

Working Memory and Intelligence: The Same or Different Constructs?

Phillip L. Ackerman, Margaret E. Beier, and Mary O. Boyle
Georgia Institute of Technology

Several investigators have claimed over the past decade that working memory (WM) and general intelligence (g) are identical, or nearly identical, constructs, from an individual-differences perspective. Although memory measures are commonly included in intelligence tests, and memory abilities are included in theories of intelligence, the identity between WM and intelligence has not been evaluated comprehensively. The authors conducted a meta-analysis of 86 samples that relate WM to intelligence. The average correlation between true-score estimates of WM and g is substantially less than unity ($\hat{\rho} = .479$). The authors also focus on the distinction between short-term memory and WM with respect to intelligence with a supplemental meta-analysis. The authors discuss how consideration of psychometric and theoretical perspectives better informs the discussion of WM–intelligence relations.

Since the 1980s, with the major theoretical and empirical developments of the construct of working memory (WM; see, e.g., Baddeley, 1986; Richardson, 1996, for reviews) as distinct from rote or span memory (which is usually referred to as *short-term memory* [STM]), several investigators have asserted that WM and intellectual abilities are highly related or identical constructs. These assertions started with demonstrations that significant correlations were found between some measures of WM and measures of comprehension (Daneman & Carpenter, 1980), and later between WM and reasoning ability (Kyllonen & Christal, 1990), and other measures, such as the SAT (e.g., Turner & Engle, 1989). Recently, several investigators have claimed that WM and general intelligence (g ; or general fluid intelligence, Gf) are essentially the same constructs. For example:

So central is the role of WM capacity in individual differences in information processing that some cognitive theorists equate WM capacity with g itself. (Jensen, 1998, p. 221)

Stauffer et al. (1996) found a correlation of + 0.995 between a factor representing general intelligence (g) and a factor representing WM. (Colom, Flores-Mendoza, & Rebollo, 2003, p. 34)¹

My colleagues and I used a structural equation modeling analysis to test this and the idea that the construct measured by WM-capacity tasks is closely associated with general fluid intelligence. . . . WM-capacity tasks measure a construct fundamentally important to higher-

level cognition. The construct is distinguishable from STM and is at least related to, maybe isomorphic to, general fluid intelligence and executive attention. (Engle, 2002, pp. 21–22)

No other cognitive factor—knowledge, speed, or learning ability—correlated with g after the working memory factor was partialled out. Thus, we have our answer to the question of what g is. It is working memory capacity. (Kyllonen, 2002, p. 433; see also Kyllonen, 1996)

However, the relationship between memory and intelligence appears to be much more complex than has been asserted by these investigators. Note that this position is not without its critics. For example, differential psychologists such as Deary (2000) and Kline (2000) have expressed substantial skepticism that WM and general intelligence are even closely linked. In this article, we evaluate these claims in the context of a meta-analysis of correlations between WM measures and intellectual ability measures. In addition, it is important to note that intelligence theory and the assessment of intelligence have both involved memory abilities over the past 110 years. Although models of WM represent relatively recent developments in the history of the science, an understanding of the construct space for individual differences in intelligence and WM benefits from a brief review of intelligence and memory ability research. Thus, we begin with a consideration of memory abilities from an intelligence assessment perspective, followed by a review of research on memory abilities and intelligence theory. Next, we briefly review memory theory and the underlying framework for asserting the overlap between WM and intelligence. A meta-analysis of 86 samples that report correlations between measures of WM and measures of intellectual abilities is then presented. A parallel set of analyses is also provided for STM and intelligence for comparison to the WM–intelligence relations. We then discuss the implications of the meta-analytic results in the context of both enduring psychometric measurement and theory issues.

Phillip L. Ackerman, Margaret E. Beier, and Mary O. Boyle, School of Psychology, Georgia Institute of Technology.

Margaret E. Beier is now at the School of Psychology, Rice University. This research was partially supported by Air Force Office of Scientific Research Grant F49620-01-0170 (Phillip L. Ackerman, principal investigator). We thank the following individuals for providing correlations from their empirical research: Scott Chaiken, Sandra Hale, Tim Salthouse, Lee Swanson, Bill Tirre, and Werner Wittmann. We also thank Joni Lakin and Sarah Molouki for their assistance in obtaining correlations and data entry and processing.

Correspondence concerning this article should be addressed to Phillip L. Ackerman, School of Psychology, MC 0170, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170. E-mail: phillip.ackerman@psych.gatech.edu

¹ It is interesting to note that in fact the original Stauffer, Ree, and Caretta (1996) article does not state this conclusion. The higher order factor that Colom et al. (2003) identified as “WM” is actually identified with four lower order factors: Processing Speed, Working Memory, Declarative Knowledge, and Procedural Knowledge, not just with WM.

Historical Background—Immediate Memory Ability

The earliest experimental studies of individual differences in “immediate memory” took place in the 1880s (e.g., Jacobs, 1887; see Whipple, 1914, 1921, for a review). Critical to later applications in the assessment of individual differences in intelligence, Jacobs (1887) noted that larger spans were found with older children compared with younger children (see Bigham, 1894; Kirkpatrick, 1894; Münsterberg & Bigham, 1894). Binet’s experiments with memory tests focused on immediate memory for sentences and for unrelated words. According to Peterson (1925),

Binet favored [memory tests] for two reasons: (1) memory involves content of the higher mental functions, not mere sensations, and (2) by means of memory tests one can indirectly study the operations and nature of such higher mental processes as discrimination, attention, and intelligence. (p. 125)

Terman’s (1916) translation and revision of the later Binet–Simon scales continued the use of a digit span test and included a sentence span test. The digit span test was given in a “forward” format for children ranging in age from 3 to 11 years, and a backward digit span test was given to children above age 7. It is useful to note that Terman (1916) suggested that “as a test of intelligence, this [backward digit span] test is better than that of repeating digits in the direct order. It is less mechanical and makes a much heavier demand on attention” (p. 208). Also, Terman noted that an effective strategy used by some more intelligent examinees was to break the sequence of numbers into groups and report the numbers separately by group (an early example of “chunking” of information; see Miller, 1956).

Since the nearly simultaneous development of the first modern omnibus intelligence tests (e.g., Binet & Simon, 1905/1961) and the first modern theory of intelligence (Spearman, 1904), there have been largely parallel developments in each field. Many intelligence measures have been developed with substantially greater attention given to criterion-related validity, as opposed to construct validity. Memory span tests have been used without substantial change in the Stanford–Binet through all its major revisions (e.g., Terman & Merrill, 1937, 1960; R. L. Thorndike, Hagen, & Sattler, 1986). Digit span tests are found in the Wechsler scales, from its earliest edition (the Wechsler–Bellevue; see Wechsler, 1944), up through the most current version—the third edition of the Wechsler Adult Intelligence Scale (WAIS–III; Psychological Corporation, 1997), and in other individual intelligence tests (see Anastasi & Urbina, 1997). The Woodcock–Johnson III refers to a simple span memory for order of digits and words as a test of Auditory Working Memory (Schrank, Flanagan, Woodcock, & Mascolo, 2002). It would be fair to say that most of the extant omnibus tests of intelligence have not fundamentally changed in assessing memory span in nearly 100 years.² Intelligence theory and basic research on intellectual abilities, in contrast, have involved substantial debate on the role of memory abilities in intelligence. We next discuss intelligence theory and memory abilities research.

Immediate Memory and Intelligence Theory

It has been said that there are as many intelligence theories as there are intelligence theorists (see, e.g., *Journal of Educational Psychology*, 1921), although there have been two major traditions.

The first tradition is the two-factor, or “g,” approach proposed by Spearman (1904). The second tradition is the group-factor approach (e.g., Kelly, 1928; Thomson, 1939) and is mostly identified with Thurstone (1938). Although there are current adherents of the Spearman tradition (e.g., Jensen, 1998), most modern theories tend to take a middle-ground approach to intelligence—such as the hierarchical model of P. E. Vernon (1950; see also Marshalek, Lohman, & Snow, 1983; Snow, Kyllonen, & Marshalek, 1984). Below, we briefly review memory abilities in the context of these theoretical perspectives.

Spearman

As noted by Carroll (1993, p. 249), Spearman had relatively little use for the construct of memory in his theory of intelligence. That is, Spearman (1927) stated that “all the available evidence indicates that *g* is exclusively involved in education *and not at all in bare retention* [italics added]” (p. 285). Later Spearman and Jones (1950) argued that insufficient evidence existed to establish memory as an ability factor. Though immediate memory does not enter into Spearman’s conceptualization of intelligence, his notion of *g* is an important construct in the current discussion. In his original presentation of the theory, Spearman (1904) presented two major tenets to the theory: (a) General Intelligence and General Discrimination (such as psychophysical measures of pitch discrimination and weight discrimination) were, for all intents and purposes, perfectly correlated; and (b) the measures with the highest *g saturation* were grades in classics and peer ratings of common sense.

Later, Spearman (1914) reviewed and reanalyzed data presented by Simpson (1912). The Simpson data included a variety of memory, verbal, reasoning, perceptual speed and perceptual judgment tests. Here, Spearman (1914) described his theoretical construct of *g* as a “*general fund of mental energy*” (p. 103). In Spearman’s (1914) reanalysis, he found that the Ebbinghaus Completion Test, when combined with other verbal and memory tests, had extremely high correlations with the general factor ($r = .95$), although tests of verbal memory had slightly higher correlations. Later, Spearman (1938) stated that *g* was well represented by individual differences in the Penrose and Raven (1936) test—later called Raven’s Progressive Matrices (Raven, Court, & Raven, 1977; see also H. R. Burke, 1958). Thus, the defining characteristics for *g* shifted from various abilities to a test of nonverbal (or

² The latest versions of the Wechsler Adult Intelligence Scale (WAIS; i.e., WAIS–III) and Stanford–Binet (SBV; see Roid, 2003) do purport to measure WM. The WAIS–III estimate of WM is predicated on existing tests (Digit Span and Arithmetic—the former is largely unchanged, and the latter has “reduced” time constraints for time-bonus points), plus one new test (Letter–Number Sequencing). In our opinion, however, the inclusion of the Arithmetic Test and the Digit Forward Test complicate, rather than simplify, the interpretation of the WM factor from the WAIS. The SBV includes two new tests that purport to assess WM (Last Word and Block Span) and also incorporate previous tests (Memory for Sentences, Memory for Digits, and Delayed Response). The new Last Word Test does conform to conceptualizations of WM, but the extant tests may not. The Block Span Test is similar to digit span, with forward and backward spans—but uses spatial problem content. As with the WAIS, generating a single WM composite from these tests results in a composite with unclear construct status—representing some confluence of span memory and WM.

spatial) inductive reasoning. In addition, the construct of g became endowed with the notion that it represented a “general fund of mental energy,” or a mental engine. However, most contemporary intelligence theorists do not adopt this particular representation of g (but see Messick, 1996), but rather assert that g is an abstraction implied by the common variance among cognitive ability tests.³ Some psychologists have taken Spearman’s notion of g and referred to it as an index of the capability of attention (Cowan, 1997) or executive process (Engle, 2002).

Group-Factor Theories and Memory

Kelly (1928) provided evidence for a common factor underlying memory span tests. When corrected for unreliability, the correlations among four memory tests ranged from .54 to .96. These tests correlated reasonably well ($r_s = .39$ to $.66$) with a General Ability factor but also had substantial residual correlations with a separate Memory factor ($r_s = .46$ to $.56$), leading Kelly (1928) to assert that Memory was a separable group factor from general intelligence. In a later review, Blankenship (1938) noted, “All of these findings indicate a definite relation between memory span and intelligence. But at the present time, results are so varying in nature that the true degree of correlation between the two is impossible to predict” (p. 17). Also, Blankenship observed that although the backward digit span was introduced in 1911, only one researcher had reported a correlation with intelligence, which was a correlation of .75 with the Army Alpha Test in a sample of prisoners (see F. D. Fry, 1931).

Thurstone (1938) included a Memory factor in his “Primary Mental Abilities,” though it was not based on span tests. Subsequently, Guilford (1956, 1967; Guilford & Hoepfner, 1971) proposed a “structure of intellect” model that expanded the number of group factors. In Guilford’s model, there were 24 separate Memory ability factors (see, e.g., Brown, Guilford, & Hoepfner, 1966; Tenopyr, Guilford, & Hoepfner, 1966).

Memory Abilities in Hierarchical Models of Intelligence

P. E. Vernon (1950) proposed a widely accepted hierarchical model of intelligence, with a g factor at the top of the hierarchy and verbal:educational and practical:mechanical abilities at the second level. He expressed doubts as to whether a rote memory factor could in fact be usefully identified separately from the other factors. Cattell (1943) did not propose a fully hierarchical theory; he introduced the concepts of Gf , which is associated with physiologically based abilities, and crystallized intelligence (Gc), which is associated with educational and experiential knowledge, as two major types of adult intelligence.

With respect to immediate memory in the Gf – Gc theoretical framework, Horn (1965) noted that a Memory Span factor loaded positively on the Gf factor ($r = .38$) and negligibly on the Gc factor ($r = -.02$). Horn (1968) reported an average memory span factor coefficient across several of his own studies as .50 with Gf and .00 with Gc (p. 249). Later additions by Horn (1989) include a major factor of “short-term acquisition and retrieval, SAR” (p. 81). Interestingly, Horn (1989) asserted that the “backward span memory test . . . is a considerably better measure of Gf and consequently a poorer measure of SAR than is forward span memory” (p. 91). It is important to note that Gustafsson (1984) and

others have asserted that Cattell’s (1943) Gf is indistinguishable from g when examined from a confirmatory factor analytic perspective.⁴

Spearman’s g in Modern Intelligence Theory

Since Spearman’s (1904) initial theory of intelligence, various investigators have embraced the broad context of the g construct (see, e.g., Jensen, 1998). Many advocates of Spearman’s g take an inductive approach by stating that g is implied by the positive correlations found among ability measures and that g is a generic representation for the general efficacy of intellectual processes. It is not possible to identify g with any single test—it must be approximated by aggregation of several highly g -saturated measures. If one selects only a single measure and identifies it as “ g ,” there is a risk of confabulating a relationship between “ g ” and some other variable, such as WM, because the association between the tests can be contaminated by test-specific variance (e.g., with the Raven, this would include spatial ability and inductive reasoning; see Babcock & Laguna, 1996).

The Received View on Memory and Intelligence

Factor Analytic Research

In the immediate memory domain, Carroll (1993) identified data sets from 117 separate samples for extensive reanalysis. Although a full review of Carroll’s (1993) work is beyond the scope of this article, he identified five lower order immediate memory factors (Memory Span, Associative Memory, Free Recall Memory, Meaningful Memory [or Memory for Ideas], and Visual Memory). He also identified “one or more higher-order memory factors” (Carroll, 1993, p. 256). Only six of the data sets reanalyzed by Carroll (1993) included measures that would be considered as WM measures rather than simple span or rote memory tests. Carroll’s (1993) reanalysis indicated that immediate memory tests tend to cluster by underlying process (e.g., associative memory, span memory) and to some degree by content (at least in the domains of verbal and spatial content). When large batteries of memory tests are administered to a single sample of participants, these factors are well-replicated and indicate a significant association with a general intellectual ability factor. Because of the difficulty in separating the overlapping content among estimates of general intelligence and span memory, it is not clear how to best characterize the association between immediate memory and general intelligence. At one extreme, the association appears to be very large (e.g., .70 or .80), such as between immediate memory and general intelligence when content overlap is not accounted for. At

³ In fact, Jensen (1998) noted explicitly that “it is wrong to regard g as a cognitive process, or as an operating principle of the mind, or as a design feature of the brain’s neural circuitry” (p. 74).

⁴ Because some WM researchers have only evaluated WM and intelligence in the context of the Raven’s Progressive Matrices or other inductive reasoning tests—purportedly exemplary measures of Gf , Gustafsson’s (1984) assertion of the equivalence between Gf and g is perhaps partly responsible for the WM researchers’ inference that $WM = Gf = g$. (However, for a contrasting approach, see Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000.)

the other extreme, when content overlap is accounted for among memory and general ability measures, the association is more modest (a mean loading of .38 for Carroll's, 1993, integrative review). Within Carroll's (1993) three-stratum theory (with *g* at the third stratum), he reported that the Memory Span factor had a mean loading on *g* of .38, with a range of loadings found from .28 to .54 (see p. 597). In the context of the range of other abilities examined, Memory Span had substantially smaller loadings on *g* than did the broad-content abilities (e.g., Spatial Visualization, .55; Quantitative Reasoning, .51; and Verbal, .49). Similar correlations have been reported in a meta-analysis by Mukunda and Hall (1992).

Short-Term Storage

A thorough review of the field of immediate memory theory is beyond the scope of this article (but see Baddeley, 1998; Cowan, 1997, for extensive reviews). Below, we outline a few salient aspects of the characteristics of STM and WM. Although we have discussed that a variety of immediate memory tests had been developed over the period from the late 1800s to the middle 1900s, the influence of information theory (Shannon & Weaver, 1949) on psychology, in the form of information-processing theories, had a significant impact on the characterization of the underlying processes of human memory. Miller (1956), in a classic review of the literature on span memory, suggested that from the metaphor of humans as communication systems, the capacity of human immediate memory was about 7 items, plus or minus 2. Subsequent theoretical and empirical investigations examined the similarities and differences among modalities for information presentation (e.g., auditory and visual), and the parameters regarding various aspects of short-term storage of information (see Lachman, Lachman, & Butterfield, 1979, for a review). The seminal model of memory proposed by Atkinson and Shiffrin (1968) included three memory "stores"—a sensory register, a short-term store, and a long-term store. As Lachman et al. (1979) noted, "Atkinson and Shiffrin represent the short-term store as a 'working memory,' by which they mean that *conscious* mental processes are performed there" (p. 221). There have been many suggested modifications to the original Atkinson and Shiffrin model of the short-term store over the past 3 decades. For current purposes, we focus on a few central aspects of the WM construct as it evolved from original considerations of the short-term store.

WM

Baddeley and his colleagues (e.g., Baddeley & Hitch, 1974) explored aspects of the STM system using tasks that were different from the standard memory span paradigm of the previous several decades. Specifically, this revised paradigm is one that involves performance of two tasks at the same time. One task may be a simple span task, but the other involves some decision making or recoding process (such as performing a digit span task while performing a reasoning test; see Baddeley, 1998). From these kinds of experiments, Baddeley (1998) concluded that immediate memory was better considered as a WM system, with a central executive and two slave systems. The central executive is conceptualized generally as concordant with the model of D. A. Norman and Shallice (1986); that is, "the central executive is involved in a

range of cognitive control processes such as planning, monitoring, and inhibition of inappropriate stimuli or responses" (Phillips & Hamilton, 2001, p. 105). WM is involved in scheduling activities of the slave systems and is involved in strategy selection and a wide variety of other tasks associated with attentional control. The first slave system is a phonological loop, which operates on "speech-based information." The second slave system is a visuospatial sketch pad, which "is responsible for setting up and manipulating visual images" (Baddeley, 1998, p. 52). In brief, then there is a general component of WM (the central executive) and content-specific components (spatial content is handled by the visuospatial sketchpad, and verbal content is handled by the phonological loop). It is not obvious whether there should be a content-specific component for numerical content—current theory and empirical research do not describe a process separate from the phonological loop for handling numerical operations. However, several studies have suggested a substantial role for WM (and the phonological loop) in performance of mental arithmetic tasks (see, e.g., Fürst & Hitch, 2000).

