al., 2002; Morrell & Park, 1993; Shah & Miyake, 1996). However, we believe there are good reasons to be skeptical of the evidence for a strong domain specificity in WMC.

First, the “domain-specific” studies have limitations related to their participants. Several studies tested relatively small, and potentially quite homogeneous, samples (i.e., students at selective universities; Daneman & Tardif, 1987; Friedman & Miyake, 2000; Handley et al., 2002; Shah & Miyake, 1996). The concern here, from a psychometric perspective, is that when participants are sampled from a restricted range of general ability, general ability can make only limited contributions to any observed correlations. Instead, the variability that is detected, even if substantial, must be attributed to primarily domain-specific skills, strategies, and abilities (e.g., Deary et al., 1996; Legree, Pifer, & Grafton, 1996). Thus, domain generality may have been underestimated in these studies. As a case in point, Shah and Miyake (1996) tested mostly undergraduates from the highly selective Carnegie Mellon University and reported low correlations between verbal and quantitative SAT scores in their samples ($r_s = .28, .45, \text{ and } .59$) relative to those seen in more comprehensive state universities with diverse student bodies (e.g., $r_s = .54, .69, \text{ and } .74$ in Turner & Engle, 1989; Engle et al., 1992; and Engle, Tuholski, et al., 1999, respectively). Such low correlations suggest a limited range of general ability within their participants and an artificially strong influence of domain-specific abilities. In fact, Shah and Miyake (1996) explicitly recognized this issue, noting that their results

may not be generalizable to more cognitively diverse samples, such as those including non-college students, young children, or elderly adults. The dissociation between spatial and verbal measures, for example, could be less apparent among such samples, in which individuals are likely to vary more widely in general ability factors. (p. 21)

As Shah and Miyake (1996) noted further, the benefits of studying ability-restricted samples is that the imposed limitation of domain-general contributions to performance makes it easier to detect and examine domain-specific contributions and mechanisms. We agree that if one’s goal is to examine the workings of domain-specific processes underlying WM span tasks, which do appear to exist, then studying a restricted range of general ability may well be a good strategy. However, if one’s goal is to assess the extent to which domain-specific and domain-general processes jointly contribute to WM span, or to any other cognitive construct, then one must study a population that has ample variability in both general and domain-specific cognitive ability. Also, in principle, domain-specific abilities could be restricted within a population and thereby be underestimated as a contributor to performance. For example, verbal ability would likely be restricted in a group of Pulitzer-winning novelists, as would spatial ability in professors of architecture, and leaping ability in professional basketball players. However, we see no reason to believe that verbal- or spatial-specific abilities would be more restricted in a selective-university population than would general cognitive ability.

A further obstacle to interpreting reported verbal–spatial WMC dissociations is that some of the verbal and spatial tasks differed dramatically in their difficulty. For example, the verbal and numerical tasks used by Daneman and Tardif (1987), in contrast to the spatial task, required higher levels of domain-specific knowledge than is typical of WM tasks (e.g., the processing component of the verbal task required generation of low-frequency words such as *sinewed*). These knowledge requirements can be inferred from the mean processing accuracies in the verbal (78%) and numerical (87%) tasks versus the spatial (96%) tasks, as well as from the fact that nonstorage versions of these processing tasks correlated just as well with verbal ability as did their processing-plus-storage counterparts. It therefore seems likely that the domain-specific demands of these tasks influenced their correlations with verbal ability, independent of their WMC demands. We find a similar problem with the tasks from Morrell and Park (1993), in which the spatial WM task produced 50% lower span scores than did the verbal WM tasks, obviously complicating the interpretation of cross-domain correlations (although Conway & Engle, 1996, demonstrated that processing-task difficulty within a given span task does not impact its correlations with ability).

A final difficulty in interpreting the dissociations between verbal and spatial WM span tasks is that they are not reliably found. The reading span and rotation span tasks developed by Shah and Miyake (1996) to measure verbal and spatial WMC, respectively, demonstrate unstable correlations with each other across studies ($r_s = .23$ and $.22$ in Shah and Miyake, 1996, and Handley et al., 2002, respectively, and $r_s = .42$ and $.04$ in Friedman and Miyake's, 2000, Experiments 1 and 2, respectively). Slightly modified versions of these tasks by Sohn and Doane (2003) correlated at $.42$. These varying correlations make it difficult to ascertain whether the tasks truly measure overlapping or nonoverlapping constructs. A related, but broader, interpretive problem with this literature is that all the studies that have reported strong dissociations between verbal and spatial WMC have used a single task to measure each construct of interest. Because all cognitive tasks reflect multiple processes, we cannot know whether the observed dissociations in these studies reflect the domain specificity of the WMC construct, or instead, the domain specificity of non-WMC related processes that also contributed to scores (i.e., task-specific sources of variance).

In addition to the ambiguities surrounding the domain-specific findings, we find four categories of data that directly suggest the generality of WMC across verbal and visuospatial domains. First, verbal WM span can sometimes predict spatial ability, and spatial WM span can sometimes predict verbal ability, with cross-domain correlations as high as those within domains (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Salthouse, Babcock, & Shaw, 1991; Salthouse & Mitchell, 1989; Salthouse, Mitchell, Skovronek, & Babcock, 1989; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; Swanson, 1996; Swanson & Howell, 2001). Second, cross-domain correlations among WM span tasks are higher than those among STM span tasks, suggesting that STM reflects more domain-specific skills, strategies, and storage abilities than does WMC (Babcock & Salthouse, 1990; Henry, 2001; Park et al., 2002; Swanson & Howell, 2001). Third, individual differences in domain-specific ability can be substantially reduced by accounting for WM span measured in a different domain (Salthouse et al., 1989; Swanson & Sachse-Lee, 2001; Wilson & Swanson, 2001). Fourth, studies using latent-variable procedures find that constructs composed of multiple verbal and spatial WM span tasks are identical or share 65% or more of their variance (Ackerman, Beier, & Boyle, 2002; Kyllonen, 1993; Law, Morrin, & Pellegrino, 1995; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Oberauer, Süß, Wilhelm, & Wittmann, 2003; Park et al., 2002; Salthouse, 1995; Süß et al., 2002; Swanson, 1996; Wilson & Swanson, 2001).

In the present study we take a latent-variable approach to the question of WMC generality and its relation to verbal- and visuospatial-reasoning abilities, and we do so because it allows more definitive conclusions about underlying constructs than do simple correlational approaches. Latent-variable procedures require multiple tasks to measure each construct, and they statistically remove the error variance associated with the individual, imperfect tasks, retaining only the variance shared among all the tasks. This shared variance represents the latent construct of interest. Construct measurement based on multiple indicators is therefore more reliable and valid than that based on a single, multiply-determined task. When using only a single task per construct, one cannot know which processes engaged by the task are responsible for observed correlations (the same problem exists with respect to interpreting experimental dissociations; see Jacoby, 1991). In this regard, then, latent-

variable techniques are analogous to the converging-operations approach to experimental research (see Salthouse, 2001), in which concepts are operationally defined by using several imperfectly correlated conditions to eliminate alternative hypotheses (Bridgman, 1945; Garner, Hake, & Eriksen, 1956). For example, constructs such as the phonological loop are validated experimentally by manipulating several task variables, such as articulatory suppression, irrelevant background speech, and word length (see Baddeley, 1986). In a similar manner, psychometric constructs such as WMC are best measured using multiple tasks that differ in some surface respects but that share theoretically critical requirements or processes. Moreover, latent-variable techniques, unlike correlational, regression, or exploratory factor analyses, permit hypothesis testing by statistically contrasting the fit of a priori theoretical models (factor structures) to the data.

Here, then, we used multiple measures of verbal and visuospatial WM span and statistically tested whether a single WMC factor, representing a domain-general construct, fit the data as well as did two domain-specific WMC factors (and if the latter provided a better fit, we tested how strongly the verbal and spatial factors were correlated). In contrast to most prior studies of verbal versus visuospatial WMC, we also included multiple measures of verbal and visuospatial STM, so that the generality of WMC could be compared with that of STM and the contributions of STM could be partialled out of the correlation between WMC and complex reasoning in the verbal and spatial domains. The inclusion of verbal and spatial STM measures also allowed us to investigate the recent finding that STM and WMC may be less separable in the spatial domain than in the verbal domain; that is, that spatial tests of STM may tap executive-control processes in ways that verbal STM tasks do not (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Oberauer, in press; Oberauer et al., 2000; Shah & Miyake, 1996).

Such findings indicate that span tasks need not have a dual-task requirement to reflect the WMC construct, which certainly complicates the field's efforts to determine what the WMC construct actually represents (see Oberauer, in press).

Also in contrast to prior work, we used verbal and visuospatial span tasks that shared a common procedure. In several verbal–spatial WMC studies (Oberauer et al., 2000; Oberauer et al., 2003; Swanson, 1996; Süß et al., 2002; Wilson & Swanson, 2001), WM tasks varied from (a) the traditional complex-span procedure to (b) Brown–Peterson-like tasks in which groups of memory items are followed by a single rehearsal-prevention task to (c) coordination tasks that require the values of memory items to be updated online per particular transformation instructions to (d) immediate recall tasks presenting a large amount of material to study and remember (such as a paragraph or street map). The benefit of such broad task-selection strategies is that the latent variable derived from these tasks will not include method-specific variance. However, the attendant costs are

1. imprecision in understanding what the latent variable represents; that is, the variance shared by WM span, Brown–Peterson, coordination, supraspan recall, and other memory tasks may not reflect the same construct that has been traditionally defined by WM span tasks;
2. difficulty in ensuring that task or method differences are appropriately balanced between verbal and spatial domains; and
3. allowance that the WM–Gf correlation merely reflects that one broadly defined general factor can predict another; that is, a strong correlation between a Gf factor and a more circumscribed WMC factor is more impressive (i.e., more compellingly reductionistic) than is a similar correlation between Gf and WMC factors that each reflect a similarly diverse array of task types.

Here we chose clarity of interpretation over breadth of measurement and modeled all the verbal and spatial WM tasks after reading span, in which the presentation of memory items is interleaved with a processing task and the lengths of the item lists vary across trials. Performance in these tasks, with either verbal or visuospatial

materials, reflects one's ability to encode, maintain, and retrieve lists of isolated stimuli in the face of a regularly occurring, highly interfering distractor task. Thus, we believed that if we found a close relation between WMC and Gf using these circumscribed tasks, it would not be because we simply substituted one poorly understood general factor for another. Yet an additional benefit of the span procedure is that we could create STM versions of each of our WM tasks that presented the same to-be-remembered stimuli but without the additional processing demand of the secondary task. We hypothesized that if WM tasks reflect primarily domain-general executive processes that are elicited by their dual-task demands, then verbal and spatial WMC should be more strongly correlated than are verbal and spatial STM, despite all tasks involving the identical storage demands.

Here we also tested a relatively broad sample of participants to maximize sensitivity to domain-general contributions to WMC. We did so by recruiting participants from a variety of university populations (i.e., from a highly selective technical institute and two comprehensive state universities) and from two community populations. Although most of our community volunteers had at least some college background, some did not, and a small proportion had not graduated from high school. Thus, with a modestly diverse sample with respect to education, culture, and race, we hoped to better gauge the domain generality of WMC than was possible in much of the previous research.

At one level, it might be unsurprising for us to find that verbal and visuospatial WMC tasks were highly correlated. After all, if the tasks are identical aside from the particular stimuli used, high correlations based on method variance alone might be expected. However, a clear demonstration of domain generality in WMC here would be important for a number of reasons. First, as we reviewed above, there are several published reports of strong verbal–spatial WMC dissociations that are influencing WM theory (from both cognitive and neuroscience perspectives), we believe inappropriately. Recent reports of WMC generality across verbal and spatial domains have appeared in specialized educational, developmental, or psychometric journals, and therefore they are unfamiliar to many experimental psychologists. Second, and somewhat counterintuitively, we predicted that verbal and spatial WM span measures would be more strongly intercorrelated, because of their general executive contributions, than would verbal and spatial STM measures, despite the fact that the STM measures are more methodologically similar to one another than are the WMC measures. Thus, we believed that a finding of substantial domain generality of WMC, but not of STM, would powerfully indicate generality at the underlying construct level rather than the surface, measurement level. Third, WM is now a central construct in cognitive theory, and therefore the WMC literature informs basic cognitive, intelligence, developmental, and neuroscience research. WMC also is increasingly influencing applied domains, particularly regarding interpersonal, clinical, and health-related issues (e.g., Brewin and Beaton, 2002; Feldman Barrett, Tugade, & Engle, in press; Finn, Mazas, Justus, & Steinmetz, 2002; Rosen, Bergeson, Putnam, Harwell, & Sunderland, 2002). Understanding what WM span tasks measure, and what the WMC construct ultimately reflects, is therefore critical to several lines of psychological inquiry. If WMC, as widely measured by verbal span tasks, is a construct relevant only to verbal–symbolic domains, this should radically affect the way theorists interpret its close relations to attention, self-control, intelligence, neuropsychological diagnoses, and executive function.

Method

Participants

Participants were recruited from undergraduate participant pools at Georgia Institute of Technology (GIT), Southern Illinois University at Edwardsville (SIUE), and the University of North Carolina at Greensboro (UNCG), as well as from the general population of the Atlanta, Georgia, and Greensboro, North Carolina, metropolitan areas. Community participants were recruited via advertisements in daily newspapers and free entertainment–news magazines and via flyers posted in public places such as post offices and supermarkets. Students from GIT, SIUE, and UNCG participated either for course credit or for cash; community volunteers from Atlanta and Greensboro participated for cash. Paid participants were compensated for each session plus a bonus for completion of the study.

The student populations across the three university testing sites are academically diverse. UNCG's average SAT score was almost identical to the national average in 2001–2002 (median verbal = 520, median math = 510; national average combined = 1,020) and GIT's scores were well above the national average (median verbal = 642, median math = 689). Only 3% of UNCG students scored above 700 on the verbal task compared with 22% of GIT students, and only 1% of UNCG students scored above 700 on the math task compared with 47% of GIT students. At SIUE, which relies on the ACT test admissions, scores were very similar to national norms: 42% of students scored below the national average of 21, and an additional 26% scored between 21 and 23.¹ Thus, our UNCG and SIUE samples likely reflected the average ability–aptitude range of U.S. college students, whereas GIT (which contributed fewer participants) likely reflected a much higher than average ability–aptitude.

Native English speakers between 18 and 35 years old were invited to participate (one 17-year-old student, with parental consent, also assented to the study). Of the 260 participants who gave written informed consent to participate, 250 completed all three sessions of the study.

Apparatus, Materials, and Procedure

At all testing sites, the memory tasks were administered with Pentium III or higher PCs with 17-in. (43.18-cm) color monitors, and were programmed with E-prime software (Beta 5 version; Schneider, Eschman, & Zuccolotto, 2002). The reasoning tasks were all presented in a paper-and-pencil format. All tasks required written responses. Instructions for all tasks were presented visually (on the computer screen or on paper) while the experimenter read them aloud and answered questions. Instructions were re-presented and reread for any participants having difficulty understanding the task. Visual examples, practice problems, or both, preceded all memory and reasoning tasks.

Tasks

Tasks were administered over the course of three sessions, each lasting 1.5–2 hr and completed 1–4 weeks apart. All participants completed the STM tasks in the first session (along with a consent form and demographic questionnaire), the WM tasks in the second session, and the reasoning tasks in the third session. The STM and WM sessions were administered individually, with participants tested in small sound-attenuated rooms, whereas the reasoning-task session was administered in groups of 1–4 participants seated around a table. Participants took a mandatory rest break approximately halfway through each session.

Separating the testing sessions along construct lines risked underestimating the correlations among constructs because of measurement error. However, our main question pertained to performance generality across verbal and visuospatial processing domains, and therefore it was most important to present verbal and visuospatial indicators of each construct within a single session. Moreover, this arrangement offered the following pragmatic benefits: (a) It familiarized participants with the simpler STM version of each span task in one session before they attempted the more complex WM version in a subsequent session, and (b) it allowed group testing of participants in the final, reasoning session.

Task order was counterbalanced within each session, with participants randomly assigned upon arrival, and therefore orders were not linked across sessions. For STM and WM sessions, counterbalancing followed a modified Latin-square design in which each task served at least once in each position across the six task orders, with the constraint that verbal and spatial tasks alternated within the session. Because of the alternation constraint, each specific task was followed by the same other task in three of the orders and by two different tasks in the two remaining orders in which the task was not completed last. For example, the word span task was followed by the arrow span task in three orders and by the matrix-span and ball span tasks in one order each. For the reasoning session, one of three pseudorandom task orders was followed in which tasks alternated among verbal, spatial, and matrix–figural domains and in which each task appeared near the beginning, middle, and end of one of the orders. All participants within the same reasoning session were tested in the same task order. See Appendix A for a listing of task orders for the three sessions.

STM Tasks

In all tasks, *set size* refers to the number of items to be recalled during each trial. Following three practice trials with a set size of 2, three trials of each set size were tested, with sets presented in a pseudorandom order that was fixed for each task; all participants saw the same items in the same order. Although most tasks used stimuli from limited pools (e.g., the Digits 1–9), memory items did not repeat within a set. Except where noted, all to-be-remembered stimuli appeared in black on a white background, centered on the computer screen, and verbal stimuli appeared in 24-point Times New Roman font. Recall was signaled by the visual presentation of a cue (???) immediately following the last memory item in a set. Participants took as much time as needed to recall each set, but they could not return to prior sets once the next set was initiated. We instructed participants to reproduce the order in which the memory items were presented. Below we present more detailed descriptions of each STM task and list the putative construct that each was designed to measure. See Figure 1 for illustrations of the spatial STM tasks.

