Journal of Experimental Psychology: General
2002, Vol. 131, No. 4, 567–589

0096-3445/02/$5.00 DOI: 10.1037//0096-3445.131.4.567

Individual Differences in Working Memory Within a Nomological Network of Cognitive and Perceptual Speed Abilities

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

It has become fashionable to equate constructs of working memory (WM) and general intelligence (*g*). Few investigations have provided direct evidence that WM and *g* measures yield similar ordering of individuals. Correlational investigations have yielded mixed results. The authors assess the construct space for WM and *g* and demonstrate that WM shares substantial variance with perceptual speed (PS) constructs. Thirty-six ability tests representing verbal, numerical, spatial, and PS abilities; the Raven Advanced Progressive Matrices; and 7 WM tests were administered to 135 adults. A nomological representation for WM is provided through a series of cognitive and PS ability models. Construct overlap between PS and WM is further investigated with attention to complexity, processing differences, and practice effects.

Memory measures have played an integral, if not central, role in the assessment of intellectual abilities since the beginning of modern intelligence assessment. Binet and Simon (1905/1961) included three tests that involved “immediate memory” in the battery of 30 or so tests that made up their measure of children’s intelligence. These tests involved (a) repetition of numbers (digit span), (b) a test of picture memory (in which the examinee is shown a set of 13 pictured objects for 30 s and then must generate a list of the items from memory), and (c) visual reproduction memory (in which the examinee studies two line drawings for 10 s and then must reproduce the patterns). For these and other tests to be useful in assessing intelligence from Binet’s perspective, two construct-validation elements must be satisfied. First, the tests must show age differentiation in performance, such that older children perform better than younger children. Second, these tests must be positively associated with other mental tests (e.g., tests of reasoning, judgment, and knowledge). Finally, the tests must be associated with the criterion. For early intellectual ability assessments, the criterion was school success.

Terman (1906) noted that a particular variant of the Ebbinghaus completion test was particularly useful for differentiating between individuals of higher or lower intelligence. In the Ebbinghaus completion test (Ebbinghaus, 1896–1897), a text passage was created with various parts missing (e.g., syllables, words), and the examinee was instructed to fill in the blanks. Such tests were later developed as *cloze* tests (Taylor, 1953) from a perspective of information theory. The completion/cloze tests were identified as excellent measures of intellectual ability (e.g., see Spearman, 1927) and have been generally considered to assess individual differences in fluency and comprehension (see Carroll, 1993). Terman’s variation, which we revisit in the present study, involves reading the complete passage to the examinee first and then presenting the written text, with various words missing, to the examinee to complete. Thus, Terman’s completion test appears to involve not only the fluency and comprehension components of the Ebbinghaus and cloze procedures but also a meaningful memory component.

By 1915, many different variations of immediate and brief memory tests had been introduced. Whipple (1914/1921) described a dozen different *rote memory* tests, including tests of objects, pictures, sentences, words, nonsense syllables, letters, numbers, and sounds. According to Whipple (1914/1921), “The more careful correlational work of the past few years demonstrates at least a fairly good degree of correspondence between immediate memory and either school standing or estimated general intelligence” (p. 194). Many modern omnibus intelligence tests (e.g., see Psychological Corporation, 1997; R. L. Thorndike, Hagen, & Sattler, 1986) include subtests of immediate memory, particularly in the form of forward and backward digit span. These subtests show relatively moderate correlations with total test scores, especially in comparison with other subtests such as information or reasoning. In the Wechsler Adult Intelligence Scales (WAIS), for example, the Digit Span test (forward and backward) correlated only .61 with full-scale IQ (Psychological Corporation, 1997). For comparison purposes, the Information Scale (which is a set of general knowledge questions) correlated .81 with full-scale IQ. For the Stanford–Binet (R. L. Thorndike et al., 1986), the Memory for Objects, Memory for Digits, and Memory for Sentences scales correlated .60, .64, and .72 with overall intelligence, respectively. In contrast, the Vocabulary scale correlated .81 with overall intelligence. In this context, it is fair to ask where the idea came from that immediate memory is synonymous with intelligence, in con-

---

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

This research was supported by Grant No. F49620-01-0170 from the Air Force Office of Scientific Research (Phillip L. Ackerman, principal investigator). We thank Andrew Conway and two anonymous reviewers for insightful comments and suggestions on a previous version of this article.

Correspondence concerning this article should be addressed to Phillip L. Ackerman, School of Psychology, Georgia Institute of Technology, Psychology Building, 274 5th Street, Atlanta, Georgia 30332-0170. E-mail: phillip.ackerman@psych.gatech.edu

trast to constructs such as vocabulary and general information (which clearly provide better estimates of overall intelligence).

Baddeley's Model of Working Memory

It has been suggested that of the current models representing immediate memory, Baddeley's (1986) may well be the most influential (American Psychological Association, 2001; see also Neath, 1998). Basic elements of Baddeley's model include (a) the implication of a common system that functions across a range of tasks and (b) the assumption that capacity is limited, whether by quantity of items to be processed or by time. In contrast to the notion of a single short-term memory store, Baddeley suggested that working memory (WM) included multiple components. The original model (Baddeley, 1986) included a central executive of limited capacity that supervised slave systems identified as the phonological loop and the visuospatial sketchpad. The phonological (or articulatory) loop is used for processing language-based information, and the visuospatial sketchpad is dedicated to visual or spatial memory (i.e., information not based on language).

The essential role of the central executive is as scheduler of the subsidiary systems, applying strategies for processing and integrating information. This introduces a critical distinction between WM and short-term memory. In WM tasks, the central executive may be engaged in additional processing other than simple rehearsal. This occurs when reasoning or learning tasks are paired with memory tasks (e.g., Baddeley & Hitch, 1974). Empirical studies have confirmed through analyses of latent variables (i.e., underlying constructs estimated from multiple observed measures) that short-term memory and WM tasks are separable and result in different relationships with verbal aptitude (Engle, Tuholski, Laughlin, & Conway, 1999) and a fluid intelligence factor (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle et al., 1999).

Subsequent to the original model, it has been suggested that the phonological loop included both a store of language-based information and an active rehearsal mechanism for refreshing information. Similarly, the visuospatial sketchpad may be conceived as including both a visual store and an active scribe (Baddeley & Logie, 1999). With the storage function assigned exclusively to the subsidiary systems, the central executive retained responsibility for monitoring and processing information relevant to the task at hand. Consequently, WM is presumed to play a critical role in all complex cognitive tasks (such as reasoning and comprehending language) that require coordination and maintenance of information.

WM and General Cognitive Ability

With Baddeley's definition of WM as a more complex construct, entailing higher executive processes, memory researchers have sought to identify the relationship of WM with other constructs in the cognitive realm. In an often-cited piece by Kyllonen and Christal (1990), it was suggested that reasoning ability and WM capacity are largely the same. From an information-processing perspective, these researchers proposed a four-sources model of cognitive function for which WM was the central component. They suggested that reasoning ability held a similar principal role in models from the psychometric abilities tradition. In a series of four large-sample studies with Air Force recruits, Kyllonen and Christal sought to evaluate the relationship between constructs of reasoning and WM as well as processing speed and general knowledge. The latent trait models that best fit the data indicated correlations of .80 to .88 for the WM and reasoning factors. The authors suggested that even though their preferred interpretation is that individual differences in reasoning performance are due to differences in WM capacity, an equally plausible argument is that WM capacity is largely determined by reasoning ability. Although the primary message from Kyllonen and Christal's study was the similarity between WM and reasoning, differences between the two constructs were noted. Specifically, reasoning ability correlated more highly with general knowledge and WM correlated with processing speed. This divergence was discussed as follows:

> This replicates results from the previous three studies, and the removal of content variance from the [Reasoning] and WM factors even strengthens the interpretation of this finding. The ability to reason, independent of the content of the reasoning, apparently has more in common with the breadth of general knowledge a person brings to the testing situation. Conversely, working-memory capacity, independent of the content being remembered, has more in common with speed of processing information. (Kyllonen & Christal, 1990, p. 425)

Several researchers have studied the Raven Advanced Progressive Matrices (called the Raven here, unless otherwise noted), thought by some to capture the essence of general fluid intelligence (or Gf, after Cattell, 1943) in conjunction with WM measures. The findings vary considerably, with seemingly different patterns based on the content of the WM test. In studies of undergraduate students, WM measures have shown correlations from .15 for the Reading Span task to .38 for the Counting Span task (Conway et al., 2002). Engle et al. (1999) reported an average correlation of .31 across three WM tests, ranging from .28 for the Reading Span task to .34 for the Operation Span task. In studies of adults representing a broader age range, the correlations of WM performance to the Raven have been reported at .20 for Reading Span and .43 for Computation Span (Jurden, 1995, ages 18–53). Babcock (1994, Study 2, ages 21–83) reported the highest relationship of a WM composite to Raven performance ($r = .55$), but in this case a 20-min time limit was imposed on Set II of the Raven, potentially introducing a speed component not present in ordinary administration. Thus, in terms of actual shared variance (the square of the raw correlation coefficient), the relatively unspeeded administration of the Raven shares between 2% and 14% of the individual-differences variance with WM measures (although the speeded administration of the Raven shares 30% of the variance with WM). All in all, this is not a very impressive communality in the context of claims that these are measures of the same construct.

Other measures of general cognitive ability have also been studied for their association with WM performance. Engle et al. (1999) reported correlations ranging from .24 to .29 between WM tests and Cattell's Culture Fair Test. The same WM measures had generally higher relations with SAT scores, particularly the Operation Span task, which correlated .49 with Verbal SAT and .46 with Quantitative SAT scores. The path coefficient between WM and Gf (derived from scores on the Raven and Cattell tests) in the best-fitting path model for this study was .59. Conway et al. (2002), using the three WM measures utilized by Engle et al., reported correlations of .28 to .37 between WM tasks and Cattell's nonverbal test of intelligence, and a path coefficient of .60 between WM and Gf (also derived from the Raven and Cattell tests). In another study using a composite of performance on Raven Set II and Horn's Leistungs-Prüf-System Reasoning test (Schweizer, 1996), errors on a WM task were inversely related to cognitive ability ($r = -.34$). In sum, the general finding is that WM performance is positively and significantly related to tasks of reasoning or fluid intelligence, but neither correlations nor path coefficients are of the magnitude (i.e., .80s) reported by Kyllonen and Christal (1990).

WM and Processing Speed

WM has sometimes been considered in relation to processing or perceptual speed, but these comparisons are less in evidence than are relations of WM to fluid intelligence. The findings of Kyllonen and Christal (1990), as discussed previously, indicated that WM was more highly related to processing speed than was a reasoning factor. For example, the path diagram presented for Kyllonen and Christal's Study 4 indicated a path coefficient of .47 between factors of WM and processing speed in contrast to a coefficient of .25 between a reasoning factor and processing speed (in the presence of a .82 path loading between reasoning and WM).

In a cross-sectional aging study reported by Salthouse and Meinz (1995), a perceptual speed score formed from a composite of letter and pattern comparison tasks predicted 8.2% of the variance in Reading Span performance, 6.2% of the variance in Computation Span performance, and 9.1% of the variance in a WM composite. A composite reaction time score for digit–digit and digit–symbol substitution speed tasks accounted for less variance in the WM measures, at 5.1% and 4.0% for Reading Span and Computation Span, respectively. Babcock (1994, Study 2) categorized processing speed tasks as having either low cognitive demand (e.g., the Line Marking task) or high cognitive demand (e.g., the WAIS Digit/Symbol substitution test). The pattern of correlations between processing speed and WM and the Raven performance differed depending on this complexity. A WM composite showed correlations of .29 and .59 with the low- and high-demand processing speed measures; similarly, Raven scores correlated at .31 and .56 with the same low- and high-demand processing speed tasks. This effort to differentiate processing speed tasks by level of complexity for simultaneous comparison to WM measures is the only such attempt we found in our review of the literature.

In a comprehensive effort to delineate the construct space of WM, Oberauer, Süβ, Schulze, Wilhelm, and Wittmann (2000) administered 23 computerized WM measures categorized by content (verbal, spatial–figural, and numeric) and function (storage and transformation, supervision, and coordination). To allow for convergent validation of the WM tasks, Oberauer et al. also administered an extensive intelligence battery to the 128 adult participants. The test battery represented functional scales of reasoning ability, speed, memorization, and creativity, with content categories equally represented in each scale.

An exploratory factor analysis of the WM task score intercorrelations resulted in three factors. The first two factors were interpretable as verbal–numerical WM (simultaneous storage/transformation and coordination functions) and spatial–figural WM (storage/transformation and coordination). The third factor had highest loadings from tasks thought to operationalize the supervisory function of WM, but all were speeded tasks, resulting in an ambiguous interpretation of the factor that could point to general speed (Oberauer et al., 2000). Factor 1 (verbal–numerical) correlated highest with the numerical and reasoning intelligence test scales (rs = .46 and .42), and Factor 2 (spatial–figural) correlated highest with reasoning, spatial, and numerical scales (rs = .56, .52, and .48, respectively). The third factor, identified as speed/supervision, related most highly with a speed scale from the intelligence tests at a correlation of .61. It is important to note that although the figural–spatial factor of WM could be distinguished from the verbal content intelligence scale ($r = .08$, *ns*), all WM factors correlated significantly with the speed scale ($rs = .31$ and .19 for the verbal–numerical and spatial factors, all *rs* significant at $p < .05$). The import of Oberauer et al.'s findings is they reinforce the WM–processing speed link observed 10 years earlier by Kyllonen and Christal (1990).