Correlations Between WM and Intelligence Measures

The first investigation of the individual-differences correlates of WM measures was provided by two small studies of reading comprehension ($N = 20$ and $N = 21$) reported by Daneman and Carpenter (1980). WM was assessed by the reading span method, which involves reading a series of sentences and later being asked to recall the last word of each sentence. The authors found a substantial correlation (.72) between the WM measure and the reading comprehension measure. However, this correlation was likely much higher than would be obtained if the measures did not share common content or method variance (i.e., both the measure of WM and the reading comprehension tests shared the same content of reading and verbal memory). Subsequent studies (see, e.g., Baddeley, 1986, for a review; see also Daneman & Merikle, 1996) that used a wider variety of WM measures reported significant but relatively smaller correlations between WM and reading comprehension.

Kyllonen and Christal (1990)

In a seminal series of empirical studies, Kyllonen and Christal (1990) addressed the question of the relationship between measures of WM, reasoning, general knowledge, and processing speed. Although the central message of their article was that "reasoning ability is little more than working memory capacity" (Kyllonen & Christal, 1990, p. 389), some different interpretations are possible. The first issue noted by the authors was that the Reasoning factor had a higher correlation with the Knowledge factor than did the WM factor (p. 425). The second issue, also noted by the authors, was that the WM factor had a higher correlation with a Processing Speed factor than did the Reasoning factor. Thus, while a strong association (" $r = .80 - .90$," Kyllonen & Christal, 1990, p. 389) was found for a factor underlying computerized measures of WM and a factor underlying computerized measures of Reasoning, the evidence was also supportive of a differentiation between these factors, on the basis of convergent and discriminant validity with General Knowledge and Processing Speed.

Engle/Kane/Conway Studies

Some subsequent investigators have largely embraced a view that WM and g represent identical constructs (e.g., Engle, 2002). Kane and Engle (2002) stated, "Briefly, we note here that WM-capacity measures, requiring a variety of different processing skills and presenting a variety of stimulus types, correlate substantially with fluid ability tasks across verbal, mathematical, and spatial domains" (p. 658). In addition, they asserted that "there are simply too many strong correlations among diverse WM-capacity tasks and diverse higher order tasks to deny that some general mechanism is involved" (Kane & Engle, 2002, p. 659).

As noted by Ackerman, Beier, and Boyle (2002), however, numerous empirical studies that have investigated the relationship between WM and g have found correlations that do not support such a strong relationship. In a brief review of the literature containing correlations between the Raven's Progressive Matrices Test and measures of WM, Ackerman et al. (2002) noted that the reported raw variable correlations by the Engle research group are generally lower than the raw variable correlations reported by Kyllonen and Christal (1990) for WM and reasoning tests (e.g., .32 to .54 in Study 1; .25 to .58 in Study 2). Specifically, Ackerman et al. (2002, p. 568) cited the Conway, Cowan, Bunting, Theriault, and Minkoff (2002) correlations between WM and Raven ranging from .15 to .38 and the Engle, Tuholski, Laughlin, and Conway (1999) correlations ranging from .28 to .34, though it is important to note that these authors based their interpretation of the strong relationship between WM and intelligence, not on these modest raw correlations, but rather on their derivation of estimated latent variables.

Although some of the quotations from various researchers are relatively current, there is an indication that, at least among some researchers, opinions regarding the isomorphic relationship between WM and g (or gf) are changing. For example, Conway, Kane, and Engle (1999) stated, "Furthermore, we suggest that WMC [working memory capacity], which reflects controlled attention capability, is the basis of gf ." More recently, they presented a revised conclusion: "In summary, several latent variable analyses suggest that WMC accounts for at least one third and perhaps as much as one half the variance in g " (p. 551), saying that "a review of the recent research reveals that WMC and g are indeed highly related, but not identical" (Conway, Kane, & Engle, 2003, p. 547).

Jurden (1995) and Babcock (1994)

Jurden (1995) reported correlations of .20 and .43 for WM and the Raven. Babcock (1994), who administered the ordinarily unspeeded Raven with a strict time limit (20 min) reported a substantially higher correlation with WM measures ($r = .55$) than did Jurden. Such results suggest that the overall relationship between WM and measures of intellectual ability may be substantially lower than that suggested by Kyllonen and Christal (1990)—and may contradict an assertion that WM and g represent the same underlying factor or highly correlated factors.

Daneman and Merikle (1996)

Daneman and Merikle (1996) conducted a meta-analysis of the correlations between memory measures (both STM and WM) and

language comprehension measures. They reported weighted average correlations among 77 independent samples. Although they excluded "extremely unskilled readers" (Daneman & Merikle, 1996, p. 425), several of the correlations included in the meta-analysis were from young children (e.g., Grade 3) and old adults. They presented separate estimated correlations for verbal and math (or numerical) memory measures, and for global tests of verbal comprehension and vocabulary and specific measures (including "making inferences," "detecting ambiguity," and "following verbal directions"). Although the global measures are clearly identifiable as markers for Gc , the specific measures are more difficult to classify, as they appear to represent a mixture of Gc and other abilities. Verbal WM and math WM tests correlated .41 and .30 with global comprehension, respectively. Simple verbal span and math span measures correlated .28 and .14 with global comprehension, respectively. For the specific comprehension measures, verbal WM and math WM correlated .52 and .48, respectively, while simple verbal span and simple math span correlated .40 and .30, respectively. Daneman and Merikle (1996) concluded that

although the results of the meta-analysis support Daneman and Carpenter's (1980) claim that measures of the combined processing and storage resources of working memory are better predictors of comprehension alone, they do not support Daneman and Carpenter's (1980) finding that verbal-storage-alone measures are not significant predictors of language comprehension. Indeed, the results of the meta-analysis show a rather respectable correlation between verbal-storage-alone measures and specific tests of comprehension. (p. 432)

Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000)

One investigation that has provided important insights into the relations between WM and intellectual abilities was reported by Oberauer et al. (2000). These investigators collected data on 23 WM tests that were created within a taxonomic framework. Specifically, the WM tests represented a sampling of stimulus content (i.e., verbal, numerical, and spatial-figural) and the underlying functions specified by current theories of WM processes (storage and transformation, supervision, and coordination). These tests were administered to a sample of 128 participants, along with a battery of 45 ability tests, which were selected from a taxonomic framework similar to that of Guilford (1967). They derived three WM factors from the 23 tests—a Verbal/Numerical WM factor (including simultaneous storage and transformation and coordination functions), a Spatial-Figural WM factor (storage and transformation and coordination), and a third factor that contained WM tests that involved supervisory functions but that were also highly speeded.

In linking the WM factors with the intellectual ability scales, Oberauer et al. (2000) found correlations between Verbal/Numerical WM factor scores and a numerical test composite of .46 and correlations between Verbal/Numerical factor scores and a reasoning test composite of .42. The Spatial-Figural WM factor scores correlated highest with the reasoning test composite (.56), the spatial test composite (.52), and the numerical test composite (.48). The Supervisory/Speed WM factor correlated at .61 with a speed test composite from the intellectual ability test battery. However, all three WM factors correlated significantly with the speed test composite. Oberauer et al.'s (2000) results further suggest that the relationship between measures of WM and intelligence may be

more complex than previously considered. That is, WM factors may have a differentiated pattern of correlations with factors of reasoning, content abilities (such as verbal, numerical, and spatial), and perceptual speed (PS). Moreover, Oberauer et al. (2000) noted that their data contradicted the assertion that simple span tests (i.e., those without the transformation component of many WM tests) represent constructs different from those measured by WM tests, though it is important to note that a single spatial STM test was administered in their study. Others have suggested that there are indeed differences between the constructs of WM and simple span tests. For example, Conway et al. (2002) reported a correlation of .82 between latent variables of WM and STM, and Engle et al. (1999) reported a correlation of .68 between these latent variables. Conway et al. (2002) asserted that WM and STM are separable because a two-factor structural equation model (WM and STM) fit the data better than a single-factor solution (WM and STM combined).

Ackerman, Beier, and Boyle (2002)

The study by Ackerman et al. (2002) further examined the relations between WM and intellectual abilities and provided an additional perspective on the specific relations between WM and PS abilities. Ackerman et al. (2002) administered 36 ability tests (including 13 PS tests representing four separable PS factors) together with 7 WM tests to a sample of 135 adults. They found that a single underlying WM factor correlated substantially with a *g* factor ($r = .70$), but the WM factor also correlated highly with a general PS factor ($r = .55$). In addition, they examined differential relations between WM, performance on the Raven test, and a *g* composite that did not include Raven test performance. The Raven test correlated .58 with a broad *g* composite, while the WM composite correlated .47 with the *g* composite. In contrast, the Raven correlated only .25 with a PS composite, whereas the WM composite was significantly more highly correlated with the PS composite ($r = .47$).

Summary and Hypotheses

Together these various studies suggest that the relationship between WM and intelligence is complex. One extreme hypothesis for this investigation is that the relationship between WM and intelligence (*g*) is unity (e.g., $\rho = 1.0$, reflecting an “isomorphic” association; Engle, 2002). A statistical representation of this hypothesis would be that the confidence intervals (CIs) for estimated true-score correlations between WM measures and measures of intelligence include 1.0. Given the extensive research showing that there are indeed significant correlations between WM measures and ability measures, rejection of a null hypothesis of a zero correlation between WM and ability is a virtual certainty, and is thus not very informative. Pending rejection of the hypothesis that the correlation between WM and intelligence is 1.0, it is useful to describe the best estimate of the relation between these constructs within a mean and CI framework. Such data will serve to provide an index of the estimated true-score relationship between these constructs.

Because there is some controversy regarding whether WM should be most highly associated with *g*, with *Gf*, or more specifically with performance on Raven’s Progressive Matrices Tests,⁵

we explore separate meta-analytic estimates for different abilities (and for the Raven specifically). In addition, consistent with results from Oberauer et al. (2000), we evaluate whether content overlap (e.g., verbal WM tests paired with verbal abilities) results in higher correlations than do cross-content pairings (e.g., verbal WM tests paired with math or spatial abilities). The hypothesis underlying these data are that higher correlations will be found for WM–ability pairings with overlapping content than with nonoverlapping content. Finally, given the Kyllonen and Christal (1990) results and the Ackerman et al. (2002) results that suggested a substantial relationship between WM test performance and speed of processing, we explore the degree of relationship between WM measures and PS abilities and between WM measures and tests of highly speeded elementary cognitive tests (ECTs; see Carroll, 1980, and Kyllonen, 1985, for extensive treatments of speeded narrow information-processing-based ability measures). For comparison purposes, we also report a meta-analysis of STM measures as correlated with ability measures.

Meta-Analysis of WM and Intellectual Abilities

Method

Literature search. Studies for possible inclusion in the meta-analysis of WM and intelligence were initially identified through a series of searches in the PsycINFO (1872–2002) database. Pairwise combinations of 22 intellectual ability search terms and 10 WM-related search terms were used. The intellectual ability search terms included the broad terms of abilities and intelligence, content and broad ability terms (*verbal, spatial, numerical, reasoning, and perceptual speed*), test names (e.g., *SAT, WAIS, Differential Aptitude Test, Nelson-Denny*), and prominent names (e.g., *Wechsler, Raven*). The WM search terms were general (*working memory*), test-specific (e.g., *operation span, computation span, listening span, reading span, ABCD order*), and names of prominent researchers in the field (e.g., *Baddeley, Salthouse, Engle*). A total of 8,698 abstracts were retrieved from the original search. A first-pass review excluded publications that involved nonhuman participants, clinical populations (e.g., brain injury, Alzheimer’s patients), children under age 13 or adults over the age of 70. At the conclusion of the first pass, 1,911 abstracts were retained for more extensive review. A second-pass review entailed a careful review of the abstracts to determine whether the samples met the exclusion criteria and whether correlations between WM and intellectual ability measures were reported. Each of the items identified as possible inclusion data sets were examined in detail. At this point, we also supplemented the list of possible inclusion publications with items from reference lists of articles and other sources on the topic of WM and intelligence, along with previous meta-analyses on related topics. We also obtained several full text doctoral dissertations from UMI (previously known as University Microfilms International, now known as the subdivision Dissertation Express). The final set of items included 57 publications (including 4 doctoral dissertations, and 1 article suggested by a reviewer that was not listed in PsycINFO, Conway & Engle, 1996).

Identification of usable correlations. After identification of studies that involved joint assessment of WM and intellectual ability measures, we attempted to obtain correlations from the published record. In several cases, raw Pearson product–moment correlations were not reported (e.g., when only factor analytic results were provided). In these cases, we

⁵ This point reflects the analysis by Babcock (1994; Babcock & Laguna, 1996) that points to specific overlap between WM and two aspects of the Raven test (rule application and the ability to manipulate geometric figures).

contacted the authors of the publications to obtain either the raw data or the correlations relating WM measures to intellectual ability measures. All of the authors we contacted for this information provided either the raw data or raw correlations. With these items, the meta-analysis was based on 86 independent samples and 9,778 participants.

Classification of intellectual ability tests. Rather than starting from an ability taxonomy (such as that by Carroll, 1993, or P. E. Vernon, 1950), for the purposes of this meta-analysis, we limited consideration to those major ability groups that would allow for a reasonable population of a matrix of ability by WM tests. Although this may be a less-than-ideal solution, examination of a list of 100 or more abilities would result in a sparse matrix and little capability to synthesize the research. Because common content (e.g., verbal, spatial, numerical) considerations were noted in the analysis of immediate memory and intelligence relations, we initially classified ability trait measures by content. Additional categories of PS were included because of previous research (e.g., Ackerman et al., 2002; Kyllonen & Christal, 1990) that identified this ability factor as a major correlate of individual differences in WM. Similarly, we segregated reasoning ability measures from other ability measures, because of the initially strong associations reported by Kyllonen and Christal (1990) for WM measures with Reasoning. We included a few additional factors (General Intelligence, measures of ECTs, and Knowledge) to address correlations between WM and *g*, to address speeded abilities, and to provide comparative data on *Gc*-related abilities, respectively. Finally, given the relative oversampling of the Raven's Standard Progressive Matrices Test and the Raven's Advanced Progressive Matrices Test in the WM–intelligence literature and the potential controversy regarding the identification of the Raven as the sine qua non of *g*, we created a separate category for the Raven tests (which would otherwise be categorized as spatial reasoning or general reasoning).

On the basis of the above framework, each unique test name was extracted from the selected publications and put on an index card, along with a short description of the test (either from the publication, or from other sources, such as test manuals or the Buros *Mental Measurement Yearbook* series). A total of 167 unique ability tests were identified. We each coded the ability tests into 1 of 12 different categories, including 5 reasoning categories (general reasoning, verbal, spatial, numerical, or Raven-specific) 3 nonreasoning content categories (verbal, spatial, numerical), speed (PS), or 3 other categories (ECTs, knowledge, and general intelligence).⁶ Tests without unanimous agreement for appropriate categorization were classified through discussion and occasionally with further reference work. The measures and their classifications are provided in the Appendix.

Classification of WM tests. The classification of WM tests was determined on the basis of surface-level features of test content. The first categorization was on the content of the test items and processes that were scored (in most cases, this is the secondary task) to indicate WM (verbal, numerical, spatial). A second categorization was made if the WM test involved simultaneous processing of different contents (e.g., a verbal primary task with a numerical secondary task).⁷ Thus, a classification of “verbal with numerical” entails recall of verbal material, with the primary task of numerical processes (such as arithmetic computation, or verification of arithmetic equations), whereas a “numerical with verbal” test involves recall of numerical items while performing a verbal primary task. WM tests with only a single classification (e.g., “verbal”) involve simultaneous processing of two tasks with the same (verbal) content. An additional category of “WM composite” was used for reported correlations based only on test performance aggregated across several WM tests. The measures and their classifications are provided in the Appendix.

Correction for attenuation due to unreliability of measures. The standard correction for unreliability of both sets of measures was applied to estimate the true-score correlation of WM and ability measures. Reliability estimates came from a variety of sources. We used the authors' reported reliability unless it was not provided, in which case we substituted a published test–retest reliability for the measure. When test–retest reliability

estimates were not available, however, estimates of internal consistency (e.g., Cronbach's α) from the same sources were used. In some cases, when these sources could not be located, or when no reliability estimates were reported, an attempt was made to locate reliability estimates reported in other research studies using these measures. Using these methods, reliability estimates were available for 60 of 88 WM (68%) variables and 124 of 167 (74%) ability measures. For those cases in which reliability estimates remained unavailable, the mean value of the reliability estimates that had been obtained for similar tests were used (done separately for WM and ability).

Aggregation of within-sample effect sizes. Several studies used multiple measures of WM within a single category or multiple ability tests within a single category. Rather than allowing a single sample to contribute more than one correlation to a specific cell of the WM \times Ability matrix of correlations—which would ignore the nonindependence of the estimated correlations—or simply choosing one of the correlations at random (thereby losing the information), we decided to use an aggregated estimate by computing the mean correlation. The mean correlation was obtained by first transforming the correlation to Fisher's *z*, computing a mean, and then transforming the mean back to an average correlation coefficient. The total number of correlations obtained from the literature was 1,103. After aggregation within categories, a total of 411 correlations remained for the meta-analysis computations.

Correlational analysis. We computed meta-analytic effect sizes using the procedure described by Hunter and Schmidt (1990). The effect size for each WM \times Ability combination was computed by weighting each correlation coefficient by sample size (we report both uncorrected correlations and correlations corrected for attenuation on both the WM and ability measures). The weighted, corrected correlations are reported as an estimate of the true-score correlation between the underlying variables, that is, an estimate of the correlation if the respective tests had perfect reliability. A further aggregation was performed across both rows and columns of the WM \times Ability pair matrix to obtain estimates of overall correlations by overarching category of ability constructs or WM test types. To perform this aggregation, we again computed average correlations when redundant samples of participants were included to avoid including nonindependent correlations. The summary aggregations are thus based on independent samples of participants.

CIs. We calculated 95% CIs for $\hat{\rho}$ using formulas provided by Hedges and Olkin (1985, p. 227).

Results

The results of the meta-analyses are shown in Table 1. The table provides a 10 (WM test type) \times 12 (intellectual ability test type) matrix of weighted mean correlations (both raw and corrected for attenuation), 95% CIs for the corrected mean correlations, and a total sample size and number of correlations (collapsed within

⁶ Tests included in the PS category were classified according to the four-factor taxonomy presented by Ackerman and Cianciolo (2000). ECTs were narrower tasks (i.e., more homogeneous in stimulus properties) with single items (i.e., only one stimulus item on a display at a time), such as category identification and meaning identity.

⁷ The designation of “primary” task and “secondary” task in WM assessments is somewhat arbitrary. However, our sense is that the constraints on the tasks are such that the typical non-WM component is the primary task, and the WM (scored) component is the secondary task. This is because a failure to perform the non-WM component (e.g., reading the sentences aloud on the sentence span task) invalidates the WM component. See, for example, the procedure outlined by Daneman and Carpenter (1980), which describes how the participants were “required to read the sentences aloud” (p. 453).