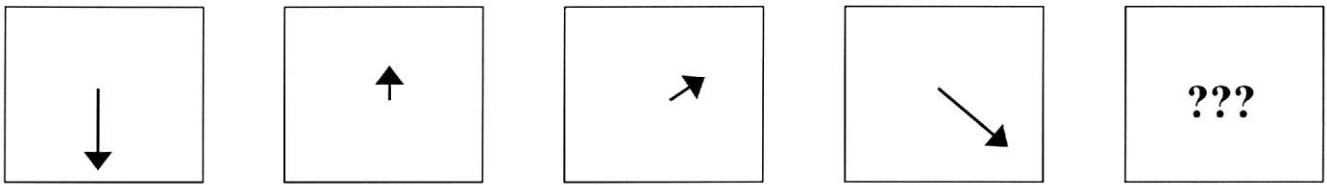
Word span (verbal STM). Participants recalled sequences of one- and two-syllable nouns that were presented in a lowercase font for 1 s each, with a 500-ms blank screen between each word. Participants named each word aloud as it appeared. Set sizes ranged from two to seven words (for 18 sets total). No word appeared more than once in the task. Response sheets presented 18 rows of seven blank spaces, with each row representing 1 set, and participants wrote the words they recalled from each set in the appropriate ordinal position, with one word in each space.

Letter span (verbal STM). Participants recalled sequences of letters that were presented in an uppercase font for 1 s each, with a 500-ms blank screen between each letter. Participants named each letter aloud as it appeared. Set sizes ranged from three to eight letters (for 18 sets total). The nine letters used were *B, F, H, J, L, M, Q, R,* and *X*, chosen to be phonologically distinct. Letters repeated across sets but not within sets, and all were used approximately equally often in the task. Response sheets presented 18 rows of eight blank spaces, with each row representing 1 set, and participants wrote the letters they recalled from each set in the appropriate ordinal position, with one letter in each space.

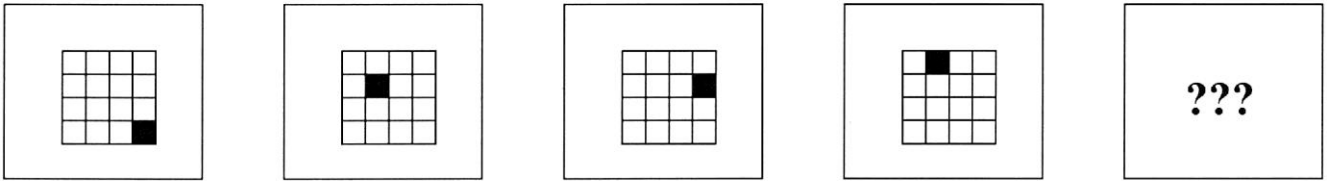
Digit span (verbal STM). Participants recalled sequences of digits (1–9) that were presented for 1 s each, with a 500-ms blank between each digit. Participants named each digit aloud as it appeared. Set sizes ranged from three to nine digits (for 21 sets total). Digits repeated across sets but not within sets, and all were used approximately equally often in the task. Response sheets presented 21 rows of nine blank spaces, with each row representing 1 set, and participants recalled the digits from each set in the appropriate ordinal position, with one digit in each space.

Arrow span (spatial STM). In a task modified from Shah and Miyake's (1996) study, participants recalled a sequence of short and long arrows radiating out from the center of the screen. Short and long arrows were 2 cm and 7.5 cm in length, respectively. Each arrow pointed at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°, creating 16 arrow size X arrow direction combinations. Each combination was used approximately equally often in the task, and no combination was repeated within a set. Each arrow was presented for 1 s followed by a 500-ms blank screen. Set sizes ranged from two to six arrows (for 15 sets total). Response sheets presented 15 rows of six circles, with each row representing 1 set. Each circle in the row consisted of an inner and outer ring and was cut by straight lines (resembling spokes in a wheel) into 45° wedges. Participants drew one arrow in each circle, radiating from the center along one of the straight lines, to correspond to the arrow in that display. The arrow was drawn pointing to the inner ring for short arrows and to the outer ring for long arrows. Participants were to reproduce the sequence of arrows in the correct serial order.

Arrow Span



Matrix Span



Ball Span

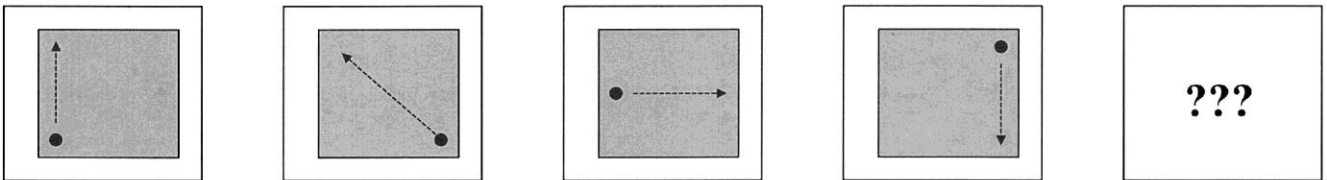


Figure 1. Illustration of the visuospatial short-term memory tasks used in the present study. Each box depicts a single item within a set of four to-be-recalled items. The question marks depict the recall cue that followed each set for every task. The broken arrows depicted in the ball span task represent the direction that the circle moved in, over the course of 1 s, as it traveled to the opposite side of the screen.

Matrix span (spatial STM). Participants recalled sequences of red-square locations within successive matrices. A sequence of 4 X 4 matrices (5 cm X 5 cm) each presented 1 of the 16 squares in red, and each appeared for 650 ms with a 500-ms interstimulus blank screen. Set sizes ranged from two to seven matrices (for 18 sets total). Red-square locations never repeated within a set; each of the 16 squares appeared in red approximately equally often in the task. Response sheets presented 18 rows of seven 4 X 4 matrices, with each row representing 1 set, and participants drew one X in each matrix corresponding to the red square in that display. As in the verbal tasks, participants attempted to reproduce the sequence of red-square locations in the correct order.

Ball span (spatial STM). Participants recalled the paths of balls moving across the screen. A gray box (approximately 20 cm X 20 cm) of 400 X 400 pixels was presented on each display. Immediately after the onset of the box, one blue-green ball (approximately 1.5 cm in diameter) appeared in one of eight locations just inside the perimeter of the box (i.e., in one of the four corners, in the middle of the top or bottom “row,” or in the middle of the leftmost or rightmost “column”). It then traveled vertically, horizontally, or diagonally to the opposite side of the box for 1 s. When the ball arrived at its destination, there was a 500-ms pause before the next ball appeared. Set sizes ranged from two to six balls per set (for 15 sets total), and participants recalled the path of each ball in the set. There were 16 possible paths of movement, and each was used approximately equally often in the task; no path was repeated within a set. Response sheets presented 15 rows of six squares, with each row representing 1 set. Each square in the row corresponded to one of the displays from the set, and it consisted of lines corresponding to all the possible movement paths. The participant drew an arrow along one of the lines in each square to indicate the path of the ball in that display. Again, participants were to reproduce the sequence of ball paths in the correct order.

WM Tasks

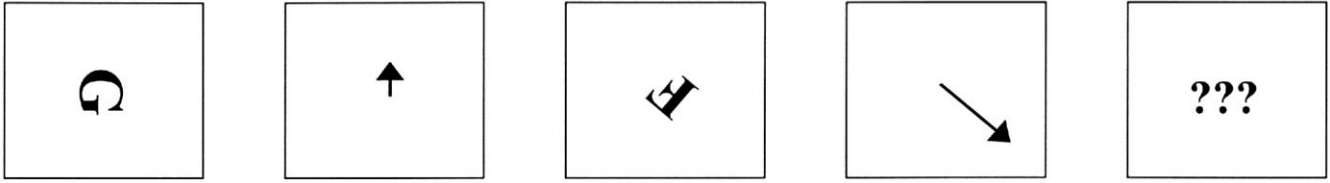
The general procedures for the WM tasks were the same as for the STM tasks, but the WM tasks presented a processing task before each of the to-be-remembered stimuli. This processing task was designed to be particularly interfering to the storage demands of the task. In all tasks, participants were told to begin the processing component as soon as it appeared on the screen (to minimize rehearsal between stimuli). With the exception of the specific words used in the operation span task, all to-be-remembered stimuli were the same as in the corresponding STM task. Below we present more detailed descriptions of each WM task, along with listing the putative construct that each was designed to measure. See Figure 2 for illustrations of each spatial WM task.

Operation span (verbal WM). In a task modified from Engle, Tuholski, et al.'s (1999) study, participants recalled words against a background arithmetic task. Each display included a math problem and a to-be-remembered word (e.g., "Is $(6 \times 2) - 5 = 7$? class") printed in a 20-point font. The arithmetic operation began with a parenthetical multiplication or division problem, followed by a number to add or subtract from the product or dividend. As soon as the equation appeared, the participant began reading the math problem aloud, then said aloud whether the equation was correct (it was correct about half the time), and then read the word aloud. For the previous example, the participant should have said, "Is six times two minus five equal to seven... yes ...class." As soon as the participant read the word, the experimenter pressed a key that blanked the screen for 500 ms, followed by either another math-word combination or the recall cue. When presented with the recall cue, the participant recalled each word from the preceding set, just as in the word span STM task. Set sizes ranged from two to five math-word problems per trial (for 12 trials total).

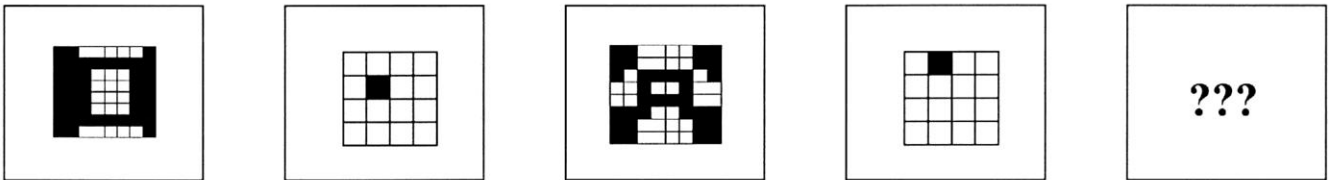
Reading span (verbal WM). In a task modified from Engle, Tuholski, et al.'s (1999) study, participants recalled letters against a background reading task. Each display included an understandable or a nonsensical sentence and a to-be-remembered letter (e.g., "We were fifty lawns out at sea before we lost sight of land. ? X") printed in a 13-point font. We selected sentences from those used by Engle, Tuholski, et al. and created nonsense versions by changing one word (e.g., *lawns* from *miles* in the example above), which came equally often from the beginning, middle, and end of the sentences. Each sentence consisted of 10–15 words ($M = 12.7$ words). As soon as the sentence appeared, the participant read it aloud, then verified aloud whether it "made sense" (it made sense about half the time), and then read the letter. In the previous example, the participant should have said, "We were fifty lawns out at sea before we lost sight of land ... no ... x." As soon as the participant read the letter, the experimenter immediately pressed a key that blanked the screen for 500 ms, followed by either another sentence-letter combination or the recall cue. When presented with the recall cue, the participant recalled each letter from the preceding set, in the order they appeared, just as in the letter span STM task. Set sizes ranged from two to five sentence-letter problems per trial (for 12 trials total).

Counting span (verbal WM). In a task modified from Engle, Tuholski, et al.'s (1999) study, participants recalled digits against a background counting task. Each display included 3–9 dark blue circles; 1, 3, 5, 7, or 9 dark blue squares; and 1–5 light green circles on a gray background (the number of blue circles, blue squares, and green circles was approximately balanced across displays in the task). The participant counted aloud the number of dark blue circles in each display, and when finished with the count, he or she repeated the total number. So, if there were 5 dark blue circles, the participant should have said, "One, two, three, four, five ... five." When the participant repeated the total count, the experimenter pressed a key that blanked the screen for 500 ms, followed by either another display or the recall cue. When presented with the recall cue, participants recalled each total from the preceding set, in the order they appeared, just as in the digit span STM task. Set sizes ranged from two to six displays per trial (for 15 trials total).

Rotation Span



Symmetry Span



Navigation Span

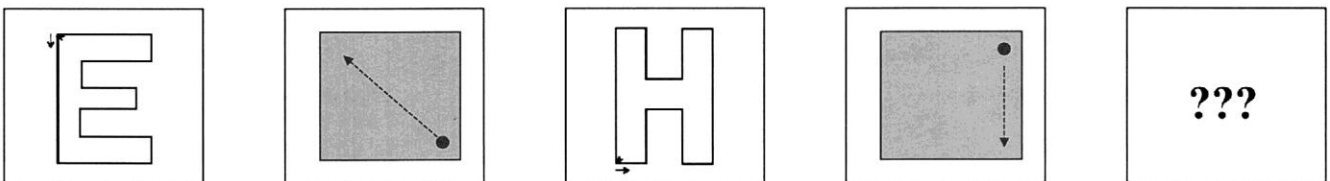


Figure 2. Illustration of the visuospatial working memory tasks used in the present study. Alternating boxes within each task depict single processing-task items and to-be-recalled items within a set of two to-be-recalled items. The question marks depict the recall cue that followed each set for every task. The broken arrows depicted in the navigation span task represent the direction that the circle moved in, over the course of 1 s, as it traveled

Rotation span (spatial WM). In a task modified from Shah and Miyake's (1996) study, participants recalled a sequence of short and long arrows radiating from the center of the screen against a background letter-rotation task. The processing task presented a normal or mirror-reversed *G*, *F*, or *R* (approximately 2 cm tall), rotated at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. The task was to mentally rotate the letter, then indicate aloud whether the letter was normal ("yes") or mirror reversed ("no"); it was normal about half the time. Immediately after this response, the experimenter pressed a key that blanked the screen for 500 ms and presented a short or long arrow (identical to those used in the arrow span STM task) for 1 s. When the arrow disappeared, another letter or the recall cue appeared. When presented with the recall cue, the participant recalled all of the arrows from the preceding displays, in the order they appeared, just as in the arrow span STM task. Set sizes ranged from two to five letter-arrow displays per trial (for 12 trials total).

Symmetry span (spatial WM). Participants recalled sequences of red-square locations within a matrix against a background symmetry judgment task. In the processing task, an 8 X 8 matrix (6 cm X 6 cm) was presented, with some squares filled in black, and participants decided whether the black-square design was symmetrical along its vertical axis; it was symmetrical about half the time. Following the participant's oral response to the symmetry display, the experimenter pressed a key that blanked the screen blank for 500 ms, followed by a 4 X 4 matrix with a filled red square (identical to those used in the matrix span STM task), presented for 650 ms. Immediately following the to-be-remembered matrix, either another symmetry display or the recall cue was presented. When presented with the recall cue, participants recalled the sequence of red-square locations in the preceding displays, in the order they appeared, just as in the matrix span STM task. Set sizes ranged from two to five symmetry-memory matrices per trial (for 12 trials total).

Navigation span (spatial WM). Participants recalled the paths of moving balls across the screen against a background task modified from Brooks's (1968) study. In the processing task, a block letter *E* or *H* was presented against a gray background, with a red asterisk and arrow in one corner of the letter. Each letter was approximately 7.5 cm tall and 5 cm wide. The participant started at the asterisk and mentally navigated in the direction of the arrow along each corner of the letter. At each corner, the participant indicated aloud whether that corner was at the top or bottom edge of the letter for a "yes" response, or not at the top or bottom edge for a "no" response. After navigating around the entire letter, the participant said "finish" to indicate arrival at the starting corner. At the word *finish* the experimenter pressed a key that erased the letter and presented a gray box (approximately 20 cm X 20 cm) of 400 X 400 pixels that presented a ball display identical to those in the ball span STM task. Immediately after the ball finished its journey across the display, another block letter or the recall cue was presented. When presented with the recall cue, participants recalled the sequence of ball paths in the preceding displays, in the order they appeared, just as in the ball span STM task. Set sizes ranged from two to five navigation-ball displays per trial (for 12 trials total).

Reasoning Tasks

Thirteen standardized tests (five verbal-reasoning, five spatial-visualization, and three figural, inductive-reasoning tasks) were timed by an experimenter using a stopwatch. Participants could not proceed to the next task until the full allotted time for each task expired. Therefore, all participants, regardless of testing site or group size, had equal time for each task and the entire session. Tasks were drawn from the following test batteries: Air Force Officer Qualifying Test (AFOQT; Berger, Gupta, Berger, & Skinner, 1990), Differential Aptitude Test (DAT; Bennett, Seashore, & Wesman, 1972), Educational Testing Service Kit of Factor-Referenced Tests (ETS; Ekstrom, French, Harman, & Derman, 1976), RAPM (Raven et al., 1998), Revised Beta Examination (3rd ed.; BETA III; Kellogg & Morton, 1999), and the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999). Below we present brief descriptions of each reasoning task, along with the putative construct that each was used to measure.

ETS Inference Test (verbal reasoning). For each item, participants read a one- to three-sentence passage about a topic and chose the one conclusion of five that could be inferred from the passage without assuming any additional information or knowledge. Following one example problem, 10 test items were presented, and participants had 6 min to complete the test. The 10 items were those constituting Part 1 of the ETS version, and they addressed topics such as fatal car accidents after dark, the endangerment of sea otters, and various cultures' uses of the Stonehenge site.

AFOQT Analogies Test (verbal reasoning). For each item, participants read an incomplete analogy and selected the one word or phrase of five that best completed the analogy. Following one example, 18 test items were presented, and participants had 5 min to complete the test. We selected a subset of items from the original test that presented relatively high frequency words in both the analogy and the response choices, so that performance would reflect primarily induction of relations among words and concepts and not vocabulary knowledge. (Our 18 items represent Items 1–6, 8–11, 13, 14, 16, 19, 20, 22, 24, and 25 from the AFOQT test.)

AFOQT Reading Comprehension (verbal reasoning). For each item, participants read a two- to six-sentence passage about a topic and completed the final sentence of the passage with one of five choices (choices ranged in length from single words to full clauses). Fourteen test items were presented, and participants had 9 min to complete the test. The 14 items were all the odd-numbered items and one even-numbered item from the AFOQT test (our items corresponded to AFOQT Items 51, 53, 54, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75). The selected passages addressed such topics as population control, sociology and group classification, filterable viruses, and drug treatment for mental disorders.