Developmental Explanations for Speed, WM, and Reasoning Relationships

In an exploration of the relations among WM, processing speed, and fluid intelligence in children and young adults, Fry and Hale (1996) proposed a cognitive developmental cascade (for a related review, see Fry & Hale, 2000). In this conception of cognitive development, age-related increases in processing speed mediate improvement in WM performance. In a study of 214 children and young adults (ages 7 to 19), Fry and Hale tested a variety of path models to represent these relations. When speed was not considered in the model, the path from age to the standard (not advanced) Raven test performance was .38, the path from age to a WM factor was .65, and the path from WM to the standard Raven was .39. However, when a speed factor was included in the path analysis, the path from age to WM reduced to .19 as the stronger path was through speed to WM. They found that over 70% of the effect of age on WM was mediated through processing speed, with a nonsignificant path from speed to Raven performance. Fry and Hale surmised that speed has no direct effect on fluid intelligence in these developmental years but that individual differences in speed directly affect WM capacity, which in turn determines fluid intelligence. At the opposite end of the developmental spectrum, Salthouse (1996) contended that age-related decreases in processing speed in later life similarly account for declines in performance on cognitive tasks, including WM measures (e.g., Salthouse & Babcock, 1991).

Processing Speed and Perceptual Speed

From an individual-differences perspective, there has been a general lack of clarity regarding different aspects of speeded processing. In the radex approach proposed by Marshalek and his colleagues (e.g., Marshalek, Lohman, & Snow; 1983; Snow, Kyllonen, & Marshalek, 1984), speeded processes are highly differentiated from one another—and most highly associated with the contents of processing (e.g., spatial, verbal, and numerical).

In contrast, Kyllonen and Christal's (1990) framework specifies that processing speed includes

> (a) encoding speed, the speed with which information makes its way from an initial percept to a representation in working memory; (b) retrieval speed, the speed with which information from one of the long-term memories is deposited in working memory; and (c) response speed, the speed of executing a motor response. (Ackerman & Kyllonen, 1991, p. 216)

According to Kyllonen, even though one can distinguish operationally among these kinds of speeded processing, psychometric measures of these processes tend to be reasonably highly correlated, which suggests a general underlying factor of processing speed (e.g., see Kyllonen, 1985).

Taxonomic Representation of Cognitive and Perceptual Speed Abilities

Over the past five decades, intellectual abilities researchers have reached a general but not quite universal consensus on the structure of cognitive abilities. Figure 1 depicts a generic form of the hierarchical representation of cognitive abilities, with general intelligence (*g*) at the top of the hierarchy. Vernon's (1950) view was that general intelligence accounts for roughly 40% of the total variance of human abilities. Broad content abilities (verbal, numerical, and spatial) constitute the second level of the hierarchy, followed by narrower subcomponents of these abilities at the next level (such as reading comprehension, lexical knowledge, and spelling). A much more detailed representation has been provided by Carroll (1993) using a three-level hierarchy. At the time of Carroll's writing, however, relatively few individual-differences investigations of WM abilities had been published. On the basis of these studies, though, Carroll was skeptical that it would be prudent to equate WM with reasoning abilities (see Carroll, 1993, pp. 646–647).

There are literally hundreds of studies that establish the existence of the broad content factors of verbal, numerical, and spatial abilities (e.g., see Carroll, 1993, for an extensive review). In contrast, there are fewer studies that have investigated perceptual speed (PS) abilities in isolation or in the context of a taxonomic representation. At the most general level, PS abilities represent basic encoding and comparison of stimuli, across a variety of different contents (Ackerman, 1988, 1990). Similar to tests of immediate memory, measures of PS abilities have been frequently incorporated into omnibus assessments of intelligence (e.g., the Digit/Symbol scale on the WAIS). Carroll (1993) suggested that there were at least two PS factors—one related to finding stimuli in isolation, the other related to comparing sets of stimuli. Along with broad content abilities, PS abilities have figured prominently as integral determinants of individual differences during the acquisition and maintenance of skilled performance (Ackerman, 1988, 1990, 1992; Ackerman & Cianciolo, 1999; Ackerman & Kanfer, 1993).

In a series of studies, Ackerman and his colleagues (Ackerman & Cianciolo, 2000; Ackerman & Kanfer, 1993) have suggested that four PS abilities could be derived. The first factor is similar to that suggested by Carroll and is dominated by tests that involve recognition of simple patterns. This factor was named PS–Pattern Recognition. The second factor is similar to the comparison factor described by Carroll. Tests that load on this factor involve scanning, comparison, and lookup processes. The factor was named PS–Scanning. A third factor is best identified as making substantial demands on immediate memory (such as the Digit/Symbol test) and was named PS–Memory. The fourth factor has been identified mostly with two tests: the Army Air Forces Dial Reading Test and the Federal Aviation Administration (FAA) Directional Headings Test (see Ackerman & Kanfer, 1993; Ackerman, Kanfer, & Goff, 1995). These tests involve both traditional PS and additional cognitive components, such as spatial ability and estimation/interpolation, and heightened memory loads. This factor was named PS–Complex. It is important to note that among all four factors of PS abilities, the speed of processing is much more central to performance than it is in the broad content abilities. Traditionally, PS tests are considered as "speed" tests, and many of the content ability tests are constructed as "power" or "level" tests (see E. L. Thorndike, Bregman, Cobb, Woodyard et al., 1926). The key to performance on the PS tests is not whether an individual can answer individual test items correctly (because all of these items could be correctly answered if time limits are removed from the test paradigm), but instead how quickly and accurately the individual can answer multiple items in sequence. It is interesting that

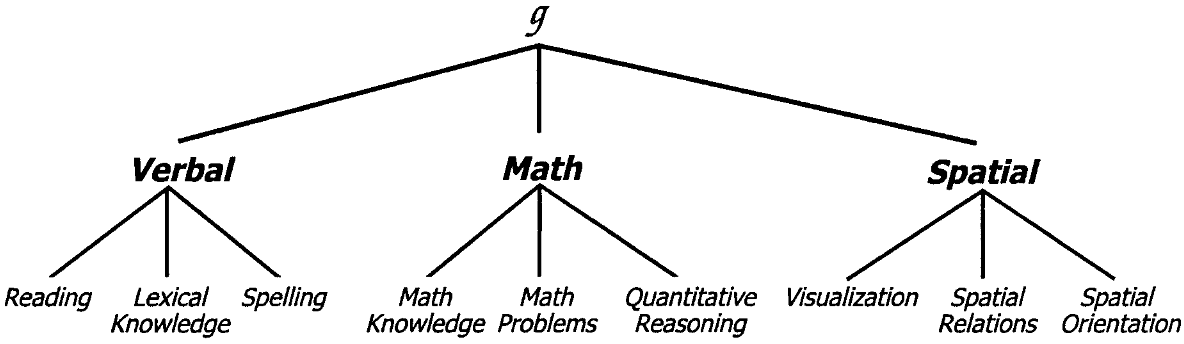

Figure 1. Depiction of a generic form of the hierarchical representation of cognitive abilities, with general intelligence (*g*) at the top of the hierarchy.

various investigators have suggested that PS abilities share relatively little common variance with estimates of *g* (e.g., see Marshalek et al., 1983). While most WM tests have diminished speed requirements, at least when compared with PS tests, it remains to be seen whether these are highly differentiated constructs, as extant theory and practice have assumed.

Present Experiment

The main goal of the present experiment is to better place WM abilities in the context of a nomological network of cognitive and PS abilities. Specifically, this investigation attempts to evaluate the claim of the univocal identification of WM ability with *g*. In addition, the investigation attempts to show that WM ability shares substantial variance with both content abilities and particular PS abilities. There are two major hypotheses (H) related to the placement of WM abilities in the context of other abilities, as follows:

> H1: WM ability will be related to a general intellectual ability but will not be univocally associated with the ability (i.e., there will be substantial variance that is *not* shared by WM and *g*).

As a corollary to H1, the relationship between WM and an estimate of *g* from broad content ability tests will *not* be substantially different from the relationship between WM and performance on the Raven test, which has frequently been mentioned as a pure measure of *g*.

> H2: WM ability will be significantly and substantially related to a general PS ability and more related to PS–Complex and PS–Memory than to PS–Pattern Recognition and PS–Scanning abilities.

Because identification of particular factors with single measures is generally considered by methodologists to be a poor approach to theory testing, a large-scale investigation is required to test these hypotheses. Specifically, an overdetermination of the broad content and PS factors is required (traditionally considered to be at least three marker tests for each underlying factor; see Mulaik, 1972). The experiment described below used at least four tests, and sometimes as many as seven tests, to identify each of the underlying factors. The assessment of the hypotheses, then, is predicated initially on evaluating the degree to which the various marker tests adequately sample the underlying factors. Once this hurdle is passed, we proceed to the evaluation of the specific hypotheses.

Method

Participants

Participants were recruited from undergraduate psychology courses and from the campus at-large at Georgia Institute of Technology (through flyers distributed at random in campus mailboxes). Inclusion criteria were that participants be native English speakers, have normal or corrected-to-normal hearing, vision, and motor coordination, and be between 18 and 30 years of age. One hundred and thirty-five adults participated. The sample had 77 men and 58 women; their mean age was 21.17 years ($SD = 1.77$, range = 18–30 years).

Apparatus

Pencil-and-paper testing was administered in a laboratory with prerecorded instructions and directions presented over a public address system. Up to 16 examinees were tested at a time. Computerized testing for the Noun-Pair tests was administered on Dell and IBM Pentium computers running MS-DOS with 17-in. monitors at individual carrels. Responses were made using the “1” and “2” keys on the numeric keypad. WM tests were administered on Dell Pentium PCs running Windows 98 in separate carrels. Test instructions and auditory stimuli were presented over headphones. The participants responded to test items using the main keyboard (letters) or the numeric keypad (numbers) on a standard IBM keyboard.

WM Tests

Seven commonly used WM tests were adapted for administration in the present study. The tests included a sampling of stimuli that represent alphabetic (words), numeric, and spatial content. Tests other than ABCD Order were composed of three trials at each set size. Each test was preceded by task instructions and at least one example, and trials within task were separated by a 2-s fixation screen.

1. *ABCD Order*. Two categories were used, with five one-syllable words in each category (e.g., the category “trees” contained member words *birch, elm, fir, oak,* and *pine*). Three study frames were displayed for 5 s each. The first frame indicated the order of two members from the same category (e.g., “The pine comes before the elm”); the second frame indicated the order of two members from the second category (e.g., “The rice comes after the beans”); and the third frame indicated the order of the categories (e.g., “The trees come before the food”). After the third study screen, an eight-choice answer screen was displayed from which participants selected the correct order of the words. Participants were allowed 15 s to enter a response (in this example, the correct order is “pine elm beans rice”). The use and ordering of category members were balanced across items, as were the variations of order (i.e., comes before, comes after, does not come before, does not come after). To increase difficulty after observing a ceiling effect in pilot testing, we used two categories and related members for Items 1–12 and two different categories and members for Items 13–24. This test was modeled after the ABCD Order test used in the Cognitive Abilities Measurement (CAM) Battery (Kyllonen, 1988). Each item was equally weighted for scoring purposes.

2. *Alpha Span*. A list of common, one-syllable words was presented auditorily at the rate of one word per second while a “Listen” screen was displayed. At the auditory “Recall” signal, participants were allowed 15 s to type in alphabetical order the first letter of each of the words presented. Set size ranged from three to eight words (18 trials total). Words were randomly selected from a stimulus pool without replacement to form trials, with the restriction that words with the same first letter were not presented together. This test was modeled after the Alpha Span task used in Oberauer et al. (2000). Credit for a perfect trial was given if all first letters were recalled in the correct order, and perfect trials were weighted by set size in computing the final score.

3. *Backward Digit Span*. Digits from 1 through 9 were presented auditorily at the rate of one digit per second while a “Listen” screen was displayed. At the auditory “Recall” signal, participants were allowed 15 s to type in reverse order the digits presented. Set size ranged from three to eight digits (18 trials total). Digits were randomly grouped to form trials, with the restriction that digits did not repeat within a trial. This test was similar to one administered by Oberauer et al. (2000), except that in our version the stimuli were presented auditorily rather than by computer display. Scoring was comparable with the Alpha Span task.

4. *Computation Span*. This test included a verification task and a recall task. Participants were allowed 6 s to verify (true or false) the accuracy of a math equation and were instructed to remember the displayed solution,

regardless of its accuracy. After the final equation of the trial was displayed, participants were prompted to remember in order each of the presented solutions from the equations (e.g., “Enter the first digit”). Each math equation included two operations using digits from 1 through 10, and the provided and actual solutions were always single-digit numbers (e.g., “$(10 / 2) - 4 = 9$”). This task was constructed as a variation of the Computation Span task used in Oberauer et al. (2000), using slightly more difficult equations derived from stimuli used in Cantor and Engle (1993). We restricted our equation elements to be no greater than 10 and limited our solutions to one-digit numbers. Equations were randomly grouped from a stimulus pool to form trials. Set size ranged from three to seven equations/solutions (15 trials total). Credit for a perfect trial was given if all digits were recalled in the correct order, and perfect trials were weighted by set size in computing the final score.

5. *Figural–Spatial Span*. This test included a primary recall task and a secondary verification task. Abstract figures were displayed on screen one at a time for 2 s in various positions along a horizontal line. A probe was displayed below one of the marked positions on the screen, and the participant was allowed 6 s to respond with 1 (*same*) or 2 (*different*) to indicate whether the probe figure matched in color and shape the figure originally presented in that position. Set size ranged from three to six figures displayed (12 trials total), with two probes per trial in the three- and four-figure sets and three probes per trial in the five- and six-figure sets. Credit for a perfect trial was given if all probes for the trial were correctly answered, and perfect trials were weighted by set size in computing the final score.

6. *Spatial Span*. For the secondary task, a 3×3 matrix containing between two and seven red Xs and one blue circle was displayed. Participants were allowed 7 s to make an odd or even judgment about the number of Xs presented in each stimulus. The recall task was to identify the pattern of blue circles formed across the stimuli, selecting from a multiple-choice response screen of four nine-cell matrices with different combinations of blue circles. Participants were allowed 7 s to provide a recognition response. Set size ranged from two to seven stimuli and blue circles in the final configuration (18 trials total). Credit for a perfect trial was given if the correct matrix of circles was identified, and perfect trials were weighted by set size in computing the final score.