Table 1
Estimated Population Correlations Between Working Memory (WM) and Ability Variables With 95% Confidence Intervals and Heterogeneity Statistics

WM measure	Reasoning								Perceptual speed	ECT	Knowledge	g	Average ability
	Verbal	Numerical	Spatial	Verbal	Numerical	Spatial	General	Raven					
Verbal	.434 (.365)	.440 (.380)	.287 (.242)	.335 (.286)	.433 (.363)	.480 (.369)	—	.312 (.249)	.237 (.203)	.600 (.539)	.409 (.357)	.350 (.263)	.390 (.329)
	41 (3,698)	15 (2,220)	8 (870)	5 (668)	3 (285)	4 (610)	—	6 (708)	9 (1,926)	2 (1,019)	4 (1,154)	7 (635)	52 (5,285)
	.41 to .46	.40 to .47	.22 to .35	.26 to .40	.33 to .52	.42 to .54	—	.24 to .38	.19 to .28	.56 to .64	.36 to .46	.28 to .42	.37 to .41
Verbal with numerical	.5466	.4041*	.1672	3.46	2.34	15.45*	—	4.29	28.96*	26.77*	1.96	7.10	—
	.392 (.317)	.458 (.373)	.461 (.394)	.345 (.290)	.510 (.411)	.452 (.340)	.376 (.308)	.356 (.274)	.321 (.263)	.488 (.402)	.244 (.196)	.407 (.291)	.406 (.327)
	18 (2,948)	10 (1,819)	1 (562)	2 (662)	2 (961)	1 (392)	1 (392)	2 (253)	7 (2,110)	3 (1,087)	6 (1,990)	6 (549)	20 (3,128)
Verbal with spatial	.36 to .42	.42 to .49	.39 to .52	.27 to .41	.46 to .56	.37 to .53	.29 to .46	.24 to .46	.28 to .36	.44 to .53	.20 to .28	.33 to .48	.38 to .44
	81.20*	34.50*	—	0.00	0.87	—	—	1.42	14.71	5.35	52.28*	11.33	—
	.404 (.333)	.153, ns (.130)	.029, ns (.025)	—	—	—	—	—	—	—	—	—	.390 (.322)
Numerical	4 (222)	1 (30)	1 (30)	—	—	—	—	—	—	—	—	—	4 (222)
	.28 to .51	—	—	—	—	—	—	—	—	—	—	—	.27 to .50
	9.05	—	—	—	—	—	—	—	—	—	—	—	—
Numerical with verbal	.267 (.217)	.414 (.338)	.321 (.266)	.370 (.304)	.471 (.372)	.563 (.415)	—	.393 (.323)	.294 (.242)	.596 (.481)	.155 (.123)	.406 (.303)	.367 (.298)
	18 (3,426)	11 (2,957)	7 (906)	7 (1,414)	4 (1,194)	4 (680)	—	6 (708)	11 (2,993)	3 (1,418)	7 (2,388)	5 (577)	31 (5,298)
	.24 to .30	.38 to .44	.26 to .38	.32 to .41	.42 to .51	.51 to .61	—	.33 to .45	.26 to .33	.56 to .63	.12 to .19	.33 to .47	.34 to .39
Numerical with spatial	42.04*	29.30*	10.01	28.86*	2.59	2.35	—	3.98	25.25*	51.90*	21.65*	13.01*	—
	.297 (.247)	.589 (.492)	.512 (.430)	.468 (.407)	.508 (.421)	.635 (.479)	.411 (.337)	.439 (.371)	.346 (.289)	.602 (.503)	.239 (.196)	.334 (.248)	.423 (.351)
	9 (2,336)	6 (2,213)	2 (168)	1 (399)	1 (399)	2 (233)	1 (392)	2 (97)	5 (1,970)	4 (1,810)	5 (1,970)	1 (160)	10 (2,396)
Spatial	.26 to .33	.56 to .62	.41 to .55	.33 to .59	.37 to .62	.55 to .71	.33 to .49	.26 to .59	.31 to .38	.57 to .63	.20 to .28	.19 to .46	.39 to .46
	43.48*	40.22*	11.28*	—	—	0.13	—	3.55	9.78	32.58*	28.47*	—	—
	9 (1,307)	7 (1,024)	2 (168)	6 (940)	4 (853)	4 (660)	—	6 (940)	4 (853)	1 (296)	5 (904)	4 (689)	11 (1,698)
Spatial with verbal	.33 to .42	.54 to .62	.39 to .62	.55 to .62	.37 to .62	.55 to .71	.33 to .49	.55 to .63	.36 to .47	.41 to .58	.32 to .43	.56 to .66	.44 to .51
	11.50	2.96	0.10	—	—	—	—	27.83*	13.45*	—	1.69	7.05	—
	.328 (.244)	.461 (.360)	.487 (.373)	.468 (.407)	.508 (.421)	.635 (.479)	.411 (.337)	.439 (.371)	.346 (.289)	.602 (.503)	.239 (.196)	.334 (.248)	.423 (.351)
Spatial with numerical	6 (719)	4 (589)	4 (460)	1 (139)	1 (139)	2 (233)	1 (392)	2 (97)	5 (1,970)	4 (1,810)	5 (1,970)	1 (160)	10 (2,396)
	.25 to .38	.39 to .52	.41 to .55	.33 to .59	.37 to .62	.55 to .71	.33 to .49	.26 to .59	.31 to .38	.57 to .63	.20 to .28	.19 to .46	.39 to .46
	6.43	15.22*	11.28*	—	—	0.13	—	3.55	9.78	32.58*	28.47*	—	—
WM composite	.455 (.388)	.653 (.562)	.487 (.421)	.468 (.407)	.508 (.421)	.635 (.479)	.411 (.337)	.439 (.371)	.346 (.289)	.602 (.503)	.239 (.196)	.334 (.248)	.423 (.351)
	3 (380)	3 (380)	3 (251)	3 (380)	3 (380)	3 (380)	3 (380)	3 (380)	3 (380)	3 (380)	3 (380)	3 (380)	3 (380)
	.37 to .53	.59 to .71	.39 to .58	.33 to .59	.37 to .62	.55 to .71	.33 to .49	.26 to .59	.31 to .38	.57 to .63	.20 to .28	.19 to .46	.39 to .46
Average WM	17.40*	0.48	8.97*	—	—	—	—	0.03	4.16	—	—	—	—
	.298 (.204)	.368 (.261)	.483 (.360)	.468 (.407)	.508 (.421)	.635 (.479)	.411 (.337)	.439 (.371)	.346 (.289)	.602 (.503)	.239 (.196)	.334 (.248)	.423 (.351)
	1 (135)	1 (135)	2 (185)	1 (135)	1 (135)	2 (185)	1 (135)	2 (185)	1 (135)	1 (296)	1 (296)	1 (94)	9 (1,073)
Average WM	.14 to .44	.21 to .50	.36 to .59	.36 to .59	.37 to .62	.55 to .71	.33 to .49	.26 to .59	.31 to .38	.57 to .63	.20 to .28	.19 to .46	.39 to .46
	—	—	0.06	—	—	—	—	—	—	—	—	—	—
	.504 (.410)	.427 (.365)	.351 (.300)	.473 (.397)	.641 (.550)	.374 (.310)	—	.374 (.310)	.287 (.242)	.419 (.358)	.618 (.519)	.614 (.450)	.476 (.388)
Average WM	1 (96)	2 (160)	2 (160)	4 (240)	1 (188)	4 (240)	4 (240)	1 (188)	3 (517)	1 (296)	4 (240)	4 (689)	11 (1,698)
	.34 to .64	.29 to .55	.19 to .47	.37 to .57	.55 to .72	.37 to .57	.37 to .57	.55 to .72	.24 to .36	.32 to .51	.54 to .68	.56 to .66	.44 to .51
	—	0.83	1.20	3.91	—	—	—	—	1.81	—	—	7.05	—
Average WM	.326 (.269)	.453 (.375)	.359 (.300)	.334 (.279)	.492 (.395)	.527 (.396)	.428 (.355)	.495 (.394)	.268 (.237)	.566 (.479)	.252 (.205)	.479 (.364)	.397 (.324)
	60 (6,640)	30 (4,787)	17 (2,224)	8 (1,514)	5 (1,246)	6 (1,166)	6 (696)	13 (1,752)	18 (4,227)	4 (1,810)	16 (3,628)	15 (1,467)	86 (9,778)
	.30 to .35	.43 to .48	.32 to .40	.29 to .38	.45 to .53	.48 to .57	.36 to .49	.46 to .53	.26 to .32	.53 to .60	.22 to .28	.44 to .52	.38 to .41

Note. For each WM measure, Line 1 contains the estimated population correlation ($\hat{\rho}$; uncorrected correlation is in parentheses). Unless otherwise noted, all correlations are significant at the .05 level ($ns =$ not significant). The first value in Line 2 of each WM measure is the number of correlations, and the total sample size is in parentheses. Line 3 contains the 95% confidence interval of the estimated population correlation, and Line 4 contains the heterogeneity statistic Q . A dash indicates that Q could not be computed with only one sample. ECT = elementary cognitive task.

* $p < .01$.

sample) upon which each mean correlation was derived, and an indication of the heterogeneity of the estimated correlations (Q ; see Hunter & Schmidt, 1990, for a discussion of this statistic). In addition, unique samples across the rows (WM test type) and columns (ability test type) were used to estimate mean correlations for each WM and ability category. An estimate of the global meta-analytically derived correlation between all 10 types of WM tests and all 12 ability test types yielded the following values: mean weighted raw correlation = .324; mean weighted correlation, corrected for attenuation = .397, 95% CI = .38 to .41. (Total sample = 9,778 participants; total number of independent samples = 86.)

Although many of the tabled estimated true-score correlations have overlapping 95% CIs, the WM tests with the strongest correlations with abilities were two types of tests that involved simultaneous processing of two different content materials, namely spatial with verbal ($\hat{\rho} = .510$) and numerical with spatial ($\hat{\rho} = .476$). The ability measures with the highest communality with WM measures were ECTs ($\hat{\rho} = .566$) and spatial reasoning ($\hat{\rho} = .527$). Relatively few different tests entered into the ECT correlations, and it may be that these were selected by investigators specifically in an attempt to demonstrate substantial overlap with WM. These results suggest a potentially promising direction for establishing the underlying components of overlap between ability measures and WM measures, but at this point it is a tenuous premise on the basis of a small amount of data.

Hypothesis 1a: WM and g are isomorphic.

At the level of g , meta-analytically derived correlations (even after correcting for unreliability of the respective tests of WM and g), do not approach an isomorphic relationship (i.e., estimated true-score correlations approaching 1.0). Instead, the maximum correlation, corrected for unreliability, between any individual pairing of WM test category (numerical with spatial) with measures of g was .614 (95% CI = .56 to .66). Even the most favorable pairing of WM and g test types yields a refutation of this hypothesis. Thus, we conclude that WM and g are not isomorphic.

Hypothesis 1b: WM and Gf are isomorphic.

Because typical tests selected as markers for Gf tend to be nonverbal (see, e.g., Horn, 1989), we examined the nonverbal (numerical and spatial reasoning) ability test categories to evaluate this hypothesis. Given the centrality some investigators place on the Raven's Progressive Matrices Test as a marker for Gf , we examined this category as well. Review of these columns of meta-analytically derived correlations similarly refutes the notion that WM and Gf are isomorphic. The most optimal pairing of WM tests (numerical with verbal) and reasoning- Gf (numerical reasoning) tests yielded an estimated true-score correlation of .634 (based on a single study), with a 95% CI of .57 to .69. Thus, we conclude that WM and Gf are not isomorphic.

Hypothesis 2: Content overlap between WM and ability measures results in significantly higher correlations than pairings with different contents (see, e.g., Shah & Miyake, 1996).

For this hypothesis, we separately examined WM tests with unique content (i.e., verbal only, numerical only, spatial only) paired with

abilities of the same content and then compared those correlations with pairings involving other content (e.g., verbal WM with numerical and spatial abilities). As can be seen in Table 1, average correlations for the spatial WM and then Numerical WM tasks with the respective content ability measures were generally, but not uniformly, significantly higher than for nonoverlapping content ability measures. For the verbal WM tasks, both verbal and numerical ability measures showed essentially equivalent correlations. Thus, although there appeared to be some increase in commonality between overlapping content of WM and ability tests, Hypothesis 2 was not decisively supported.

Hypothesis 3: We hypothesized that the relationship between WM measures and speed abilities (PS and ECTs) would be as large as the correlation between WM and other intellectual abilities.

For this comparison, we compared the mean correlation and 95% CI across the WM tests for PS ability and for ECT separately against the mean correlations for spatial, verbal, and numerical content abilities. The estimated true-score correlation between PS and WM was .268 (95% CI = .26 to .32), and the estimated true-score correlation between ECT and WM was .566 (95% CI = .53 to .60). In contrast, the average estimated true-score correlations for spatial, verbal, and numerical abilities with WM were .359, .326, and .453, respectively. The average PS ability correlations with WM did not exceed those of the content abilities, but the average ECT test correlations did significantly exceed the correlations for all of the content abilities and even exceeded the average true-score correlation between the g measures and WM measures (.479, 95% CI = .44 to .52). The estimated true-score correlation for WM and ECT significantly exceeded that of all other abilities, except for spatial reasoning, for which there was a nonsignificant advantage to the ECT measures.

An illuminating contrast is provided by comparing the overall correlations between WM measures and Raven's Progressive Matrices Test ($\hat{\rho} = .495$) and the overall correlations between the WM measures and ECT measures ($\hat{\rho} = .566$). These two estimates of common variance are significantly different from one another (given nonoverlapping 95% CIs), indicating that WM measures are significantly more highly correlated with narrow measures of information-processing speed and accuracy than they are with a measure most closely identified with nonverbal reasoning or Gf .

Discussion

At the level of individual variables and composite variables (which have been found to provide as good an estimate of latent factors as optimally weighted factor scores, especially in samples smaller than 300 participants; see R. L. Thorndike, 1986), we can confidently reject the assertion that WM and g are isomorphic to one another. With respect to the subsidiary hypotheses, there were mixed results indicating that content overlap between WM and ability measures results in greater commonality. However, even with a limited database of ECT measures, it appears that narrow speeded information-processing tests have, on average, higher correlations with WM than do broader content and PS abilities.

Moderator Analysis

The presence of numerous significant heterogeneity (Q) statistics in the meta-analysis suggests that there may be salient moderators of the WM–intelligence relations. Although the number of independent samples in many categories is relatively small, and thus limits the power of such analyses, we considered two moderators that have been suggested as potentially important factors,⁸ namely, the ages of the participant samples under consideration and the speed requirements in the WM test administration. Each of these moderator analyses are discussed in turn below.

Age as a Moderator of WM–Intelligence Relations

To test the moderating effect of age, three age classifications were used. Young samples were defined as up to 30 years of age (under 30 group), participants older than 30 were classified as older samples (over 30 group), and a sample was coded as “mixed” if a range of ages spanning 30 was jointly presented. Because of the lower cell frequencies for this moderator analysis (in comparison to the main meta-analysis), WM categories were combined to yield only verbal, numerical, and spatial content, but the same ability categories were retained. The final age moderator analysis included 410 raw correlations for the under 30 group, 115 correlations for the mixed group, and 13 correlations for the over 30 group. For other aspects of this analysis, we followed the procedures in the original meta-analysis. The results of the moderator analysis are shown in Table 2. Note that the power to detect differences in this analysis are limited by the comparatively small number of correlations for the over 30 group (2.4% of the total number of correlations).

There were five instances of nonoverlapping 95% CIs in the 33 possible pairwise comparisons. Four of these instances represented higher correlations for the mixed or over 30 groups in comparison with the under 30 groups, and the other was a lower correlation for the mixed age group compared with the under 30 group (verbal WM with spatial ability). Correlations between WM and PS were responsible for two of these different cell contrasts. This finding is consistent with the extant literature showing substantial decline in PS abilities with increasing adult age (see, e.g., Salthouse, 1994), indicating that in mixed age groups, there may be increased correlations between WM and PS as a function of the concomitant changes in both PS and WM with age. At the average correlation level (“Average ability” column in Table 2), there is no consistent indication that WM correlations are higher for mixed age or over 30 participants in comparison to under 30 participants.

Pacing of WM Tests as a Moderator of WM–Intelligence Relations

To evaluate the effects of speed in test administration, the WM tests were coded as participant paced or experimenter paced. Participant-paced studies were defined as those in which advancing to the next stimulus was under the participant’s volitional control. For tests given this coding, presentation rates for stimuli were neither fixed nor time limited. Most tests of this sort were administered individually, such as the typical procedure for reading span (the participant reads aloud a sentence from an index card or computer screen, and the next stimulus is presented upon

completion). Studies were classified as experimenter paced if either stimulus presentation or response times were fixed or time limited. Many, but not all, computerized tests were included in this category. Several study–test combinations could not be classified with reasonable confidence. As a result, data from five samples were partially or completely excluded from the speed moderation analysis. The WM study–test combination and raw correlation results were distributed as follows: experimenter paced, 327 correlations; participant paced, 171 correlations. Like the age moderator analysis, WM categories were reduced to main content of verbal, numerical, and spatial, and the same ability categories were retained. The results of this moderator analysis are provided in Table 3.

There were 10 instances of nonoverlapping 95% CIs in the 23 possible pairwise comparisons, 8 of which indicated lower correlations between WM and intelligence measures for the experimenter-paced conditions, and 2 of which showed higher correlations for the experimenter-paced conditions, though in 5 of these comparisons there were only one or two independent samples underlying the cells. Greater confidence in the pattern of results is found at the average ability level (see the far right column of Table 3), which shows nonoverlapping CIs for Verbal WM and Numerical WM cells (though not for the Spatial cells). In the case of verbal WM and numerical WM, larger WM–intelligence correlations were found for participant-paced conditions than for experimenter-paced conditions, though in both cases the differences were not very large in terms of magnitude of the differences (e.g., .433 vs. .368 for verbal WM, .449 vs. .384 for numerical WM). From this analysis, it would be fair to conclude that participant-paced WM tests tend to share slightly more variance with intelligence measures. However, it is important to point out that given the small number of cell entries, it was not possible to perform a parallel analysis for speeded versus power intellectual ability tests. Such an analysis could support or contradict the notion that greater correlations are found when WM measures and intellectual ability tests share method variance. Note that in the meta-analysis sample of studies, only about 5% of the tests were power tests, the rest being speeded tests, meaning that additional empirical studies would be required to evaluate this issue.

WM Latent-Variable Analysis

There has been some speculation that the isomorphism between WM and g may occur at the level of latent common factors (e.g., Engle et al., 1999). Evaluation of this proposition is not readily addressed within a meta-analytic framework, though there have been recent attempts to apply structural equation modeling (SEM) to meta-analytic correlation matrices (see, e.g., Viswesvaran & Ones, 1995). Because several assumptions underlying SEM are violated under these circumstances, such as unequal numbers of observations underlying the correlations, the results with this kind of analysis should be taken with some skepticism (see, e.g., M. J. Burke & Landis, 2003, for a discussion of threats to validity of such analyses). However, given the importance of this proposal, it is useful to consider what such analyses might reveal about the underlying relationships among WM and ability variables. To

⁸ We thank three anonymous reviewers for these suggestions.