Remote Associates Test (verbal reasoning). For each item, participants generated a word that was somehow semantically related to three remotely associated words provided (e.g., *basket, room, foot*; answer is *ball*). Following 2 practice items, 20 test items were presented and participants had 8 min to complete the test. We selected items to represent a broad range of difficulty according to norms provided by Kihlstrom (n.d.). The

normed items from this Web site were taken from the original Remote Associates Test (Mednick, 1962; Mednick & Mednick, 1967) and from a version of the test by Bowers, Regehr, Balthazard, and Parker (1990). We presented items in order of increasing difficulty (normative probability of error for each item was .15, .20, .25, .30, .40, .40, .45, .50, .50, .55, .60, .60, .65, .70, .70, .70, .75, .80, .80, and .90, respectively).

ETS Nonsense Syllogisms Test (verbal reasoning). For each item, participants read two absurd premises followed by a conclusion and indicated whether the conclusion represented good or poor reasoning if the premises were assumed to be true (e.g., “All bugs are frogs. No frogs have feet. Some bugs have feet.”). Participants completed 4 practice items and 15 test items, with 4 min allotted to the test items. The 15 items were those constituting Part 1 of the ETS version.

DAT Space Relations Test (spatial visualization). Each item presented the illustration of a piece of paper that could be folded along its edges to make a three-dimensional shape. Participants selected the appropriate illustrated shape out of four choices. Following 2 practice items, 18 test items were presented, and participants had 8 min to complete the test. (Our 18 items represented Items 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, and 55 from the original test.) Twelve items presented monochromatic items in which the response choices differed only in shape, 3 items presented dichromatic items in which the response choices differed only in their shading, and 3 items presented dichromatic items in which the response choices differed in both shape and shading.

AFOQT Rotated Blocks Test (spatial visualization). Each item presented the illustration of an irregularly shaped, three-dimensional block at some angle of rotation. Participants selected from among five illustrated choices the one block that, despite a different orientation, represented the same shape as the target block. Following 3 practice items, 10 test items were presented, and participants had 8 min to complete the test. (Our 10 items represent Items 332–334, 336–338, 340–342, and 344 from the AFOQT.)

ETS Surface Development Test (spatial visualization). Each item presented the illustration of a piece of paper that, when folded, would make the illustrated three-dimensional shape presented next to it. From five to eight edges of the unfolded paper were marked with letters, and five of the edges of the shape were marked with numbers. Participants determined which of the papers’ lettered edges corresponded to each of the shape’s numbered edges. The test presented five paper–shape pairs, and each had five numbered edges for which responses were required (yielding 25 responses for the test). Following 1 practice item, participants had 6 min to complete 25 test items. Our items represented the first five of six paper–shape pairs constituting Part 1 of the ETS version.

ETS Form Board Test (spatial visualization). Each item presented a drawn set of five geometric two-dimensional shapes that could be put together in some combination to form a two-dimensional geometrical figure at the top of the page. Participants indicated which two to five of the pieces were necessary to form the target figure by marking them with a plus sign (and marking unnecessary pieces with a minus sign). After completing 2 practice item sets, participants had 4 min to complete 12 test items. Six of the item sets corresponded to one target figure, and the other 6 corresponded to a different target figure. Our items represented those in Part 1 of the ETS version.

ETS Paper Folding Test (spatial visualization). Each item presented illustrations of a square piece of paper being folded one to three times and then having one or two holes punched through it. Participants selected one of five illustrated squares that would represent the positions of the punched holes if the piece of paper were unfolded. Following 1 practice item, 10 test items were presented, and participants had 4 min to complete the test. Our items represented those in Part 1 of the ETS version.

RAPM, Set II (figural inductive reasoning–Gf). Each item presented a pattern of eight black-and-white figures arranged in a 3 X 3 matrix with one figure missing. Figures ranged from simple geometrical shapes to complex textured patterns. Participants selected the one figure (of eight) presented below the matrix that would best

complete the pattern. Following 2 practice items (from the RAPM Set I), participants had 10 min to complete 18 test items that increased in difficulty. Our items represented the odd-numbered items from the RAPM Set II.

WASI Matrix Reasoning (figural inductive reasoning–Gf). Each item presented a pattern of colored figures and most were arranged in a matrix with one figure missing. Nine items presented 2 X 2 matrices, two items presented 3 X 3 matrices, two items presented a missing piece from a continuous wallpaper-like design, and one item presented a missing piece from a linear sequence of five figures. Figures ranged from geometrical shapes to textured patterns. Participants selected the one colored figure (of five) presented below the matrix that would best complete the pattern. Following 2 practice items, participants had 7 min to complete 14 items that increased in difficulty. Our items represented Items 14, 16, 18, 20, 22, 24, 26, 28, 30, 31, 32, 33, 34, and 35 from the WASI.

BETA III Matrix Reasoning (figural inductive reasoning–Gf). Each item presented a pattern of three novel, black-and-white figures arranged in a 2 X 2 matrix with one figure missing. Figures ranged from geometrical shapes to textured patterns. Participants selected the one figure (of five) presented to the right of the matrix that would best complete the pattern. Following 2 practice items, participants had 10 min to complete 20 test items that increased in difficulty. Our items represented the 20 final items (Numbers 6–25) of the BETA III test.

Data Scoring

All STM and WM span tasks were scored in the following manner, which is different from the weighted method we have traditionally used (e.g., Engle, Tuholski, et al., 1999; Turner & Engle, 1989). Within each set, an item was scored as correct only if it was recalled in correct serial position (e.g., for a set *J—M—F* in letter span, *M* would have to be recalled in the second position to be correct). The number of correct items within each set was then converted into a proportion-correct score, and then the mean proportion-correct score was computed over all sets in the task for the participants' task score. Thus, the score for each span task reflected the average proportion correct across the different set sizes. Of the four different scoring methods we considered, including the one we have previously used, this one produced the most normally distributed scores (all methods produced highly correlated scores, with *r*s between .88 and .99 among the scoring methods for each individual task). In contrast with the present method, the weighted method we have used in the past typically produces positively skewed distributions. There, sets that are recalled correctly earn a score equivalent to the size of the set (e.g., a set of two earns 2 points; a set of five earns 5 points), and therefore large sets contribute disproportionately to the total score; that is, correctly recalling one or two of the Size 5 sets adds many more points to the total score than do one or two of the Size 2 sets.

As is conventional, accuracy on the processing component of the WM span tasks did not enter into the score (e.g., Engle, Tuholski, et al., 1999; Turner & Engle, 1989). For completeness, however, in Appendix B we report the descriptive statistics for the processing data; accuracy was near ceiling for all tasks. Moreover, for all WM span tasks the correlation between processing accuracy and recall score was positive (as presented in Appendix C, all *r*s were in the .15–.35 range), and therefore there was no evidence that participants traded off processing accuracy for increased recall.

For all reasoning tasks, we assigned 1 point for every correct item and then divided the total number of points by the number of items for a proportion-correct final score. The exception to this rule was the surface-development task, in which each of the five items was given a proportion-correct score based on the five responses that were required for each item; the total score for the task was the mean proportion-correct score across the five items.

Data Screening

Two steps were taken to screen the data for possible outliers. First, scores for any task that were more than 3.5 standard deviations from the mean were defined as univariate outliers. There were only 11 such values out of the 5,900 in the data set, and each was replaced with a value corresponding to 3.5 standard deviations from the appropriate mean. Second, Cook's *D* statistic (Cook, 1977) was computed for each variable to identify possible

multivariate outliers. This statistic reflects the extent to which a participant's value for one variable influences the relations among all of the other variables. A Cook's D greater than 1 is considered indicative of a possible outlier (Cook, 1977); no data met this criterion.

Statistical Procedure

We report several fit statistics for our confirmatory factor analyses and structural equation analyses, as is typically recommended (e.g., Kline, 1998). The chi-square statistic reflects whether there is a significant difference between the reproduced and observed covariance matrixes. Therefore, the chi-square is sometimes called a *badness-of-fit* statistic, because nonsignificant chi-square values are desirable. However, when moderate-to-large sample sizes are used, even a slight difference between the reproduced and observed covariance matrixes can result in a significant chi-square statistic. Thus, chi-square to degrees of freedom ratios less than 2 are often used informally to indicate acceptable fit, and so we report these values along with the chi-square statistics. In addition, we report the confirmatory fit index (CFI), normed fit index (NFI), and nonnormed fit index (NNFI), which are less sensitive to sample size. Each reflects the proportion of the observed covariance matrix explained by the model, and hence, *goodness of fit*. Finally, we report the standardized root-mean-square residual (SRMR), which reflects the average squared difference between the observed and reproduced covariances. CFI, NFI, and NNFI values of greater than .90 and SRMR values less than .05 are generally considered indicative of an acceptable model fit (Kline, 1998).

Results

All data from 10 participants were dropped from analyses because they indicated on a demographic questionnaire that they did not meet one or more of the participation criteria (see Method). In addition, all data were dropped from 4 participants who had missing values (most commonly, a page missing from one of the reasoning tasks). Thus, the final sample consisted of 236 participants. Demographic data for these participants were as follows:

There were 158 women and 78 men. Mean age was 21.7 years ($SD = 4.6$ years). Of these participants, 4% were Asian/Pacific Islanders, 31% were Black (African/Caribbean descent), 1% were Hispanic, 1% were Native American/Alaskan Native, and 65% were White (European/Middle Eastern descent); 4 participants did not provide racial background information. Percentages do not sum to 100 because participants had the option of endorsing as many categories as were applicable. With regard to education, 97% of the participants had a high school diploma (3 did not respond to this question), 78% were enrolled in college at the time of participation (3 did not respond to this question), and 16% had a college diploma (4 did not respond to this question). One hundred eighty of the participants were students enrolled at GIT, SIUE, or UNCG ($Ns = 62, 48$ and 70 , respectively), and 56 were paid community volunteers (29 from the Atlanta, Georgia, area, and 27 from the Greensboro, North Carolina, area).

Descriptive statistics for the memory and reasoning variables (after treatment of outliers) are displayed in Table 1. Cronbach's coefficient alpha was derived for each measure as an index of internal consistency. For all tasks, these coefficient alphas were computed at the level of individual items. For the span tasks, all item scores reflected proportion correct for the set. For the reasoning tasks, all item scores were binary, correct versus incorrect (the latter including omission and commission errors), except for the ETS Surface Development Test, in which each of the five items was given a proportion-correct score based on the five individual responses required of each. The coefficient alphas for the memory and reasoning tasks were generally high (near or above .80), indicating adequate reliability. The only measures with reliability substantially lower than .70 were the ETS Nonsense Syllogisms Test and the WASI; however, because these two measures showed zero-order correlations of similar magnitudes to our other tasks, they were retained. All of the variables were approximately normally distributed, with skewness values less than 2 and kurtosis values less than 4 (see Kline, 1998). Furthermore, the normalized Mardia coefficient (2.75) for all 25 memory and reasoning variables indicated that there was not a high degree of multivariate nonnormality in the data, where values less than 3 are generally considered acceptable (Romeu & Ozturk, 1993).

Table 1
Descriptive Data and Reliability Estimates (Cronbach's α) for All Memory and Reasoning Measures

| Task | <i>M</i> | <i>SD</i> | Range | Skew | Kurtosis | α |
|-----------|----------|-----------|----------|-------|----------|----------|
| OpeSpan | .65 | .17 | .06-.98 | -.45 | -.02 | .80 |
| ReadSpan | .67 | .16 | .14-1.00 | -.32 | -.10 | .78 |
| CouSpan | .69 | .15 | .23-1.00 | -.26 | -.34 | .77 |
| NavSpan | .48 | .18 | .03-.96 | .11 | -.40 | .83 |
| SymmSpan | .47 | .20 | .00-.98 | .27 | -.54 | .86 |
| RotaSpan | .61 | .18 | .16-.97 | -.29 | -.40 | .82 |
| WordSpan | .78 | .11 | .40-.99 | -.42 | .14 | .81 |
| LettSpan | .78 | .10 | .51-.99 | -.30 | -.26 | .79 |
| DigSpan | .80 | .10 | .47-1.00 | -.39 | .27 | .79 |
| BallSpan | .58 | .12 | .16-.90 | .01 | .33 | .75 |
| ArrwSpan | .65 | .15 | .13-.98 | -.67 | .86 | .84 |
| MatxSpan | .51 | .15 | .05-.89 | -.15 | -.09 | .86 |
| Inference | .61 | .22 | .00-1.00 | -.36 | -.54 | .68 |
| Analogy | .56 | .22 | .06-1.00 | -.13 | -.51 | .81 |
| ReadComp | .52 | .26 | .00-1.00 | .15 | -1.01 | .83 |
| RemoAsso | .46 | .17 | .00-.85 | -.21 | -.13 | .71 |
| Syllogism | .54 | .17 | .13-1.00 | .26 | -.17 | .60 |
| SpaceRel | .57 | .21 | .06-1.00 | -.24 | -.48 | .81 |
| RotaBlock | .37 | .23 | .00-1.00 | .48 | -.34 | .68 |
| SurfDev | .60 | .25 | .00-1.00 | -.10 | -.99 | .83 |
| FormBrd | .39 | .24 | .00-1.00 | .37 | -.71 | .79 |
| PapFold | .60 | .24 | .00-1.00 | -.26 | -.60 | .77 |
| Ravens | .55 | .15 | .06-.89 | -.59 | .75 | .68 |
| WASI | .69 | .14 | .20-1.00 | -.49 | .55 | .55 |
| BETAIII | .86 | .13 | .37-1.00 | -1.61 | 3.53 | .70 |

Note. $N = 236$ for all tasks. OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; WordSpan = word span; LettSpan = letter span; DigSpan = digit span; BallSpan = ball span; ArrwSpan = arrow span; MatxSpan = matrix span; Inference = Educational Testing Service Kit of Factor-Referenced Tests (ETS)—Inference Test; Analogy = Air Force Officer Qualifying Test (AFOQT)—Analogies Test; ReadComp = AFOQT Reading Comprehension; RemoAsso = Remote Associates Test; Syllogism = ETS Nonsense Syllogisms Test; SpaceRel = Differential Aptitude Test—Space Relations Test; RotaBlock = AFOQT Rotated Blocks Test; SurfDev = ETS Surface Development Test; FormBrd = ETS Form Board Test; PapFold = ETS Paper Folding Test; Ravens = Ravens Advanced Progressive Matrices; WASI = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning; BETAIII = Revised Beta Examination (3rd ed.)—Matrix Reasoning.

Correlations among all the variables are displayed in Table 2, where it can be seen that they were uniformly positive (most $r_s > .40$). Nonetheless, there was evidence in the data for two aspects of construct validity: convergent validity and discriminant validity (Campbell & Fiske, 1959); that is, measures of one putative construct tended to correlate more strongly with each other, indicative of convergent validity, than with measures of other constructs, indicative of discriminant validity. Note that correlations of the magnitude found here are extremely well suited for latent-variable analyses, because they are not so high as to indicate complete redundancy in the measures, but they are not so low as to indicate potentially spurious associations among the variables.

Is WMC a Unitary Construct?

The correlations among the WMC tasks that crossed stimulus domains appear to be stronger than those among the STM tasks, suggesting greater domain generality of WMC. This impression was strengthened by an exploratory factor analysis of all our tasks (principal axis, promax rotation), which produced a very clear six-factor solution with a single domain-general WMC factor and two domain-specific STM factors (along with verbal-, spatial-, and figural/inductive-reasoning factors).²

However, to examine the generality hypothesis more directly and rigorously, we performed two series of confirmatory factor analyses to statistically compare the fit of competing models: a domain-general model consisting of a single WMC factor and a domain-specific model consisting of separate but correlated verbal and spatial WMC factors. In the first series of analyses we allowed task-specific error to correlate when it statistically improved the fit of the model (correlated errors—or correlated residuals—reflect variance that is

shared between pairs of tasks but not with the other tasks constituting the latent variable). In the second series we were more conservative and did not include error correlations in the models.

In all subsequent figures, which illustrate our latent-variable models, circles represent the latent variables and rectangles represent the observed, or manifest, variables from which the latent variables are derived. Numbers on double-headed-arrow paths between latent variables represent the correlation between them; these values can be squared as an indication of shared variance between latent variables. Numbers on single-headed-arrow paths between latent variables (path coefficients) can be interpreted in the same way as semipartial correlations; that is, they can be squared to indicate the proportion of variance in the criterion construct that is uniquely accounted for by the predictor construct. Numbers on paths leading only to manifest variables represent the factor loadings for those tasks.

The one-factor WMC model, depicted in Figure 3A, consisted of a single factor with loadings for all of the WMC variables, as well as residual correlations for operation span/reading span, operation span/counting span, symmetry span/rotation span, and navigation span/rotation span. Inclusion of correlated errors was based on three criteria. First, we selected the four pairs of tasks for which the correlated errors would make a priori theoretical sense. Operation span and reading span were the only WMC tasks that involved words (the memoranda in the former and the processing-task stimuli in the latter); these two tasks were also the only ones to present simultaneously a string of stimuli for the processing component to the left of the stimulus for the storage component. Operation span and counting span were the only WM tasks that involved numbers, and so their shared error variance could reflect some aspect of numerical processing that was not involved in the other WMC tasks. Rotation span and symmetry span were the only WM tasks that involved static (as opposed to dynamic) visuospatial memoranda. Finally, navigation span and rotation span were the only WM tasks that required participants to record their written responses as arrows; in rotation span, participants drew the to-be-recalled arrows they had seen, and in navigation span, participants drew arrows to correspond to the movements of the to-be-recalled ball trajectories they had seen. We then tested whether the correlation between each of these pairs of error terms was significant at $p < .05$, and they all were (in the order mentioned above, the correlations were .45, .17, .31, and .22, respectively). Finally, inclusion of each correlated error term had to significantly improve the fit for the one-factor WM model by a chi-square difference test ($p < .05$). All did: For operation span/reading span, X^2 difference (1, $N = 236$) = 41.46; for operation span/counting span, X^2 difference (1, $N = 236$) = 4.68; for symmetry span/rotation span, X^2 difference (1, $N = 236$) = 13.92; for navigation span/rotation span, X^2 difference (1, $N = 236$) = 6.28.