7. *Word-Sentence Span*. This test included a sentence verification task and a recall task. Participants were first presented with a common word to study for 2 s for later recall (e.g., *cross* or *train*). Participants were then asked to verify (true or false) the correctness of a sentence displayed for a maximum of 6 s. Recall words and verification sentences alternated through the trial, at the end of which participants were prompted to recall and enter the first two letters of each recall word in the order presented. The test was modeled after one included in the CAM battery (Kyllonen, 1988). Each sentence contained between five and eight words and was of medium length as compared with similar tasks that require recall of the last word of the sentence (see, e.g., Lehto, 1996; Oberauer et al., 2000; Turner & Engle, 1989). Sentences were selected to be easily judged as true or false and require only common knowledge—for example, “A canoe is powered by gasoline.” Recall words and verification sentences were randomly grouped from a stimulus pool to form trials. Set size ranged from two to six words (15 trials total). Credit for a perfect trial was given if the first two letters of all study words were recalled in the correct order, and perfect trials were weighted by set size to compute a final score.

Cognitive Ability Tests

Nineteen cognitive ability tests were selected to serve as markers for three content ability factors: Verbal, Numerical, and Spatial. In addition, we also administered the Raven test—a test that is frequently considered (though not universally so, see Burke, 1958) to univocally represent general intelligence (*g*).

Verbal ability. The Verbal factor includes the following seven tests.

1. *Vocabulary* (Educational Testing Service [ETS] Kit; Ekstrom, French, Harman, & Derman, 1976). This is a classic vocabulary test. Individuals are presented with a word and must choose the word that most closely matches it. This test has two 6-min parts, with 24 items in each part.

2. *Similarities* (Multidimensional Aptitude Battery [MAB]; Jackson, 1985). This is a test of verbal knowledge. Each item presents two words, and participants must select the option that best describes how the two words are alike (e.g., “How are a car and a bicycle alike?”). This test has one part with a 7-min limit and 28 items.

3. *Comprehension* (MAB; Jackson, 1985). This is a test of common cultural knowledge (e.g., “What does it mean if someone is called ‘penny-wise and pound foolish’?”). Each item asks for the correct response to, or the rationale behind, everyday situations, cultural conventions, or practices. This test has one part with a 7-min limit and 34 items.

4. *Word Beginnings* (ETS Kit; Ekstrom et al., 1976). This is a test of verbal fluency. In each test part, participants are given three letters (e.g., “*str*”) and are asked to produce as many words that begin with these letters as time allows. This test has two parts with a time limit of 3 min for each part.

5. *Cloze* (Ackerman, Beier, & Bowen, 2000). Preparation of the Cloze test starts with a text passage of about 250 words in length. Following the technique originated by Taylor (1953), a “structural” (Ohnmacht, Weaver, & Kohler, 1970) Cloze test was constructed. This entailed leaving the first and last sentences of the passage intact, and deleting every 5th word (regardless of its grammatical or contextual relationship), starting with the second sentence. These words were replaced with an underlined blank 10 spaces long. Participants were instructed to read through the passage and fill in the blanks with the words that best fit into the sentence. If participants did not know the exact words, they were instructed to guess. Two points were given for the actual missing word, and 1 point was given for words that fit the gist of the paragraph (and were grammatically correct in the context of the text). Initial administration of the Cloze test had one passage, with a time limit of 10 min. (Two alternative forms of the Cloze tests were administered later in the study.)

6. *Completion* (Ackerman et al., 2000). A procedure identical to that of developing the Cloze test was used to develop the Completion test. The Completion test differs from the Cloze test in administration (based on a design introduced by Terman, 1906). Specifically, participants were instructed to listen to the passage read in its entirety (over the public address system), without looking at the Completion test form. After the passage was read, participants were shown the Completion test form and instructed to fill in as many of the missing words as possible. If they did not remember the exact words, participants were instructed to guess. Scoring for the Completion test was identical to scoring on the Cloze test. Initial administration of the Completion test had one passage, with a time limit of 8 min, after the reading of the passage. (Two alternative forms of the Completion tests were administered later in the study.)

7. *Reading Comprehension* (Nelson–Denny Reading Test; Brown, Fishco, & Hanna, 1993). This test consists of text passages, each followed by a series of multiple-choice questions. There were a total of seven passages and 38 questions in this one-part test, with a time limit of 20 min.

Numerical ability. The Numerical factor includes the following seven tests.

1. *Number Series* (Thurstone, 1962). This is a test of inductive reasoning in which a series of numbers is provided (where the series has been generated by an unstated rule), and the next number in the series is to be identified. The test had one part of 20 items, with a time limit of 4 min.

2. *Problem Solving* (test created by D. F. Lohman; see Ackerman & Kanfer, 1993).[1] This is a test of math word problems. The test had one part of 15 items, with a time limit of 5 min.

3. *Math Knowledge* (Lohman; see Ackerman & Kanfer, 1993). This is a wide-range test of mathematical knowledge, from simple computation to algebra, geometry, and other advanced topics. The test had one part of 32 items, with a 12-min time limit.

4. *Subtraction and Multiplication* (ETS Kit; Ekstrom et al., 1976). This test consists of alternating rows of subtraction (two-digit numbers) and multiplication (two-digit number multiplied by a one-digit number) problems. Two parts with 60 items each were administered in a speeded format with a time limit of 2 min each.

5. *Necessary Facts* (Lohman; see Ackerman & Kanfer, 1993). This is a problem-solving test that does not actually require solution of the problem. Instead, participants must determine whether sufficient information is presented in the problem for a solution to be calculated or what information is missing. Two parts of this test with 10 items and a time limit of 5½ min each were administered.

6. *Arithmetic* (Cureton & Cureton, 1955). This test presents a series of math problems requiring a variety of operations such as adding two fractions and reducing to the lowest term. The most difficult problems require more than one operation. Two parts of 10 items each were administered with a time limit of 4 min for each part.

7. *Numerical Approximation* (locally developed). This test was modeled after the numerical approximation test described in Guilford and Lacey (1947; Test No. CI706A). Each problem requires that the examinee arrive at an estimated answer and then choose from among five possible answers. This test had two parts of 20 items each, with a short time limit (4½ min/part) to discourage exact computations.

Spatial. The Spatial factor includes the following five tests.

1. *Paper Folding* (Lohman; see Ackerman & Kanfer, 1993). This test is an adaptation of other classic tests of the same name (e.g., see Ekstrom et al., 1976). Two parts with 12 items each and a time limit of 6 min/part were administered.

2. *Spatial Analogy* (test created by P. Nichols; see Ackerman & Kanfer, 1993). This is a standard four-term analogical reasoning test, using figural stimuli. The test had one part of 30 items and a time limit of 9 min.

3. *Cube Comparisons* (ETS Kit; Ekstrom et al., 1976). Items in this test illustrate a pair of six-sided cubes, displaying three sides of each cube. Each side is defined as having a different design, letter, or number. For each pair, the task is to determine whether the blocks could be the same or must be different, based on possible rotations and constancy of the markings. This test had two parts, with 21 items in each part and a time limit of 3 min/part.

4. *Verbal Test of Spatial Ability* (Lohman; see Ackerman & Kanfer, 1993). This is a test of image generation and manipulation. Participants are asked to close their eyes and imagine the items described verbally. Then they are asked a multiple-choice question about the items in the image. This test had one part of 24 items and is experimenter-paced. Each item takes about 10 s for the item presentation and 20 s of allowed response time. Total completion time is 12 min.

5. *Spatial Orientation* (Lohman; see Ackerman & Kanfer, 1993). This is a test of three-dimensional visualization. Participants are required to imagine a block figure, as seen from a different perspective. Two 10-item parts of this test were administered, with a time limit of 2½ min/part.

Raven Advanced Progressive Matrices (I + II). The Raven Advanced Progressive Matrices I and II (Raven, Court, & Raven, 1977) are tests of inductive reasoning. Participants are given an item that contains a figure (with three rows and columns) with the lower right-hand entry cut out, along with eight possible alternative solutions. Participants choose the solution that correctly completes the figure (across rows and columns). This test had two parts: a brief Part I (12 items and a 5-min limit) and a longer Part II (36 items and a 40-min time limit).

PS Tests

Based on previous taxonomic research that has established four major factors of PS ability (e.g., see Ackerman & Cianciolo, 2000), we selected 16 PS tests to serve as markers for four PS factors: PS–Scanning, PS–Pattern Recognition, PS–Memory, and PS–Complex. Except where indicated, the tests were locally developed (Ackerman & Cianciolo, 2000),[2] and the initial administration of each test included three separate alternate form parts, with durations of 1.5–2 min/part.

PS–Scanning. The PS–Scanning factor consists of the following four tests.

1. *Name Comparison.* In this test, participants identify identical or mismatched name pairs.

2. *Number Sorting.* In this test, participants find the largest of five large numbers.

3. *Number Comparison.* In this test, participants identify identical or mismatched number pairs.

4. *Noun-Pair.* In this test, participants do a visual lookup of word pairs, variably mapped (15 blocks of 18 trials = 270 trials). For an extensive discussion of this task, see Ackerman and Woltz (1994).

PS–Pattern Recognition. The PS–Pattern Recognition factor consists of the following five tests.

1. *Finding a and t.* In this test, participants scan for instances of "a" and "t" in text passages (passages were in Italian).

2. *Mirror Reading.* In this test, participants find target words written in mirrored text.

3. *Summing to 10.* In this test, participants circle pairs of numbers if they sum to 10.

4. *Finding ∈ and ¥.* This test is the same as Finding a and t, except the text was random symbols.

5. *Canceling Symbols.* In this test, participants scan a page for a single target figure among other simple target figures.

PS–Memory. The PS–Memory factor consists of the following four tests.

1. *Naming Symbols.* In this test, participants write in single-letter code for five different simple figures.

2. *Divide by 7.* In this test, participants circle two-digit numbers if they are exactly divisible by 7.

3. *Coding.* In this test, participants look up and circle a letter or number code for common words.

4. *Digit/Symbol.* In this test, participants put numbers next to symbols corresponding to a lookup key.

PS–Complex. The PS–Complex factor consists of the following three tests.

1. *Dial Reading.* This test of perceptual encoding, memory, and speed was modeled after one portion of the Dial and Table Reading Test designed by the U.S. Army Air Forces Aviation Psychology Research Program (Guilford & Lacey, 1947); one part of the test was administered.

2 and 3. *Directional Headings—Part I and Part II.* This test of memory, perceptual encoding, and learning was modeled after a test designed by the FAA Civil Aeromedical Institute (see Cobb & Mathews, 1972). Participants are given items that include a directional letter, arrow, and degree heading (e.g., S ↑ 180). They must decide the direction implied by these indicators or indicate that conflicting information is presented in the item. In the first part, a conflict is defined as *any* mismatch of indicators. In Part II, the more complex, a conflict exists only if *two* or more indicators have a mismatch. Two parts of the test were administered.

---

[1] Extensive details and examples of items from these tests are provided in the Appendix (pp. 429–432) of Ackerman and Kanfer (1993).

[2] Extensive details and examples of items from these tests are provided in Figure 5 (pp. 269–272) of Ackerman and Cianciolo (2000).

Procedure

The study took place over five 3-hr sessions, totaling 15 hr. The first four sessions were separated by at least 24 hr, and no more than 48 hr. Session 5 was completed 2 weeks after the conclusion of Session 1. Each session included some amount of paper-and-pencil ability testing. Breaks of 6 min were given after every hour of testing. Sessions 1 and 2 included initial testing of all of the paper-and-pencil PS tests, interspersed with content ability tests. WM tests were administered during Session 2, with embedded 5 min breaks after approximately each 20 min of testing. The initial noun-pair testing was performed in Session 3. Sessions 3, 4, and 5 were devoted to retesting of the PS tests and additional content tests. As part of a larger study, during Sessions 2 and 4, additional measures unrelated to the present study were administered. Participants were remunerated $150, or a combination of course credit and cash, for their participation, not contingent on performance.

Table 1

Descriptive Statistics for Ability Tests: Number of Items, Maximum Possible Score, Means, Standard Deviations, and Reliability Estimates

Test	No. of items	Maximum possible score	*M*	*SD*	r_{xx}
Vocabulary	48	48	21.32	6.46	.80[a]
MAB Similarities	34	34	26.36	3.17	.49[a]
MAB Comprehension	28	28	21.62	2.82	.51[a]
Word Beginnings	Open-ended unlimited		29.01	8.58	.68[b]
Nelson–Denny Comprehension	38	38	32.57	3.98	.58[a]
Cloze	37[c]	74[d]	41.54	7.63	.77[a]
Completion	44[c]	88[d]	61.75	10.74	.82[a]
Number Series	20	20	11.43	2.52	.66[a]
Problem Solving	15	15	5.29	2.36	.55[a]
Math Knowledge	32	32	24.11	5.28	.83[a]
Subtraction and Multiplication	120	120	53.40	17.19	.90[b]
Necessary Facts	20	20	10.10	4.10	.70[a]
Arithmetic	20	20	8.37	4.14	.65[b]
Math Approximation	40	40	19.53	5.57	.76[b]
Paper Folding	24	24	14.79	5.91	.84[a]
Spatial Analogy	30	30	19.80	5.11	.84[a]
Verbal Test of Spatial Ability	24	24	13.93	4.75	.74[a]
Spatial Orientation	20	20	9.00	3.69	.61[a]
Cube Comparison	42	42	25.59	8.74	.90[a]
Raven I + II	48	48	37.41	5.71	.83[a]
ABCD Order	24	24	16.35	5.20	.86[a]
Alpha Span	18	99[e]	39.77[f]	18.56	.83[g]
Backward Digit Span	18	99[e]	48.94[f]	16.98	.69[g]
Computation Span	15	75[e]	44.74[f]	15.43	.78[g]
Figural–Spatial	12	54[e]	24.34[f]	9.23	.40[g]
Spatial Span	18	81[e]	50.71[f]	14.23	.68[g]
Word-Sentence Span	15	60[e]	25.55[f]	11.89	.73[g]
Name Comparison	300	300	83.16	18.00	.86[g]
Number Sorting	117	117	47.90	10.17	.87[g]
Number Comparison	300	300	81.60	14.43	.86[g]
Noun Pair (Reaction Time)	75	0	2,459.81	348.34	.91[a]
Naming Symbols	720	720	303.51	58.87	.91[g]
Factors of 7	775	775	140.58	42.39	.92[g]
Coding	264	264	111.78	24.27	.87[g]
Digit/Symbol	660	660	255.05	53.92	.97[g]
Finding a and t	393	393	141.74	30.07	.96[g]
Mirror Reading	300	300	130.30	24.14	.83[g]
Summing to 10	779	779	191.19	31.50	.92[g]
Finding ϵ and ¥	336	336	116.84	26.32	.94[g]
Canceling Symbols	217	217	131.41	27.49	.96[g]
Dial Reading	72	72	28.15	7.02	.72[h]
Directional Headings I	120	120	53.30	15.27	.71[h]
Directional Headings II	120	120	43.50	12.40	.77[h]

Note. MAB = Multidimensional Ability Battery.