Table 2
Age Moderator Analysis: Estimated Population Correlations Between Working Memory (WM) and Ability Variables With 95% Confidence Intervals

WM measure	Reasoning										Average ability		
	Verbal	Numerical	Spatial	Verbal	Numerical	Spatial	General	Raven	Perceptual speed	ECT		Knowledge	<i>g</i>
Verbal													
Under 30	.394 (.329) 41 (4,848)	.430 (.366) 20 (3,189)	.411 (.348) 5 (875)	.364 (.313) 2 (701)	.492 (.400) 5 (1,246)	.421 (.324) 3 (625)	.376 (.308) 1 (392)	.312 (.246) 4 (522)	.248 (.210) 10 (2,935)	.541 (.472) 4 (1,810)	.271 (.230) 8 (2,667)	.387 (.288) 10 (804)	.396 (.330) 46 (5,353)
Mixed	.37 to .42 .411 (.337)	.40 to .46 .609 (.514)	.35 to .46 .264 (.225)	.30 to .43 .311 (.259)	.45 to .53 1 (128)	.35 to .48 2 (377)	.29 to .46 .549 (.413)	.23 to .39 1 (52)	.21 to .28 3 (583)	.51 to .57 1 (181)	.24 to .31 1 (181)	.33 to .45 1 (127)	.37 to .42 .365 (.301)
Over 30	.7 (700) .35 to .47 .523 (.435)	1 (128) .49 to .71	4 (557) .18 to .34	4 (529) .23 to .39	1 (134) .23 to .52	2 (377) .47 to .62	1 (392) .384 (.330)	1 (52) -04 to .48 .384 (.330)	3 (583) 1 (102)	1 (52) 1 (102)	1 (181) .26 to .51	1 (127) .06 to .39	12 (1,606) .32 to .41 .379 (.322)
Numerical													
Under 30	.270 (.220) 20 (3,974)	.486 (.397) 15 (3,688)	.378 (.293) 3 (269)	.328 (.270) 2 (701)	.486 (.385) 4 (1,194)	.559 (.416) 2 (531)	.411 (.337) 1 (392)	.517 (.403) 9 (1,330)	.298 (.244) 11 (3,257)	.597 (.492) 4 (1,810)	.203 (.162) 10 (3,043)	.514 (.378) 7 (953)	.383 (.312) 25 (4,759)
Mixed	.24 to .30 .356 (.289)	.46 to .51 .519 (.482)	.30 to .48 .322 (.272)	.26 to .39 .411 (.388)	.44 to .53 1 (128)	.50 to .61 3 (541)	.33 to .49 .563 (.418)	.48 to .56 .501 (.430)	.27 to .33 436 (.357)	.57 to .63 1 (52)	.17 to .24 2 (345)	.47 to .56 1 (80)	.36 to .41 .399 (.327)
Over 30	.5 (552) .28 to .43 .523 (.420)	1 (128) .38 to .64	5 (677) .25 to .39	5 (713) .35 to .47	1 (128) .38 to .64	3 (541) .50 to .62	1 (392) .508 (.421)	.26 to .68 .454 (.390)	3 (583) .37 to .50	1 (52) .237 (.210)	2 (345) .24 to .43	1 (80) .56 to .79	12 (1,695) .36 to .44 .418 (.350)
Spatial													
Under 30	.324 (.252) 6 (645)	.508 (.410) 4 (515)	.478 (.380) 5 (436)	.468 (.407) 1 (139)	.508 (.421) 1 (139)	.635 (.479) 2 (233)	.463 (.328) 2 (184)	.360 (.296) 1 (296)	.315 (.252) 4 (652)	.478 (.387) 1 (296)	.592 (.470) 1 (94)	.424 (.340) 10 (1,050)	.37 to .47 .500 (.398)
Mixed	.25 to .39 .366 (.289)	.44 to .57 .554 (.438)	.40 to .55 .568 (.460)	.33 to .59	.37 to .62	.55 to .71	.34 to .57	.26 to .46	.24 to .38	.38 to .56	.44 to .71	.37 to .47 1 (128)	.37 to .47 1 (128)
Over 30	1 (128) .21 to .51	1 (128) .42 to .66	1 (128) .44 to .68										.36 to .62

Note. For each WM measure, Line 1 contains the estimated population correlation ($\hat{\rho}$); uncorrected correlation is in parentheses). Unless otherwise noted, all correlations are significant at the .05 level. The first value in Line 2 of each WM measure is the number of correlations, and the total sample size is in parentheses. Line 3 contains the 95% confidence interval of the estimated population correlation. No data were found for the Spatial WM—over 30 category. ECT = elementary cognitive task.

Table 3

Pacing (Speed) Moderator Analysis: Estimated Population Correlations Between Working Memory (WM) and Ability Variables With 95% Confidence Intervals

WM measure	Reasoning										ECT	Knowledge	g	Average ability	
	Verbal	Numerical	Spatial	Verbal	Numerical	Spatial	General	Raven	Perceptual speed						
Verbal															
	P	.455 (.385)	.483 (.424)	.433 (.372)	.643 (.522)	.638 (.500)	.108 (.080)	.330 (.246)	.163 (.135)	.594 (.537)	.430 (.375)	.431 (.315)	.433 (.363)		
	E	.30 (2.636)	10 (1.476)	2 (2.48)	1 (100)	1 (52)	1 (94)	2 (253)	3 (1,024)	2 (1,019)	4 (1,251)	8 (560)	33 (2,868)		
Numerical															
	P	.42 to .49	.44 to .52	.33 to .53	.57 to .71	.44 to .76	-.10 to .30	.21 to .44	.10 to .22	.55 to .63	.38 to .47	.36 to .50	.40 to .46		
	E	.357 (.294)	418 (.347)	369 (.313)	342 (.291)	486 (.385)	506 (.387)	376 (.308)	306 (.256)	446 (.370)	188 (.152)	238 (.192)	368 (.305)		
Spatial															
	P	.33 to .39	.38 to .46	.32 to .42	.29 to .39	.44 to .53	.46 to .56	.29 to .46	.48 to .54	.40 to .50	.14 to .23	.13 to .34	.34 to .39		
	E	21 (3,237)	11 (2,053)	8 (1,380)	4 (1,078)	4 (1,194)	4 (908)	1 (392)	3 (403)	3 (1,087)	5 (1,809)	2 (287)	26 (4,021)		
Paced															
	P	.337 (.280)	.620 (.519)	.433 (.372)	.643 (.522)	.634 (.510)	.572 (.425)	.411 (.337)	.429 (.349)	.351 (.290)	.603 (.503)	.459 (.328)	.449 (.370)		
	E	11 (2,425)	7 (2,122)	2 (248)	2 (284)	1 (399)	2 (556)	1 (392)	2 (253)	6 (2,111)	4 (1,810)	7 (2,206)	14 (2,829)		
Elementary															
	P	.30 to .37	.59 to .65	.314 (.257)	.304 (.251)	.471 (.372)	.549 (.409)	.33 to .49	.32 to .53	.31 to .39	.57 to .63	.23 to .31	.42 to .48		
	E	.270 (.218)	446 (.361)	6 (774)	4 (1,078)	4 (1,194)	3 (516)	524 (.411)	293 (.238)	614 (.496)	178 (.141)	553 (.410)	384 (.309)		
Advanced															
	P	.24 to .30	.42 to .48	.24 to .37	.24 to .35	.42 to .51	.49 to .61	.48 to .56	.48 to .56	.26 to .32	.14 to .22	.50 to .60	.36 to .41		
	E	19 (3,772)	13 (3,289)	6 (774)	4 (1,078)	4 (1,194)	3 (516)	9 (1,343)	9 (2,745)	3 (1,418)	7 (2,516)	6 (829)	32 (5,902)		
Expert															
	P	.238 (.186)	.592 (.470)	.690 (.500)	.468 (.407)	.508 (.421)	.598 (.464)	.479 (.345)	.322 (.250)	.358 (.298)	.489 (.400)	.592 (.470)	.352 (.266)		
	E	2 (130)	1 (94)	1 (94)	1 (139)	1 (139)	1 (139)	2 (183)	4 (652)	1 (296)	1 (296)	1 (94)	2 (130)		
Total															
	P	.07 to .40	.45 to .57	.43 to .56	.33 to .59	.37 to .62	.48 to .70	.36 to .59	.25 to .39	.26 to .46	.40 to .57	.44 to .71	.19 to .50		
	E	.354 (.277)	5 (643)	6 (564)	1 (139)	1 (139)	1 (139)	479 (.345)	322 (.250)	358 (.298)	489 (.400)	592 (.470)	437 (.352)		
Total															
	P	.28 to .42	.45 to .57	.43 to .56	.33 to .59	.37 to .62	.48 to .70	.36 to .59	.25 to .39	.26 to .46	.40 to .57	.44 to .71	.19 to .50		
	E	5 (643)	6 (564)	6 (564)	1 (139)	1 (139)	1 (139)	479 (.345)	322 (.250)	358 (.298)	489 (.400)	592 (.470)	437 (.352)		

Note. For each WM measure, Line 1 contains the estimated population correlation ($\hat{\rho}$; uncorrected correlation is in parentheses). Unless otherwise noted, all correlations are significant at the .05 level. The first value in Line 2 of each WM measure is the number of correlations, and the total sample size is in parentheses. Line 3 contains the 95% confidence interval of the estimated population correlation. ECT = elementary cognitive task; P = participant paced; E = experimenter paced.

Table 4

Correlation Matrix of Working Memory (WM) and Ability (AB) Measures Used in the Structural Equation Models

WM/AB measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. WM: verbal														
2. WM: verbal with numerical	.526													
3. WM: numerical	.429	.447												
4. WM: numerical with verbal	.504	.556	.521											
5. WM: numerical with spatial	.419	.594	.420	.569										
6. WM: spatial	.343	.423	.344	.495	.402									
7. AB: verbal	.365	.317	.217	.247	.308	.244								
8. AB: numerical	.380	.373	.338	.492	.469	.360	.361							
9. AB: spatial	.242	.394	.266	.402 ^a	.430	.373	.236	.417						
10. AB: reasoning—verbal	.286	.290	.304	.377 ^a	.406 ^a	.407	.211	.294	.423					
11. AB: reasoning—numerical	.363	.411	.372	.510	.459 ^a	.421	.300	.402	.377	.328				
12. AB: reasoning—spatial	.369	.340	.415	.420	.446 ^a	.479	.248	.447	.377	.453	.430			
13. AB: Raven	.249	.274	.323	.371	.459	.324	.328	.514	.542	.170	.430 ^a	.430 ^a		
14. AB: knowledge	.357	.196	.123	.196	.300	.314	.479	.360	.236 ^a	.251	.211	.299	.460	

Note. N used for the analysis is the harmonic mean of all samples included, 456.

^a Imputed value.

perform these analyses, we reexamined the studies providing data for the meta-analysis to obtain ability–ability correlations and WM–WM correlations, in order to have a complete estimated correlation matrix.

We used the procedure outlined by Viswesvaran and Ones (1995) for conducting SEM with meta-analytic data. The meta-analytic correlation matrix and the intercorrelations between WM and ability measures were used as the basis for the SEM analysis. Measures with more than one third of total cells in the correlation matrix empty, PS, and ECT abilities were not included in the analysis. Thus the original matrix of 10 WM measure types and 12 abilities was reduced to 6 WM measure types and 8 abilities. Eight of the 105 values in the resulting correlation matrix were imputed by examining the patterns of correlations between similar measures (see, e.g., Viswesvaran & Ones, 1995). The correlations used for this analysis can be found in Table 4.

Two sets of analyses were conducted with LISREL (Jöreskog & Sörbom, 2001) to examine the relationships among abilities and WM. First, an exploratory factor analysis was conducted to help define the factor structure of the ability and WM variables. The exploratory factor analysis solution did not reveal a clear simple structure concordant with the extant abilities literature, though it did preserve separable WM and other ability factors. WM and three other first-order factors were identified, though the identification of Spatial, Reasoning, and Verbal abilities was somewhat more tentative, given the underdetermination of these factors with high-quality markers and the confluence of different tests underlying the ability categories. We used the exploratory factor analysis as the basis for a confirmatory factor analysis (CFA) model specification and added a second-order general ability factor (g). In this model, we conceptualized WM as a first-order factor, aligned with Carroll's (1993) hierarchical model of abilities. The resulting model can be seen in Figure 1. Model fit was adequate, $\chi^2(69, N = 456) = 273.12, p < .01$, root mean square error of approximation (RMSEA) = .078, normed fit index (NFI) = .89. As can be seen in the model, the relationship between g and WM is large (.89), but it is on par with the relations between g and spatial ability (.86) and is smaller than the relationship between g and reasoning ability (.96). Also of note in the figure is that the structure coefficient

between g and verbal ability is low relative to the others (.65) and relative to the extant literature on intelligence. This may be a function of the underidentification of the verbal factor relative to the other factors in the model (e.g., the reasoning factor), given the undersampling of Gc abilities in the WM–intelligence literature.

For the second analysis, we examined the assertion that WM and g represent the same construct by conducting two CFAs. The first model assessed whether a single factor (g) could represent both WM and the intellectual ability measures. Structure coefficients for the ability and WM measures were fixed on the basis of their relationships with the first-order factors in the previous analysis.⁹ The resulting model provided an adequate fit, yielding the following results: $\chi^2(85, N = 456) = 337.85, p < .01$, RMSEA = .080, NFI = .87. A second model specified separate factors for g and WM; g was identified by all the intellectual ability measures, and WM was identified by all the WM measures. The fit of this model was better than the first model, $\chi^2(83, N = 456) = 293.00, p < .01$, RMSEA = .072, NFI = .89, but was still not very good in traditional terms of SEM analyses (Bentler & Bonnett, 1980). A chi-square difference test comparing the two models was conducted, yielding a significant result, $\chi^2(2, N = 456) = 44.85, p < .01$. That is, a model with only a single higher order g factor provided a significantly worse fit than a model with separate factors of WM and g . The model with separate WM and g factors is shown in Figure 2.

Although the results from the first analysis might be interpreted as indicating that the WM factor has as high a loading on g as the other ability factors, such a conclusion should be tempered by a comparison with more extensive modeling of the g loadings of various abilities. That is, in Carroll's (1993) analysis, the median loading of Inductive Reasoning on g was .57, for Spatial Relations

⁹ Loadings were fixed in this confirmatory framework to further test the factor loadings identified in the previous analysis and to make the solution easier to interpret and more stable. Fixing the loadings did not significantly change the relationship between g and WM in the model. When loadings were freely estimated, the relationship between g and WM was .47; when loadings were fixed, it was .50.

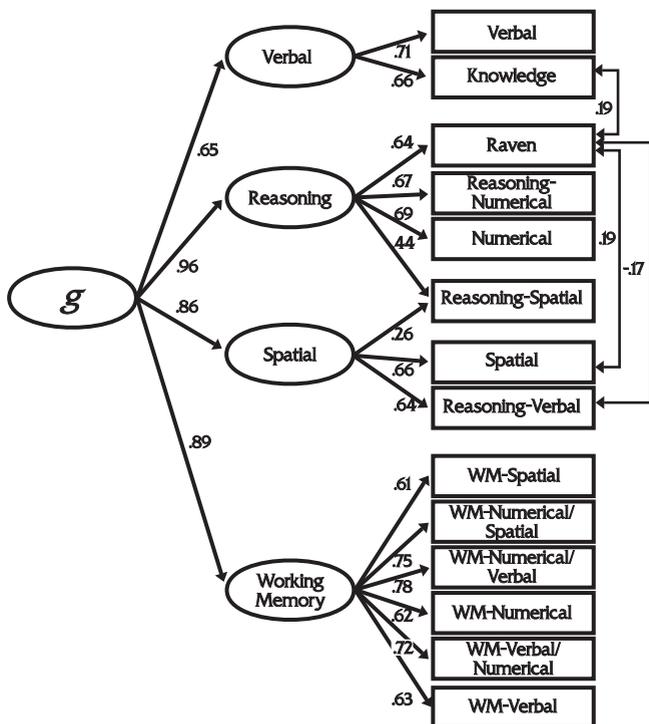


Figure 1. A hierarchical factor analytic representation of *g* and first-order factors, based on a structural equation model of meta-analytically derived correlations among ability and working memory (WM) measures.

the loading was .40, for Quantitative Reasoning the loading was .51, for Lexical Knowledge (Verbal) the loading was .37, and for Memory Span, the loading on *g* was .36. In the current SEM analysis, all of these loadings of the first-order factors on *g* were

larger (by more than .3 to .4) than what the extant literature supports, suggesting that the true associations between these latent variables and *g* may be substantially lower than that shown in the current SEM analysis. There are two highly salient reasons for the discrepancy between the current SEM results and those of Carroll's (1993) analysis. First, Carroll (1993) did not adjust the raw correlations in his data sets for attenuation. Although such corrections are common in meta-analytic research, there is some controversy about whether these corrections are appropriate, because they "confound issues of validity and reliability" and because "equating true scores with construct scores is logically inconsistent with the classical test theory model itself" (Boorsboom & Mellenbergh, 2002, p. 505). Second, because Carroll (1993) did not use SEM procedures, but instead used classic principal axis factoring, followed by Schmid-Leiman rotations, the correlations he reported are not confounded by the degree of model fit inherent in SEM results reported here.

Discussion

As we noted earlier, SEM with meta-analytically derived correlations (i.e., where different correlations are based on different measures, different samples, and different numbers of participants) is not optimal from a psychometric perspective. Taking the SEM results with some degree of skepticism is thus in order. Nonetheless, neither SEM solution points to an indication that WM and *g* are isomorphic to one another. We were able to reject a model in which WM measures and ability measures were captured by a single factor in favor of a model with separate *g* and WM factors—though the estimated correlation between WM and *g* latent factors in this model was .50. The initial model (see Figure 1) seems to provide a representation that is more concordant with the existing abilities literature, given the presence of both a higher order

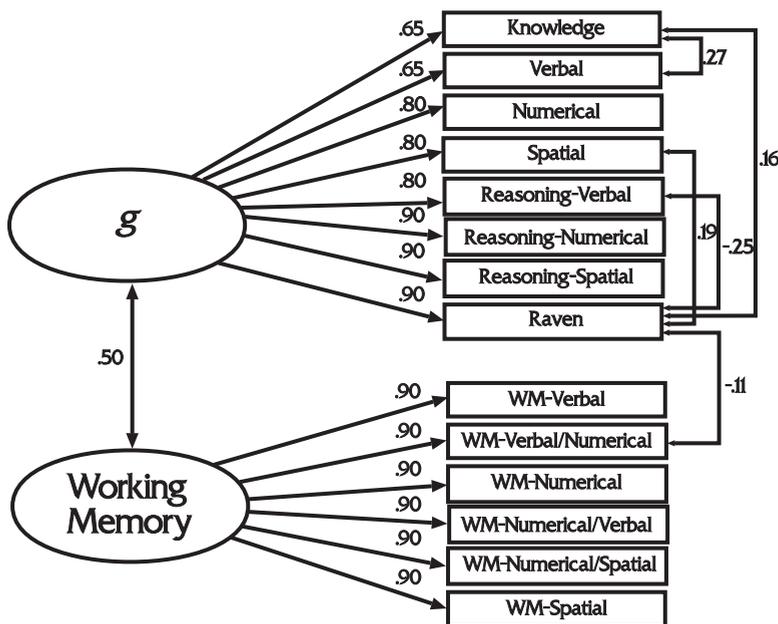


Figure 2. A nonhierarchical representation of *g* and working memory (WM) factors, based on a structural equation model of meta-analytically derived correlations among ability and WM measures.

general factor and lower order group factors. However, the estimates of the correlations between lower order factors and the higher order g factor appear to be overestimated, at least in comparison to extant research in the domain of intelligence (e.g., Carroll, 1993). It would not be unreasonable to tentatively propose that WM is a factor that has as much contribution to the general factor of intelligence as the content and reasoning factors do, but in this context it has little more common variance than Spatial ability has with g , and has less common variance than Reasoning does with g .