Here, the fit of the one-factor model was very good: $X^2(5, N = 236) = 9.87$, $X^2:df = 1.97$, CFI = .99, NFI = .98, NNFI = .98, SRMR = .02. Also illustrated in Figure 3A, the two-factor model consisted of two correlated factors, with a WMC–verbal factor derived from the three verbal WM tasks and WMC–spatial factor derived from the three spatial WM tasks. Although the two-factor model appeared to fit the data slightly better than did the one-factor model, $X^2(4, N = 236) = 6.75$, $X^2:df = 1.69$, CFI = 1.00, NFI = .99, NNFI = .99, SRMR = .01, the difference in chi-square values for the two models was nonsignificant by a chi-square difference test, which is a rather liberal test of relative model fit, X^2 difference (1, $N = 236$) = 3.12, $p > .05$. Thus, the hypothesis for a domain-general WMC construct could not be rejected. Moreover, and probably more important than the null significance test, even with an imposed separation between WMC–verbal and WMC–spatial constructs, the correlation between them was near 1.0 (.93, with a 95% confidence interval of .91—.95). A domain-general WMC accounts most parsimoniously for these data.

However, in our second set of analyses (shown in Figure 3B) we did not allow error terms to correlate, because the inclusion of these correlated residuals, although justified, affected the pattern of results. Here, the one-factor model did not provide a particularly good fit to the data, $X^2(9, N = 236) = 62.35$, $X^2:df = 6.93$, CFI = .94, NFI = .93, NNFI = .89, SRMR = .05, whereas the two-factor model did, $X^2(8, N = 236) = 18.11$, $X^2:df = 2.26$, CFI = .99, NFI = .98, NNFI = .98, SRMR = .03. Moreover, the two-factor model fit significantly better than did the one-factor model, X^2 difference (1, $N = 236$) = 44.24, $p < .05$, suggesting some domain specificity in WMC. We note, however, that the correlation between the verbal and spatial WMC factors was .84 in this model (with a

95% confidence interval of .80—.87), indicating that verbal and visuospatial WMC shared about 70% of their variance. By this more conservative analysis, then, WM span tasks appear to be determined jointly by domain-general (attentional) processes along with domain-specific (storage, coding, rehearsal) processes but primarily by a domain-general mechanism. Taken together, our two analyses indicate that verbal and visuospatial WMC share 70%-85% of their variance and represent relatively little variance of domain-specific origin.

Table 2
Correlation Matrix for Proportion-Correct Measures From All Memory and Reasoning Tasks

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | | | |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
| 1. Opespan | — | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2. ReadSpan | .73 | — | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3. CouSpan | .59 | .55 | — | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4. NavSpan | .57 | .57 | .57 | — | | | | | | | | | | | | | | | | | | | | | | | | |
| 5. SymmSpan | .55 | .60 | .49 | .65 | — | | | | | | | | | | | | | | | | | | | | | | | |
| 6. RotaSpan | .59 | .59 | .58 | .69 | .71 | — | | | | | | | | | | | | | | | | | | | | | | |
| 7. WordSpan | .62 | .55 | .36 | .33 | .46 | .43 | — | | | | | | | | | | | | | | | | | | | | | |
| 8. LettSpan | .57 | .55 | .42 | .36 | .47 | .46 | .72 | — | | | | | | | | | | | | | | | | | | | | |
| 9. DigSpan | .59 | .59 | .42 | .31 | .44 | .43 | .67 | .72 | — | | | | | | | | | | | | | | | | | | | |
| 10. BallSpan | .45 | .46 | .43 | .50 | .63 | .60 | .39 | .48 | .38 | — | | | | | | | | | | | | | | | | | | |
| 11. ArrwSpan | .51 | .51 | .42 | .57 | .64 | .66 | .46 | .49 | .39 | .65 | — | | | | | | | | | | | | | | | | | |
| 12. MatrxSpan | .40 | .42 | .39 | .47 | .60 | .61 | .38 | .44 | .41 | .41 | .43 | .35 | .45 | .42 | — | | | | | | | | | | | | | |
| 13. Inference | .42 | .38 | .25 | .28 | .42 | .42 | .43 | .46 | .40 | .43 | .45 | .42 | .61 | .69 | .61 | — | | | | | | | | | | | | |
| 14. Analogy | .43 | .42 | .24 | .28 | .39 | .45 | .43 | .42 | .39 | .40 | .47 | .39 | .61 | .69 | .61 | .69 | — | | | | | | | | | | | |
| 15. ReadComp | .46 | .47 | .28 | .33 | .40 | .42 | .44 | .47 | .41 | .41 | .43 | .35 | .65 | .65 | .65 | .65 | .37 | — | | | | | | | | | | |
| 16. RemoAsso | .28 | .28 | .06 | .11 | .22 | .18 | .28 | .25 | .25 | .25 | .19 | .26 | .18 | .29 | .43 | .37 | .44 | .29 | — | | | | | | | | | |
| 17. Syllogism | .29 | .30 | .17 | .22 | .27 | .28 | .30 | .35 | .25 | .33 | .26 | .22 | .42 | .44 | .44 | .44 | .29 | .42 | .29 | — | | | | | | | | |
| 18. SpaceRel | .30 | .36 | .21 | .41 | .47 | .51 | .23 | .30 | .24 | .51 | .55 | .51 | .42 | .55 | .48 | .32 | .42 | .42 | .42 | .68 | — | | | | | | | |
| 19. RotaBlock | .32 | .30 | .19 | .39 | .41 | .43 | .23 | .24 | .17 | .46 | .47 | .45 | .35 | .42 | .37 | .28 | .42 | .34 | .45 | .70 | .58 | — | | | | | | |
| 20. SurfDev | .33 | .38 | .26 | .42 | .46 | .51 | .28 | .39 | .32 | .49 | .55 | .48 | .41 | .53 | .49 | .34 | .45 | .45 | .70 | .62 | .61 | .61 | — | | | | | |
| 21. FormBrd | .31 | .33 | .27 | .45 | .46 | .46 | .19 | .31 | .22 | .50 | .47 | .52 | .41 | .47 | .47 | .25 | .41 | .65 | .62 | .41 | .65 | .62 | .61 | — | | | | |
| 22. PapFold | .26 | .32 | .20 | .42 | .42 | .45 | .25 | .35 | .22 | .47 | .47 | .45 | .38 | .47 | .39 | .26 | .42 | .67 | .55 | .63 | .61 | .61 | .61 | .61 | — | | | |
| 23. Ravens | .32 | .35 | .25 | .34 | .39 | .47 | .25 | .33 | .26 | .53 | .54 | .42 | .38 | .49 | .42 | .35 | .34 | .60 | .48 | .59 | .48 | .57 | .57 | .57 | .57 | — | | |
| 24. WASI | .34 | .38 | .29 | .38 | .39 | .43 | .35 | .40 | .33 | .47 | .51 | .48 | .42 | .51 | .44 | .44 | .31 | .53 | .48 | .49 | .44 | .44 | .44 | .44 | .44 | .44 | — | |
| 25. BETAI | .44 | .44 | .30 | .42 | .44 | .49 | .43 | .43 | .38 | .48 | .48 | .47 | .45 | .53 | .44 | .38 | .39 | .58 | .43 | .48 | .45 | .45 | .45 | .45 | .45 | .45 | .63 | — |

Note. $N = 236$ for all tasks. Opespan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; WordSpan = word span; LettSpan = letter span; DigSpan = digit span; BallSpan = ball span; ArrwSpan = arrow span; MatrxSpan = matrix span; Inference = Educational Testing Service Kit of Factor-Referenced Tests (ETS)—Inference Test; Analogy = Air Force Officer Qualifying Test (AFOQT)—Analogies Test; ReadComp = AFOQT Reading Comprehension; RemoAsso = Remote Associates Test; Syllogism = ETS Nonsense Syllogisms Test; SpaceRel = Differential Aptitude Test—Space Relations Test; RotaBlock = AFOQT Rotated Blocks Test; SurfDev = ETS Surface Development Test; FormBrd = ETS Form Board Test; PapFold = ETS Paper Folding Test; Ravens = Ravens Advanced Progressive Matrices; WASI = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning; BETAI = Revised Beta Examination (3rd ed.)—Matrix Reasoning.

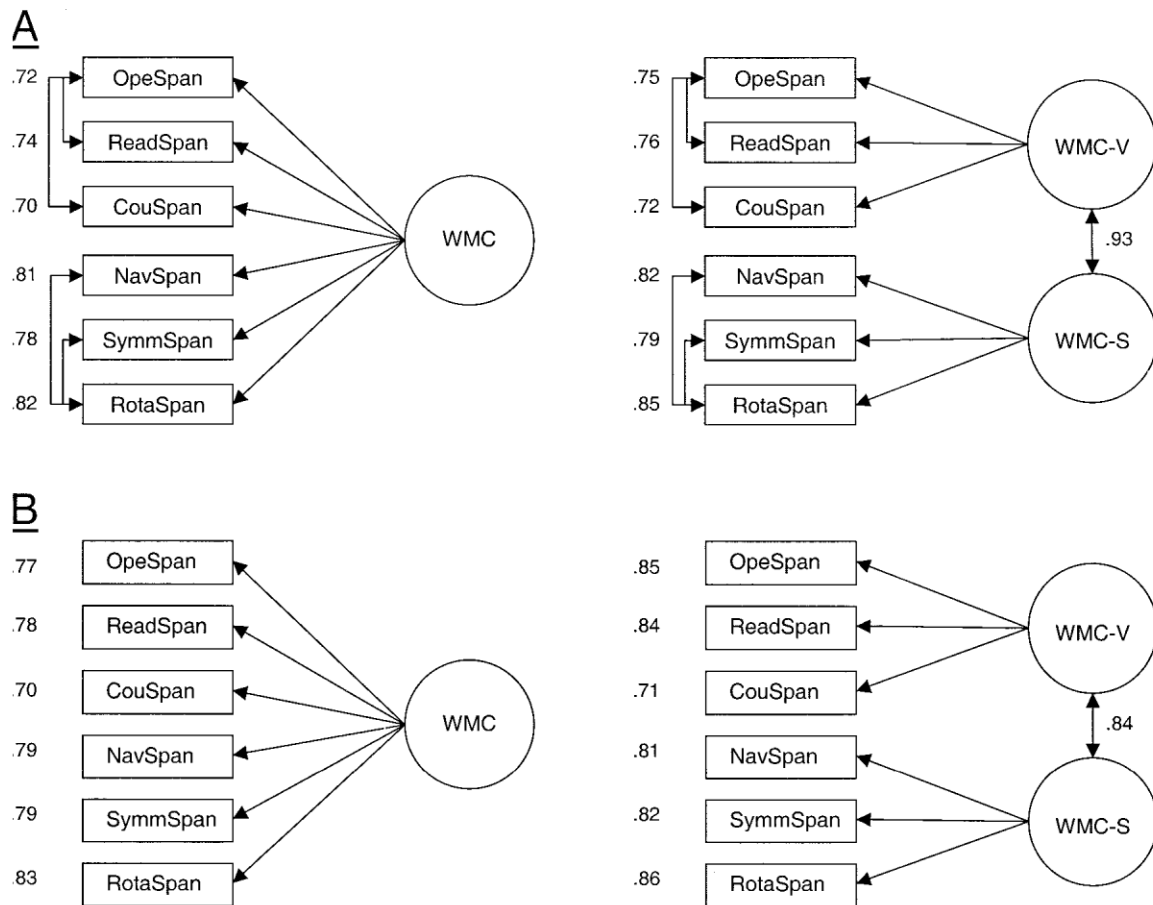


Figure 3. A: Path model for confirmatory factor analysis consisting of a single working memory capacity (WMC) factor versus two domain-specific factors. Paths connecting manifest variables (rectangles) to each other represent correlated error terms added to the model. B: Path model for confirmatory factor analyses contrasting one-factor versus two-factor models but with no correlated errors. In both panels, paths connecting latent variables (circles) to each other represent the correlations between the constructs, and numbers appearing next to each manifest variable represent the loadings for each task onto the latent variable. OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; WMC-V = verbal working memory capacity; WMC-S = spatial working memory capacity.

The Relation Between WMC and STM

We next tested for the relations among the WMC and STM constructs in the verbal and visuospatial domains. Our inclusion of both STM and WMC versions of the memory span tasks allowed us to test for the connection between the STM and WMC constructs across stimulus domains within a single study. We predicted that verbal and spatial WMC factors would be more strongly correlated (i.e., more domain general) than would verbal and spatial STM factors, because of the greater executive contribution to WMC tasks and because this was consistent with our exploratory factor analysis. A second important issue we explored pertains to the distinction between WMC and STM in the verbal versus visuospatial domains. Recall that evidence from verbal tasks suggests that WMC and STM constructs are strongly correlated but separable and that verbal WMC is a far better predictor of complex cognitive ability than is verbal STM capacity (Conway et al., 2002; Daneman & Merikle, 1996; Engle, Tuholski, et al., 1999). In contrast, some recent evidence from visuospatial tasks suggests that WMC and STM constructs are extremely highly correlated, if not inseparable, and that spatial STM and WMC are equally strong predictors of complex cognitive ability (Miyake et al., 2001; Oberauer, in press; Shah & Miyake, 1996).

By confirmatory factor analyses, the best-fitting model for our 12 span tasks is illustrated in Figure 4, and the fit was very good, $X^2(48, N = 236) = 77.49$, $X^2:df = 1.61$, CFI = .98, NFI = .96, NNFI = .98, SRMR = .03 (no correlated errors were included in this, or in subsequent, models). We compared the fit of this four-factor model

to five others: (a) a single-memory factor with no distinction between WMC and STM or between verbal and visuospatial tasks, (b) a two-factor model with a distinct and domain-general WMC and STM, (c) a two-factor model with a verbal STM separate from a WMC factor consisting of all the other verbal and spatial span tasks, (d) a three-factor model with a domain-general WMC and two domain-specific STMs, and (e) a three-factor model with a verbal STM, verbal WMC, and a single spatial-memory factor.

The four-factor model indicates that verbal and spatial WMC constructs were separable, as were the verbal and spatial STM constructs. Note that, as in our prior confirmatory analysis, the verbal and spatial WMC constructs again shared 70% of their variance. Verbal and spatial STM, in contrast, shared only 40% of their variance, suggesting that, as measured by span tasks, STM is more domain specific than is WMC. We statistically confirmed that verbal and spatial WMC were more strongly correlated than were verbal and spatial STM by fixing the path between the WMC factors to be equal to the path between the STM factors, which resulted in a significant decrease in model fit, X^2 difference(1, $N = 236$) = 64.25, $p < .05$. Thus, as defined by span tasks, WMC is a more domain-general construct than is STM.

The finding that WMC is a more general construct than is STM is important because the verbal and spatial STM tasks were actually more similar to each other, methodologically and procedurally, than were the corresponding WMC tasks. All the STM tasks presented verbal or spatial items individually at a fixed rate. In contrast, the verbal and spatial WMC tasks differed in the manner and timing of stimulus presentation. The spatial tasks presented each processing-task stimulus until response, interleaved with each individual memorandum presented for a fixed duration. The verbal tasks, however, presented the processing and memory stimuli together, and all remained exposed until the participant completed the processing task and identified the memory stimulus aloud; in fact, in the counting span task, the processing and memory stimuli were completely integrated. Thus, this aspect of method variance not only worked against our finding a unitary WMC factor across verbal and spatial domains, but it also worked against our finding WMC to be more domain general than is STM. The very strong correlations between verbal and spatial WMC that we observed, then, represent a powerful demonstration of the generality of the WMC construct across stimulus domains.

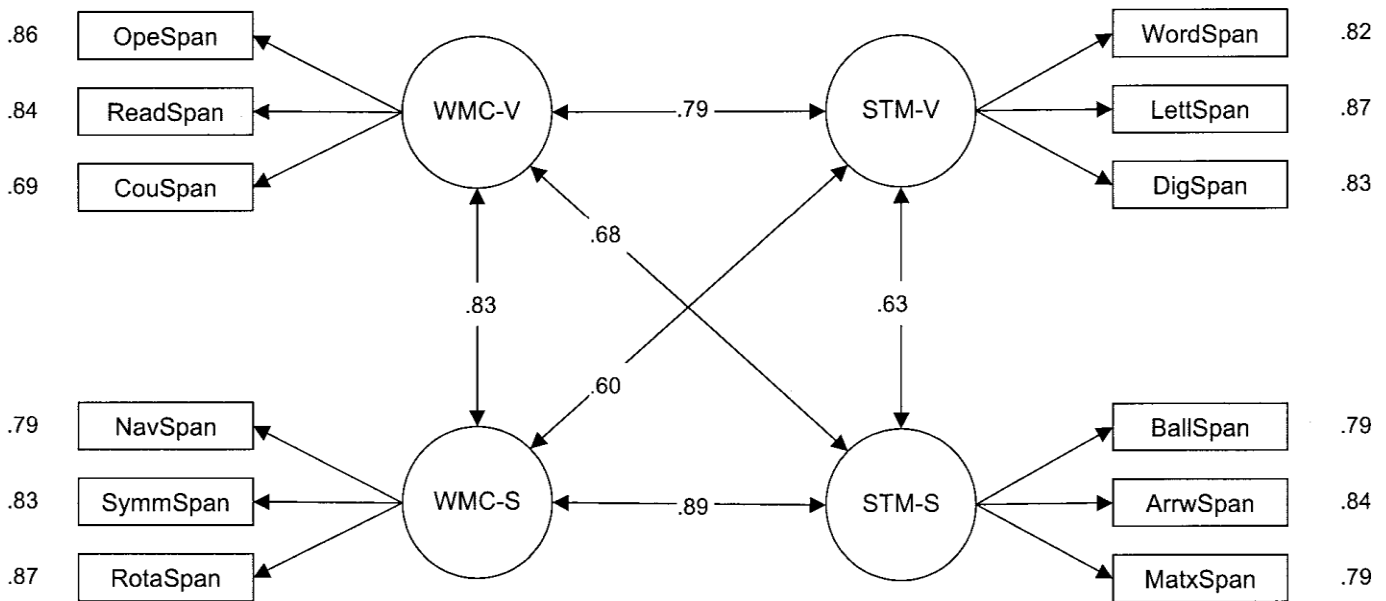


Figure 4. Best-fitting measurement model for all 12 memory span tasks. All paths are statistically significant at $p < .05$. Paths connecting latent variables (circles) to each manifest variable represent the correlations for each task onto the latent variable, and numbers appearing next to each manifest variable represent the loadings for each task onto the latent variable. OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; WMC-V = verbal working memory capacity; STM-V = verbal short-term memory; WordSpan = word span; LettSpan = letter span; DigSpan = digit span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; WMC-S = spatial working memory capacity; STM-S = spatial short-term memory; BallSpan = ball span; ArrwSpan = arrow span; MatxSpan = matrix span.