[a] Reliability estimate computed as Cronbach's α. [b] Reliability estimate based on Part 1/Part 2 correlation, corrected for total test length by Spearman–Brown prophecy formula. [c] In these tests, 2 points per item are possible. [d] Maximum of 2 points possible for each item. [e] Maximum score based on number of trials, weighted by set size. [f] Score based on perfectly recalled trials weighted by set size. [g] Reliability estimate based on average correlation between three test parts, corrected for total test length by Spearman–Brown prophecy formula. [h] Test–retest (same form) reliability.

Results

Overview

To provide an adequate test of the two main hypotheses, we need to scaffold the analyses. There are seven parts to the Results section. (a) The descriptive statistics of the tests are reviewed first to assess the presence of ceiling or floor effects and to establish the reliabilities of the individual measures. (b) The underlying factor structure of the broad content tests is evaluated to verify that accurate estimates are made of Verbal, Numerical, and Spatial factors and to confirm the presence of a higher order *g* factor. (c) The underlying factor structure of the PS ability measures is evaluated to provide for estimates of the four hypothesized PS factors and a general PS factor. (d) The WM tests are analyzed, with the expectation that a single general WM factor is found. (e) The two hypotheses are evaluated in a single model that includes a *g* factor, the content ability factors, a general PS factor and the constituent PS factors, and the WM factor. (f) We attempt to disentangle the complex relations among PS abilities and WM. (g) Finally, we address an issue relevant to separating test process from test content in the context of a differentiation between Cloze and Completion tests and the WM factor.

Descriptive Statistics

For each of the 43 ability tests, descriptive statistics are presented in Table 1. The table indicates the total number of items and the total possible score (which may be different from the total number of items, either when items are not dichotomously coded as right or wrong only or—as in the case of several of the WM tests—the total score is a weighted function of the number of items scored correctly). Means, standard deviations, and reliabilities are also reported. Generally speaking, reliabilities computed on multiple test parts (e.g., test–retest or alternate form) were the highest (range $= .40$ to $.97$, $M = .84$), whereas internal consistency reliabilities (Cronbach's α) were lower, especially for tests that had heterogeneous content (range $= .49$ to $.91$, $M = .76$). Such values do not indicate that these tests have low reliability, because item homogeneity and reliability are confounded in internal consistency computations. (For example, standard IQ test batteries, such as the WAIS, have relatively low internal consistency reliabilities but have high test–retest reliabilities, congruent with the underlying foundation of the measures as assessing *broad* traits; see Wechsler, 1944.) Intercorrelations among the 43 ability tests are presented in Table 2. The measures in Table 2 clearly show a positive manifold (i.e., the ubiquitous finding of positive correlations among ability measures). Only a small proportion (54 of 861, or 6%) of these correlations are negative (mainly between relatively unspeeded content ability measures and highly speeded PS ability measures). Only two of the negative correlations exceeded an $\alpha = .05$ criterion (the correlation between the Canceling Symbols test and both the Nelson–Denny Comprehension test [$r = –.210$] and the Math Approximation test [$r = –.176$]). The ability intercorrelations are consistent with the extant literature, in the sense that small negative correlations are only found between highly speeded tests on the one hand and complex content tests on the other hand (e.g., see Carroll, 1993).

Content Abilities

The initial analysis of content abilities was an exploratory factor analysis, used to evaluate how well each of the content ability tests served as a marker variable for the three ability factors. A Humphreys and Montanelli (1975) parallel analysis was conducted, indicating three clear factors underlying the correlation matrix. A varimax rotation was performed to maximize the simplicity of the variable loadings. The resulting solution clearly supported the identification of Verbal, Numerical, and Spatial factors. However, one test had relatively poor identification with the intended factor (Word Beginnings), and two of the tests had notable promiscuity; that is, the Problem Solving and Necessary Facts tests loaded significantly on both the Verbal and Spatial factors. Because the factors were themselves overdetermined (with more than three variables per factor), we decided to drop these three tests from further analyses, thereby increasing the coherence of each of the three content factors.

Subsequent to discarding the three tests, we subjected the remaining variables to a confirmatory factor analysis (CFA) using LISREL Version 8.0 (Jöreskog & Sörbom, 1993). We specified each variable to be associated only with its respective content ability factor as identified in the exploratory factor analysis. Thus, there were three factors specified: Verbal, Numerical, and Spatial Ability. The fit of this model was unsatisfactory[3]: $\chi^2(101, N = 135) = 193.75, p < .01$, RMSEA $= .083$, CFI $= .87$. Modification indices revealed that the model fit could be improved by allowing the Verbal test of Spatial Abilities to load on the Verbal factor in addition to the Spatial factor. Because this test has a significant verbal component (as the name suggests), we felt that this loading made theoretical sense and allowed it. A path was also indicated between the Spatial factor and the Number Series test. Because the

---

[3] A significant chi-square index of fit can be an indication of poor fit of the model to the data. However, the chi-square distribution is greatly affected by sample size (the larger the sample, the more likely the chi-square will be significant). Therefore, proponents of the structural equation modeling approach have suggested several different indices of fit be used to augment the evaluation of fit provided by the chi-square statistic (Byrne, 1998). The root-mean-square error of approximation (RMSEA) and the comparative fit index (CFI) were chosen to augment the evaluation of fit for the analysis included here. The RMSEA asks how well the hypothesized model would fit the population covariance matrix with unknown or optimally chosen parameter values (Byrne, 1998). Use of the RMSEA is recommended because it appears to be sensitive to misspecification of the model (Hu & Bentler, 1998; MacCallum & Austin, 2000). In addition, the commonly used guidelines for interpretation of model fit associated with the RMSEA generally support appropriate conclusions about model quality (Hu & Bentler, 1998). These guidelines, as described by Byrne, are that RMSEA values between 0 and .05 indicate very good fit, but values up to .08 can represent reasonable errors of approximation in the population. Values greater than .10 indicate poor fit.

The CFI is in a group of fit indices that compares the fit of the hypothesized model against some standard (e.g., a null model). The CFI was designed to address the bias to underestimate fit shown by the normed fit index (Bentler & Bonett, 1980) in smaller samples (i.e., samples under 200; Bentler, 1990). According to Bentler (1990) the CFI is the best comparative fit index to use for sample sizes such as the one in this study ($N = 135$). CFI values of .90 or above indicate adequate fit to the data (Bentler, 1992; Byrne, 1998).

Table 2
Correlations Among Tests

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Vocabulary	—																
2. MAB Similarities	.515	—															
3. MAB Comprehension	.386	.387	—														
4. Word Beginnings	.297	.270	.134	—													
5. Nelson–Denny Comprehension	.257	.379	.347	.120	—												
6. Cloze	.484	.437	.285	.441	.314	—											
7. Completion	.403	.561	.368	.403	.495	.504	—										
8. Number Series	.247	.218	.189	.360	.172	.345	.220	—									
9. Problem Solving	.204	.375	.393	.172	.299	.263	.253	.376	—								
10. Math Knowledge	.062	.223	−.032	.213	.051	.211	.148	.291	.297	—							
11. Subtraction and Multiplication	.050	.024	−.019	.243	.198	.153	.157	.257	.109	.149	—						
12. Necessary Facts	.276	.362	.419	.183	.280	.344	.269	.412	.413	.284	.020	—					
13. Arithmetic	.180	.167	.077	.224	.181	.198	.133	.324	.312	.387	.602	.202	—				
14. Math Approximation	.217	.231	.088	.312	.190	.311	.227	.428	.461	.597	.334	.405	.504	—			
15. Paper Folding	.301	.374	.336	.148	.257	.300	.313	.381	.396	.164	.032	.607	.230	.300	—		
16. Spatial Analogy	.142	.186	.166	.248	.116	.175	.184	.470	.303	.238	.050	.468	.231	.306	.578	—	
17. Verbal Test of Spatial Ability	.271	.411	.389	.182	.405	.332	.343	.327	.400	.234	.089	.552	.294	.434	.580	.452	—
18. Spatial Orientation	.152	.138	.208	.164	−.004	.070	.125	.265	.284	.051	−.041	.354	.116	.251	.451	.405	.424
19. Cube Comparison	.098	.190	.250	.182	.093	.136	.060	.323	.343	.183	−.059	.507	.119	.238	.503	.573	.477
20. Raven I + II	.212	.226	.245	.150	.193	.292	.246	.314	.353	.165	−.008	.506	.248	.354	.562	.565	.504
21. ABCD Order	.211	.223	.255	.180	.352	.235	.335	.326	.201	.078	.129	.344	.186	.149	.411	.412	.416
22. Alpha Span	.042	.236	.037	.298	.224	.163	.239	.243	.126	.133	.201	.219	.238	.110	.300	.306	.308
23. Backward Digit Span	−.002	.042	.023	.096	.129	.100	.187	.188	.063	.007	.097	.212	.180	.096	.309	.322	.285
24. Computation Span	.211	.247	.078	.200	.238	.113	.284	.117	.028	.186	.277	.164	.283	.229	.211	.249	.216
25. Figural–Spatial	.133	.203	.197	.122	.174	.224	.218	.204	.074	.129	.008	.227	.100	.099	.366	.212	.326
26. Spatial Span	.144	.271	.163	.217	.204	.164	.265	.287	.286	.167	.190	.301	.284	.308	.393	.378	.373
27. Word-Sentence Span	.174	.307	.181	.277	.299	.298	.352	.254	.080	.036	.150	.303	.202	.082	.367	.320	.321
28. Name Comparison	.213	.276	.123	.242	.188	.248	.331	.204	.056	−.066	.480	.061	.277	−.001	.036	.097	.136
29. Number Sorting	.083	.118	.059	.245	.233	.168	.303	.382	.135	.077	.487	.228	.401	.161	.207	.322	.311
30. Number Comparison	−.030	.135	.058	.085	.137	.004	.207	.161	.045	−.049	.409	.164	.269	−.050	.059	.173	.213
31. Noun-Pair[a]	.112	.202	.133	.212	.319	.153	.312	.236	.082	.004	.263	.279	.234	.112	.215	.311	.331
32. Naming Symbols	.072	.163	.105	.206	.199	.176	.267	.205	.137	.152	.242	.182	.233	.038	.208	.264	.315
33. Factors of 7	.056	.112	−.067	.315	.080	.211	.127	.338	.074	.237	.561	.010	.390	.343	−.044	.038	.055
34. Coding	.072	.234	.083	.306	.279	.274	.261	.179	.024	.039	.276	.159	.200	−.039	.266	.197	.237
35. Digit/Symbol	.107	.225	.157	.211	.285	.247	.213	.246	.114	.058	.213	.213	.210	.042	.267	.170	.315
36. Finding a and t	−.143	−.021	−.092	.150	−.015	.072	.087	.125	−.160	−.107	.265	−.003	.056	−.109	.081	.147	.092
37. Mirror Reading	.002	.084	.110	.197	.168	.153	.193	.197	−.063	−.120	.210	.203	.083	−.085	.314	.380	.241
38. Summing to 10	−.104	.045	−.029	.224	.075	.098	.196	.295	.002	−.030	.446	.096	.218	.010	.085	.103	.175
39. Finding ϵ and ¥	−.157	.055	−.007	.189	.090	.035	.142	.112	−.075	−.112	.256	.107	.143	−.095	.147	.232	.143
40. Canceling Symbols	−.067	.059	−.039	.185	−.210	.036	.045	.006	−.053	−.040	−.019	.039	.046	−.176	.196	.236	.088
41. Dial Reading	.186	.184	.277	.147	.307	.210	.212	.462	.345	.196	.155	.550	.304	.352	.483	.445	.493
42. Directional Headings I	.108	.189	.230	.288	.217	.163	.273	.364	.305	.062	.201	.401	.300	.161	.429	.424	.479
43. Directional Headings II	.074	.183	.123	.220	.302	.161	.249	.266	.213	.129	.242	.348	.362	.180	.277	.329	.434

Note. Correlations greater than .17 are significant at $p = .05$, correlations greater than .22 are significant at $p = .01$. MAB = Multidimensional Ability Battery.

[a] The Noun-Pair test scores have been multiplied by −1.0 to express positive associations of performance as positive correlations.

Number Series test required the identification of patterns within a series of numbers (similar to the identification of patterns required in spatial tests), we allowed for this loading as well. A correlated residual was indicated between the Arithmetic test and the Numerical Approximation test that we attributed to the speeded nature of these tests as well as the similarity of the operations used. The model with these modifications is shown in Figure 2. The model fit was quite good—$\chi^2(98, N = 135) = 129.15$, $p < .05$, RMSEA = .049, CFI = .96—suggesting that we had adequately identified three underlying ability factors with the tests used in this study.