A Reappraisal of WM–Intelligence Relations in Light of Extant Empirical Data

We feel confident in asserting that the results of the meta-analysis and subsequent analyses support a conclusion that WM is not isomorphic to g , Gf , Reasoning, or any other group factor of intelligence. The results of the content overlap comparisons are somewhat less conclusive (see, e.g., Shah & Miyake, 1996), in that it appears that overlapping test contents for WM and ability measures do tend to inflate correlations between tests. However, it is not a large or significant effect. Ultimately, though, determining that WM and intelligence are related but not isomorphic constructs raises a fundamental question about whether there is any information to be gleaned from these correlations. Below we try to address this question.

Part of the reason we were initially skeptical about the WM– g isomorphism claim is that high-quality estimates of general intelligence (or g) cannot be derived only from a single nonverbal reasoning scale (such as the Raven). Rather, high-quality estimates of g are generated from the average across multiple tests of differing formats, contents, and processes. As R. L. Thorndike (1994) noted, “It is my strong belief that the best measure of g —and consequently broadly effective prediction—will stem from a diverse set of cognitive tasks, chosen to call for relational thinking in a diversity of contexts” (p. 154). By focusing attention on a single test to assess g , most investigations have not tested WM against a robust estimate of g . The few investigations that correlated WM measures with an estimate of general intelligence indicated relatively modest correlations between them (see Table 1). These correlations were, however, larger than most of the other WM–ability correlations, and equal to only numerical, reasoning (including Raven), and ECT correlations.

A more fundamental question has to do with the construct-level discussion of both WM and g . There are two basic propositions from the theoretical literature: (a) that WM measures directly assess “executive control,” or attentional capacity (see, e.g., Baddeley, 2002; Marsh & Hicks, 1998) and (b) that g is a representation of attentional capacity (at least as seen from the perspective of some researchers; see discussions by Lohman, 1996; Messick, 1996). If these two constructs have an estimated true-score correlation of .479, we can be confident in rejecting the conjunction of these two propositions. Logically, this can be resolved if either or both of the propositions are false. We consider each of these issues in turn.

Is WM the same as executive control and attentional capacity? According to Baddeley (2002), “The third component of the working memory framework, the central executive, was initially conceived in the vaguest possible terms as a limited capacity pool of

general processing resources” (p. 89).¹⁰ There is a substantial literature (see Kane & Engle, 2002, for a review) supporting the notion that individual differences in WM are related to performance on several different tasks that require executive control and controlled attention. For example, Bleckley, Durso, Crutchfield, Engle, and Khanna (2003) concluded, “Our results add to the growing body of literature supporting the equating of WM and controlled attention” (p. 888). Note that the body of evidence does not suggest that individual differences in WM completely account for, or even mostly account for, the variance in individual differences in performance on these various tasks. Rather, when extreme groups of participants are examined or when nonclinical participants and clinical samples are contrasted, the differences between groups on task performance are significantly different from zero.

Thus, individual differences in WM are related to, but do not explain, performance¹¹ on these various tasks of executive control and attentional capacity, except to the degree that a *ceteris paribus* qualification is invoked. Moreover, WM measures are significantly associated with measures of speed (e.g., PS and ECTs), indicating that some of the variance underlying WM measures is not common to unspeeded aspects of attention. Whether speed is a fundamental property of WM is debatable (see, e.g., Baddeley, 1998; Cowan, 1997), but it is also a debatable property of g (see, e.g., Sternberg, 1986). Finally, it is useful to note that there is controversy regarding the “unitary, limited-capacity central executive” (Lehto, 1996, p. 29; see also Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), at least as far as adolescents are concerned. Ultimately, it may be that WM is related to executive control and attentional capacity, but the construct does not account for all of the individual differences in these capabilities.

On a more conceptual level, there are aspects of WM measurement that suggest WM does not adequately sample the range of executive control and attention. The paradigm for WM assessment requires storage and transformation, supervision, and coordination processes. This paradigm clearly taxes the information-processing system in many ways, but it represents only one source of attention—namely, divided attention. It may not adequately sample the other kinds of attentional processes that individuals engage in,

¹⁰ Baddeley (2002) went on to review the relationship between the central executive and focused attention, divided attention, and switching attention. He concluded that “the capacity to focus available attentional capacity is clearly an important feature of the executive” (p. 90) but that various studies “appear to argue for a separable executive capacity to divide attention” (Baddeley, 2002, p. 90). Finally, “the question of whether task switching should be regarded as an executive process, or perhaps a range of processes, remains to be decided” (Baddeley, 2002, p. 91).

¹¹ An example from the abilities literature may be helpful in understanding this point. Many variables are correlated with general intelligence—such as socioeconomic status (SES). Jensen (1998) reported adult correlations between .50 and .70 for SES and IQ (a general intelligence score). However, the fact that SES correlates with general intelligence does not “explain” individual differences in intelligence in a causal sense. The explanation and causal arrow could just as easily go in the opposite direction (i.e., intelligence causes SES differences), or it is possible that some other variable or any number of variables act to bring about differences in both SES and intelligence. For a more in-depth discussion of correlations between experimental and differential constructs, see Underwood (1975).

such as focused attention or selective attention (Kahneman, 1973; but see, e.g., Lustig & Hasher, 2002, for a more complex account of WM tasks that includes inhibition and proactive interference). Although we are not aware of any research that has documented the typical attentional demands on an individual engaged in intellectual processing, it may be that focused and selective attention play a more central role in determining success or failure on typical real-world tasks. The typical WM test paradigm includes stimuli that are highly familiar to the participant. Administration of such stimuli ensures that individual differences in knowledge play a diminished role in performing the WM task. However, there is a concomitant effect of minimizing any influence of higher order intellectual functions that depend on knowledge and prior skills. To the degree that knowledge and skills are integral for determining individual differences in executive or attentional control (see, e.g., Ferguson, 1956), such as has been represented in the situated cognition literature (see, e.g., Greeno, 1998), these are largely untapped by WM tests. We suggest that measures of WM do not completely account for individual differences in executive control or attention.

As to consideration of g as a fund of mental energy and/or attentional capacity, many intelligence theorists agree that this is not a useful conceptualization. Even some of the strongest adherents of Spearman's theory (e.g., Jensen, 1998) do not endorse this view. P. E. Vernon's (1950) representation of g , though sharing some characteristics with Spearman's view, is that g accounts for about 20%–40% of the variance in all human abilities. The Raven's test tends to correlate with a composite general intelligence score of about .60 (e.g., Ackerman et al., 2002), indicating only 36% shared variance between the nonverbal reasoning test and general intelligence. Carroll's (1993, 1996) three-stratum theory of intelligence provides for eight different second-level ability factors that together map on to general intelligence. The positive manifold among all intellectual ability tests implies the presence of the general factor, and the higher correlations between various sets of tests by content or process implies the second-order factors.

However, although some second-order factors have higher loadings on the general intelligence factor (e.g., Gf) than others (e.g., PS), none of the factors has primacy for the determination of general intelligence. Thus, contrary to Gustafsson's (1984) assertion, Gf does not appear to be the same as g . As noted by Krueger and Spearman (1907), the Ebbinghaus Completion Test, which is mostly a verbal fluency test (see Ackerman, Beier, & Bowen, 2000), has a higher correlation with g than any other test. Carroll's (1993) analysis of Gustafsson's models noted several difficulties with the basis of the $gf = g$ assertion, but Carroll (1993) pointed out that the issue was not conclusively settled. However, individual differences in a wide range of other intellectual activities (especially in the domain of knowledge) also are important components of the overall general intelligence factor. Moreover, factors of attention or executive control have been notoriously difficult to find or replicate. Davies, Jones, and Taylor (1984) reported that

with respect to individual differences, there is little or no support for a general time-sharing ability or for a general vigilance ability; and the existence of a general selective-attention ability seems unlikely. Instead, a number of specific abilities, linked to specific task variables and information-processing demands, appear to be involved to varying degrees in the performance of all of these tasks. (p. 432)

Carroll's (1993) culling of the individual-differences literature largely supported this conclusion.

If g is not the same as executive attention, what about Gf ? Both Hebb (1942) and Cattell (1943) argued that Gf (or Intelligence A) is an aspect of intelligence identified mostly with attention and what current researchers would call *executive control*. However, as numerous researchers have pointed out, although executive control is necessary for performance on tests that load highly on Gf , it is not sufficient to account for performance on such tests (e.g. relative novelty of content and operations is one important determinant of performance). In addition, given the replicable separation of Gf from general visualization ability (see Carroll, 1993; Horn, 1965), many tests that load highly on General Visualization ability (but less so on Gf), such as rotation of the Shepard-Metzler figures, involve a substantial degree of executive control. Thus, executive control may be necessary for performance on Gf tests, but it is neither sufficient for Gf performance nor univocally identified with Gf . Ultimately, the intelligence literature is generally unresponsive of the notion that g , Gf , or any subsidiary ability can be univocally equated with executive control or controlled attention. Thus, we assert that the available data indicate that neither WM nor g (nor Gf for that matter) are isomorphic with executive control or attention. The difficulty here is that positive manifold among ability measures makes it difficult, if not impossible, to demonstrate causal determinants of any single ability or higher level ability construct, without first demonstrating discriminant validity with other ability constructs, something that has not been accomplished in this research domain (see, e.g., Campbell & Fiske, 1959).

Meta-Analysis of STM and Intellectual Abilities

To place the WM–ability meta-analysis in a larger context, we conducted a meta-analysis of STM measures with ability measures. As we noted earlier, there are hundreds of studies that correlate STM measures with abilities (e.g., given the numerous studies that report on correlations among the WAIS subtests or Stanford–Binet subtests), and there have been previous meta-analyses of memory-for-order tests and abilities (Daneman & Merikle, 1996; Mukunda & Hall, 1992). Mukunda and Hall's (1992) meta-analysis excluded studies prior to 1976 and included several studies that fall outside of the inclusion criteria used for our WM meta-analysis (e.g., unpublished data, studies of preschool and young [e.g., first grade] children) and did not differentiate between ability content. Daneman and Merikle's (1996) meta-analysis only considered general and specific comprehension abilities and also included data from young children (Grade 3) and elderly participants. To provide a relevant comparison between STM and WM correlations with abilities, we sought to obtain a representative sample of STM–ability correlations, with the same inclusion criteria as were used in the WM meta-analysis.

The main issues to be addressed by the STM meta-analysis are whether STM correlations with intellectual abilities differ in magnitude or pattern when compared with the WM correlations with intellectual abilities. We assess the same basic issues (e.g., correlation with g and content differentiation), though there were insufficient data available to assess the correlations between STM measures and ECT tests. We also address the similarity and dif-

ferences of correlations between STM–ability and WM–ability pairings.

Method

Literature search. Studies for possible inclusion in the meta-analysis of STM and intellectual abilities were initially identified through review of three sources for citations to appropriate data sources: Carroll’s (1993) extensive reanalysis of ability data sets, Mukunda and Hall’s (1992) meta-analysis of memory-for-order correlations with abilities, and Daneman and Merikle’s (1996) meta-analysis of WM and language comprehension. In addition, we also obtained any STM and ability correlations that were reported in the articles collected for the WM meta-analysis. The list of possible studies was expanded by a PsycINFO search of each of the terms *immediate memory* and *span memory* in combination with *intelligence*, which generated an additional 659 documents to review. Following the same inclusion criteria as the WM–intelligence meta-analysis, the final set of correlations that were input to the meta-analysis included data from 5,440 participants distributed in 49 independent samples.

Classification of STM tests. The STM tests were classified using the same content analysis that was applied to the WM tests (i.e., verbal, spatial, numerical), with the exception that the STM tests had only a single content, consistent with the single-task paradigm inherent in these measures. In all other details, the STM meta-analysis followed the same procedure that was used for the WM meta-analysis. The measures and their classifications are provided in the Appendix.

Results

The results of the meta-analysis of STM measure correlations with intellectual abilities are shown in Table 5. The table provides a 4 (STM test type) \times 12 (intellectual ability test type) matrix of weighted mean correlations (both raw and corrected for attenuation), 95% CI, total sample size, and total number of sample-collapsed correlations upon which each mean correlation was computed. The estimate of the overall meta-analytically derived correlation across all types of STM measures and ability measures, corrected for attenuation was .260 (95% CI = .23 to .28). (Total sample = 5,440 participants; total number of independent samples = 49.)

As with the WM meta-analysis, the estimated population correlation between STM measures and measures of *g* did not include 1.0 in the CI ($\hat{\rho}$ = .347, 95% CI = .30 to .39). A comparison with the analogous entry in the WM meta-analysis (see Table 1) also indicates that the estimated population correlation between STM and *g* was significantly smaller than the estimated population correlation between WM and *g* ($\hat{\rho}$ = .479), although the raw weighted correlations were not as different as the correlations corrected for attenuation (.281 for STM, .364 for WM), indicating that the respective reliabilities for the WM measures tended to be lower than that for the STM measures. Meta-analytic correlations between the Raven and STM measures, although predicated on a smaller sample of studies and participants (6 studies with a total of 500 participants for STM vs. 13 studies with a total of 1,752 participants for WM), were significantly smaller for STM ($\hat{\rho}$ = .207) in comparison with the correlations between WM and the Raven ($\hat{\rho}$ = .495).

Evaluation of common-content correlations (e.g., verbal STM with verbal ability) were generally consistent with the notion that content commonality increased correlations. All three sets of STM

correlations (verbal, numerical, spatial) were larger with similar ability content, though only the numerical STM value was significantly larger than the correlations across discrepant content abilities (verbal, spatial). These results were largely similar to the pattern of correlations found with the WM measures.

There were no studies found that reported correlations between STM and ECT measures, thus it is not possible to evaluate the differentiation of STM and WM measure relations with ECTs. However, correlations between STM measures and PS ability ($\hat{\rho}$ = .233) were smaller than the correlations between WM measures and PS ability ($\hat{\rho}$ = .268), but not quite significantly so, given the small overlap between their respective CIs.

STM Latent-Variable Analysis

To examine the relationship between STM and *g*, we used a procedure that was parallel to the SEM analyses of WM and *g*. The original matrix of 11 ability measures and 4 STM measures was reduced to 6 ability measures and 3 STM measures because of missing data and the exclusion of PS abilities. The correlation matrix used for this analysis is shown in Table 6. Two of the 45 values in the correlation matrix were imputed as described earlier. The first analysis was a CFA with a second-order general ability factor (*g*) and four first-order factors representing content abilities (verbal, numerical, spatial, and STM). The model fit was good, $\chi^2(21, N = 471) = 70.93, p < .01, RMSEA = .070, NFI = .95$. The model is shown in Figure 3. As can be seen in the figure, the structure coefficient between STM and *g* (.51) is smaller than the relationship between the other content abilities and *g* and smaller than the coefficient between *g* and WM shown in Figure 1. This analysis was followed by a two-factor model that examined the relationship between *g* and STM, which is shown in Figure 4. As in the previous WM analysis, structure coefficients for the ability and WM measures were fixed on the basis of their relationships with the second-order factor in the previous analysis. The fit of this model was adequate, $\chi^2(29, N = 471) = 101.28, p < .01, RMSEA = .071, NFI = .92$. As can be seen in the figure, the correlation between *g* and STM (.49) is roughly the same as that found for *g* and WM (.50) shown in Figure 2. This analysis reiterates our meta-analytic findings in regard to the relations among WM, STM, and *g*. That is, STM may not be as highly related to *g* as is WM, but the relationship between STM and *g* is substantial.

Discussion

The meta-analysis of STM and intellectual abilities we conducted provides an overall correlation between STM and intellectual abilities that is not substantially different from that of the partially overlapping set of correlations reported by Mukunda and Hall (1992). They reported a weighted average correlation of .203, and our meta-analysis found a weighted average correlation of .221. However, our analysis separately reports correlations with *g* measures (mean r = .281), and we corrected for attenuation of both STM and ability measures. We also provide separate mean correlations by ability type and STM content. For STM and verbal abilities (the closest content factor to general language comprehension), the Daneman and Merikle (1996) average correlation

Table 5
Estimated Population Correlations Between Short-Term Memory (STM) and Ability Variables With 95% Confidence Intervals and Heterogeneity Statistics

STM measure	Reasoning								Perceptual speed	Knowledge	g	Average ability
	Verbal	Numerical	Spatial	Verbal	Numerical	Spatial	General	Raven				
Verbal	.351 (.288)	.276 (.227)	.171 (.143)	.302 (.248)	.323 (.264)	.307 (.225)	.347 (.289)	.180 (.138)	.221 (.174)	.465 (.384)	.340 (.317)	.288 (.236)
	21 (2,165)	13 (1,713)	4 (990)	1 (265)	2 (695)	1 (265)	2 (317)	2 (255)	4 (879)	1 (253)	9 (1,013)	26 (2,708)
	.31 to .39	.23 to .32	.11 to .23	.19 to .41	.21 to .43	.19 to .41	.24 to .44	.06 to .30	.16 to .28	.36 to .56	.28 to .39	.25 to .32
Numerical	53.55*	29.41*	3.97	—	18.49*	—	1.58	1.09	7.95	—	44.79*	—
	.197 (.166)	.275 (.235)	.141 (.123)	.297 (.247)	.286 (.226)	.340 (.251)	—	.244 (.219)	.187 (.157)	.287 (.231)	.322 (.262)	.224 (.187)
	18 (2,125)	17 (2,435)	9 (1,530)	3 (493)	2 (523)	2 (429)	—	4 (247)	8 (1,185)	2 (215)	2 (238)	26 (2,875)
Spatial	.16 to .24	.24 to .31	.09 to .19	.21 to .38	.20 to .36	.25 to .42	—	.12 to .36	.13 to .24	.16 to .40	.20 to .43	.19 to .26
	51.03*	48.48*	4.65	20.05*	0.01	6.66*	—	3.01	9.55	0.17	2.81	—
	.155 (.132)	.236 (.195)	.293 (.246)	.165 (.135)	.178 (.144)	.378 (.276)	—	.158, <i>ns</i> (.134)	.284 (.225)	.186 (.150)	.174, <i>ns</i> (.140)	.240 (.200)
Composite	5 (1,149)	5 (1,149)	6 (1,316)	1 (265)	1 (718)	1 (265)	—	3 (147)	3 (939)	1 (718)	1 (61)	10 (1,524)
	.01 to .21	.18 to .29	.24 to .34	.05 to .28	.11 to .25	.27 to .48	—	-.01 to .31	.22 to .34	.11 to .26	-.08 to .41	.19 to .29
	1.73	53.65*	38.68*	—	—	—	—	1.91	4.20	—	—	.501 (.410)
Average STM	.486 (.400)	.391 (.320)	—	—	—	—	—	—	—	—	—	—
	1 (62)	1 (62)	—	—	—	—	—	—	—	—	—	1 (62)
	.26 to .66	.16 to .58	—	—	—	—	—	—	—	—	—	.29 to .67
Average STM	.263 (.218)	.292 (.244)	.210 (.177)	.298 (.246)	.263 (.215)	.381 (.281)	.347 (.289)	.207 (.174)	.233 (.187)	.264 (.215)	.347 (.281)	.260 (.214)
	37 (4,142)	28 (3,570)	14 (2,752)	3 (493)	4 (1,495)	2 (429)	2 (317)	6 (500)	14 (2,561)	4 (1,186)	12 (1,291)	49 (5,440)
	.23 to .29	.26 to .32	.17 to .25	.21 to .38	.21 to .31	.30 to .46	.24 to .44	.12 to .29	.20 to .27	.21 to .32	.30 to .39	.23 to .28

Note. For each STM measure, Line 1 contains the estimated population correlation ($\hat{\rho}$; uncorrected correlation is in parentheses). Unless otherwise noted, all correlations are significant at the .05 level (*ns* = not significant). The first value in Line 2 of each STM measure is the number of correlations, and the total sample size is in parentheses. Line 3 contains the 95% confidence interval of the estimated population correlation, and Line 4 contains the heterogeneity statistic Q . A dash indicates that Q could not be computed with only one sample. ECT = elementary cognitive task.