With respect to the issue of spatial STM, our data do not replicate the equivalence of visuospatial WMC and STM observed by others (Miyake et al., 2001; Oberauer, in press; but see Park et al., 2002). Although these constructs were very strongly correlated in our data, they appeared to be only slightly more strongly correlated than were verbal WMC and STM. Indeed, fixing these two paths to be equivalent did not significantly reduce the fit of the model, X^2 difference(1, $N = 236$) = 1.87, $p > .15$. Here, then, spatial STM was no more WMC-like than was verbal STM. Indeed, one could argue that the spatial STM and WMC tasks were more methodologically similar to one another than were the verbal tasks, and therefore their slightly higher correlations may in part reflect greater shared method variance. More specifically, the spatial STM and WMC tasks presented the identical memoranda (with identical presentation durations), and in the WMC tasks these memoranda were presented in isolation from the processing-task stimuli. In contrast, the verbal WMC tasks presented the memoranda along with the processing-task stimuli, and therefore the presentation rates for STM and WMC stimuli were different from one another. Moreover, the verbal tasks presented somewhat different stimuli from one another: Word span and operation span presented different to-be-remembered words, and whereas digit span presented the Digits 1-9, counting span memoranda ranged from only 3-9. Thus, the lack of a difference here in the relation between WMC and STM in the verbal and spatial domains is especially telling because differences in method variance worked toward our finding such a difference.

The Contribution of Executive Attention to Fluid Reasoning

Our confirmatory factor analyses indicated that WMC is primarily domain general, with verbal and spatial WMC factors sharing 70%–85% of their variance, depending on the model. In contrast, verbal and spatial STM are more dissociable by stimulus domain, sharing only 40% of their variance. Our next set of analyses assessed how strongly WMC/executive attention and STM storage/rehearsal abilities related to both Gf and domain-specific reasoning capabilities in verbal and visuospatial realms. To do this we used a series of structural equation models.

Across all the models of this series, we used the same factor structure for the reasoning tasks. Our goal was to assess simultaneously the relations between immediate memory and both Gf and domain-specific reasoning. We therefore used a bifactor, or nested-factor, model for the reasoning tasks (e.g., Gustafsson & Balke, 1993; Jensen & Weng, 1994), in which all 13 reasoning variables loaded onto a Gf factor. Gf thus reflected variance shared among all the verbal-, spatial-, and matrix-reasoning tasks. At the same time, the five verbal tasks were loaded onto a reasoning—verbal factor (REA-V), and the spatial tasks were loaded onto a reasoning—spatial factor (REA-S). These factors each represented the variance shared among the domain-specific tasks that was not shared with the other tasks in the battery. The three figural-matrix tasks (i.e., RAPM, WASI, and BETA III) were loaded onto Gf only, because these variables were hypothesized to reflect decontextualized (i.e., nonverbal and nonspatial) fluid reasoning. The model provided a very good fit, $X^2(55, N = 236) = 63.33$, $X^2:df = 1.15$, CFI = 1.00, NFI = .96, NNFI = .99, SRMR = .03. Thus, with a measurement model established, we next examined the relations among the memory and reasoning constructs.

Our prediction was that the shared variance among WMC tasks, representing domain-general executive attention, would strongly predict Gf. We tested this prediction in several ways. Unfortunately, we could not use our four-factor model of WMC and STM tasks discussed above because in structural equation modeling, as in regression, one cannot build interpretable models when the predictors are strongly intercorrelated (recall that verbal and spatial WMC, and spatial WMC and STM, were each correlated above .80). In such cases of multicollinearity, once the contribution of one of the predictors has been factored out, there is little variance left in the criterion task to be accounted for by the other predictors; in structural models that suffer from multicollinearity, the standardized coefficients (correlations) among some constructs often end up being larger than 1.0 or smaller than -1.0 and therefore are uninterpretable (as, in fact, they were for our data).

We found two solutions to this multicollinearity problem. In the first, we proceeded as if WMC were a unitary, domain-general construct as indicated by our first set of models. Although this strategy is non-optimal, given our findings that verbal and spatial WMC are not entirely domain general and that each possess some unique

variance, our primary interest here was not to test a model of WMC structure but to test for the relation between what is shared among WMC tasks and what is shared among fluid reasoning tasks. In the second solution, we isolated the shared and unique variance between verbal and spatial WMC and tested which of these had the strongest link to reasoning. Fortunately, both analysis strategies yielded comparable results.

In the first model we used the unitary, domain-general model comprised of the six WM span tasks depicted in Figure 3B. We then created a structural model that added paths from WMC to Gf, REA-V, and REA-S. The model, depicted in Figure 5, provided a good fit: $\chi^2(139, N = 236) = 245.41$, $\chi^2:df = 1.77$, CFI = .96, NFI = .91, NNFI = .95, SRMR = .05. Most important, we found a strong positive effect of WMC on Gf (.64). This correlation is identical to the one found in a model in which we allow correlated residuals among our WMC tasks, as depicted in Figure 3A, and therefore it does not vary with the fit of the unitary WMC factor to the data. Moreover, this estimate is very much in line with prior findings of correlations between WMC and Gf in the .60 range, despite our defining both constructs much more broadly than it had been in this prior work (Conway et al., 2002; Engle, Tuholski, et al., 1999). Research using a variety of span and reasoning tasks, with a variety of participant populations, is clearly converging on the idea that WMC is closely related to Gf.

WMC also had weaker, but significantly positive, effects on both REA-V (.31) and REA-S (.24). Thus, the variance shared by verbal and visuospatial WMC tasks, reflecting domain-general WMC, not only predicted general reasoning capability but also had predictive validity for residual, domain-specific reasoning in both the verbal and spatial spheres.

The WMC construct, defined by the common variance to verbal and visuospatial WM tasks, has a strong link to Gf and weaker (but significant) links to verbal-specific and spatial-specific reasoning. These findings do not tell us, however, what specific aspects of WM are responsible for these correlations. As noted in our introduction, we have proposed that WM span tasks are multiply determined by both domain-general executive-attention processes and domain-specific coding and storage (or STM) processes (Engle, Kane, & Tuholski, 1999; Engle, Tuholski, et al., 1999). Moreover, findings from latent-variable studies using verbal span tasks indicate that the executive component of WM, rather than the storage—STM component, carries the weight in the correlation between WM span and general fluid ability (Conway et al., 2002; Engle, Tuholski, et al., 1999). Our next analysis therefore addressed the relative contribution of executive-attentional, verbal storage—STM, and visuospatial storage—STM processes to the relation between immediate memory and reasoning.

Structural equation modeling was used to answer this question, first by creating a new predictor-side model for the memory tasks, including all the STM and WM tasks. As illustrated in Figure 6, our solution to the multicollinearity problem described previously was a nested-factor structure, similar to that created for the reasoning tasks. We extracted an executive attention factor (Exec Attn) simultaneously with verbal storage (Storage—V) and spatial storage (Storage—S) factors. The logic guiding specification of this model was that no WMC or STM task provides a pure measure of either domain-general executive attention or domain-specific storage and rehearsal. Instead, WMC tasks capture domain-general executive attention primarily but also domain-specific storage, whereas STM tasks capture domain-specific storage primarily but also domain-general executive attention. This logic was clearly supported by the strong correlations among all of our WMC and STM factors in the four-factor model, which indicated substantial shared variance among these constructs.

Our nested-factor model therefore consisted of an Exec Attn factor, with loadings from all 12 memory variables, reflecting the domain-general, “executive” variance shared by all the WM and STM span tasks. The model also consisted of domain-specific factors, with loadings from the six verbal and six spatial tasks on the Storage—V and Storage—S factors, respectively. Thus, from each variable we partialled variance hypothesized to reflect domain-general executive attention and variance hypothesized to reflect storage—rehearsal—coding processes specific to either verbal or spatial stimuli. This model provided a very good fit, $\chi^2(42, N = 236) = 88.84$, $\chi^2:df = 2.12$, CFI = .98, NFI = .95, NNFI = .96, SRMR = .06. Thus, our memory data could be modeled by assuming domain-general processes common to all tasks and domain-specific processes specific to the tasks within each domain.

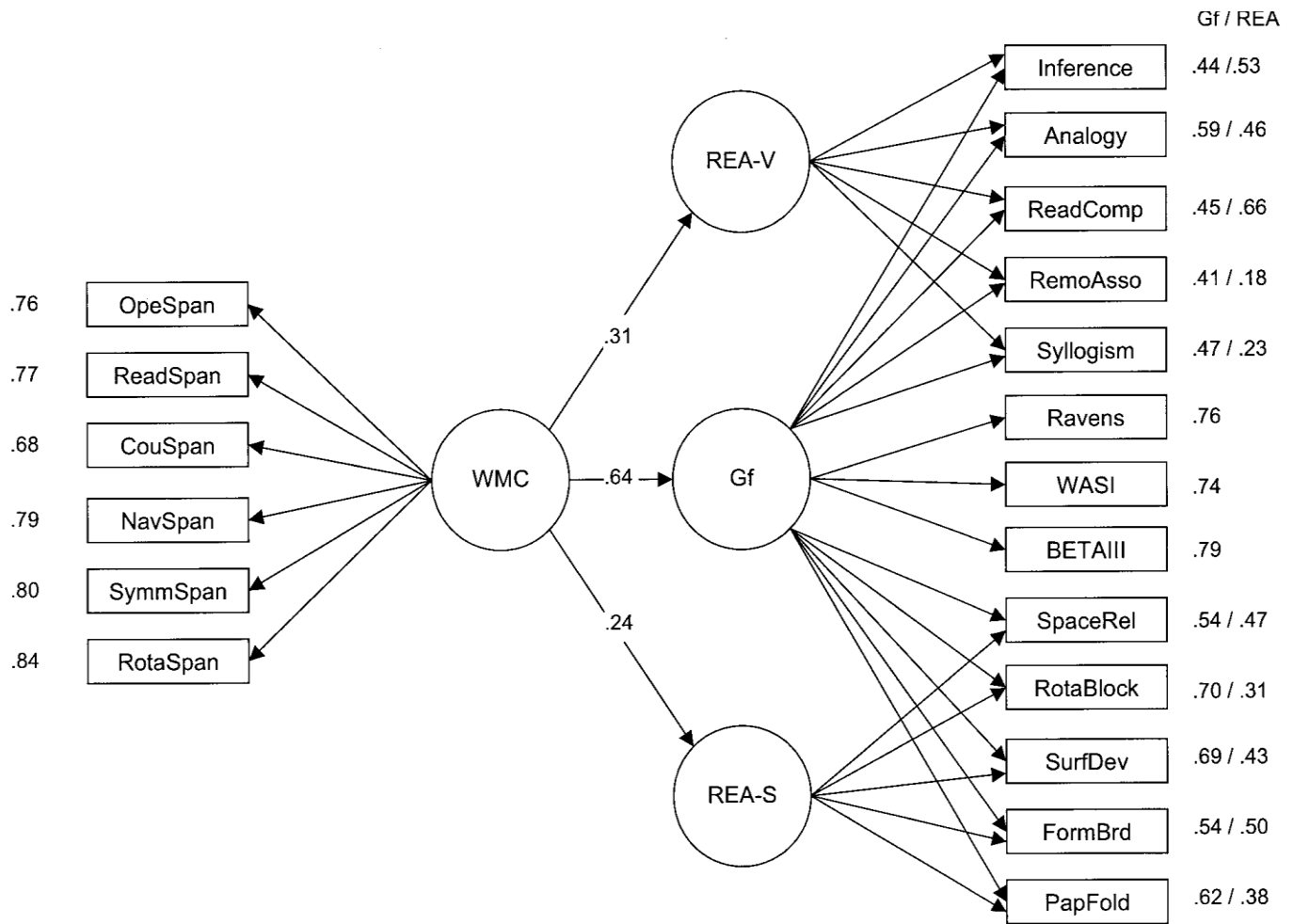


Figure 5. Path model for structural equation analysis of the relation between working memory capacity and reasoning factors. All paths are statistically significant at $p < .05$. The numbers in the general fluid intelligence (Gf) column represent the factor loadings for each reasoning task onto the Gf factor; the numbers in the reasoning (REA) column represent the simultaneous factor loadings for each reasoning task onto either the REA-V or the REA-S factor, with the verbal tasks loading onto REA-V and the spatial tasks loading onto REA-S (matrix reasoning tasks were loaded only onto the Gf factor). Paths connecting latent variables (circles) to each other are analogous to semipartial correlations between these constructs. OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; WMC = working memory capacity; REA-V = reasoning-verbal factor; REA-S = reasoning-spatial factor; Inference = Educational Testing Service Kit of Factor-Referenced Tests (ETS)—Inference Test; Analogy = Air Force Officer Qualifying Test (AFOQT)—Analogies Test; ReadComp = AFOQT Reading Comprehension; RemoAsso = Remote Associates Test; Syllogism = ETS Nonsense Syllogisms Test; Ravens = Ravens Advanced Progressive Matrices; WASI = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning; BETAIII = Revised Beta Examination (3rd ed.)—Matrix Reasoning; SpaceRel = Differential Aptitude Test—Space Relations Test; RotaBlock = AFOQT Rotated Blocks Test; SurfDev = ETS Surface Development Test; FormBrd = ETS Form Board Test; PapFold = ETS Paper Folding Test.

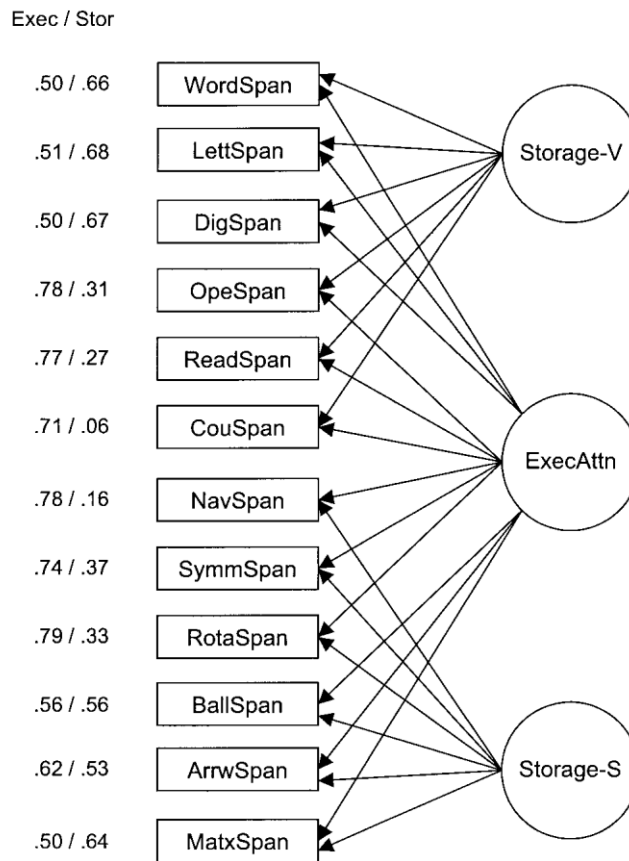


Figure 6. Measurement model for all 12 memory span tasks. The numbers in the executive (Exec) column represent the factor loadings for each span task onto the executive attention (ExecAttn) factor; the numbers in the storage (Stor) column represent the simultaneous factor loadings for each reasoning task onto either the verbal storage–rehearsal (Storage-V) or the spatial storage–rehearsal (Storage-S) factor, with the verbal span tasks loading onto Storage-V and the spatial span tasks loading onto Storage-S. WordSpan = word span; LettSpan = letter span; DigSpan = digit span; OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; BallSpan = ball span; ArrwSpan = arrow span; MatxSpan = matrix span.

An important question is whether we have any direct evidence that the common variance among span tasks reflects executive rather than storage processes. Indeed, the structural models for WM span tasks presented by both Engle, Tuholski, et al. (1999) and Conway et al. (2002) show that the variance common to WMC and STM constructs relates less strongly to Gf than does the residual variance from WMC. This has been interpreted to mean that the common variance between WMC and STM reflects primarily storage, and the residual WMC variance reflects primarily executive-control processes. However, here we have both rational and empirical justification for labeling the common variance as executive and the domain-specific variance as storage.

First, we tested a less restricted sample of participants than is typical, and therefore one might expect executive processes to become more important, for even STM performance, to individuals with lower ability (see Engle, Kane, & Tuholski, 1999; Engle, Tuholski, et al., 1999). Moreover, our storage–STM tasks were not limited to word stimuli, as in the Engle, Tuholski, et al. (1999) and Conway et al. (2002) studies; they also presented digits, letters, or novel spatial memoranda, some of which may have required more executive processes for successful maintenance and recall. Finally, given that at the level of simple storage and rehearsal our verbal and visuospatial STM tasks were substantially dissociable (as strongly predicted by the cognitive and neuroscience literatures), it seems unlikely that the shared variance among verbal and visuospatial span tasks would reflect common (nonexecutive) storage or rehearsal mechanisms.