PS Abilities

Similar to the method for analyzing the content abilities, we performed an initial exploratory factor analysis on the PS tests. The parallel analysis indicated four clear factors underlying the correlation matrix. A varimax rotation was obtained, and the individual variables were evaluated for promiscuous loadings. Three variables were identified that did not clearly load uniquely on the expected factors (i.e., Noun-Pair, Factors of 7, and Summing to 10). These tests were thus dropped from further analyses, leaving at least three tests for each of the four PS factors.

The remaining variables were subjected to a CFA using a procedure similar to that used for the ability factors, allowing each indicator to load on only one of the four PS factors (PS–Scanning, PS–Complex, PS–Pattern Recognition, and PS–Memory). The fit of this model was again only moderate to poor: $\chi^2(59, N = 135) = 156.16$, $p < .01$, RMSEA = .11, CFI = .89. Modification indices revealed that model fit would be improved by allowing the Finding

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
—																									
.467	—																								
.521	.527	—																							
.235	.354	.355	—																						
.232	.285	.379	.417	—																					
.243	.380	.370	.407	.530	—																				
.078	.214	.241	.408	.343	.304	—																			
.238	.179	.318	.266	.273	.232	.178	—																		
.327	.366	.378	.351	.439	.453	.317	.123	—																	
.185	.226	.225	.460	.453	.416	.538	.135	.377	—																
−.005	.017	−.013	.137	.236	.056	.180	.023	.095	.195	—															
.238	.236	.247	.310	.394	.244	.318	.228	.346	.282	.529	—														
.074	.182	.087	.193	.351	.214	.255	.061	.207	.237	.676	.623	—													
.189	.215	.216	.382	.441	.319	.295	.119	.321	.391	.450	.564	.500	—												
.145	.172	.182	.259	.365	.329	.243	.266	.179	.278	.244	.403	.345	.445	—											
−.065	−.079	−.113	.050	.092	−.066	.158	.051	.066	.053	.284	.386	.174	.216	.234	—										
.130	.097	.192	.273	.259	.123	.247	.280	.155	.288	.303	.432	.329	.421	.447	.246	—									
.109	.126	.243	.244	.294	.235	.197	.372	.207	.217	.117	.296	.228	.351	.653	.240	.549	—								
.007	.138	−.021	.045	.189	.082	.061	.078	.052	−.005	.485	.395	.486	.316	.268	.358	.348	.145	—							
.071	.284	.227	.315	.239	.265	.141	.082	.115	.249	.431	.452	.493	.441	.331	.154	.396	.250	.594	—						
.082	.119	−.005	.153	.307	.231	.097	.122	.152	.159	.369	.511	.532	.436	.416	.455	.295	.358	.558	.417	—					
.056	.266	.031	.216	.260	.248	.194	−.010	.144	.203	.290	.360	.521	.418	.375	.314	.327	.285	.706	.618	.530	—				
.078	.099	−.012	−.053	.171	.103	−.097	.008	.055	.075	.073	.144	.194	.219	.327	.169	.293	.207	.535	.398	.394	.602	—			
.425	.506	.503	.351	.385	.373	.144	.303	.395	.245	.128	.475	.327	.431	.353	.060	.189	.283	.130	.239	.246	.259	.085	—		
.480	.493	.347	.347	.398	.331	.169	.189	.321	.278	.242	.532	.436	.470	.286	.099	.335	.273	.321	.370	.462	.401	.332	.569	—	
.377	.359	.316	.298	.362	.288	.182	.163	.305	.249	.246	.561	.452	.407	.373	.165	.352	.310	.293	.375	.500	.332	.258	.472	.792	—

a and t and the Mirror Reading tests to load on the PS–Scanning factor in addition to the PS–Pattern Recognition factor. These paths seemed logical in that both of these tests use verbal material (letters or words) familiar to users of standard English in the same way that all indicators for the PS–Scanning factor do. The other indicators of PS–Pattern Recognition (i.e., Finding ∈ and ¥ and Canceling Symbols) do not include verbal content. Modification indices also revealed that the model fit would be improved by allowing the Number Sorting test to load on the PS–Complex factor. This loading could be a function of the complexity of the Number Sorting test as compared with the other indicators for the PS–Scanning factor (Name Comparison and Number Comparison tests). The Name Comparison and Number Comparison tests require comparison of only two alternatives, whereas the Number Sorting test requires comparison of five alternatives. The revised model fit was adequate: $\chi^2(56, N = 135) = 123.79$, $p < .01$, RMSEA = .095, CFI = .92; the model is shown in Figure 3. As can be seen in the figure, even though there is substantial commonality among the PS factors (correlations in the .40s and .50s), four PS factors are clearly identified in the model.

WM Abilities

A parallel analysis of a correlation matrix of the seven WM ability tests indicated one underlying factor. Analysis of the loadings of each test on the WM factor indicated that, while they

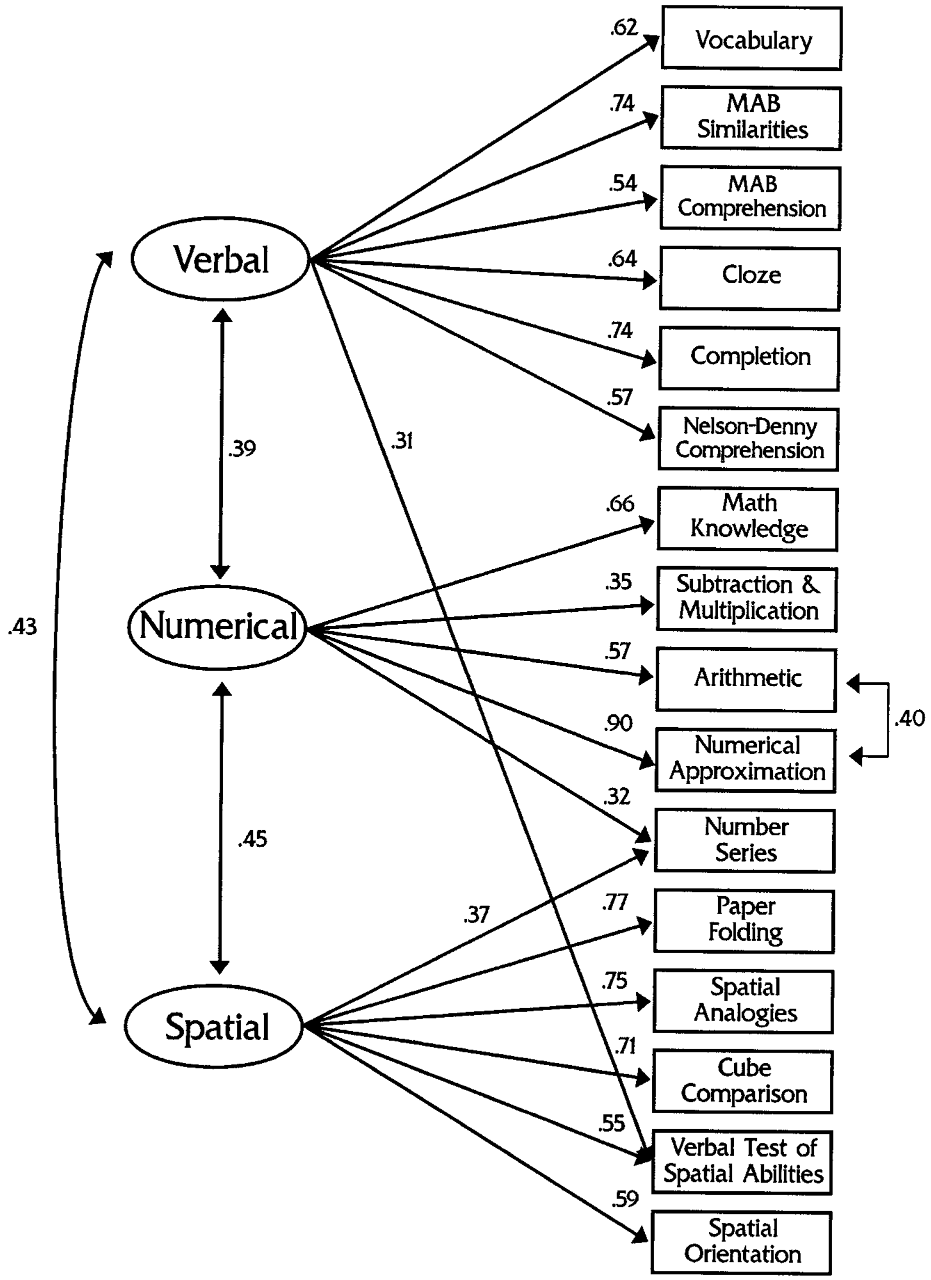

Figure 2. Confirmatory factor analysis of content ability factors. MAB = Multidimensional Aptitude Battery.

differed considerably in their respective associations with the factor, all of the tests had a significant loading (>.30) on the factor. Thus, we retained all seven of the WM tests for subsequent analysis.

A CFA was also performed with the WM tests, in which all of the WM tests loaded on one WM factor. The model fit was quite good, resulting in a nonsignificant chi-square statistic at the $p = .05$ level: $\chi^2(14, N = 135) = 21.57$, $p = .088$, RMSEA = .064, CFI = .97. Modification indices indicated that the model fit could be improved by allowing a correlated residual between the Word Sentence and Computation Span tests. Due to the similarity of the processes used for both of these tests, a correlated residual was added. The resulting model fit was excellent: $\chi^2(13, N = 135) = 10.56$, $p = .65$, RMSEA = .00, CFI = 1.0. The model is shown in Figure 4.

Where Does the Raven Fit In?

Numerous investigators have identified performance on the Raven test as the sine qua non of intelligence (or at least fluid

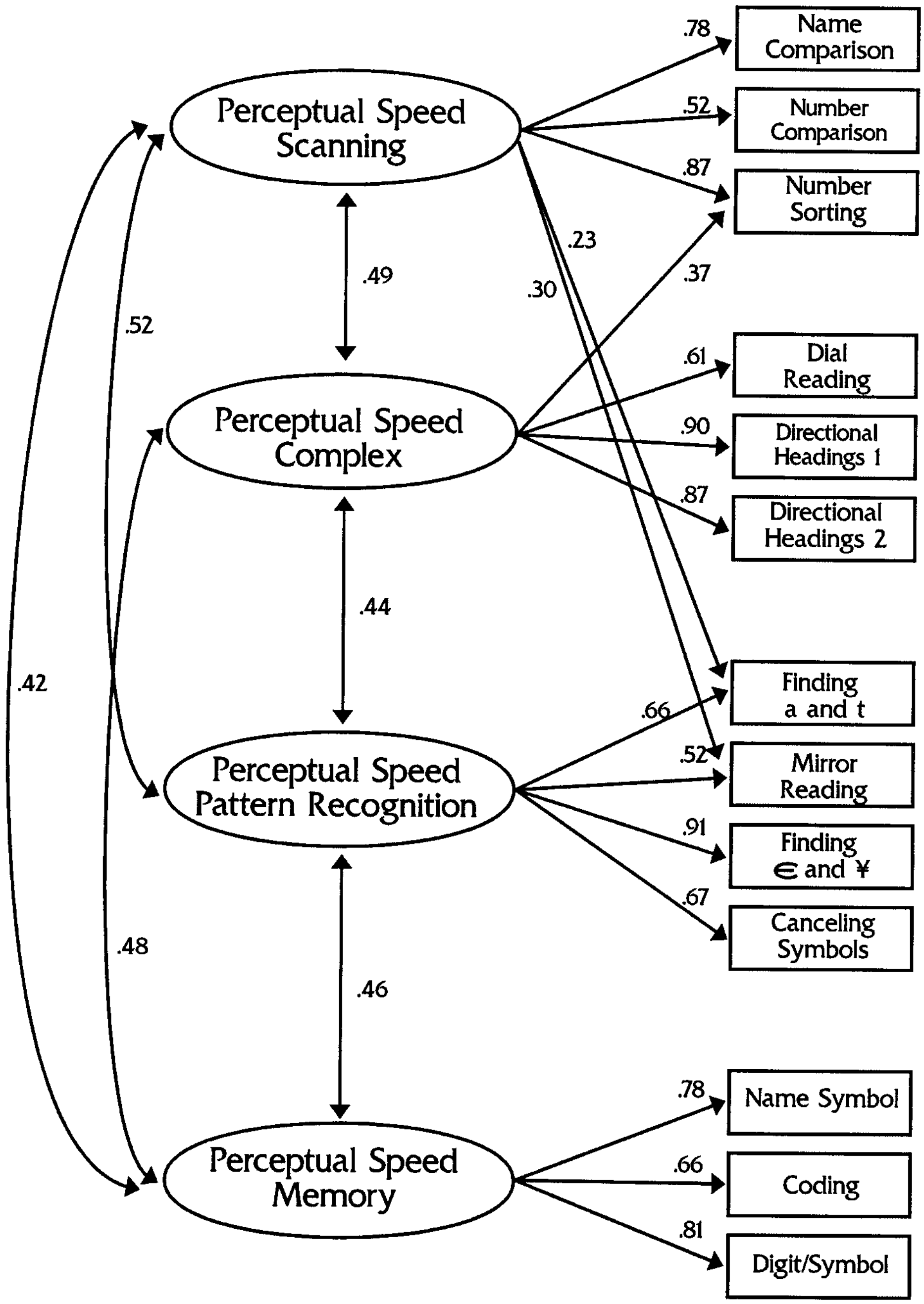

Figure 3. Confirmatory factor analysis for perceptual speed ability factors.

intelligence), since the earliest citations to the test (e.g., Deary & Stough, 1996; Jensen, 1998; Spearman, 1938). Others have been more skeptical about the univocal identification of performance on the Raven test with *g* (e.g., Burke, 1958) and even more skeptical on the utility of the Raven for predicting real-world performance in educational and vocational situations (e.g., Ackerman, 1999b). Nonetheless, there has been sort of a circularity implicit in much recent work associating WM measures, the Raven, and intelligence, where WM is assumed to represent something essential about *g*, and *g* is indexed by Raven performance. Indeed, Carpenter, Just, and Shell (1990) proposed that "the speed of any specific inference process is unlikely to be a major determinant of goal management [on the Raven]" (p. 429). Such an inference is contradicted by earlier research, such as that of Sternberg and Gardner (1983) and similar findings in more recent investigations (e.g., Babcock, 1994; Salthouse, 1996) that show significant associations between processing speed measures and Raven performance.