* $p < .01$.

Table 6
Correlation Matrix of Short-Term Memory (STM) and Ability (AB) Measures Used in the Structural Equation Models

STM/AB Measure	1	2	3	4	5	6	7	8	9
1. AB: verbal									
2. AB: numerical	.370								
3. AB: spatial	.290	.300							
4. AB: numerical reasoning	.500	.600	.380						
5. AB: Raven	.410	.570	.510	.430 ^a					
6. AB: knowledge	.540	.370	.430	.500	.430 ^a				
7. STM: verbal	.288	.227	.143	.264	.138	.384			
8. STM: numerical	.166	.235	.123	.226	.219	.231	.470		
9. STM: spatial	.132	.195	.233	.144	.134	.150	.020	.280	

Note. *N* used for the analysis is the harmonic mean of all samples included, 471.

^a Imputed value.

was .28, whereas our estimate was slightly lower (.218, $\hat{\rho} = .263$). The marginally larger correlation found by Daneman and Merikle can most likely be attributed to their more liberal inclusion criteria (e.g., children and older adults).

A comparison of STM–ability correlations with WM–ability correlations indicates that WM measures tend to correlate more highly with most of the ability categories (excluding reasoning—verbal and knowledge ability categories, for which there are overlapping CIs). In the aggregate, the difference between an average corrected correlation for WM and intellectual ability measures ($\hat{\rho} = .397$) and between STM and intellectual ability measures ($\hat{\rho} = .260$) is equivalent to the difference of 15.8% shared variance between WM and intellectual abilities and 6.8% shared variance between STM and intellectual abilities. Whether this is considered a large difference or a small difference from a theoretical perspective probably depends on the individual researcher’s general orientation toward the larger issue of the centrality of WM for explaining individual differences in intelligence.

General Discussion and Enduring Issues

General Discussion

Across the meta-analyses and SEM analyses presented in this article, two findings appear to be the most salient. The first is that contrary to assertions by various researchers, the relationship between WM and *g* (or *Gf*) is not isomorphic. In addition, the relative lack of convergent validity between WM and general ability measures is probably of a magnitude that would clearly result in a challenge to substitutability for either research or practical purposes. At the individual level (e.g., a single test or measure), WM measures do not show uniformly greater convergence with either tests of content (e.g., verbal, spatial, numerical) or with reasoning abilities, but rather they appear to share greater variance with narrow measures of elementary information-processing tasks, though this is based on a relatively small sample of measures. The second finding pertains to what is not present in the analyses. That is, despite a wide review of the literature concerning WM and intellectual abilities, there are many cells of the matrix of WM and

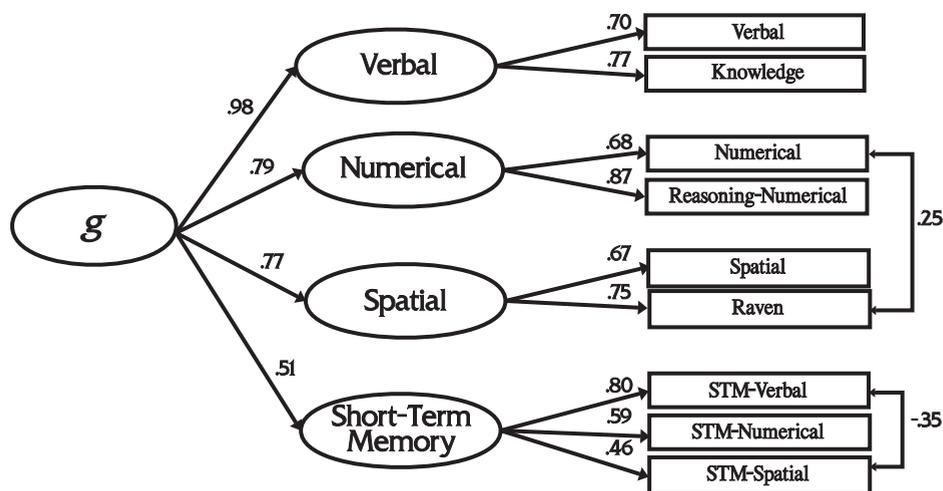


Figure 3. A hierarchical factor analytic representation of *g* and first-order factors. Confirmatory factor analysis based on meta-analytically derived correlations among ability and short-term memory (STM) measures.

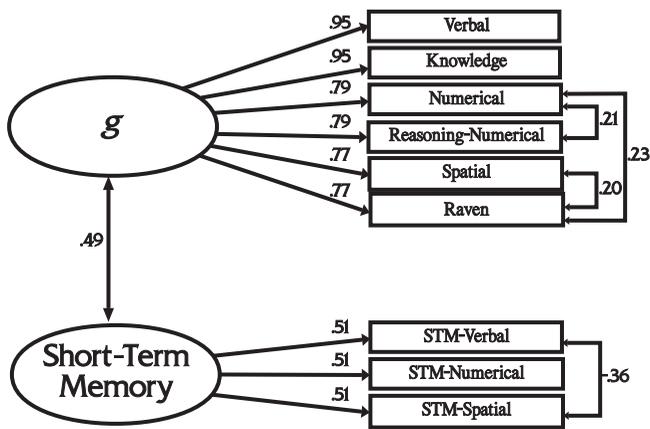


Figure 4. A nonhierarchical representation of g and short-term memory (STM) factors based on meta-analytically derived correlations among ability and STM measures.

ability measures that are either unpopulated or only sparsely populated. Thus, much of what one can say about the relations between WM constructs and ability constructs is substantially limited by insufficient data. Because there is clearly much more work that needs to be done to ultimately place WM in the nomothetic network of ability constructs, it is important to outline some of the key psychometric issues that have limited the informational value of prior and current research on the topic. We explore a few of these issues below.

Enduring Psychometric Issues Pertaining to WM–Intelligence Relations

Although we conducted a meta-analysis that adjusted correlations for unreliability of the respective measures, consideration of reliability and other psychometric issues may help illuminate the underlying relationship between WM and intelligence measures. At a very general level, examining the relations between tasks developed from an information-processing perspective and those developed from a differential perspective entails some risks of psychometric artifacts (see, e.g., Kyllonen, 1985, for a discussion of information-processing-inspired difference scores in individual-differences research). Below we consider four broad issues that are related to the degree of overlap found between WM and intelligence measures: reliability, difficulty, dual-task methodology, and the principle of aggregation.

Reliability. One of the key considerations when correlating two variables is the respective reliabilities of each variable, because reliability sets an upper bound on validity. When one or both variables have low reliability, the resulting low correlation between measures can erroneously suggest a small amount of shared variance. For example, consider the Digit Span subtest of the WAIS (which represents a composite of forward digit span and backward digit span). The correlation between the Digit Span subtest and total WAIS composite score was reported to be .56. Taken at face value, the correlation represents a shared variance with total WAIS score of 31.4%. However, the Digit Span subtest had the lowest test–retest reliability of any of the other WAIS subtests—it was reported by Wechsler (1958; see also Derner,

Aborn, & Canter, 1950) to range from .66 (for adults ages 25–54) to .71 (for adults ages 18–19). If we use the reliability of the full-scale WAIS (.97) and the reliability of the Digit Span subtest (.71 for young adults) and apply the formula to correct for attenuation, the estimated true-score correlation between the Digit Span subtest and the overall WAIS composite is .675, which corresponds to a shared variance of 45.5%—a difference of potential importance to construct overlap considerations. It is interesting that this value exceeds all of the meta-analytically derived corrected correlations between WM measures and estimates of general intelligence (see Table 1), though we have not adjusted the meta-analytic correlations for possible restriction of range in talent (something that would ordinarily boost correlations when college student participant samples are subject to prior selection on intellectual abilities).

However, correcting correlations for unreliability does have drawbacks. First, there are several different forms of reliability (see, e.g., R. L. Thorndike, 1947), and the appropriate estimate for any particular purpose is not always available. As Cronbach (1990) noted, in the physical sciences, the ideal estimate of reliability would be between two identical measurements, such as the measurement of weight with two identical scales. In psychology, such an assessment is not generally appropriate, because the examinee will be changed by the experience of taking the test—something not characteristic of measuring an inanimate object. Almost all of the estimates of reliability we obtained for the meta-analysis were measures of internal consistency (e.g., alpha). However, these estimates confound test reliability with item homogeneity (see, e.g., Ackerman & Humphreys, 1991). For current purposes, a more suitable estimate of reliability for WM tests would be a delayed test–retest, alternate-form reliability. This would allow for an estimate of the “general and lasting” (R. L. Thorndike, 1947, p. 102) variance that most researchers would identify as the WM construct, in contrast to the temporary and/or specific variance, which would be appropriately placed in the error category. Unfortunately, although researchers have collected such data on many existing intelligence and ability tests, similar estimates of reliability for WM tests are lacking (but see Klein & Fiss, 1999, for a small-sample [$N = 33$] study on this topic). The difficulty that this lack of data poses is not just in the lack of accuracy for computing estimated true-score correlations between WM and ability measures, it is a problem for those who wish to place the WM construct on an even footing with extant intelligence measures. Temporal stability and common rank ordering of individuals with alternate-form assessments of the same construct are requisite criteria to establish the usefulness of WM as a trait.

Test difficulty. The relationship between the difficulty of tests and their correlations is substantially more complicated than the issue of reliability. In classical test theory, difficulty for each item is represented as the proportion of examinees passing or failing. (In the more complex item response theory, difficulty is indexed by multiple parameters—but this is a level of analysis that is only practical for tests with large samples of examinees and large samples of test items—something not characteristic of the WM assessment field).

For many unspeeded ability tests, it is relatively straightforward to compute the distribution of item difficulties. With this information, test developers can refine their measures to either provide maximal discrimination accuracy at one or another level (e.g., by

skewing the distribution of the item difficulties toward one end of the distribution or the other). Alternatively, the developer can select items with a uniform frequency distribution, so that equal discrimination accuracy is obtained throughout the ability range. In practice, operational tests are often skewed to match the purpose of the test and the range of abilities in the examinee population (e.g., the original Stanford–Binet scales did not make discriminations beyond the “above average” level because they were primarily designed to detect those individuals with mental retardation). One reason this issue is important is that a mismatch between the respective item difficulty distributions of two variables will result in attenuated correlations between the two variables, that is, correlations closer to zero (for discussion of this issue, see, e.g., Carroll, 1945; Ferguson, 1941). Moreover, in a group of several tests, differences and similarities of test difficulty may yield “difficulty factors” or spurious covariance—which are, in some sense, factors or covariance associated with the method of testing rather than the process construct that is the ostensible target of testing.

When considering memory tests, insufficient data are available to obtain a precise determination of item difficulty distributions (i.e., the frequency distribution of item passing rates). Although it is in theory possible to obtain such information from traditional span measures (such as forward digit span and backward digit span), traditional measures are far too short (containing only a dozen or so items) to provide detailed statistics (see Brogden, 1946, for relevant validity indices for 9-item and 18-item tests of different item difficulty distributions). Nonetheless, it is apparent that the backward digit span, with the same number of digits to be recalled, is a more difficult test overall than forward digit span, simply because the overall scores are lower when the tests are administered to a single norming sample. Addressing item difficulty distributions with WM tests is much more complicated than with simple span measures because most WM tests involve two simultaneous tasks (where failure can occur on either task, see below) and because the composite span measure is rarely presented as the number of items that can be successfully recalled (the typical score is a total number of items recalled perfectly across trials of different length). The essence of this discussion is that task difficulty differences undoubtedly affect the correlations both among different memory tasks and between memory tasks and ability tests, but to an unknown degree. Future research could help resolve these issues by (a) using a far larger number of items in individual WM tests and (b) providing information on the item difficulty distributions of particular WM tests, so that the effects of item difficulty distributions on correlations with ability measures can be determined directly.

Dual-task and secondary-task methodology. With the exception of the backward span task, nearly all modern WM tasks involve simultaneous performance of two tasks. For example, in the reading span task, the participant reads a series of sentences aloud (the primary task) and then is asked to report the last word in each sentence (the secondary task). From a psychometric perspective, far too little information is available that provides an account of how overall performance should be assessed. Instead, the typical procedure (see, e.g., Daneman & Carpenter, 1980) is to focus only on the secondary task. A review of the literature indicates that participants are not specifically instructed about task priorities, or, perhaps this information about the procedure is too

trivial to consider mentioning in the Method section. Typically, performance on the primary task is not taken into account, and only data from the secondary task are reported (but see Waters & Caplan, 1996, for a more in-depth examination of task components). When primary task performance is taken into account, it is typically used as an exclusion criterion (e.g., participants with a less-than-threshold level of performance on the primary task are excluded from analysis). Investigators do not know, for example how two individuals who obtain the same score on the secondary task but differ by 5% accuracy on the primary task might differ in overall capabilities. Such issues have been discussed in the wider dual- and secondary-task literature (e.g., Ackerman, 1984; Ackerman, Schneider, & Wickens, 1984; Ackerman & Wickens, 1982; Navon & Gopher, 1979). Empirical research is clearly needed that addresses how changes in component-task emphasis affect WM test performance, both within and between individuals. However, from an individual-differences perspective, a preferred procedure would be to include separate component measures in a regression or correlational analysis so that their separate and interactive influences could be directly assessed (see, e.g., Bayliss, Jarrold, Gunn, & Baddeley, 2003).

The principle of aggregation, the bandwidth-fidelity dilemma, and Brunswik symmetry. The principle of aggregation (see Rush-ton, Brainerd, & Pressley, 1983) indicates that, *ceteris paribus*, multiple measures of the same trait will yield a more “stable and representative estimator” (Rushton et al., 1983, p. 18) than single behavior measures. The principle is similar to the classical test theory notion that errors of measurement are randomly distributed. When aggregating across multiple measures, errors will cancel one another, leading to a closer estimation of the underlying true score. The principle of aggregation provides a substantial foundation for omnibus measures of intellectual ability (e.g., WAIS, Stanford–Binet), in that these measures provide a single aggregated score across many different correlated subscales. However, aggregation of this sort has a downside, which is known as the *bandwidth-fidelity dilemma* (see, e.g., Cronbach, 1990). That is, when a fixed amount of time is available for testing, there must be a trade-off between the bandwidth (i.e., the construct breadth) and the fidelity of measurement (i.e., precision). Thus, in omnibus intelligence tests, the overall IQ score is remarkably stable from one testing occasion to the next, but individual subscales provide much less stability and reliability. In contrast, an abstract reasoning test like the Raven Progressive Matrices achieves high fidelity but much lower bandwidth than an omnibus intelligence test. The only means for obtaining high fidelity and high bandwidth would be to substantially increase testing time. One should keep in mind when assessing the correlations between individual ability tests and other measures both the degree of aggregation that is taking place and the trade-offs between bandwidth and fidelity of assessment.

For memory tests, the principle of aggregation and the bandwidth-fidelity problem represent heightened concerns in comparison to intelligence assessment. The historical research strategy in experimental psychology (see, e.g., Cronbach, 1957; see also Hilgard, 1987, for a more extensive review) has been to develop tasks that are nearly univocally associated with a particular underlying mental process or operation. As such, the trend has been to develop tasks with both high fidelity and low bandwidth (e.g., the classic memory search paradigm of the Sternberg task). Tasks with

high bandwidth have been especially vexing to researchers who are attempting to develop and test taxonomies of elementary information processes (see, e.g., Carroll, 1980; Newell & Simon, 1972). WM tasks, as a general rule, are difficult to parse into low-bandwidth components, especially when simultaneous performance on the secondary tasks is taken into account. Although this point is controversial, Cowan, Towse, and their colleagues noted that “understanding WM-span tasks has been difficult in part because performance depends on both specific skills that differ by domain and general skills that cross domains” (Cowan et al., 2003, p. 114; see also, e.g., Hutton & Towse, 2001).¹²

Because STM span tests have low bandwidth, tests of such processes (even when given with an adequate number of trials to attain reasonable reliability levels) have relatively low correlations with tests of high bandwidth. Correlating a digit span test with an omnibus intelligence test represents what Wittmann and Süß (1999) have called a *lack* of Brunswik symmetry—that is, a mismatch between the breadth of the predictor (the digit span test as representing a narrow construct) and the criterion (the overall intelligence score, which is a highly broad construct). Because multiple processes are operating in WM tasks, performance indicators have higher bandwidth in WM tasks than in STM span tasks. Thus, on the basis of Brunswik symmetry considerations alone, one should expect that there will be higher correlations found between WM tasks and omnibus measures of intelligence than between STM tasks and omnibus measures of intelligence. Aggregating across multiple WM measures (especially when they are designed within a facet-based approach that adequately samples across constructs that are not central to the definition of WM, such as test content; see Humphreys, 1962) will provide a more robust assessment of WM-*g* relations, and it might result in a larger association between the measures. If content commonality is responsible for some of the observed WM-*g* covariance, however, there might be a somewhat lower association between the measures.

In general, there is a strong relationship between a task’s level of complexity and the correlation between scores on that task and measures of general intelligence (see, e.g., Marshalek et al., 1983). Whether task complexity and bandwidth are more generally isomorphic has yet to be convincingly demonstrated, but the extant data suggest that this may have some verisimilitude (see, e.g., Larson, Merritt, & Williams, 1988, for one demonstration of this phenomenon).

Conclusions and Remaining Issues

Tests of immediate memory have been an integral part of intelligence assessment and intelligence theory since the late 1800s and early 1900s. Few omnibus intelligence tests lack an assessment of immediate memory, though tests of such abilities range from simple forward digit span to memory for sentences and other complex materials. Over the past century, there have been many investigations of the relations among various kinds of immediate memory (e.g., associative memory, simple span, memory for meaningful material) and relations between these memory measures and intellectual abilities. There is substantial support (see, e.g., Carroll, 1993) for several different factors of immediate

memory and for their moderate correlations with general intelligence, or *g*.

Starting in the mid-1970s, experimental psychologists have revised their perspectives on the architecture of the information-processing system (e.g., Baddeley & Hitch, 1974), in particular shifting an earlier focus on short-term storage, or STM systems, into a focus on WM. Investigations of individual differences in WM have led several investigators to propose that WM has key importance for understanding the nature of individual differences in intelligence. The movement to relate WM to intellectual abilities started with Daneman and Carpenter’s (1980) study, which indicated significant correlations between measures of verbal WM and reading comprehension. Later, Kyllonen and Christal (1990) demonstrated substantial overlap between measures of WM and reasoning abilities, leading to a claim that “reasoning ability is (little more than) working memory capacity” (p. 389), though it should be noted that this assertion was provided with both an exclamation point and a question mark. In the decade that followed the Kyllonen and Christal article, claims of overlap between WM and intelligence have increased in frequency and in scope of the argument, resulting in the strongest assertion that WM is the same as *Gf* or “isomorphic” to *g*.