Of most importance, however, the factor loadings of each task in the model empirically supported our interpretation of the factors. As illustrated in Figure 6, the WM tasks in each domain loaded higher onto the Exec Attn factor than did the corresponding STM tasks: For example, the verbal WM tasks loadings were .78, .77, and .71, whereas the verbal STM task loadings were .51, .50, and .50. Thus, although the Exec Attn factor represents variance shared by WM and STM tasks, it is more heavily weighted to the WM–executive tasks than to the STM–storage tasks. Conversely, for the domain-specific storage factors, the STM tasks within each domain had higher loadings than did the WM tasks: For example, the verbal STM task loadings were .68, .67, and .66 on Storage–V, whereas the verbal WM task loadings were .31, .27, and .06. Again, these domain-specific storage factors represent variance common to STM and WM tasks, but here they are more heavily weighted to the STM–storage processes than to WM–executive processes.³

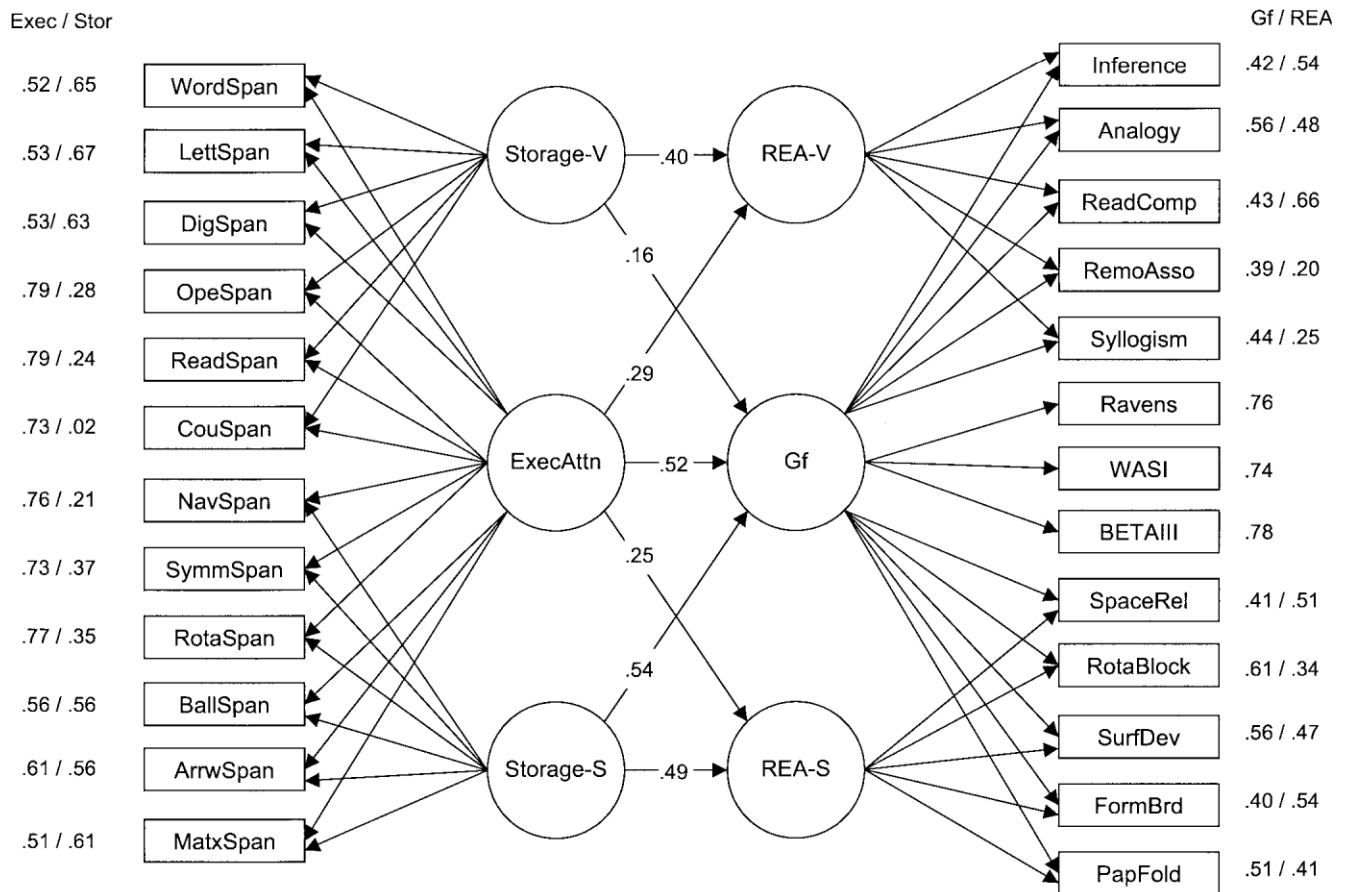


Figure 7. Path model for structural equation analysis of the relation between memory span and reasoning factors. All paths are statistically significant at $p < .05$. The numbers in the executive (Exec) column represent the factor loadings for each memory span task onto the executive attention (ExecAttn) factor; the numbers in the storage (Stor) column represent the simultaneous factor loadings for each memory span task onto either the verbal storage (Storage-V) or the spatial storage (Storage-S) factor, with the verbal span tasks loading onto Storage-V and the spatial span tasks loading onto Storage-S. The numbers in the general fluid intelligence (Gf) column represent the factor loadings for each reasoning task onto the Gf factor; the numbers in the reasoning (REA) column represent the simultaneous factor loadings for each reasoning task onto either the REA-V or the REA-S factor, with the verbal tasks loading onto REA-V and the spatial tasks loading onto REA-S (matrix reasoning tasks were loaded only onto the Gf factor). Paths connecting latent variables (circles) to each other are analogous to semipartial correlations between the constructs. WordSpan = word span; LettSpan = letter span; DigSpan = digit span; OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span; BallSpan = ball span; ArrwSpan = arrow span; MatxSpan = matrix span; REA-V = reasoning–verbal factor; REA-S = reasoning–spatial factor; Inference = Educational Testing Service Kit of Factor-Referenced Tests (ETS)—Inference Test; Analogy = Air Force Officer Qualifying Test (AFOQT)—Analogies Test; ReadComp = AFOQT Reading Comprehension; RemoAsso = Remote Associates Test; Syllogism = ETS Nonsense Syllogisms Test; Ravens = Ravens Advanced Progressive Matrices; WASI = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning; BETAI = Revised Beta Examination (3rd ed.)—Matrix Reasoning; SpaceRel = Differential Aptitude Test—Space Relations Test; RotaBlock = AFOQT Rotated Blocks Test; SurfDev = ETS Surface Development Test; FormBrd = ETS Form Board Test; PapFold = ETS Paper Folding Test.

We tested for the relations among our executive, storage, and reasoning constructs by specifying paths from Exec Attn to Gf, REA-V, and REA-S. We also included paths from Storage—V to Gf and REA-V and from Storage—S to Gf and REA-S. The model, depicted in Figure 7, provided a very good fit, $X^2(246, N = 236) = 338.45$, $X^2:df = 1.38$, CFI = .98, NFI = .92, NNFI = .97, SRMR = .05. Consistent with our previous model that linked a unitary WMC factor to reasoning, Exec Attn had a strong effect on Gf (.52) and weaker effects on both REA-V and REA-S, again near .30 (.29 and .25, respectively). There was obviously close agreement in the relations to reasoning between the Exec Attn factor here and the unitary WMC factor in our prior model. It is therefore reasonable to infer that in these data, the variance common to STM and WMC tasks reflects executive-attention processes rather than simple storage (as suggested also by the tasks' factor loadings). Moreover, this executive-attention construct is closely related to general fluid ability and is additionally predictive, albeit to a lesser degree, of verbal- and spatial-specific reasoning.

The Contributions of Domain-Specific STM to Fluid Reasoning

Storage—V had a strong positive effect on REA-V (.40) but a weaker one on Gf (.16). This is consistent with prior research showing that verbal storage—STM processes account for unique variance in verbal ability beyond that accounted for by executive—WMC processes (e.g., Cantor, Engle, & Hamilton, 1991; Cowan et al., 2003; Engle, Tuholski, et al., 1999) but not for substantial variance associated with general, fluid aspects of ability (Conway et al., 2002; Engle, Tuholski, et al., 1999). Storage—S showed a different pattern of relations to reasoning, with strong positive effects on both REA-S (.49) and Gf (.54). Thus, spatial-storage processes accounted not only for aspects of spatial reasoning that were independent of Gf and executive attention but also for a sizeable proportion (30%) of Gf variance that was not shared with executive attention. This finding is consistent with others (Groeger, Field, & Hammond, 1999; Miyake et al., 2001; Oberauer, in press; Shah & Miyake, 1996) in which spatial STM—storage tasks are strong predictors of general cognitive ability and perhaps as strong as are spatial and verbal WMC tasks.

Our replication of this intriguing finding is inconsistent with our failure above to find that spatial WMC and STM factors were more closely related to each other than were verbal WMC and STM factors. That is, spatial storage appears to be acting like a general, executive factor in predicting Gf, but spatial STM measures (reflecting primarily storage) did not have an especially strong relation to spatial WMC, relative to the relations among the verbal memory constructs. We therefore wondered whether our strong Gf/spatial-storage findings might be, in part, an artifact of the particular Gf factor we extracted in this study. That is, a strong link between spatial storage and Gf factors may have arisen here because our Gf factor was biased toward spatial ability. In fact, our exploratory factor analysis indicated that one of our verbal tasks, nonsense syllogisms, had a stronger spatial than verbal loading (see also Ford, 1995), and it also indicated that the RAPM, a decontextualized measure, had a nontrivial spatial loading (see also Burke, 1958). Moreover, some items from the other matrix-reasoning tasks, the BETA III and WASI, also subjectively seem to be visuospatial in character. Finally, the spatial-reasoning tasks generally had higher loadings on the Gf factor—dominated by the matrix tasks—than did the verbal tasks (in the model from Figure 5, all five spatial tasks and only one verbal task have a Gf loading greater than .50; in the model from Figure 7, three spatial tasks and only one verbal task have a loading greater than .50).

To test the idea that spatial storage may not behave as domain generally with a less spatially biased Gf factor, we used the nested-factor memory span model to predict three Gf factors across different models, progressing from more to less spatially determined factors. In the first structural model, we defined Gf as a latent variable composed of the three matrix-reasoning tasks, and the effects on Exec Attn, Storage—S, and Storage—V were .55, .54, and .17, respectively. Spatial storage and executive attention were clearly strong, unique predictors of Gf here, whereas verbal storage was not. Next, we defined Gf in a more balanced way, as a latent variable composed of three verbal tasks (Analogies, Reading Comprehension, and Inferences) and three spatial tasks (Space Relations, Rotated Blocks, and Paper Folding). Here, the effects on Gf were .57, .47, and .24, respectively. Spatial storage again accounted for unique Gf variance but not quite as much as in the prior model; of importance, the Gf correlation with executive attention did not change. In our final model we defined the Gf factor in a verbally biased way, composed of the three verbal tasks used above in addition to the Remote

Associates Test, a verbal task that nonetheless loaded most strongly with the matrix tasks in our exploratory factor analysis (indicative of its inductive-reasoning requirement). The correlations with the memory factors were .51, .29, and .36, respectively. Although it is noteworthy that spatial storage still predicted unique Gf–verbal variance here, it is also important that its shared variance with Gf was reduced relative to the prior models (and numerically weaker than that between verbal storage and Gf). Note, again, that the shared variance between executive attention and Gf remained largely unchanged, regardless of how Gf was defined across the models. Thus, although spatial storage–rehearsal processes do appear to be more domain general in their predictive power than are corresponding verbal processes, they are not as general or as consistently powerful as the executive-attention processes that are captured by memory span tasks.

General Discussion

The present study addressed four primary questions. First, is WMC a general construct, despite the fact that verbal and visuospatial STM are typically dissociable? Second, how closely are WMC and STM related to each other in the verbal versus spatial domains? Third, how strongly is WMC related to both domain-general and domain-specific aspects of fluid reasoning? Fourth, do the executive-attention and domain-specific storage–rehearsal processes involved in both WMC and STM span tasks differentially predict domain-general and domain-specific aspects of fluid reasoning?

To answer these questions, we tested 236 young adults from diverse university and community populations in verbal and visuospatial span tests of WMC, STM, and fluid reasoning. We then used confirmatory factor analysis and structural equation modeling to test hypotheses about the nature of the WMC and STM constructs and their relations to Gf, verbal reasoning, and spatial visualization. In short, we found, first, that the underlying constructs measured by verbal and visuospatial span tasks were very highly correlated, sharing 70%–85% of their variance, indicating that these WM span tasks are primarily measures of a domain-general capability and only secondarily measures of domain-specific abilities, skills, and so forth. In contrast, verbal and spatial measures of STM had less shared variance than unique variance (with 40% shared), despite the fact that the verbal and spatial STM tasks were structurally more similar to one another than were the verbal and spatial WMC tasks. Second, WMC was similarly closely related to STM in both the verbal and visuospatial domains. As in prior work with verbal tasks (Engle, Tuholski, et al., 1999), WMC and STM were separable, regardless of stimulus domain, but they did share substantial variance with one another. Third and fourth, WMC was strongly correlated with domain-general fluid reasoning, as in prior work (Engle, Tuholski, et al., 1999), and this correlation was driven primarily by the domain-general processes that are captured by span tasks.

The Domain Generality of WMC

The answer to the question of whether WMC is domain general is a strong, but qualified, yes. The answer is a strong yes because our best-fitting model from confirmatory factor analysis indicated WMC to be unitary across verbal and visuospatial span tasks. The yes is qualified, however, by the fact that our confirmatory models that did not include correlated error terms indicated that separable—but highly correlated—verbal and visuospatial WMCs best accounted for the data. One can make reasonable arguments for or against including correlated errors in structural models, therefore we would most confidently and conservatively say that verbal and spatial WMC share at least 70% of their variance, and therefore WM span measures capture primarily domain-general variance in cognitive ability. Indeed, other recent latent-variable studies have similarly pointed to either a unitary WMC across verbal and spatial domains or distinguishable verbal and spatial WMCs that share 70%–90% of their variance (Ackerman et al., 2002; Kyllonen, 1993; Law et al., 1995; Oberauer et al., 2000, 2003; Park et al., 2002; Salthouse, 1995; Swanson, 1996). STM span, in contrast, appears to tap more domain-specific storage and rehearsal capabilities (although not entirely specific, as verbal and spatial STM shared 40% of their variance), which fits with the experimental and neuropsychological data that have been used to support fractionated WM models such as Baddeley’s (1986, 2000).

Our conservative conclusion is consistent with a view of WMC in which domain-specific storage and rehearsal processes contribute to WM span performance, but the dominant, domain-general, and attentional–executive aspect of WMC drives the correlations between WM span and general cognitive-ability measures (see Engle &

Kane, 2004; Engle, Kane, & Tuholski, 1999). In addition, our behavioral findings are highly consistent with conclusions recently drawn from the cognitive neuroscience literature. In short, research using single-cell, neuropsychological and neuroimaging techniques suggests strongly that similar dorsolateral prefrontal cortex (dPFC) regions (and single cells within these dPFC regions) are involved in active maintenance of verbal and nonverbal material, despite early studies suggesting domain specificity in dPFC circuitry; what differs by domain are the particular posterior brain regions that are involved in memory maintenance in concert with dPFC regions (for reviews, see Kane & Engle, 2002; Postle & D'Esposito, 2000).

As discussed at length in our introduction, however, the idea of a primarily domain-general WMC conflicts with some empirical demonstrations of strong dissociations among verbal and spatial WMC measures (Daneman & Tardif, 1987; Friedman & Miyake, 2000; Handley et al., 2002; Morrell & Park, 1993; Shah & Miyake, 1996). We believe that those dissociations are best explained by their reliance on cognitively restricted, high-ability participant samples, which limited the potential contribution of domain-general mechanisms to span variability. As well, their reliance on a single measure of verbal and visuospatial WMC left ambiguous whether the dissociations were driven by the intended WMC constructs or by measurement error, that is, by other non-WMC contributors to performance. Our view regarding measurement is that future studies of verbal and spatial WMC should use multiple tasks for each construct, as we have here, to minimize such ambiguities.

Here we may test our claims about participant sampling of cognitive ability, post hoc, by contrasting the domain generality of WMC and STM across different subsamples of our participants. Using factor scores derived from the three figural–matrix reasoning tests in our battery (RAPM, WASI, and BETA III), we created a Gf measure for each participant, and labeled the top 25% of Gf scorers as *high ability* ($n = 58$). For controls, we then randomly generated two additional, mutually exclusive groups, of 58 participants each, from the entire pool of 236 participants. Finally, on the basis of prior findings that restricted ranges of high ability are more likely to yield patterns of domain specificity than are equally restricted ranges of low ability (for review, see Jensen, 1998), we also examined the data from the bottom 25% of our Gf scorers, or those of *low ability* ($n = 59$). Our assessment of domain generality proceeded by simply taking the mean score for each participant across the three tests of each memory construct (verbal WM, spatial WM, verbal STM, spatial STM) and correlating these mean construct scores.