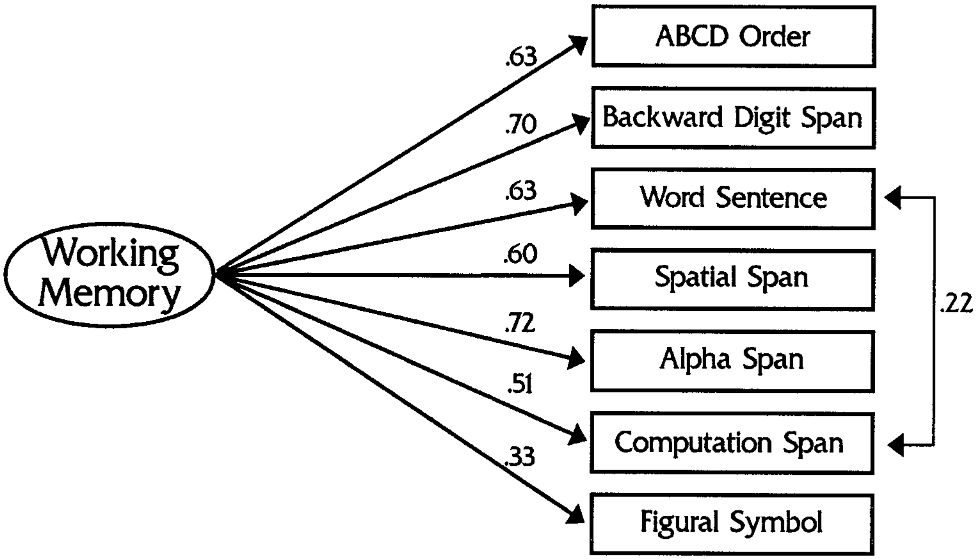

Figure 4. Confirmatory factor analysis for working memory ability.

Table 3 provides correlations among all of the composite ability measures, along with the Raven test. To better locate WM performance in the context of the nomological network of ability constructs, it is informative to compare the correlations among the WM composite, performance on the Raven, and the seven ability factors derived in the earlier analyses. The correlations are presented in Table 4, along with a test for the difference between the respective correlations for the WM composite and the Raven test. The first thing to notice in the table is that the Raven and WM correlate only moderately ($r = .475$). In fact, both variables have higher correlations with the Spatial Ability composite than they do with each other. However, the Raven test has a significantly larger correlation with Spatial Ability than does the WM composite (consistent with Burke's, 1958, analysis that identified the Raven more with spatial abilities than with general intelligence). Although both WM and the Raven were significantly correlated with Verbal and Math abilities, no significant differences were noted in their respective correlations.

Consistent with the corollary to Hypothesis 1, WM and Raven have virtually identical correlations with an aggregated general ability composite (g) from the verbal, math, and spatial composites (.562 and .584, respectively). Similarly, both WM and Raven correlated moderately with the PS–Complex ability. In contrast, WM correlated significantly higher than the Raven with the other three PS factors, suggesting that the implicit speed requirements of the WM tests are more central to WM than they are to the Raven

Table 3
Correlations Among Composite Measures

Test	1	2	3	4	5	6	7	8	9	10	11
1. Verbal	—										
2. Math	.316**	—									
3. Spatial	.398**	.382**	—								
4. *g*	**.752****	**.745****	**.781****	—							
5. PS–Pattern Recognition	.046	.061	.268**	.164	—						
6. PS–Scanning	.255**	.347**	.232**	.368**	.520**	—					
7. PS–Memory	.322**	.261**	.319**	.398**	.439**	.424**	—				
8. PS–Complex	.315**	.408**	.638**	.599**	.404**	.504**	.430**	—			
9. General PS composite	.271**	.358**	.457**	.477**	**.759****	**.803****	**.751****	**.764****	—		
10. Raven	.332**	.301**	.696**	.584**	.074	.117	.253**	.432**	.251**	—	
11. Working Memory	.379**	.352**	.549**	.562**	.229**	.386**	.457**	.479**	.477**	.475**	—

Note. Correlations shown in boldface are part–whole correlations (i.e., correlations between these measures and their constituent composites—Verbal, Math, and Spatial for *g*, and PS–Pattern Recognition, PS–Scanning, PS–Memory, and PS–Complex, with General PS Composite). *g* = general intelligence; PS = perceptual speed.
** $p < .01$.

Table 4

Correlations Between Ability Composites, Working Memory (WM), and Raven Scores

Test	WM	Raven	*t*(diff)
Verbal	.379**	.332**	0.58
Math	.352**	.301**	0.61
Spatial	.549**	.696**	−2.28*
g composite (Verbal + Math + Spatial)	.562**	.584**	−0.31
PS–Pattern Recognition	.229**	.074	1.78*
PS–Scanning	.386**	.117	3.25**
PS–Memory	.457**	.253**	2.53**
PS–Complex	.479**	.432**	0.60
General PS composite (PS–Pattern Recognition + PS–Scanning + PS–Memory + PS–Complex)	.477**	.251**	2.84**
Working Memory		.475**	
Raven	.475**		

Note. *t*(diff) is the *t* test for the difference between correlations ($df = 132$). The correlation between the *g* composite and the General PS composite = .477**. PS = perceptual speed.
* $p < .05$. ** $p < .01$.

(especially when the Raven is administered in a relatively unspeeded fashion [5 min for the 12-item Part I, and 40 min time limit for 36-item Part II of the test, in contrast to the shortened time limit in the Babcock, 1994 studies]). In general, it appears that the WM composite is a promiscuous variable. WM correlates significantly with all of the content ability factors, the PS ability factors, and the Raven. The highest correlations are found with Spatial and PS–Complex factors. The Raven is much more clearly associated with the Spatial Ability factor, though it also is associated with the PS–Complex factor. A composite PS ability illustrates that the overall relationship between the WM composite and PS is significantly greater than the relationship between the Raven and PS. The correlation between the WM composite and the overall *g* composite is slightly larger (but not significantly so) than that of the PS composite and *g* (*r*s = .562 and .477, respectively), $t(132) = 1.15$, *ns*.

Putting It All Together

Although Table 4 provides an initial confirmation of the two main hypotheses (i.e., H1: that WM will be related to general intelligence, but not univocally so; and H2: that WM will be significantly and substantially related to a general PS factor), the raw correlations reported do not take account of multicollinearity among the ability measures. For example, there is a correlation of .477 between the PS composite and the *g* composite, indicating that these composites have substantial common variance. One method to statistically control for this common variance among the measures would be to calculate a series of partial correlations. However, a more unified approach is to test models with simultaneous linear equations (e.g., via a LISREL analysis). Such a method greatly diminishes the number of specific statistical tests and also allows for the assessment of the degree of relative fit between models. We decided to adopt this approach to evaluate the validity of the two hypotheses.

A structural equation model was created incorporating the content ability, PS, and WM factors. Because the number of tests used to identify each factor was large relative to the sample size, composites were created for the Verbal, Numerical, Spatial, PS–Complex, PS–Scanning, PS–Pattern Recognition, and PS–Memory factors. These factor composites consisted of the sum of the unit weighted *z* scores of the indicators for each factor as described earlier in the confirmatory factor analyses. Those indicators loading on more than one factor were included in the composite of the factor with the largest loading. The model is basically a confirmatory factor analysis with a general ability factor (*g*) determined by the verbal, numerical, and spatial composites and the Raven, the WM factor determined by the WM tests, and a PS factor determined by the PS composites. In the model, the PS–Complex composite was allowed to load on both the PS and *g* factors because its constituent tests require more cognitive load than other PS tests (e.g., see Ackerman & Kanfer, 1993; Ackerman et al., 1995). To allow for a test of our second hypothesis (that WM ability would be significantly and substantially related to a general PS ability), we also allowed the WM factor to correlate with both the *g* and PS factors. The fit of this model was only moderate: $\chi^2(85, N = 135) = 148.01$, $p < .01$, RMSEA = .07, CFI = .91. The modification indices indicated that the fit could be improved by allowing correlated residuals between the Numerical ability and the PS–Scanning composites and the Spatial ability and the PS–Pattern Recognition composites. These correlated residuals seemed logical because of the similarity of subject matter of the tests included in each composite (i.e., the PS–Scanning factor consisting of numerical information, the PS–Pattern Recognition factor consisting of tests requiring recognition of symbols). The resulting model fitted well, $\chi^2(83, N = 135) = 129.77$, $p < .01$, RMSEA = .06, CFI = .93, and can be seen in Figure 5.

As a direct test of Hypothesis 2, a chi-square difference test was conducted between the model in Figure 5 and the identical model *without* a path between the WM composite and the PS factor. Eliminating this path significantly worsened the fit of the model, $\chi^2(1, N = 135) = 26.56$, $p < .001$, providing evidence that WM is indeed significantly and substantially related to a general PS ability factor, even after accounting for its relationship to general ability.

Disentangling PS and WM Relations

Demonstrating that PS abilities and WM are substantially related to one another is but one step in the path toward an improved understanding of the nomological connections between the two constructs. Although the present study was not specifically designed to accomplish this goal, the large number and variety of PS tests administered provide a basis for several different approaches to investigate the constructs. Below we assess PS–WM relations through four different approaches: partial correlations, practice effects, analysis by PS complexity, and analysis by PS stimulus–response consistency.

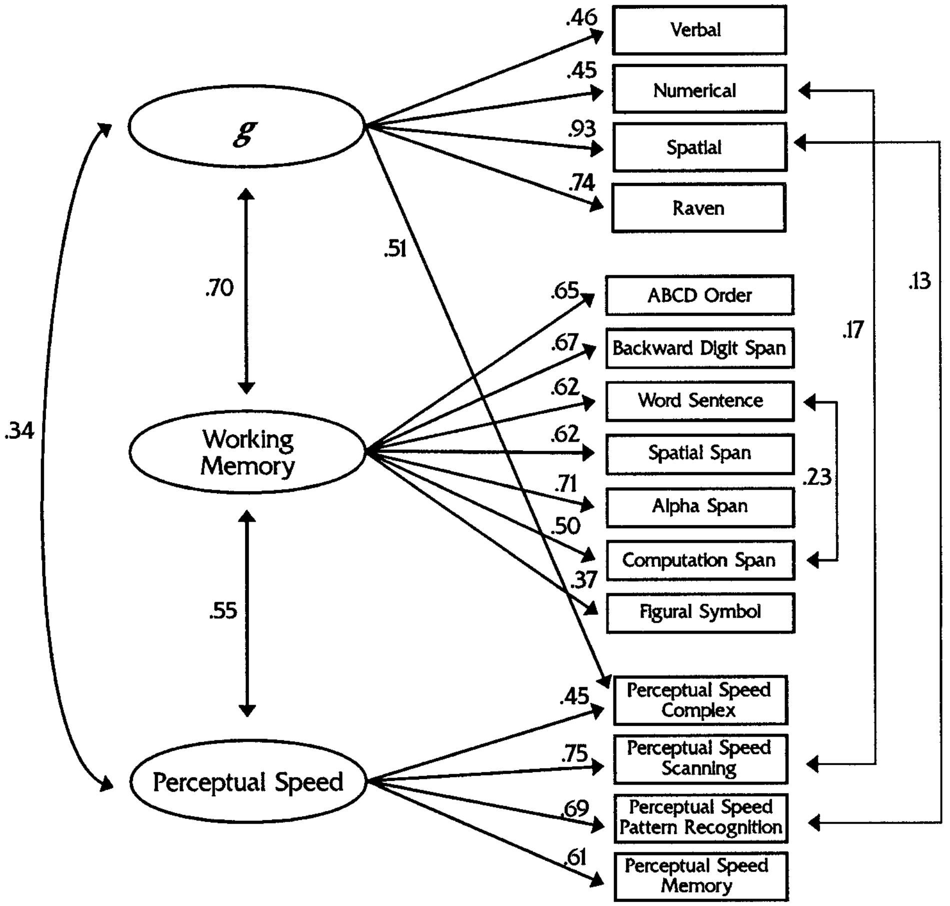

Figure 5. Confirmatory factor analysis of verbal, numerical, and spatial ability, Raven, working memory, and perceptual speed (PS) abilities. g = general ability. Verbal, Numerical, Spatial, PS–Complex, PS–Scanning, PS–Pattern Recognition, and PS–Memory indicators are composite measures.

Partial correlations. One way to examine the common and unique variance of PS abilities, WM, and g is through a set of partial correlations. Partial correlations are used to statistically control the variance in one variable, and then estimate the association between the other variables. At a simplified level, the correlation (squared) between, for example, PS–Pattern Recognition and WM, with g partialed out, yields an estimate of the common variance between PS–Pattern Recognition and WM, that is not also held in common with g. In this fashion, we can evaluate which PS factors share the most variance with WM that is not also shared with g. (Note that in the context of claims that WM is the same as Spearman's g, any significant partial correlation between a PS ability and WM violates "hierarchical order"; see Thomson, 1939, for an extended discussion of hierarchical order.) Also, it is possible to evaluate the correlations of PS abilities with g—this time partialing out the influence of WM. These statistics illuminate which aspects of PS abilities are most associated with g but are independent of WM ability.