The meta-analysis reported in this article clearly demonstrates that WM measures are significantly correlated with measures of intellectual abilities, in terms of broad content abilities (verbal, numerical, and spatial), with general and specific content-based reasoning abilities, with PS and ECTs, with knowledge abilities, and with *g*. However, even when the measures are corrected for unreliability, in no case did the estimated true-score correlations between WM and ability exceed a value of .653, indicating a maximum shared variance of 42.6% (note that this particular correlation was based on only three studies, with a total sample size of 380 participants). On average (weighted by sample size), WM tests correlated .364 with measures of *g*. Correcting for unreliability of WM and *g* measures increased the average correlation between WM and *g* to .479, yielding a shared variance of 22.9%.

In contrast, the highest average correlations were found between WM tests and narrow speeded ECTs (e.g., Sentence Verification Test, Consonant-Vowel Test). Other research (e.g., Carroll, 1980; Kyllonen, 1985) has indicated that ECTs are not highly correlated with *g*, and in hierarchical theories of intelligence, ECTs tend to be placed at the lowest order factors because they are low in complexity (see Snow et al., 1984, for a discussion of the complexity continuum).

The results demonstrate that WM is not the same thing as *g*. In some sense, the claim of isomorphism between WM and *g* appears to be an example of the “jingle fallacy”—that is, the fallacy that words that are accorded the same meaning may not in fact refer to

¹² Incidentally, the need to determine the underlying abilities in evaluation of individual differences in dual-task designs was explicitly highlighted by Ackerman et al. (1984), concerning the determination of a single ability to do more than one task at a time (called a *timesharing ability*). The logic of determining the nature of a timesharing ability is the same as that needed to separate individual task component abilities from a general WM ability when WM is assessed by primary-, secondary-, or dual-task methods.

the same underlying constructs (E. L. Thorndike, 1904; see also Kelly, 1927). In the current context, theoretical descriptions of WM as an index of executive control and attention sound very much like Spearman's (1938) conceptualization of *g* as a fund of general mental energy.¹³ The overlapping descriptions between these constructs may very well have influenced an overarching conceptualization to see WM and *g* as the same thing, even when the data supporting their interrelationship are modest at best.

A review of Table 1 illustrates another factor that may be partly responsible for this state of affairs. Although 86 separate samples of individuals were obtained from the literature, a relatively small number of samples (15, or 17% of the total) actually included assessments of *g*. In the absence of direct estimates of *g*, some researchers may have attempted to overgeneralize from a modest sampling of ability measures (especially in the domain of ECTs) to a general factor of intelligence.

A determination that WM and *g* are not isomorphic is only one part of the overall picture that emerges from the current analysis. It is clear that WM measures are related to intellectual abilities, though at a far more modest level than unity. Given the ubiquitous finding that intellectual ability measures (even those that are very narrow or highly speeded) are positively correlated with one another, a hypothesis of no relations between WM tests and intellectual abilities would not have been a viable conjecture. Rather, it is useful to determine whether the relationship goes beyond something conjectured by E. L. Thorndike (1940), that is, that superiority in one trait implies superiority in other traits. Meehl (1990) and Lykken (1991) have referred to this as "the crud factor," namely that in social science, "everything is correlated with everything, more or less" (Meehl, 1990, p. 123).

In some areas of psychology, the threshold (*r*) for the crud factor appears to be about .30. The overall estimated true-score correlation between WM measures and ability measures significantly exceeds this value ($\hat{\rho} = .397$), whereas the analogous correlation for STM and ability measures ($\hat{\rho} = .260$) does not. Thus, on the one hand, WM measures do appear to correlate more highly with intellectual abilities than would be expected from background noise alone or if WM was measuring something no more central to intelligence than STM. On the other hand, the meta-analysis results suggest that WM measures do not show substantial discriminant validity—meaning that they correlate significantly and substantially with many different abilities, rather than with one or two key abilities. The only ability domain to show substantially different correlations with WM are those measured by narrow ECTs.¹⁴ To the degree that this is a robust finding, WM may be better placed with the lower order cognitive abilities rather than with the higher order abilities, such as broad content or reasoning. Such a placement of WM in a three-stratum model would be consistent with the overarching model of Carroll (1993) and would also be consistent with the complexity continuum model of Snow and his colleagues (Marshalek et al., 1983; Snow et al., 1984).

In the context of his theory of intelligence, Spearman (1927) proposed the theorem of "the indifference of the indicator" (p. 197). That is, "for the purpose of indicating the amount of *g* possessed by a person, any test will do just as well as any other, provided only that its correlation with *g* is equally high" (Spearman, 1927, p. 197). He went on to say, "Indeed, were it worth while, tests could be constructed which had the most grotesque

appearance, and yet after all would correlate quite well with all the others" (Spearman, 1927, p. 198). Observers might think that some tests of WM are grotesque indeed, in that they often involve tasks that have little in the way of face validity to examinees (at least when it comes to the kinds of intellectual tasks that people do on a daily basis). Ultimately, though, the question is posed as to whether degree of overlap between WM measures (or latent factors) and intelligence yields information that has value added beyond the indifference of the indicator. At this point, and on the basis of the meta-analyses and SEM results, we are reluctant to suggest that the case has been made for WM in terms of informing intelligence theory beyond the common variance found among other ability measures.

Resolution of the question of how and how much WM and intelligence are related ultimately requires additional research. In our opinion, the issue cannot be ultimately settled until studies are conducted that provide multiple tests of a wide range of ability factors (e.g., reasoning, spatial, verbal, numerical, PS), multiple tests of WM in each of the different content domains (verbal, numerical, spatial), separate measures of component tasks in any primary or secondary task, WM tests that do not depend on time-sharing performance, and an adequate sample. The rule of thumb used by factor analysis theorists is three high-validity marker tests for each factor, and 10 times the number of participants as number of variables—though with a college sample, which has restricted range of talent, additional participants would be needed. If such a study were to be conducted, it would presumably shed a great amount of light on the overarching question about the relationship between WM and intelligence. However, even this ambitious design would only serve to demonstrate convergent validity (see, e.g., Campbell & Fiske, 1959). Additional measures, such as information-processing tasks that are not presumed to relate to either WM or intellectual ability, would be needed to provide the necessary discriminant validity among WM and intelligence.

¹³ Incidentally, Spearman (1927, p. 263) discussed a study of tasks that he called *concentrative attention* (e.g., "tapping and adding, separately") and *diffusive attention* (e.g., "tapping and adding at the same time")—which correspond in some manner to focused and divided attention in modern terms. However, Spearman (1927) concluded that both kinds of tests were substantially related to *g*, but he did not identify either kind of test as isomorphic with *g*.

¹⁴ An anonymous reviewer suggested that "one would expect that WM is positively correlated to all cognitive abilities (to a degree corresponding to their *g* loading)." Given that ECT abilities have the lowest ability loadings on *g*, and broad content and reasoning abilities have the highest loadings on *g* (see, e.g., Carroll, 1980, 1993; Snow et al., 1984), the current results strongly support an assertion of a lack of discriminant validity for WM.

References

References marked with an asterisk (*) indicate studies included in the working memory meta-analyses. References marked with a dagger (†) indicate studies included in the short-term memory meta-analysis.

Ackerman, P. L. (1984). Multivariate evaluation of workload, resources, and performance in dual task data. *Proceedings of the 28th Annual Meeting of the Human Factors Society*, 210–214.

- Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2000). Explorations of crystallized intelligence: Completion tests, cloze tests and knowledge. *Learning and Individual Differences: A Multidisciplinary Journal in Education, 12*, 105–121.
- *Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General, 131*, 567–589.
- Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied, 6*, 259–290.
- Ackerman, P. L., & Humphreys, L. G. (1991). Individual differences theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 223–282). Palo Alto, CA: Consulting Psychologists Press.
- Ackerman, P. L., Schneider, W., & Wickens, C. D. (1984). Deciding the existence of a time-sharing ability: A combined theoretical and methodological approach. *Human Factors, 26*, 71–82.
- Ackerman, P. L., & Wickens, C. D. (1982). Methodology and the use of dual and complex-task paradigms in human factors research. *Proceedings of the 26th Annual Meeting of the Human Factors Society*, 354–358.
- †Anastasi, A. (1930). A group factor in immediate memory. In R. S. Woodworth (Ed.), *Archives of psychology* (Vol. 15, No. 120). New York: Columbia University.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Prentice Hall.
- †Anderson, V. A. (1939). Auditory memory span as tested by speech sounds. *American Journal of Psychology, 52*, 95–99.
- 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, remembering, and forgetting: Proceedings of the second conference* (pp. 89–195). New York: New York Academy of Sciences.
- *Babcock, R. L. (1994). Analysis of adult age differences on the Raven's Advanced Progressive Matrices test. *Psychology and Aging, 9*, 303–314.
- *Babcock, R. L., & Laguna, K. D. (1996). An examination of the adult age-related differences on the Raven's Advanced Progressive Matrices: A structural equations approach. *Aging, Neuropsychology, and Cognition, 3*(3), 187–200.
- Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A. (1998). *Human memory: Theory and practice* (Rev. ed.). Needham Heights, MA: Allyn & Bacon.
- Baddeley, A. (2002). Is working memory still working? *European Psychologist, 7*, 85–97.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 8. Advances in research and theory* (pp. 47–89). New York: Academic Press.
- *Baddeley, A., Logie, R., & Nimmo-Smith, I. (1985). Components of fluent reading. *Journal of Memory & Language, 24*, 119–131.
- *Baker, L. (1985). Working memory and comprehension: A replication. *Bulletin of the Psychonomic Society, 23*, 28–30.
- †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.
- Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.
- *Bichsel, J., & Roskos-Ewoldsen, B. (1999). Imaginal discovery, working memory, and intelligence. *Journal of Mental Imagery, 23*(1–2), 17–34.
- Bigham, J. (1894). Memory. *Psychological Review, 1*, 453–461.
- Binet, A., & Simon, T. (1961). New methods for the diagnosis of the intellectual level of subnormals (E. S. Kite, Trans.). In J. J. Jenkins & D. G. Paterson (Eds.), *Studies of individual differences: The search for intelligence* (pp. 90–96). New York: Appleton-Century-Crofts. (Reprinted from *L'Année Psychologique, 11*, 191–244; original work published 1905)
- Blankenship, A. B. (1938). Memory span: A review of the literature. *Psychological Bulletin, 35*, 1–25.
- Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review, 10*, 884–889.
- Boorsboom, D., & Mellenbergh, G. J. (2002). True scores, latent variables, and constructs: A comment on Schmidt and Hunter. *Intelligence, 30*, 505–514.
- Brogden, H. E. (1946). Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. *Psychometrika, 11*, 197–214.
- Brown, S. W., Guilford, J. P., & Hoepfner, R. (1966). *A factor analysis of semantic memory abilities: Reports from the psychological laboratory, University of Southern California, 37*. Los Angeles: University of Southern California, Department of Psychology.
- Burke, H. R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *Journal of Genetic Psychology, 93*, 199–228.
- Burke, M. J., & Landis, R. S. (2003). Methodological and conceptual challenges in conducting and interpreting meta-analyses. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 287–309). Mahwah, NJ: Erlbaum.
- †Burns, N. R., Nettelbeck, T., & Cooper, C. J. (1999). Inspection time correlates with general speed of processing but not with fluid ability. *Intelligence, 27*, 37–44.
- 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.
- Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika, 10*, 1–19.
- Carroll, J. B. (1980). *Individual difference relations in psychometric and experimental cognitive tasks* (Tech. Rep. No. 163). Chapel Hill: University of North Carolina, The L. L. Thurstone Psychometric Laboratory.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. (1996). A three-stratum theory of intelligence: Spearman's contribution. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 1–17). Hillsdale, NJ: Erlbaum.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153–193.
- *Chaiken, S. R., Kyllonen, P. C., & Tirre, W. C. (2000). Organization and components of psychomotor ability. *Cognitive Psychology, 40*, 198–226.
- *Cherry, K. E., & LeCompte, D. C. (1999). Age and individual differences influence prospective memory. *Psychology and Aging, 14*, 60–76.
- *Cherry, K. E., & Park, D. C. (1993). Individual difference and contextual variables influence spatial memory in younger and older adults. *Psychology and Aging, 8*, 517–526.
- †Chiang, A., & Atkinson, R. C. (1976). Individual differences and interrelationships among a select set of cognitive skills. *Memory & Cognition, 4*, 661–672.
- †Christal, R. E. (1958). Factor analytic study of visual memory. *Psychological Monographs: General and Applied, 72*(13, Whole No. 466).

- *Cianciolo, A. T. (2001). *Unitary or differentiated ability constructs for describing performance? Investigating individual differences and task characteristics*. Unpublished doctoral thesis, Georgia Institute of Technology, Atlanta.
- †Cohen, R. L., & Sandberg, T. (1980). Intelligence and short-term memory: A clandestine relationship. *Intelligence, 4*, 319–331.
- Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences, 34*, 33–39.
- *†Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.
- *Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4*, 577–590.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (1999). Is Spearman's G determined by speed or working memory capacity? *Psychology, 10*, 74. Retrieved July 11, 2002, from <http://psycprints.ecs.soton.ac.uk/archive/0000709/>
- 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.
- Cowan, N. (1997). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- *†Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., et al. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132*, 113–132.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- *†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., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1–18.
- *Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General, 130*, 208–223.
- 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 re-examined. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 491–508). London: Erlbaum.
- Davies, D. R., Jones, D. M., & Taylor, A. (1984). Selective- and sustained attention tasks: Individual and group differences. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 395–447). Orlando, FL: Academic Press.
- Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.
- †Dempster, F. N., & Cooney, J. B. (1982). Individual differences in digit span, susceptibility to proactive interference, and aptitude/achievement test scores. *Intelligence, 6*, 399–416.
- Derner, G. F., Aborn, M., & Canter, A. H. (1950). The reliability of the Wechsler-Bellevue subtests and scales. *Journal of Consulting Psychology, 14*, 172–179.
- †Diehl, M., Willis, S. L., & Schaie, K. W. (1995). Everyday problem solving in older adults: Observational assessment and cognitive correlates. *Psychology and Aging, 10*, 478–491.
- *†Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Work knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80*, 465–472.
- *Dobson, S. H. (1998). *Precursors of adult age differences in abstract and practical problem solving*. Unpublished doctoral thesis, University of South Carolina, Columbia.
- *Earles, J. L. (1996). Adult age differences in recall of performed and nonperformed items. *Psychology and Aging, 11*, 638–648.
- *Earles, J. L. K., Connor, L. T., Smith, A. D., & Park, D. C. (1997). Interrelations of age, self-reported health, speed, and memory. *Psychology and Aging, 12*, 675–683.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.
- *Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 972–992.
- *†Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology, 82*, 799–804.
- *†Engle, R. W., Tuoholski, 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.
- Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika, 6*, 323–330.
- Ferguson, G. A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology, 10*, 121–131.
- *†Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237–241.
- Fry, F. D. (1931). The correlation of the reverse audito-vocal digit span with the general intelligence and other mental abilities of 308 prisoners of the Eastern State Penitentiary of Pennsylvania. *Psychological Clinic, 19*, 156–164.
- Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition, 28*, 774–782.
- †Garrett, H. E. (1928). The relation of tests of memory and learning to each other and to general intelligence in a highly selected group. *Journal of Educational Psychology, 19*, 601–613.
- *Gilbert, D. K. (1995). *The acquisition and utilization of spatial mental models for young and older adults*. Unpublished doctoral thesis, University of Memphis, Memphis, TN.
- *Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading, 2*, 247–269.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*, 5–26.
- Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin, 53*, 267–293.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guilford, J. P., & Hoepfner, R. (1971). *The analysis of intelligence*. New York: McGraw-Hill.
- Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.
- *Hambrick, D. Z., & Engle, R. W. (2001). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*, 339–387.
- Hebb, D. O. (1942). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society, 85*, 275–292.

- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hilgard, E. R. (1987). *Psychology in America: A historical survey*. San Diego, CA: Harcourt Brace Jovanovich.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois, Urbana.
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*, 242–259.
- Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 61–116). New York: Freeman.
- Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist*, *17*, 475–483.
- †Hundal, P. S., & Horn, J. L. (1977). On the relationships between short-term learning and fluid and crystallized intelligence. *Applied Psychological Measurement*, *1*, 11–21.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Hutton, U. M. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory*, *9*, 383–394.
- †Ichikawa, S. (1983). Verbal memory span, visual memory span, and their correlations with cognitive tasks. *Japanese Psychological Research*, *25*, 173–180.
- †Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, *108*, 151–181.
- Jacobs, J. (1887). Experiments on “prehension.” *Mind*, *12*, 75–79.
- †Jensen, A. R. (1964). Individual differences in learning: Interference factor (Cooperative Research Project No. 1867). Berkeley: University of California, Institute of Human Learning.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jöreskog, K., & Sörbom, D. (2001). LISREL (Version 8.51) [Computer software]. Lincolnwood, IL: Scientific Software International.
- Journal of Educational Psychology. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 123–275.
- *Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology*, *87*, 93–102.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- 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.
- †Karlin, J. E. (1942). A factorial study of auditory function. *Psychometrika*, *7*, 251–279.
- †Kelley, H. P. (1964). Memory abilities: A factor analysis. *Psychometric Monographs*, No. 11.
- Kelly, T. L. (1927). *Interpretation of educational measurements*. Yonkers-on-Hudson, NY: World Book.
- Kelly, T. L. (1928). *Crossroads in the mind of man: A study of differentiable mental abilities*. Stanford, CA: Stanford University Press.
- *†Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, *16*, 312–322.
- Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychological Review*, *1*, 602–609.
- Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory test. *Behavior Research Methods, Instruments, & Computers*, *31*, 429–432.
- Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). New York: Routledge.
- Krueger, F., & Spearman, C. (1907). Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten [The correlation between different intellectual efficiencies]. *Zeitschrift für Psychologie (Leipzig)*, *44*, 50–114.
- Kyllonen, P. C. (1985). *Dimensions of information processing speed* (Tech. Rep. No. AFHRL-TP-84–56). Brooks Air Force Base, TX: Air Force Systems Command.
- *Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-scores model. *Journal of General Psychology*, *120*, 375–405.
- Kyllonen, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49–75). Hillsdale, NJ: Erlbaum.
- Kyllonen, P. C. (2002). *g*: Knowledge, speed, strategies, or working-memory capacity? A systems perspective. In R. J. Sternberg & E. L. Gigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 415–445). Mahwah, NJ: Erlbaum.
- *Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.
- Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.
- *†La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1118–1133.
- *Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A “worst performance” analysis of individual differences. *Intelligence*, *14*, 309–325.
- *Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence*, *12*, 131–147.
- *Larson, G. E., & Saccuzzo, D. P. (1989). Cognitive correlates of general intelligence: Toward a process theory of *g*. *Intelligence*, *13*, 5–31.
- Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology*, *49(A)*, 29–50.
- Lohman, D. F. (1996). Spatial ability and *g*. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Hillsdale, NJ: Erlbaum.
- †Lunneborg, C. E. (1977). Choice reaction time: What role in ability measurement? *Applied Psychological Measurement*, *1*, 309–330.
- Lustig, C., & Hasher, L. (2002). Working memory span: The effect of prior learning. *American Journal of Psychology*, *115*, 89–101.
- Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology: Vol. 1. Matters of public interest* (pp. 3–39). Minneapolis: University of Minnesota Press.
- Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 339–349.
- Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107–127.
- *†Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory for text. *Journal of Educational Psychology*, *75*, 314–318.
- †McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ-R technical manual*. Itasca, IL: Riverside.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*, 108–141.
- Messick, S. (1996). Human abilities and modes of attention: The issue of stylistic consistencies in cognition. In I. Dennis & P. Tapsfield (Eds.),