The high-ability group showed markedly greater domain specificity in immediate memory than did the other groups. For high-ability participants, verbal and spatial WMC shared 32.5% of their variance, compared with 56.3%, 47.6%, and 54.8% for the two random groups and the low Gf group, respectively. Verbal WMC and verbal STM shared 32.5% of their variance for high-ability participants compared with 42.3%, 49.0%, and 44.9%, for the other groups, respectively; spatial WMC and spatial STM shared 41% compared with 57.8%, 64.0%, and 53.3%, respectively; and finally, verbal and spatial STM shared only 8.4%, compared with 30.3%, 34.8%, and 18.5%, respectively. Despite the fact that our high-ability group represented only the 75th percentile of ability and not a more extreme group, these informal, post hoc analyses nonetheless suggest strongly that research studies examining high-ability participants will underestimate the generality of both WMC and STM. However, whether researchers choose to test a relatively broad or relatively restricted cognitive sample should depend on the specific research question. If one is not interested in the domain-general contributions to WMC, for example, in trying to determine what particular domain-specific processes are involved in WMC and how they work, it may be preferable to test participants from a narrow band of general cognitive ability, in which domain-specific contributions to performance are most easily detected. In contrast, to make broad claims about the relative contributions of domain-general versus domain-specific processes to WMC, we strongly recommend testing participants from a broad range of general ability so that general contributions are not underestimated.

The Relation Between WMC and STM

Consistent with prior work in the verbal domain (Conway et al., 2002; Engle, Tuholski, et al., 1999), the WMC and STM constructs were highly correlated but separable for both our verbal- and visuospatial-task data. These findings are consistent with the view

that simple and complex span tasks require some shared executive and storage–rehearsal processes but that WM span tasks require additional executive processes to deal with their dual-task demands, whereas STM tasks may elicit additional rehearsal processes that are disrupted by the secondary task in WM span. Our findings are somewhat incompatible with recent work in spatial memory, however, in which spatial STM has been found to be isomorphic with spatial WM; in these studies, the addition of a secondary processing task did not seem to affect what the spatial span task was fundamentally measuring (Miyake et al., 2001; Oberauer, in press; Oberauer et al., 2000). Although, in our data, the correlation between STM and WMC was numerically stronger in the spatial than in the verbal domain (.89 vs. .79, respectively), this difference was not significant. How should we reconcile these discrepant findings?

Park et al. (2002) also failed to differentiate the WMC–STM relation across verbal and spatial domains, and therefore ours is not the first failure to replicate. In the Park et al. data, the correlation between spatial WMC and STM latent variables was, again, not significantly larger than that between verbal WMC and STM (.84 and .80, respectively). A most straightforward explanation would be that both the Park et al. study and the present study had insufficient power to detect a difference between these correlations. However, both studies tested many more participants than did the “positive” studies, and both also had the power to differentiate verbal from spatial WMC and verbal from spatial STM (ours also had the power to distinguish the correlation between the WMC factors from that between the STM factors). Indeed, it is probably more accurate to argue that the improved power of our studies is what allowed us to distinguish spatial WMC from STM where others had failed to do so. We see no obvious flaws in the studies finding a special link between spatial STM and WMC, nor important differences in the kinds of tasks that were used to measure immediate memory in the “positive” and “negative” studies, and so it seems that further work will be required to determine whether some particular conditions elicit closer similarity between spatial WMC and STM than between verbal WMC and STM, or whether such findings are just not easily replicable.

WMC as a Mechanism of Gf

To what extent is the domain-general aspect of WMC important to domain-general fluid reasoning ability and to domain-specific aspects of verbal reasoning and spatial visualization? We answered this question via two structural equation models, in which WMC was defined either as a latent variable composed of the common variance among 6 verbal and spatial WM span tasks or as a latent variable composed of the common variance among 12 verbal and spatial WM and STM span tasks. In both models, we derived a Gf factor from all 13 of our reasoning tasks, a verbal-reasoning factor representing the variance common to verbal-reasoning tasks that was not shared with the nonverbal tasks, and a spatial-reasoning factor representing the variance common to our spatial visualization tasks that was not shared with the nonspatial tasks.

Our estimates for the correlations between WMC/executive attention and Gf were .64 and .52 in the two models, respectively, suggesting a substantial—but not perfect—relation between WMC and Gf. Stated differently, WMC accounted for approximately 30%–40% of unique variance in fluid intelligence. These correlations are impressively close to those derived from other studies using verbal WM span tasks as the sole measures of WMC (.60 and .59, respectively, in Conway et al., 2002, and Engle, Tuholski, et al., 1999), despite the fact that here our span measures were more diverse and our derivation of Gf was based on a larger and broader sample of fluid reasoning tasks. Although it may be premature to declare $.60 \pm .05$ a “magical number” in the quest to link WM processes to the psychological foundation for fluid intelligence, we are struck by the limited variation around this estimate. Indeed, even researchers who have used task types other than span to measure WMC have found relations of a very similar magnitude (.70 and .65, respectively, in Ackerman et al., 2002, and Süß et al., 2002). An obvious exception to this trend is Kyllonen (1993), who reported path coefficients of .99 between WM and general intelligence (*g*) factors in several analyses. These estimates are suspect, however, because of not only their status as outliers but also the way that *g* was measured. Kyllonen used only two tasks to define the *g* latent variable, a mathematics–algebra test and a test involving mathematical word problems, and both had substantial surface similarity to the quantitative tasks that Kyllonen used as part of the WM latent variable.

We therefore argue that although WMC is strongly related to Gf, and may be among the critical sources of general fluid ability (i.e., the positive manifold among reasoning tests), it is probably unwise to claim WMC to be the cognitive mechanism of Gf. If it were, and WMC and Gf actually reflected a single construct, then one would expect correlations between WMC and Gf factors to be closer to the .85–.95 range rather than in the .55–.65 range. Moreover, other investigators have found that WMC and Gf are differentially related to other constructs, which should not occur if WMC equals Gf. For example, whereas Kyllonen and Christal (1990) found WMC and reasoning ability to be virtually inseparable, WMC was more closely related to speed measures than was reasoning, and reasoning was more closely related to domain knowledge than was WMC. On the whole, then, we argue that WMC is an important contributor to Gf, but the two constructs are probably not isomorphic (see also Ackerman et al., 2002; Conway, Kane, & Engle, 2003; Stankov, 2002).

Regardless of the absolute strength of the relation between WMC and Gf, however, two important questions remain. First, one may reasonably ask why we have assumed the causal pathway to run from WMC to Gf rather than the reverse. That is, although we have claimed that WMC is an underlying mechanism of Gf, others have argued that, instead, higher intelligence leads to better WM task performance (Plomin & Spinath, 2002). Second, given that cognitive psychology and neuroscience have yet to fully understand the WM system and individual differences therein (or, in the extreme, that the WM construct remains vaguely and inconsistently defined) and that WM span tasks are similar to some tests already within psychometric IQ batteries, one could argue that the reduction of Gf to WMC is illusory reductionism, and that we have, instead, only renamed Gf rather than explained it (see Deary, 2000, 2001, 2002; Plomin & Spinath, 2002).

In fact, there is nothing in the present data to demand that WMC variation causes Gf variation and not vice versa. Our data were obviously correlational, and therefore they cannot indicate direction of causation (indeed, the statistical fits of our models would not change by reversing the WMC–Gf paths). Nonetheless, we believe that there are good reasons to align the causal arrow as we have and, moreover, that the (partial) alignment of Gf with WMC represents a valid theoretical advance.

First, the WMC construct and the tasks that measure it derive from a rich theoretical and empirical grounding in basic cognitive, developmental, comparative, clinical, and neuroscience research. Although different theoretical models of WM are currently driving research programs (Andrade, 2001; Miyake & Shah, 1999a; Richardson et al., 1996), these theories are actually in broad agreement in many details about the architecture, functions, and biological substrates of WM (see Kintsch, Healy, Hegarty, Pennington, & Salthouse, 1999; Miyake & Shah, 1999b; but see Nairne, 2002). Thus, the WM construct is much more tractable at the present time than is psychometric Gf.

Second, our attentional view of WMC is supported by findings that high-level skills and strategies do not impact the correlations between WMC and complex cognition (Conway & Engle, 1996; Engle et al., 1992; Turley-Ames & Whitfield, 2003) and that individual differences in WMC predict differences in performance of very low-level attention tasks such as dichotic listening, Stroop, and antisaccade, in which higher order intellectual processes are unlikely to intervene (for reviews, see Conway & Kane, 2001; Engle & Kane, 2004). In the antisaccade task, for example, participants are simply required to move their eyes away from an attention-attracting visual stimulus. Just how high-level intelligence differences would lead to performance differences in such a simple task is unclear to us. On the whole, then, we view the claim that Gf relies in part on domain-general WMC processes to reflect a valid and important theoretical step forward in the study of cognitive abilities.

Moreover, our data suggest that the predictive power of WMC is not limited to domain-general reasoning ability. We unexpectedly found that our WMC/executive-attention factors also predicted reasoning capabilities specific to the verbal and visuospatial domains (i.e., the verbal- and visuospatial-reasoning processes that were not shared with each other). In our two models, the single WMC factor and the executive-attention factor produced path coefficients with verbal reasoning of .31 and .29, respectively, and with visuospatial reasoning of .24 and .25, respectively. Thus, WMC factors derived from the variance shared by verbal and visuospatial

tasks—thereby representing domain-general ability—modestly predicted the verbal-specific and visuospatial-specific reasoning abilities that were independent of Gf (i.e., WMC accounted for approximately 10% of the variance in domain-specific reasoning). We speculate that the functions represented by these correlations are related to the role of WMC in domain-specific learning. WMC is important to the accumulation of knowledge within a particular domain, such as vocabulary or computer programming (Daneman & Green, 1986; Shute, 1991), as well as to the application of that knowledge (Hambrick & Engle, 2002). Although we assembled our reasoning tasks to reflect more fluid problem-solving ability than crystallized knowledge, performance on at least some of our tasks (such as verbal analogies or block rotation) may have benefited from a degree of acquired domain knowledge or task-specific experience. If so, this domain-specific knowledge would be reflected by the domain-specific reasoning factors, and this may have driven their correlation with domain-general WMC.

The Domain Specificity of STM Storage–Rehearsal Abilities and Their Relation to Reasoning

A venerable literature suggests that verbal and visuospatial measures of STM are empirically dissociable. Dual-task studies indicate that verbal secondary tasks, such as articulatory suppression, produce a greater decrement in verbal than in spatial serial recall, and vice versa (e.g., Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Logie, Zucco, & Baddeley, 1990). Moreover, studies of neurological patients indicate dissociable verbal and visuospatial deficits resulting from damage to different brain areas (e.g., De Renzi & Nichelli, 1975; Vallar & Baddeley, 1984). Finally, individual-differences studies sometimes find no correlation between verbal and visuospatial STM measures (e.g., Babcock & Salthouse, 1990; Bayliss et al., 2003; Groeger et al., 1999).

There are some interesting exceptions to these general rules, however. First, experimental work shows that when memory for order is critical to both verbal and visuospatial STM tasks (in addition to, or instead of, memory for items), verbal STM may be substantially disrupted by visuospatial secondary tasks (Morris, 1987; Quinn, 1994), and visuospatial STM may be significantly disrupted by verbal secondary tasks and irrelevant speech (e.g., Jones, Farrand, Stuart, & Morris, 1995; Klauer & Stegmaier, 1997). These findings suggest that aspects of verbal and visuospatial rehearsal, particularly regarding maintenance of order information, may be domain general and related to executive attention. Second, correlations among verbal and visuospatial STM tasks are sometimes substantial, with r s between .30 and .70 (Bayliss et al., 2003; Chuah & Maybery, 1999; Park et al., 2002; Swanson & Howell, 2001). Indeed, the cross-domain correlations among individual verbal and spatial STM tasks in the present study ranged from approximately .40 to .50, and at the latent construct level these factors correlated at .63. Verbal and visuospatial STM therefore do not represent entirely independent constructs. We think it is worth noting here that, generally speaking, experimental or neurological dissociations among measures do not necessarily imply the stochastic independence of their underlying processes or mechanisms. If one assumes, as we do, that no cognitive task is process pure (e.g., Jacoby, 1991) and that all WMC and STM tasks are jointly determined by (at least) domain-general executive processes and domain-specific storage and rehearsal processes, then any empirical dissociation may reflect only the specificity of the storage and rehearsal processes and not the contribution of the general executive.

Nonetheless, an interesting distinction between verbal and visuospatial storage–rehearsal processes arose in our final model, which addressed the dissociation of executive-attention and storage–rehearsal aspects of WM. While removing the general, executive variance from verbal and visuospatial storage tasks (which was nonnegligible), we found that the residual variance from only spatial storage–rehearsal factors was domain general in predicting fluid reasoning. Whereas verbal storage–rehearsal processes were correlated strongly with verbal reasoning and very weakly with general ability (see also Cantor et al., 1991; Engle, Tuholski, et al., 1999), spatial storage–rehearsal processes predicted both spatial reasoning and general fluid ability strongly. As we noted earlier, these findings are consistent with others demonstrating a close link between “simple” spatial storage and higher order cognitive capabilities such as executive function (Miyake et al., 2001), complex reasoning and Gf (Oberauer, in press; Shah & Miyake, 1996), and free recall of either verbal or visuospatial materials (Park et al., 2002). However, we also found that when our Gf factor was defined in a less spatially biased manner, spatial storage had a substantially weaker correlation with it. Although the correlation between

spatial storage and a verbally biased Gf factor was significant and larger than the correlation between verbal storage and a spatially biased Gf factor, it was clear that spatial storage did not have the broad predictive utility that the domain-general executive-attention factor had.

There are further anomalies in the small literature on spatial STM that will require future work to resolve. In Miyake and Shah's studies (Miyake et al., 2001; Shah & Miyake, 1996), for example, spatial STM span tasks correlated strongly with math SAT scores, spatial-visualization composites, and executive-function composites, and they did so as strongly as did spatial WM span tasks. However, these spatial STM and WM tasks did not always account for the same variance in these complex cognitive measures. In the Miyake et al. (2001) study, a latent-variable model with a single spatial WM–STM factor fit the data as well as one that separated the WM and STM constructs (the latter were correlated at .86), and this single spatial memory construct accounted for 35% of the variance in an executive-function construct composed of Tower of Hanoi and random generation tasks. Here, then, spatial WM and STM tasks appeared to measure a single construct related to executive attention. In contrast, Shah and Miyake (1996) found patterns more similar to those in the present study. Their spatial STM task correlated with spatial ability as strongly as did their spatial WM task, but a partial-correlation analysis showed that each task contributed primarily unique, rather than shared, variance to spatial ability. In our data, spatial storage–rehearsal correlated strongly with Gf over and above the variance accounted for by our executive-attention factor, suggesting that spatial STM and WM were each independently tapping different aspects of general ability.

Thus, although there is growing evidence that spatial STM span tasks measure more than just a simple, passive storage capability, it is unclear what they actually measure. The uninteresting hypothesis that spatial STM tasks lend themselves to verbal coding or rehearsal strategies is unlikely to be correct, because verbal STM tasks, which are maximally influenced by these strategies, do not strongly predict Gf. Instead, two alternative hypotheses seem somewhat viable. One is that high scores on spatial STM tasks may reflect some (nonverbal) strategic capabilities, such that those individuals who are better able to generate novel, successful strategies (e.g., spatial chunking) to solve these unfamiliar tasks are also more highly intelligent (see Shah & Miyake, 1996). Alternatively, spatial STM tasks may be somewhat purer measures of executive attention than are verbal STM tasks because spatial STM tasks preclude the use of phonological rehearsal and the reliance on semantic–lexical information at recall, both of which may act as suppressor variables in the relation between verbal STM span and general ability. That is, without the well-practiced strategies people have developed for maintaining linguistic information in verbal tasks, participants in spatial STM tasks must rely solely on attentional–executive processes to maintain target information. Although plausible and consistent with the larger visuospatial rehearsal literature (e.g., Awh, Jonides, & Reuter-Lorenz, 1998; Klauer & Stegmaier, 1997), this hypothesis cannot explain why, in our structural model, the link between spatial storage and Gf was independent of the executive-attention factor derived from WM and STM tasks. Obviously, more work is required to determine the potentially important abilities and processes that underlie the spatial STM span task. In closing, however, we reiterate our own conclusion that spatial storage processes do not appear to be as general in their predictive power as are the executive-attention processes that contribute to performance in WMC span tasks.

What Is the Nature of WMC?

The fact that our conclusions about the relation between spatial STM and Gf change with the task composition of Gf may call into question the generality of our more critical findings. That is, might the domain generality of WMC or STM look quite different if we defined them with a different collection of tasks? Although results from factor analyses are always bound by the particular measures that are included, we think it highly unlikely that our conclusions are specific to our tasks. First, we constructed our span battery to reflect a broad variety of verbal and spatial stimuli and secondary-processing tasks, eliminating any bias of a particular stimulus or processing type. Second, our conclusions were drawn from latent variables, and not individual tasks, and therefore it would take several of our span tasks “behaving badly” to substantially alter the results. Third, the generality of WMC is suggested by other latent-variable studies, from multiple laboratories and using tasks quite different from ours, as we previously reviewed at length (and, prior work shows strong enough links

among span and other varieties of WMC tasks to suggest that our findings would generalize even to nonspan WMC tasks; see Engle, Tuholski, et al., 1999; Oberauer et al., 2000, 2003). Fourth, our WMC and STM span tasks presented almost identical memory stimuli within each domain, and therefore the greater generality of WMC versus STM is a serious challenge to explain by measurement error alone.

Moreover, we should consider what the alternative findings could possibly look like. For WMC to be more domain specific than we found, either our verbal tasks must have leant themselves to spatial encoding–rehearsal–storage, or our spatial tasks must have leant themselves to verbal encoding–rehearsal–storage; that is, our verbal or spatial span tasks (or both) must have been biased toward the opposite domain. A spatial bias in our verbal tasks is extremely unlikely given how similar our tasks were to those that have been extensively demonstrated to predict verbal abilities and given that none of our verbal tasks had any spatial components. The latter possibility, that our spatial span measures allowed verbal storage or strategy, is a ubiquitous problem in research on spatial cognition, and therefore is much more likely to be relevant here. However, if our spatial tasks actually suffered verbal contamination, this would have been more obvious in our STM tasks than in our WM tasks, and therefore our STM constructs should have shown more domain generality than did our WMC constructs. That is, STM span, without secondary-processing tasks, should better allow for the engagement of verbal strategies than should WMC span. In fact, our STM tasks were primarily domain specific, and our WMC tasks, which made complex coding and rehearsal strategies very difficult because of their dual-task demands, were primarily domain general, markedly more so than the STM tasks.