The raw and partial correlations for the four PS factors, WM, and g are shown in Table 5. PS–Pattern Recognition has the smallest raw correlations with WM and g. After partialing, there is no significant variance overlap between PS–Pattern Recognition and WM (with g partialed out) or between PS–Pattern Recognition and g (with WM partialed out). Therefore, even though the tests that make up this particular PS factor (Name Comparison, Number Comparison, and Number Sorting) are relatively commonly used as prototypical PS measures, they show no interesting common variance to WM, independent of their association with g. PS–Scanning, in contrast, shows significant common variance with WM, independent of g, and significant common variance with g, independent of WM. However, these are not very differentiated in magnitude, and as such do not provide useful construct differentiation information. Comparison between the partial correlations for PS–Memory and PS–Complex composites do provide more useful information. For the PS–Memory factor, there is substantially more common variance with WM, when g is partialed, than there is with g, with WM partialed. The PS–Complex composite shows the opposite pattern—that is, more common variance with g, when WM is partialed out, in comparison with WM, with g partialed. This pattern suggests that a closer look is appropriate for the distinguishing characteristics of the PS–Memory tests and the PS–Complex tests.

Table 5
Raw and Partial Correlations Between Perceptual Speed (PS) Composites, Working Memory (WM), and the g Composite

	Raw r		Partial r	
Test	WM	g	$r_{PS,WM.g}$	$r_{PS,g.WM}$
PS–Pattern Recognition	.229**	.164	.168	.045
PS–Scanning	.386**	.368**	.236**	.202*
PS–Memory	.457**	.398**	.311**	.198*
PS–Complex	.479**	.599**	.225**	.459**

Note. For calculation of significance for raw correlations, $df = 132$; for partial correlations, $df = 130$.
* $p < .05$. ** $p < .01$.

On the one hand, the three tests making up the PS–Memory composite (Digit/Symbol, Naming Symbols, and Coding) all have the same general processing requirements. In each test, a key is presented to the examinee showing the associative mapping of stimuli to responses. Then, the examinee is presented with a series of stimulus prompts and a set of corresponding blanks. In addition to the common processes of encoding and responding, participants who can first remember the location of the stimuli in the key arrays, and then those participants who can memorize the stimulus–response associations quickly, are at a distinct advantage in overall test performance. The tests making up the PS–Complex factor are more varied in processing requirements (Directional Headings I, Directional Headings II, and Dial Reading). In the Directional Headings tests, the examinee must integrate verbal (e.g., North) and spatial information (e.g., ↑ and 90°), and determine whether they provide consistent or inconsistent information. In the Dial Reading test, examinees must iteratively look up and occasionally interpolate information on a graphical display. What is common to these tests is the spatial content and the iterative or sequential nature of comparisons that need to be made in answering the questions. The spatial content of the tests is clear from their respectively higher correlations with Spatial tests than with Verbal or Numerical tests. However, even though both tests have high commonality with *g*, they also have high commonality with WM. It is only from the perspective of the partial correlations that it is clear that they share more variance uniquely with *g* than with WM. Together, these partial correlation results suggest that the defining characteristics of PS–WM commonality, which is not attributable to their common associations with *g*, are in the context of *associative learning*.

Practice effects. One of the defining characteristics of PS ability tests is the presence of large performance improvements with practice (e.g., see Ackerman, 1988, 1990). Practice on PS tests generally reduces the association between the tests and *g*, a fact that is generally attributed to the following factors: (a) learning the general strategy for test performance and refining general encoding and responding procedures, and (b) memorizing stimulus–response associations when possible. If the associations between PS abilities, WM, and *g* become differentiated with PS test practice, it may be possible to further illuminate the common and unique aspects of the PS abilities that are associated with WM.

Because we repeated the administration of the PS tests over a 2-day to 2-week delay period, it is possible to evaluate whether the effects of practice differentially affect the correlations between PS abilities on the one hand and WM and *g* composites on the other hand. It should be noted that although each test allowed for as many as a couple of hundred responses to target stimuli, we did not assess whether asymptotic performance levels were reached on these tests. Table 6 provides the initial and postpractice correlations between the four PS factor composites and both *g* and WM composites, along with tests of the differences in correlations over practice. As expected, there was a general reduction in association between PS composites and both *g* and WM, although significant reductions in correlations were only found for PS–Complex for *g* and for PS–Pattern Recognition for WM. The PS–Pattern Recognition composite, while having the smallest association with WM in the first administration, showed the only significant reduction in

Table 6
Correlations Between Perceptual Speed (PS) Abilities, Working Memory (WM), and g Composites Over PS Practice; Tests of Differences in Correlations

	r_{PS}				r_{PS}			
Test	1st admin with g	2nd admin with g	t(diff)	d	1st admin with WM	2nd admin with WM	t(diff)	d
PS–Pattern Recognition	.164	.116	1.52	.26	.229**	.174*	1.77*	.31
PS–Scanning	.368**	.331**	0.86	.15	.386**	.361**	0.59	.10
PS–Memory	.398**	.347**	1.28	.22	.457**	.452**	0.13	.02
PS–Complex	.599**	.471**	3.00**	.52	.479**	.403**	1.63	.28
Correlations between 1st and 2nd administrations of PS measures								
PS–Pattern Recognition	.932**							
PS–Scanning	.860**							
PS–Memory	.876**							
PS–Complex	.814**							

Notes. For t tests, $df = 132$. d = Cohen's d effect size statistic; d values of .20–.49 are considered small effects and .50–.79 are medium-sized effects (see Cohen, 1988).
* $p < .05$. ** $p < .01$.

association with WM after practice on the PS tests. The PS–Complex composite showed the only significant reduction in association with *g* after practice, and a small but nonsignificant reduction in association with WM after practice. The PS–Memory composite was the most robust in pre–post practice correlations with WM, while showing a small, nonsignificant decline in correlations with *g*. Although we are not asserting the validity of a null hypothesis of "no change" in association between PS–Memory and WM ability over practice, the demonstrated stability (especially in the context of a small decline in correlations with *g*) of these correlations, taken together with the previous analysis of partial correlations, again suggests an important and stable commonality between the measures of associative learning and the WM construct.

Analysis by PS complexity. If WM represents the efficacy of a central executive (e.g., see Engle, 2002), it is reasonable to expect that as PS test complexity increases, the correlation between the PS tests and WM will concomitantly increase. However, complexity of tasks is notoriously difficult to directly assess (e.g., see Wood, 1986), and a common surrogate for complexity has been reaction time or completion time. That is, complexity of test items can be roughly equated with the amount of time it takes to answer a single item (e.g., see Kyllonen, 1985). Obviously, this approach has limited generalizability, in that it is most applicable for tests with high-frequency targets and not at all applicable for tasks with low-frequency targets, such as vigilance tasks. To approach the relationship between PS tests and WM, we recomputed PS test performance in terms of response time (RT), which is identified as the average time to complete a single item on a PS test. It is important to note that this involves a reciprocal transformation (e.g., total test time divided by the total number of correct responses) and is a nonlinear transformation of the typical attainment scores for ability tests (for an extensive discussion of the ramifications of this transformation on ability test scores, see Ackerman, 1987). For the 16 separate PS tests, the means (in RT), standard deviations, and respective correlations with both WM and *g* composites are provided in Table 7.

At the outset, it is notable that the Dial Reading test is a clear outlier, in terms of both mean and between-individual standard deviation estimates, in comparison with the other 15 PS tests. For example, a comparison between Dial Reading and the other PS tests shows the mean Dial Reading performance to be nearly 12 standard deviations above the overall PS test mean RTs ($z = 11.90$). As such, we have excluded the Dial Reading test from the computations of correlations with WM and *g*. Correlations between mean PS test performance and the two ability composites confirm what is apparent from a visual review of the table entries. That is, there is a weak (nonsignificant) relationship between the average RT on the PS tests and their respective associations with WM ($r = .279$). There is a somewhat larger correlation between RT on the PS tests and *g* ($r = .408$), but this too is not statistically significant, due in part to the small number of degrees of freedom ($df = 13$). Only the association between the absolute magnitude of individual-differences dispersion on the PS tests (*SD*) and *g* yielded a significant correlation ($r = .516$, $p < .05$), which is consistent with the concept that higher *g*-loaded tests tend to induce a greater magnitude of individual-differences variance. It should be kept in mind that RT is an imperfect surrogate measure of complexity. Nonetheless, from these results we can conclude that complexity of the PS tests is not strongly related to the magnitude of association of test performance with WM.

Table 7

Perceptual Speed Test Performance in Terms of Reaction Time (RT), and Correlations With Working Memory (WM) and g

Test	Mean RT	*SD*	r_{WM}	r_g
Naming Symbols	923	181	.375**	.294**
Digit/Symbol	1,109	252	.380**	.329**
Summing to 10	1,450	233	.256**	.204*
Factors of 7	2,110	683	.088	.291**
Mirror Reading	2,148	430	.264**	.243**
Canceling Symbols	2,148	464	.062	.047
Noun-Pair (RT)	2,460	348	.468**	.378**
Coding	2,529	546	.287**	.278**
Finding a and t	2,653	568	.147	.073
Finding ∊ and ¥	3,244	773	.302**	.140
Name Comparison	3,406	796	.248**	.308**
Number Comparison	3,410	595	.326**	.221**
Directional Headings I	3,711	1,302	.361**	.464**
Directional Headings II	4,479	1,287	.325**	.452**
Number Sorting	5,879	1,199	.445**	.410**
Dial Reading	18,288	5,380	.430**	.542**
Overall averages	3,756	4,077		
Averages without Dial Reading	2,777	1,303		

Note. For correlations, all have been multiplied by −1.0, so that positive correlations indicate a positive association between the speed of processing and scores on the ability composites. Mean RTs with working memory, $r = .279$; Mean RTs with *g*, $r = .408$; standard deviations with working memory, $r = .139$; standard deviations with *g*, $r = .516^*$. *g* = general intelligence composite; RT mean and *SD* provided in millisecond units.
$^* p < .05$. $^{**} p < .01$.

Analysis by PS stimulus–response consistency. To address the notion that the most central component of WM is controlled processing (e.g., see Kane, Bleckley, Conway, & Engle, 2001), it is possible to evaluate the degree of association between PS tests and WM, dependent at least to some degree on the underlying consistency of stimulus–response mappings of test items. For this analysis, as in the analysis of complexity issues, we no longer use the previously derived PS factors bur rather the individual tests. The identification of specific PS tests and their underlying mappings is discussed in detail in Ackerman and Cianciolo (2000). Consistent with the theory of automatic and controlled processing (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), we identify variable stimulus–response mappings as those most associated with controlled processing, and those with consistent stimulus–response mappings as most associated with allowing development of automatic processing. Table 8 shows the means, standard deviations, effect size for change over PS test administrations, and the respective correlations with the WM composite over administrations, broken down by the consistency of stimulus–response mappings of the PS tests.

There are three central aspects of the table to note. First, each of the PS tests indicated a significant and substantial practice effect over administrations (with effect sizes ranging from medium to very large), though the consistently mapped PS tests had substantially larger mean gains in performance (average *d* across

Table 8
Perceptual Speed (PS) Test Means, Standard Deviations Over Two Administrations, and Correlations With Working Memory

Test	1st PS admin		2nd PS admin			1st PS admin, r_{WM}	2nd PS admin, r_{WM}
	M	*SD*	*M*	*SD*	*d*		
Variably mapped tests							
Name Comparison	83.16	18.00	95.21	20.23	0.932	.194*	.207*
Number Sorting	47.90	10.17	50.85	13.53	0.561	.453**	.495**
Number Comparison	81.60	14.43	98.54	20.94	1.145	.327**	.201*
Noun-Pair (RT)	2,459.81	348.34	2,096.85	297.60	1.644	.468**	.391**
Dial Reading	28.15	7.02	35.47	7.98	1.316	.460**	.427**
Consistent, but requiring conjunctive searches							
Directional Headings I	53.30	15.27	69.43	19.54	1.165	.420**	.312**
Directional Headings II	43.50	12.40	50.97	14.84	0.762	.387**	.360**
Finding a and t	141.74	30.07	162.04	33.26	1.268	.115	.067
Finding ∊ and ¥	116.84	26.32	142.78	29.15	1.674	.258**	.242**
Consistently mapped tests, ordered by size of memory/target set (in parentheses)							
Coding (14)	111.78	24.27	144.70	28.80	1.696	.338**	.358**
Factors of 7 (14)	140.58	42.39	176.95	45.82	1.568	.087	.122
Summing to 10 (10)	191.19	31.50	226.83	36.45	1.827	.268**	.248**
Digit/Symbol (9)	255.05	53.92	294.30	66.94	1.232	.378**	.408**
Mirror Reading (9)	130.30	24.14	146.64	23.40	1.428	.269**	.214*
Naming Symbols (5)	303.51	58.87	380.96	64.71	1.916	.414**	.392**
Canceling Symbols (1)	131.41	27.49	150.23	28.26	1.485	.080	.044

Note. admin = administration; WM = Working Memory; *d* = Cohen's *d;* RT = reaction time.
* $p < .05$. ** $p < .01$.

tests = 1.59 and 1.12, for consistently mapped and variably mapped PS tests, respectively). Second, for both the variably mapped and consistently mapped tests, there was substantial heterogeneity in initial correlations with WM. It is clear that the consistency of mapping for the PS tests is not univocally associated with the underlying communality of PS with WM. Third, although the initial mean correlations for variably mapped PS tests were higher than for consistently mapped PS tests (mean *r*s = .385 and .266, respectively), changes in correlations with WM across practice were minimal for both types of tests (second administration mean $r = .350$ and .260, respectively). A separate category of PS tests with consistent but conjunctive searches showed results in between those of the variably mapped and consistently mapped PS tests initially but a slightly larger decline in correlations with WM over practice (average $r = .299$ for the first administration, .253 for the second administration).