- Human abilities: Their nature and measurement* (pp. 77–96). Hillsdale, NJ: Erlbaum.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- *†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.
- Mukunda, K. V., & Hall, V. C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence*, *16*, 81–97.
- Münsterberg, H., & Bigham, J. (1894). Memory. *Psychological Review*, *1*, 34–38.
- Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, *86*, 214–255.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1–18). New York: Plenum Press.
- *†Norman, S., Kemper, S., & Kynette, D. (1992). Adults’ reading comprehension: Effects of syntactic complexity and working memory. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *47*, P258–P265.
- *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.
- Penrose, L. S., & Raven, J. C. (1936). A new series of perceptual tests: Preliminary communication. *British Journal of Medical Psychology*, *16*, 97–104.
- *Perlow, R., Jattuso, M., & Moore, D. D. W. (1997). Role of verbal working memory in complex skill acquisition. *Human Performance*, *10*, 283–302.
- Peterson, J. (1925). *Early conceptions and tests of intelligence*. Yonkers-on-Hudson, New York: World Book.
- Phillips, L. H., & Hamilton, C. (2001). The working memory model in adult aging research (pp. 101–125). In J. Andrade (Ed.), *Working memory in perspective*. New York: Psychology Press.
- Psychological Corporation. (1997). *WAIS-III/WMS-III technical manual*. San Antonio, TX: Author.
- †Puckett, J. M., & Kausler, D. H. (1984). Individual differences and models of memory span: A role for memory search rate? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 72–82.
- *Radvansky, G. A., Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults’ use of mental models as representations for text materials. *Psychology and Aging*, *5*, 209–214.
- †Rasile, D. A., Burg, J., & Burreight, R. G. (1995). The relationship between performance on the Gordon Diagnostic System and other measures of attention. *International Journal of Psychology*, *30*, 35–45.
- Raven, J. C., Court, J. H., & Raven, J. (1977). *Raven’s Progressive Matrices and Vocabulary Scales*. New York: Psychological Corporation.
- Richardson, J. T. E. (1996). Evolving concepts of working memory. In J. T. E. Richardson et al. (Eds.), *Working memory and human cognition* (pp. 3–30). New York: Oxford University Press.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition, technical manual*. Itasca, IL: Riverside.
- *†Rosen, V. M., & Engle, R. W. (1997). Forward and backward serial recall. *Intelligence*, *25*, 37–47.
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*, 18–38.
- *Salthouse, T. A. (1992). Why do adult age differences increase with task complexity? *Developmental Psychology*, *28*, 905–918.
- Salthouse, T. A. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology*, *30*, 240–259.
- *Salthouse, T. A., Babcock, R. L., Mitchell, D. R. D., Palmon, R., & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence*, *14*, 187–230.
- *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.
- †Saunders, D. R. (1959). On the dimensionality of the WAIS battery for two groups of normal males. *Psychological Reports*, *5*, 529–541.
- Schrank, F. A., Flanagan, D. P., Woodcock, R. W., & Mascolo, J. T. (2002). *Essentials of WJ III Cognitive Abilities Assessment*. New York: Wiley.
- *†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.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Simpson, B. R. (1912). Correlations of mental abilities. *Teachers College, Columbia University Contribution to Education*, No. 53.
- *Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition*, *20*, 539–548.
- Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47–103). Hillsdale, NJ: Erlbaum.
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.
- Spearman, C. (1914). The theory of two factors. *Psychological Review*, *21*, 101–115.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- Spearman, C. E. (1938). Measurement of intelligence. *Scientia Milano*, *64*, 75–82.
- Spearman, C., & Jones, L. L. W. (1950). *Human ability: A continuation of “The abilities of man.”* London: Macmillan.
- *†Stankov, L. (1994). The complexity effect phenomenon is an epiphenomenon of age-related fluid intelligence decline. *Personality and Individual Differences*, *16*, 265–288.
- Stauffer, J. M., Ree, M. J., & Caretta, T. R. (1996). Cognitive-components tests are not much more than g: An extension of Kyllonen’s analysis. *Journal of General Psychology*, *123*, 193–205.
- Sternberg, R. J. (1986). Haste makes waste versus a stitch in time? A reply to Vernon, Nador, and Kantor. *Intelligence*, *10*, 265–270.
- Tenoppy, M. L., Guilford, J. P., & Hoepfner, R. (1966). *A factor analysis of symbolic memory abilities: Reports from the psychological laboratory, University of Southern California*, 38. Los Angeles: University of Southern California, Department of Psychology.
- Terman, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence*. Boston: Houghton Mifflin.

- Terman, L. M., & Merrill, M. A. (1960). *Stanford–Binet Intelligence Scale: Manual for the third revision, Form L–M*. Boston: Houghton Mifflin.
- Thomson, G. H. (1939). *The factorial analysis of human ability*. Boston: Houghton Mifflin.
- Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York: Science Press.
- Thorndike, E. L. (1940). *Human nature and the social order*. New York: Macmillan.
- Thorndike, R. L. (1947). *Research problems and techniques (Report No. 3). Army Air Forces aviation psychology program research reports #3*. Washington, DC: U.S. Government Printing Office.
- Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior*, 29, 332–339.
- Thorndike, R. L. (1994). *g. Intelligence*, 19, 145–155.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford–Binet Intelligence Scale: Fourth Edition*. Itasca, IL: Riverside.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, 1.
- *Tirre, W. C., & Pena, C. M. (1993). Components of quantitative reasoning: General and group ability factors. *Intelligence*, 17, 501–521.
- *Tucker, P., & Warr, P. (1996). Intelligence, elementary cognitive components, and cognitive styles as predictors of complex task performance. *Personality and Individual Differences*, 21, 91–102.
- *Tuholski, S. W. (1997). *Examining the relationships among attentional resources, working memory, and fluid intelligence*. Unpublished doctoral thesis, University of South Carolina, Columbia.
- *Turner, M. L., & Engle, R. W. (1986). Working memory capacity. *Proceedings of the Human Factors Society*, 30, 1273–1277.
- *Turner, M. L., & Engle, R. W. (1989). Is working memory task dependent? *Journal of Memory and Language*, 28, 127–154.
- Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, 30, 128–134.
- *Verguts, T., & De Boeck, P. (2002). On the correlation between working memory capacity and performance on intelligence tests. *Learning and Individual Differences*, 13, 37–55.
- †Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, 7, 53–70.
- Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.
- Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48, 865–885.
- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49(A), 51–79.
- Wechsler, D. (1944). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.
- Whipple, G. M. (1914). *Manual of mental and physical tests: Pt. 1. Simpler processes* (2nd ed.). Baltimore: Warwick & York.
- Whipple, G. M. (1921). *Manual of mental and physical tests: Pt. 2. Complex processes* (2nd ed.). Baltimore: Warwick & York.
- †Wilding, J., Valentine, E., Marshall, P., & Cook, S. (1999). Memory, IQ, and examination performance. *Educational Psychology*, 19, 117–132.
- †Williams, T. H. (1975). The Wechsler scales: Parents and (male) children. *Journal of Educational Measurement*, 12, 119–128.
- Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 77–108). Washington, DC: American Psychological Association.

Appendix

List of Tests and Tasks by Category for Meta-Analysis

Ability Tests	Vocabulary: Synonyms	Picture Recognition: WJ-R
<i>Verbal Ability</i>	Vocabulary: K-BIT	Shape Series
Air Forces Reading Ability Test	Vocabulary: SAT	Space Relations Test
American National Adult Reading Test	Vocabulary: WAIS-R	Spatial ability composite
Anagram Solution	Word Beginnings: ETS kit	Spatial Content: BIS
Classify Words: GMAT	Word Knowledge: ASVAB	Spatial Orientation
Cloze	Writing Fluency: WJ-R	Spatial Relations
Completion	Writing Sample: WJ-R	Spatial Relations: WJ-R
Comprehension: WAIS-R	<i>Numeric Ability</i>	Surface Development
Dictation: WJ-R	123 Symbol Reduction	Verbal Test of Spatial Ability
English: ACT	Addition	Visual Closure: WJ-R
English Composition Test: CEEB	Applied Problems: WJ-R	<i>Reasoning Ability—Verbal</i>
Fluency (word)	Arithmetic: Subtraction and Division	Abstract Problem Solving
Following Directions: GMAT	Arithmetic Reasoning: ASVAB	Abstract Reasoning: Shipley-Hartford Scale
Homonym Matching	Arithmetic Reasoning Test	Induction—Verbal
Incomplete Words: WJ-R	Arithmetic Test	Inference Test
Information: WAIS-R	Arithmetic: WAIS-R	Inferences From Written Passages: GMAT
Japanese Anagram	Calculation: WJ-R	Integrative Reasoning
Listening Comprehension	CEEB Mathematical Achievement Test	Integrative Verbal Reasoning
Listening Comprehension: WJ-R	(Level 1)	Letter Series
Meaning Identity (word similarities)	Math: ACT	Letter Sets: ETS kit
Multiple Choice Vocabulary Test	Math Knowledge	Miller Analogies Test
Multiple verbal measures	Math Knowledge: ASVAB	Nonsense Syllogisms: ETS kit
Multivariate Ability Battery Comprehension	Multiple numerical measures	Three-Term Series
Multivariate Ability Battery Similarities	Necessary Arithmetic Operations	Verbal Analogies: GMAT
Nelson-Denny Reading Comprehension	Necessary Facts	Verbal Reasoning: EAS
Nelson-Denny Reading Rate Test	Number Reduction	<i>Reasoning Ability—Numeric</i>
Number of Meanings for Sentences	Number Series: GMAT	Arithmetic Reasoning
Opposites for Words: GMAT	Number Triplets	Induction—Quantitative
Oral Vocabulary: WJ-R	Numeric Content: BIS	Number Series
Paragraph Comprehension: ASVAB	Numerical Ability: EAS	Number Series—Nonstandard
Passage Comprehension: WJ-R	Numerical Approximation	Number Sets
Peabody Picture Vocabulary Test	Numerical Reasoning: EAS	Quantitative Concepts: WJ-R
Picture Vocabulary	Problem Solving	<i>Reasoning Ability—Spatial</i>
Proofing: WJ-R	Quantitative: SAT	Diagramming Relationships
Reading: ACT	Quantitative: SCAT	Geometric Analogies
Reading Comprehension	Subtraction and Multiplication	Induction—Spatial
Reading Vocabulary: WJ-R	<i>Spatial Ability</i>	K-BIT Matrices
Similarities: WAIS-R	Block Design	Matrices
Sound Blending: WJ-R	Block Design: WAIS-R	Matrices: IPAT
Synonyms: GMAT	Classify Figures: IPAT	Spatial Analogies
TWSE-SAT	Clocks Test	Topology: IPAT
Verbal Analogies	Coordinate Reading	<i>Reasoning Ability—Raven</i>
Verbal Analogies: WJ-R	Cube Assembly	Raven's Advanced Progressive Matrices
Verbal Comprehension: EAS	Cube Comparison	Raven Standard Progressive Matrices
Verbal Content: BIS	EAS Space Visualization	Raven Standard Progressive Matrices—
Verbal: SAT	Figure Series: IPAT	Nonstandard
Verbal: SCAT	Hidden Figures	<i>Reasoning Ability—General</i>
Vocabulary: Ammon's Quick Test	Maze	Analysis—Synthesis
Vocabulary: Antonyms	Mental Rotation	Analysis—Synthesis: WJ-R
Vocabulary: Educational Testing Service	Minnesota Paper Form Board Test	Concept Formation: WJ-R
(ETS) kit	Multiple spatial measures	Gf composite
Vocabulary: General Aptitude Test Battery	Object Assembly: WAIS-R	Multiple Reasoning
Vocabulary: Jastak & Jastak's short form of	Paper Folding	Reasoning Ability Composite
the Wechsler Adult Intelligence Scale	Paper Form Board Test	
Vocabulary: Mill Hill	Pattern Comprehension	
Vocabulary: Nelson-Denny subtest	Picture Arrangement: WAIS-R	
Vocabulary: Shipley-Hartford Scale	Picture Completion: WAIS-R	

Perceptual Speed

Cancelling Symbols
 Clerical Ability
 Clerical Name
 Clerical Number
 Code Learning
 Coding
 Coding Speed: ASVAB
 Cross Out
 Cross Out: WJ-R
 Dial and Table Reading
 Dial Reading
 Digit and Letter Copying
 Digit Symbol Substitution
 Digit Symbol: WAIS-R
 Directional Headings
 Discrimination (RT)
 Even-Odd
 Factors of 7
 Finding a and t
 Finding yen and set
 Identical Pictures Test
 Larger-Smaller
 Letter Comparison
 Lexical Decision
 Mirror Reading
 Multiple perceptual speed measures
 Name Comparison
 Name Symbol
 Noun Pair (RT)
 Number Comparison
 Number Facts
 Number Sorting
 Numerical Operations: ASVAB
 Pattern Comparison
 Perceptual Speed Composite
 Sequential Figure Matching
 Simultaneous Figure Matching
 Simultaneous String Matching
 Speed of Identification
 Sum to 10
 Visual Matching: WJ-R

Elementary Cognitive Tasks (ECTs)

AB grammatical reasoning
 Arrow grammatical reasoning
 Category identification
 Consonant-vowel
 Meaning identity
 Multiple ECT measures
 Semantic relations verification

Knowledge

Auto and shop information: ASVAB
 Aviation information
 Current events
 Electrical information
 Electronics info: ASVAB
Gc composite
 General knowledge
 General knowledge/information

General mechanics
 General science: ASVAB
 Humanities: WJ-R
 Knowledge survey
 Meaning identity
 Mechanical comprehension: ASVAB
 Mechanical principles
 Multiple knowledge measures
 Science: SAT
 Science: WJ-R
 Social Studies: WJ-R
 Tool functions

g—General Intelligence

ACT: Composite
 AH4
 American Council on Education
 Psychological Exam
 Armed Forces Qualifying Test
 BIS—Aggregate
 Cattell Culture Fair Test
 Cognitive ability composite
 Fluid intelligence (*gf*) composite
 General Certificate of Secondary Education
 Otis IQ
 Henmon-Nelson Test IQ
 K-BIT composite
 SAT—Total
 Swedish DBA Group Test Battery
 Thorndike Intelligence Exam
 Wonderlic Personnel Test

*Working Memory (WM) Tests**Verbal*

ABCD Order
 Alpha Span
 Backward Letter Span
 Listening Span
 Reading Span
 Sentence Span
 Sentence-Word
 Verbal Coordination
 Verbal Span
 Verbal WM
 Verification: Word

Verbal With Numeric

Alphabet Recoding
 Operation Span
 Operation: Word
 Word Span

Verbal With Spatial

Rotation: Word
 Size Judgment Span

Numerical

Backward Digit Span
 Computation Span
 Digit Span
 Math Span
 Mental Arithmetic

Mental Math
 Operation: Digit
 Quantitative WM
 Random Generation

Numerical With Verbal

ABC Assignment
 Complex Span
 Continuous Paired Associates
 Digit Span/Concurrent Task
 Sentence: Digit

Numeric With Spatial

Complex Span—Visuospatial: Verbal
 Counting Span
 Memory Updating
 Mental Counters
 Sequential Memory
 Slots
 Star Counting

Spatial

Complex Span
 Figural-Spatial
 Matrix Memory
 Memory Updating
 Paper Folding
 Pattern Transformation
 Rotation: Arrow
 Spatial Coordination
 Spatial Integration
 Spatial Span (tic-tac-toe)
 Spatial WM
 Synthesis
 Tracking
 WM: Figures

Spatial With Verbal

Complex Span
 Location Span/Concurrent Task
 Spatial Span (letter rotation)
 Verbal Cubes
 Verification: Arrow

Spatial With Numeric

Spatial Span (x and o)
 Spatial WM

*Short-Term Memory (STM) Tests**Verbal STM*

Auditory Fusion Memory Span
 Auditory Memory Span (speech sounds)
 Forward Letter—Dissimilar
 Forward Letter—Similar
 Memory for Instructions
 Memory for Sentences
 Nonsense Syllable Span
 Simple Letter Span
 Simple Word Recall (High)
 Simple Word Span
 STM—Fixed word pool

STM—Fixed: Auditory Suppression
 STM—Unlimited Word Pool
 STM—Unlimited Word Pool: Auditory
 Suppression
 Story Recall
 Visual Fusion Memory Span
 Visual Letter Span
 WJ-R Memory For Sentences
 WJ-R Memory For Words
 Word Recall

Forward Digit Span
 Immediate Digit Span
 Probed Serial Recall
 Repeated Digit Span
 Simple Visual Number Span
 Slow Running Memory
 Visual Forward Digit Span

Face Recognition
 Immediate Color Form Span
 Matrix-Dot Span
 Picture Detail Recall
 Position in Succession
 Position Memory
 Position Recall
 Rey-Osterrieth Complex Visual Figure
 Simple Arrow Span
 Simple Figure Span
 Simple Location Span

Spatial STM

Color Memory
 Color Recall
 Corsi Span Task
 Delayed Color Form Span

Composite

Digit Span-Word Span Composite

Numeric STM

Delayed Digit Span
 Fast Running Memory

Note. GMAT = Graduate Management Admission Test; WAIS-R = Wechsler Adult Intelligence Scale—Revised; WJ-R = Woodcock-Johnson—Revised; CEEB = College Entrance Examination Board; ASVAB = Armed Services Vocational Aptitude Battery; TWSE = Test of Standard Written English; EAS = Employee Aptitude Survey; BIS = Berlin Intelligence Structure; SCAT = School and College Ability Test; K-BIT = Kaufman Brief Intelligence Test; IPAT = Institute for Personality and Ability Testing; *Gf* = fluid intelligence; RT = response time; *Gc* = crystallized intelligence.

Received July 10, 2003

Revision received June 2, 2004

Accepted June 2, 2004 ■

ORDER FORM

Start my 2005 subscription to *Psychological Bulletin!*

ISSN: 0033-2909

_____ \$87.00, **APA MEMBER/AFFILIATE** _____
 _____ \$180.00, **INDIVIDUAL NONMEMBER** _____
 _____ \$465.00, **INSTITUTION** _____
In DC add 5.75% / In MD add 5% sales tax _____
TOTAL AMOUNT ENCLOSED \$ _____

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
 PSYCHOLOGICAL
 ASSOCIATION

SEND THIS ORDER FORM TO:

American Psychological Association
 Subscriptions
 750 First Street, NE
 Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.
 TDD/TTY (202) 336-6123.
 For subscription information, e-mail:
subscriptions@apa.org

Send me a FREE Sample Issue
 Check enclosed (make payable to APA)
Charge my: VISA MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

 Signature (Required for Charge)

BILLING ADDRESS:

Street _____

City _____ State _____ Zip _____

Daytime Phone _____

E-mail _____

SHIP TO:

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____ *BULA15*