Thus, the findings from the present study strongly support the view that WM span tasks are jointly determined by domain-general executive-attention functions primarily and domain-specific storage and rehearsal processes secondarily, where the domain-general executive contributions are critical to the general predictive utility of WM span (Engle & Kane, 2004; Engle, Kane, & Tuholski, 1999). We believe that these executive-attention functions reflect a capability to maintain information in an active, easily accessible state, whether that information corresponds to a list of several external stimuli or a single goal for action. Moreover, such active maintenance is particularly important in the presence of interference, which disrupts rapid retrieval of information from LTM, and in the blocking of distraction and competing response tendencies (Engle & Kane, 2004; Kane & Engle, 2002, 2003).

The fact that WMC is not dissociable along domain-specific lines leaves strong process-specific views of WMC untenable (e.g., Daneman & Tardif, 1987; Shah & Miyake, 1996); it also conflicts with the related, and more extreme, position that there are no dedicated WM processes, functions, or mechanisms that are independent of task-specific skills acquired through experience (MacDonald & Christiansen, 2002). From their connectionist approach to understanding language comprehension, MacDonald and Christiansen (2002) concluded that WM span tasks, such as reading span, predict language-comprehension measures only because they involve language ability themselves, not because they tap an independent WM function. Span tasks and comprehension tasks are simply different measures of the same thing—language skills. Moreover, individual differences in these skills result from differences in language exposure or knowledge and biological factors affecting precision of phonological representations. By this view, WMC cannot be dissociated from verbal ability, for it is not an independent construct. Our findings that verbal WM span tasks are very strongly correlated with, if not indistinguishable from, spatial WM span tasks and that the common variance among verbal and spatial span tasks predicts general and specific aspects of reasoning may seem to contradict the view that verbal WM span measures reflect only verbal ability. However, MacDonald and Christiansen (2002) claimed that their view allows for correlations among verbal and spatial WM tasks by positing similar underlying skills acquired through experience:

To the extent that the biology and experience underlying language comprehension skills and the skills used in navigation through space are similar, or to the extent that particular spatial and linguistic [WM] tasks have similar components, performance on linguistic and spatial [WM] tasks will correlate with one another. (p. 50)

We find this claim to be problematic, for its acceptance requires that some specific experience and skill underlies remembering words while verifying equations, remembering letters while judging the sensibility of sentences, remembering arrow positions while mentally rotating letters, and remembering ball-movement paths while mentally navigating around the perimeter of objects. Moreover, this particular skill, developed through a common verbal–spatial experience, happens to be the same as that required by such diverse reasoning tasks as reading comprehension, figural matrix completing, and mental rotation of three-dimensional shapes. In our view, the claim for the existence of such a general “skill” can only make sense if one reinterprets it as a developed ability to maintain stimulus representations (verbal or visuospatial) in the presence of interference from prior experiences and attention shifts to and away from the target stimuli—in other words, a domain-general WMC that is independent of any one particular processing domain.

Notes:

¹ The 2001–2002 data we report for the SAT and ACT were taken from *Barron’s Profiles of American Colleges* (Barron’s Educational Series, 2003).

² Detailed results from exploratory factor analyses are available from Michael J. Kane.

³ Our nested-factor model here is actually less inconsistent with those of Conway et al. (2002) and Engle, Tuholski, et al. (1999) than it first appears. Conway et al. (2002) reported briefly that in a model in which all STM and WMC tasks loaded onto a common factor (WMC) and the STM tasks alone constituted a residual storage factor (somewhat similar to our present analysis), only the correlation between WM and Gf was significant. Thus, when STM is a residual factor rather than WM, it is the common WM variance that predicts intelligence. Furthermore, a reanalysis of the Engle, Tuholski, et al. (1999) data (Andrew R. A. Conway, personal communication, June 12, 2003) yields the same result, with the common (WM + STM) factor correlating .52 with Gf and the residual STM factor correlating —.08 with Gf. Thus, the variance common to WMC and STM tasks can reflect executive-attention variance, depending on the particulars of the remaining model structure (i.e., whether the residual variance is drawn from STM or WMC tasks).

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.
- Andrade, J. (2001). *Working memory in perspective*. Hove, East Sussex, England: Psychology Press.
- 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* (Vol. 2, pp. 89–95). New York: Academic Press.
- Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 780–790.
- Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and Aging*, *5*, 421–428.
- Baddeley, A. D. (1986). *Working memory*. London: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
- Barron’s Educational Series. (2003). *Barron’s profiles of American colleges* (25th ed.). Hauppauge, NY: Author.
- Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*, 71–92.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1972). *Differential Aptitude Test: Space Relations, Form 1*. New York: Psychological Corporation.

- Berger, F. R., Gupta, W. B., Berger, R. M., & Skinner, J. (1990). *Air Force Officer Qualifying Test (AFOQT) Form P: Test manual* (AFHRL-TR89-56). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Bowers, K. S., Regehr, G., Balthazard, C. G., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, *22*, 72–110.
- Brewin, C. R., & Beaton, A. (2002). Thought suppression, intelligence, and working memory capacity. *Behavior Research and Therapy*, *40*, 923–930.
- Bridgman, P. W. (1945). Some general principles of operational analysis. *Psychological Review*, *52*, 246–249.
- Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*, 349–369.
- Burke, H. R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *Journal of Genetic Psychology*, *93*, 199–228.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*, 229–246.
- Carpenter, P. A., & Just, M. A. (1988). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Erlbaum.
- Carpenter, P. A., Miyake, A., & Just, M. A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology*, *46*, 91–100.
- Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology*, *73*, 7–44.
- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*, 1086–1095.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.
- Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, *4*, 577–590.
- Conway, A. R. A., & Kane, M. J. (2001). Capacity, control and conflict: An individual differences perspective on attentional capture. In C. Folk & B. Gibson (Eds.), *Attraction, distraction and action: Multiple perspectives on attention capture* (pp. 349–372). Amsterdam: Elsevier Science.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547–552.
- Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*, 15–18.
- Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J., et al. (2003). Children's working memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*, 113–132.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 561–584.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.
- Daneman, M., & Tardif, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 491–508). Hillsdale, NJ: Erlbaum.

- Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.
- Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental– differential approach. *Trends in Cognitive Sciences*, 5, 164–170.
- Deary, I. J. (2002). *g* and cognitive elements of information processing: An agnostic view. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 151–181). Mahwah, NJ: Erlbaum.
- Deary, I. J., Egan, V., Gibson, G. J., Austin, E., Brand, C. R., & Kellaghan, S. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, 23, 105–132.
- De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 357–376). Cambridge, MA: MIT Press.
- De Jong, R. (2001). Adult age differences in goal activation and goal maintenance. *European Journal of Cognitive Psychology*, 13, 71–89.
- De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex*, 11, 341–354.
- Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.
- Ekstrom, R. B., French, J. W., Harman, M. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 972–992.
- Engle, R. W., & 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* (pp. 145–199). New York: Academic Press.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge University Press.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Feldman Barrett, L., Tugade, M. M., & Engle, R. W. (in press). Individual differences in working memory capacity and dual-process theories of mind. *Psychological Bulletin*.
- Finn, P. R., Mazas, C. A., Justus, A. N., & Steinmetz, J. E. (2002). Early-onset alcoholism with conduct disorder: Go/no go learning deficits, working memory capacity, and personality. *Alcoholism: Clinical & Experimental Research*, 26, 186–206.
- Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, 54, 1–71.
- Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, 129, 61–83.
- Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, 63, 149–159.
- Groeger, J. A., Field, D., & Hammond, S. M. (1999). Measuring memory span. *International Journal of Psychology*, 34, 359–363.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407–434.
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, 44, 339–384.
- Handley, S. J., Capon, A., Copp, C., & Harper, C. (2002). Conditional reasoning and the Tower of Hanoi: The role of spatial and verbal working memory. *British Journal of Psychology*, 93, 501–518.

- Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, 9, 233–247.
- Henson, R. (2001). Neural working memory. In J. Andrade (Ed.), *Working memory in perspective* (pp. 151–173). Hove, East Sussex, England: Psychology Press.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R., & Weng, L.-J. (1994). What is a good g? *Intelligence*, 18, 231–258.
- Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1008–1018.
- Jonides, J., Reuter-Lorenz, P. A., Smith, E. E., Awh, E., Barnes, L. L., Drain, M., et al. (1996). Verbal and spatial working memory in humans. *Psychology of Learning and Motivation*, 35, 43–88.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130, 169–183.
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, 9, 637–671.
- Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70.
- Kellogg, C. E., & Morton, N. W. (1999). *Revised Beta Examination* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Kihlstrom, R. (n.d.). Retrieved November 2000 from http://socrates.berkeley.edu/~kihlstrm/remote_associates_test.htm [now at <http://socrates.berkeley.edu/~kihlstrm/RATest.htm>]
- Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F., & Salhouse, T. A. (1999). Models of working memory: Eight questions and some general issues. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 412–441). New York: Cambridge University Press.
- Klauer, K. C., & Stegmaier, R. (1997). Interference in immediate spatial memory: Shifts of spatial attention or central-executive involvement? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 50(A), 79–99.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *Journal of General Psychology*, 120, 375–405.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence*, 14, 389–433.
- Law, D. J., Morrin, K. A., & Pellegrino, J. W. (1995). Training effects and working memory contributions to skill acquisition in a complex coordination task. *Learning and Individual Differences*, 7, 207–234.
- Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence*, 23, 45–57.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale, NJ: Erlbaum.
- Logie, R. H., Zucco, G., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, 75, 55–74.
- Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition*, 3, 294–301.
- Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, 130, 199–207.

- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.
- Mednick, S. A., & Mednick, M. T. (1967). *Examiner's manual: Remote Associates Test*. Boston: Houghton Mifflin.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.
- Miyake, A., & Shah, P. (1999a). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.
- Miyake, A., & Shah, P. (1999b). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442–481). New York: Cambridge University Press.
- Morrell, R. W., & Park, D. C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging*, *8*, 389–399.
- Morris, N. (1987). Exploring the visuo-spatial scratch pad. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *39(A)*, 409–430.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*, 214–255.
- Oberauer, K. (in press). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Understanding and measuring intelligence*. Thousand Oaks, CA: Sage.
- Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045.
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299–320.
- Plomin, R., & Spinath, F. M. (2002). Genetics and general cognitive ability (*g*). *Trends in Cognitive Sciences*, *6*, 169–176.
- Postle, B. R., & D'Esposito, M. (2000). Evaluating models of the topographical organization of working memory function in frontal cortex with event-related fMRI. *Psychobiology*, *28*, 132–145.
- Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Author.
- Quinn, J. G. (1994). Towards a clarification of spatial processing. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *47(A)*, 465–480.
- Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Richardson, J. T. E., Engle, R. W., Hasher, L., Logie, R. H., Stoltzfus, E. R., & Zacks, R. T. (1996). *Working memory and human cognition*. New York: Oxford University Press.
- Roberts, R. J., Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology*, *12*, 105–126.
- Romeu, J. L., & Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. *Journal of Multivariate Analysis*, *46*, 309–334.
- Rosen, V. M., Bergeson, J. L., Putnam, Harwell, A., & Sunderland, T. (2002). Working memory and apolipoprotein E: What's the connection? *Neuropsychologia*, *40*, 2226–2233.
- Salthouse, T. A. (1995). Differential age-related influences on memory for verbal-symbolic and visual-spatial information? *Journal of Gerontology: Psychological Sciences*, *50(B)*, P193–P201.
- Salthouse, T. A. (2001). A research strategy for investigating group differences in a cognitive construct: Application to ageing and executive processes. *European Journal of Cognitive Psychology*, *13*, 29–46.

- Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging, 6*, 118–127.
- Salthouse, T. A., & Mitchell, D. R. D. (1989). Structural and operational capacities of integrative spatial ability. *Psychology and Aging, 4*, 18–25.
- Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 507–516.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychological Software Tools.
- 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.
- Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research, 7*, 1–24.
- Sohn, Y. W., & Doane, S. M. (2003). Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory & Cognition, 31*, 458–466.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan.
- Stankov, L. (2002). *g*: A diminutive general. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 19–37). Mahwah, NJ: Erlbaum.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
- Süß, H.-M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—And a little bit more. *Intelligence, 30*, 261–288.
- Swanson, H. L. (1996). Individual and age-related differences in children's working memory. *Memory & Cognition, 24*, 70–82.
- Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology, 93*, 720–734.
- Swanson, H. L., & Sachse-Lee, C. (2001). A subgroup analysis of working memory in children with reading disabilities: Domain-general or domain-specific deficiency? *Journal of Learning Disabilities, 34*, 249–263.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- Toth, J. P., & Hunt, R. R. (1999). Not one versus many, but zero versus any: Structure and function in the context of the multiple memory systems debate. In J. K. Foster & M. Jelicic (Eds.), *Memory: Systems, process, or function?* (pp. 233–272). Oxford, England: Oxford University Press.
- Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition, 28*, 341–348.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40*, 385–398.
- Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*, 446–468.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.
- Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior, 23*, 151–161.
- Wickens, C. D. (1980). The structure of processing resources. In R. Nickerson & R. Pew (Eds.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.
- Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities, 34*, 237–248.

Appendix A

Orders in Which Tasks Were Performed With Sessions

| Session and task | Order | | | | | f |
|------------------|-------|----|----|---|---|---|
| | a | b | c | d | e | |
| Session 1 | | | | | | |
| WordSpan | 1 | 2 | 3 | 4 | 5 | 6 |
| BallSpan | 2 | 3 | 6 | 5 | 4 | 1 |
| DigSpan | 3 | 4 | 5 | 6 | 1 | 2 |
| MatxSpan | 4 | 5 | 2 | 1 | 6 | 3 |
| LettSpan | 5 | 6 | 1 | 2 | 3 | 4 |
| ArrwSpan | 6 | 1 | 4 | 3 | 2 | 5 |
| Session 2 | | | | | | |
| OpeSpan | 1 | 2 | 3 | 4 | 5 | 6 |
| NavSpan | 2 | 3 | 6 | 5 | 4 | 1 |
| CouSpan | 3 | 4 | 5 | 6 | 1 | 2 |
| SymmSpan | 4 | 5 | 2 | 1 | 6 | 3 |
| ReadSpan | 5 | 6 | 1 | 2 | 3 | 4 |
| RotaSpan | 6 | 1 | 4 | 3 | 2 | 5 |
| Session 3 | | | | | | |
| Inference | 1 | 9 | 4 | | | |
| PapFold | 2 | 8 | 12 | | | |
| Analogy | 3 | 12 | 10 | | | |
| SpaceRel | 4 | 10 | 1 | | | |
| BETAIII | 5 | 11 | 8 | | | |
| RemoAsso | 6 | 7 | 2 | | | |
| RotaBlock | 7 | 13 | 5 | | | |
| Syllogism | 8 | 5 | 13 | | | |
| SurfDev | 9 | 2 | 7 | | | |
| Ravens | 10 | 1 | 11 | | | |
| ReadComp | 11 | 3 | 6 | | | |
| FormBrd | 12 | 4 | 9 | | | |
| WASI | 13 | 6 | 3 | | | |

Note. WordSpan = word span; BallSpan = ball span; DigSpan = digit span; MatxSpan = matrix span; LettSpan = letter span; ArrwSpan = arrow span; OpeSpan = operation span; NavSpan = navigation span; CouSpan = counting span; SymmSpan = symmetry span; ReadSpan = reading span; RotaSpan = rotation span; Inference = Educational Testing Service Kit of Factor-Referenced Tests (ETS)—Inference Test; PapFold = ETS Paper Folding Test; Analogy = Air Force Officer Qualifying Test (AFOQT)—Analogies Test; SpaceRel = Differential Aptitude Test—Space Relations Test; BETAIII = Revised Beta Examination (3rd ed.)—Matrix Reasoning; RemoAsso = Remote Associates Test; RotaBlock = AFOQT Rotated Blocks Test; Syllogism = ETS Nonsense Syllogisms Test; SurfDev = ETS Surface Development Test; Ravens = Ravens Advanced Progressive Matrices; ReadComp = AFOQT Reading Comprehension; FormBrd = ETS Form Board Test; WASI = Wechsler Abbreviated Scale of Intelligence, Matrix Reasoning.

Appendix B

Descriptive Data and Reliability Estimates (Cronbach's α) for Processing Accuracy in the Working Memory Span Tasks

| Task | <i>M</i> | <i>SD</i> | Skew | Kurtosis | α^a |
|----------|----------|-----------|-------|----------|------------|
| OpeSpan | .97 | .05 | -4.94 | 34.29 | .75 |
| ReadSpan | .98 | .03 | -2.75 | 10.25 | .51 |
| CouSpan | .98 | .03 | -1.63 | 2.89 | .48 |
| NavSpan | .98 | .06 | -5.68 | 44.71 | .78 |
| SymmSpan | .97 | .05 | -3.03 | 11.43 | .70 |
| RotaSpan | .98 | .04 | -2.38 | 6.40 | .61 |

Note. *N* = 234 (data missing from 2 participants). OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span. ^a The α computation is based on the proportion of processing items within each memory set that were solved correctly.

Appendix C

Correlation Between Recall Score and Processing Accuracy for Each of the Working Memory Span Tasks

| Task | <i>r</i> |
|----------|----------|
| OpeSpan | .30 |
| ReadSpan | .19 |
| CouSpan | .33 |
| NavSpan | .18 |
| SymmSpan | .27 |
| RotaSpan | .19 |

Note. *N* = 234 for all tasks (data missing from 2 participants). OpeSpan = operation span; ReadSpan = reading span; CouSpan = counting span; NavSpan = navigation span; SymmSpan = symmetry span; RotaSpan = rotation span.

Received February 10, 2003
 Revision received December 16, 2003
 Accepted January 27, 2004 ■