Process Distinctions and WM

A separate analysis of the Cloze and Completion test data with the WM composite illustrates both the promise and pitfalls of validating process distinctions in an individual-differences paradigm (e.g., see Underwood, 1975). From a process perspective, the fundamental distinction between the Cloze and Completion tests administered in this study was that the Cloze test involves only comprehension and fluency, whereas the Completion test also involves memory (because the participants heard each passage read aloud immediately prior to completing the Completion test). A straightforward expectation is that, *ceteris paribus* (i.e., everything else being equal), scores on the Completion test should correlate more highly with WM ability than the Cloze test. In the first administration of each test, Cloze and Completion test performance correlated .26 and .37 with WM, respectively. Even though the two tests are substantially intercorrelated ($r = .50$), the difference is marginally significant with a one-tailed *t* test, $t(132) = 1.37$, $p = .09$. For the second administration of Cloze and Completion tests, the correlations and differences were of a slightly smaller magnitude, just enough smaller to yield a nonsignificant difference (*r*s = .29 and .37 with WM, and .56 with each other), $t(132) = -0.97$, *ns*. The third administration, 2 weeks after the first tests, had fewer participants, yet yielded a similar set of results (*r*s = .19 and .26 with WM, .57 with each other), $t(122) = -0.85$, *ns*. Aggregating across all three repetitions, with a constant sample, demonstrates the principle of aggregation (i.e., the correlations between the Cloze and Completion tests increase as the reliabilities of the individual measures increase with the length of the test). For the aggregated data, the correlations for Cloze and Completion composite scores with WM were .31 and .40, respectively, but the correlation between Cloze and Completion composites increased to .73. The difference between correlations with WM is only marginally significant with a one-tailed test, $t(122) = 1.47$, $p = .07$.

For the individual tests, the average amount of difference in shared variance between the Cloze test and WM and the Completion test and WM was 4.93%, a difference that was marginally significant in the present study of over 100 participants, and clearly a difference that will not be significant in smaller samples. Thus, even though these data generally support the proposition that two tests differing in the implied involvement of WM determinants of performance also differ in their associations with WM, they also show that such differences turn out to be relatively small (especially in comparison with the other ability determinants of performance on the respective tests). As noted by Ackerman (1986, 1999a, 2000), the *content* of a test (in this case, verbal content) is much more frequently found to be the major determinant of individual differences in test performance, resulting in a diminished involvement of the underlying processes.

Discussion

WM and *g*

Over the past hundred years or so, there have been substantial shifts in both the operationalization and the definition of the construct of general intelligence, or *g*. Initially, Spearman identified course grades in the classics (e.g., Greek and Latin) and peer ratings of "common sense" as the domains that had the highest *g* saturation (Spearman, 1904). Later, Spearman modified this specification to note that the Ebbinghaus completion test (similar to the Cloze and Completion tests used in this study) had the highest *g* saturation (Krueger & Spearman, 1907). From this orientation, it was reasonable that such measures of *g* would be closely identified with omnibus tests of intelligence, such as the Stanford–Binet, which uses a modified completion test as one scale.

From a construct orientation, Spearman identified *g* as a "mental engine" (Spearman, 1914) that could be used for a wide range of intellectual tasks. Much later, and without substantive empirical justification, Spearman changed course and identified *g* as synonymous with performance on the test developed by Penrose and Raven (see Spearman, 1938), which was later developed into the Raven's Progressive Matrices Test. The main conceptual and operational advantage of this test was that it was nonverbal and, as such, could be (at least theoretically) administered to individuals from widely differing backgrounds and cultures. Subsequent investigators, many of them adherents of Spearman's later views of intelligence, have selected the Raven as a univocal operationalization of intelligence or at least fluid intelligence, Gf (Cattell, 1943; see, e.g., Deary & Stough, 1996; Jensen, 1998).

There are two problems inherent in this approach. First, if the Raven is not an exemplary measure of general intelligence (or even Gf), any corroborations between experimental measures (such as WM) and Raven performance are apt to miss important variance associated with intelligence and result in a distortion of construct validity. Second, given that the Raven shows much lower correlations with real-world learning and achievement behaviors than do omnibus IQ measures (such as the Stanford–Binet and Wechsler tests), validation against the Raven will potentially miss criterion-related validity as well.

From a construct validation perspective, then, measures of WM come up short in comparison with the Raven (e.g., in the present study, the correlation between the WM composite and Raven performance was .475, although some restriction of range is apparent in most such studies with college and university students). One clear reason for the lack of equivalence between WM and Raven performance is that WM tests were significantly more highly related to PS abilities ($r = .477$) in comparison with the Raven ($r = .251$). Such results are consistent with other studies that have shown WM to be correlated with various individual measures of PS and similar processing speed measures (e.g., see Kyllonen & Christal, 1990).

It is certainly reasonable to argue that speed of processing is important to intelligence (e.g., see E. L. Thorndike et al., 1926), in which case WM measures, in the aggregate, turn out to be essentially equivalent markers to the Raven for general intelligence (*r*s = .562 and .584, respectively, in the present study), though WM is more highly associated with speed than is the Raven. The problem with the association between WM and speed, however, is that it undermines current theories of WM that identify the underlying construct as "closely associated with general fluid intelligence" and "maybe isomorphic to, general intelligence and executive function" (Engle, 2002, pp. 21–22). In other words, current theoretical perspectives fail to adequately account for the strong but differential associations of WM with different aspects of speed of processing.

Perceptual Speed, Processing Speed, WM, and *g*

Several investigators have speculated on the underlying causal mechanisms for the previously demonstrated communality among measures of WM and *g*. Investigators such as Kail and Salthouse (1994) and Jensen (1998) have suggested that processing speed limits WM and that processing speed is the causal mechanism underlying WM and *g*. Others, such as Cowan (1998), have suggested that it is much more difficult to establish the direction of the causal arrow (e.g., that WM capacity sets limits on processing speed or vice versa), while still others suggest that WM capacity is indeed the underlying causal factor (e.g., Conway, Kane, & Engle, 1999).

In light of the present study and extant data on human abilities, however, we find each of these positions to fall short of an adequate explanation. There are several reasons for this assertion, as follows. As demonstrated in the present study, processing speed (or perceptual speed) represents a much more heterogeneous category of human individual differences than can be captured by a single underlying factor construct. In this study (and previously, in Ackerman & Cianciolo, 2000), four differentiable factors of PS abilities were found, all of which share significant common variance, but none of which can be univocally associated with the term *processing speed*. In addition, three of the different PS abilities share significant common variance with WM that is not, at the same time, shared variance with *g*—again inconsistent with a single common underlying cause explanation. At the same time, the PS tests with the shortest response times, which would be consistent with a more basic conceptualization of processing speed, do not have the highest correlations with WM. Indeed, prior research suggests that the most basic speed measures (often called psychomotor speed tests) have relatively minimal correlations with general intelligence, indicating also that basic processing speed is not the limiting factor for general abilities either (e.g., see Ackerman, 1988, 1990; Ackerman & Cianciolo, 1999, 2000).

The equivalence of WM with an underlying construct of controlled attention (e.g., Kane et al., 2001) was not supported by the analysis of PS ability tests broken out by the consistency of stimulus–response mappings or by changing relations of PS test performance and WM over PS test practice. The PS ability that is most highly associated with WM, independent of *g* (PS–Memory), has as its common construct individual differences in the speed of associative learning (e.g., the Digit/Symbol test). Although associative learning is a key construct in general learning and intellectual performance, it is a much narrower construct than controlled attention.

In the final analysis, we have demonstrated that the construct identification of individual differences in WM is substantially more complex than has been previously identified in the experimental literature. There is little doubt that WM measures are significantly associated with measures of general intelligence. However, we have shown that by including a representative sample of PS ability measures, WM is not univocally associated with general intelligence but also is substantially associated with PS abilities. These results are not inconsistent with extant data (e.g., Kyllonen & Christal, 1990), yet they point to a different interpretation of WM. Mainly, these data call into question the notion that WM and reasoning, WM and Gf, or WM and *g* are the same things. These results also strongly point to the danger of failing to include measures of relevant variables when attempting to explore the underlying meaning of individual differences in WM. Research that examines only correlations between WM measures and the Raven test (or similar tests such as Cattell's Culture Fair Test) reveal only one part of the more complex underlying determinants of WM performance. These approaches may continue to perpetuate a myth about the meaning of individual differences in WM. As a construct, WM is clearly important to the attempt to understand the human information-processing system. However, when considering individual differences in abilities, there is considerably more to the story than is provided by measuring performance on a single test of nonverbal or spatial reasoning, or an undifferentiated sampling of processing speed or PS measures.

References

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10,* 101–139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3–27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 883–901.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77,* 598–614.

Ackerman, P. L. (1999a). Adult intellectual development: Trait complexes and the world of work. In *Proceedings of the British Psychological Society Organizational Psychology* (pp. 1–5). London: British Psychological Society.

Ackerman, P. L. (1999b). Traits and knowledge as determinants of learning and individual differences: Putting it all together. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 437–460). Washington, DC: American Psychological Association.

Ackerman, P. L. (2000). A reappraisal of the ability determinants of individual differences in skilled performance. *Psychologische Beiträge, 42,* 4–17.

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., & Cianciolo, A. T. (1999). Psychomotor abilities via touchpanel testing: Measurement innovations, construct, and criterion validity. *Human Performance, 12,* 231–273.

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., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology, 78,* 413–432.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1,* 270–304.

Ackerman, P. L., & Kyllonen, P. C. (1991). Trainee characteristics. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 193–229). West Sussex, England: Wiley.

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, and volition. *Journal of Educational Psychology, 86,* 487–515.

American Psychological Association. (2001). Award for distinguished scientific contributions: Alan D. Baddeley. *American Psychologist, 56,* 849–864.

Babcock, R. L. (1994). Analysis of adult age differences on the Raven's Advanced Progressive Matrices test. *Psychology and Aging, 9,* 303–314.

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

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

Baddeley, A., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and control* (pp. 28–61). New York: Cambridge University Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the *Bulletin. Psychological Bulletin, 112,* 400–404.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

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. (Original work published 1905 in *L'Année Psychologique, 11,* 191–244)

Brown, J. I., Fishco, V. V., & Hanna, G. H. (1993). *Nelson-Denny Reading Test.* Chicago: Riverside.

Burke, H. R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *Journal of Genetic Psychology, 93,* 199–228.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS,* *and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1101–1114.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431.

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

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Cobb, B. B., & Mathews, J. J. (1972). *A proposed new test for aptitude screening of air traffic controller applicants* (No. FAA-AM-72-18). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163–183.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (1999). Is Spearman's *g* determined by speed or working memory capacity? Book review of Jensen on Intelligence-*g*-factor. *Psycoloquy, 10*(074). Retrieved May 14, 2002 from http://cogprints.ecs.soton.ac.uk/cgi/psyc/newpsy?10.074

Cowan, N. (1998). What is more explanatory: Processing capacity or processing speed? *Behavioural and Brain Sciences, 21,* 835–836.

Cureton, E. E., & Cureton, L. W. (1955). *The Multi-Aptitude Test.* New York: Psychological Corporation.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51,* 599–608.

Ebbinghaus, H. (1896–1897). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern [On a new method for testing mental abilities and its use with school children]. *Zeitschrift für Psychologie und Psysiologie der Sinnesorgane, 13,* 401–459.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7,* 237–241.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54,* 1–34.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Army air forces aviation psychology program research reports: Printed classification tests* (Report No. 5). Washington, DC: U.S. Government Printing Office.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Humphreys, L. G., & Montanelli, R. G., Jr. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10,* 193–206.

Jackson, D. N. (1985). *Multidimensional Aptitude Battery.* London, Ontario, Canada: Sigma Assessment Systems.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Erlbaum.

Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology, 87,* 93–102.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86,* 199–225.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

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* (No. AFHRL-TP-84-56). Brooks Air Force Base, TX: Air Force Systems Command.

Kyllonen, P. C. (1988). *Cognitive Abilities Measurement (CAM) Battery* (Version 4.0). Unpublished computer program.

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

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology, 49A,* 29–50.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–127.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Neath, I. (1998). *Human memory: An introduction to research, data, and theory.* Pacific Grove, CA: Brooks/Cole.

Oberauer, K., Süβ, 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.

Ohnmacht, F. W., Weaver, W. W., & Kohler, E. T. (1970). Cloze and closure: A factorial study. *Journal of Psychology, 74,* 205–217.

Psychological Corporation. (1997). *WAIS–III/WMS–III technical manual.* San Antonio, TX: Author.

Raven, J. C., Court, J. H., & Raven, J. (1977). *Raven's Progressive Matrices and Vocabulary Scales.* New York: Psychological Corporation.

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

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences, 50B,* 297–306.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

Schweizer, K. (1996). Level of encoding, preattentive processing and working-memory capacity as sources of cognitive ability. *Personality and Individual Differences, 21,* 759–766.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

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 nature of "intelligence" and the principles of cognition.* New York: MacMillan.

Spearman, C. E. (1938). Measurement of intelligence. *Scientia Milano, 64,* 75–82.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in deductive reasoning. *Journal of Experimental Psychology: General, 112,* 80–116.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30,* 415–433.

Terman, L. M. (1906). Genius and stupidity. *Pedagogical Seminary, 13,* 307–373.

Thomson, G. H. (1939). *The factorial analysis of human ability.* London: Houghton Mifflin.

Thorndike, E. L., Bregman, E. O., Cobb, M. V., Woodyard, E., and the Staff of the Division of Psychology of the Institute of Educational Research of Teachers College, Columbia University. (1926). *The measurement of intelligence.* New York: Teachers College, Columbia University.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford–Binet Intelligence Scale* (4th ed.). Itasca, IL: Riverside.

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

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 38,* 127–154.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30,* 128–134.

Vernon, P. E. (1950). *The structure of human abilities.* New York: Wiley.

Wechsler, D. (1944). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Whipple, G. M. (1921). *Manual of mental and physical tests. Part 1: Simpler processes; Part 2: Complex processes* (2nd ed.). Baltimore: Warwick & York. (Original work published 1914)

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37,* 60–82.

Received March 21, 2002
Revision received August 7, 2002
Accepted August 7, 2002 ■