Available online at www.sciencedirect.com

SCIENCE DIRECT®

Intelligence 34 (2006) 79–120

INTELLIGENCE

The criterion validity of tasks of basic cognitive processes☆

Dasen Luo[a,*], Lee A. Thompson[b], Douglas K. Detterman[b]

[a] *Department of Psychology, Indiana University of Pennsylvania, PA 15705, United States*
[b] *Case Western Reserve University, United States*

Received 14 August 2004; received in revised form 21 December 2004; accepted 23 December 2004
Available online 22 February 2005

Abstract

The present study evaluated the criterion validity of the aggregated tasks of basic cognitive processes (TBCP). In age groups from 6 to 19 of the Woodcock-Johnson III Cognitive Abilities and Achievement Tests normative sample, the aggregated TBCP, i.e., the processing speed and working memory clusters, correlate with measures of scholastic achievement as strongly as the conventional indexes of crystallized intelligence and fluid intelligence. These basic processing aggregates also mediate almost exhaustively the correlations between measures of fluid intelligence and achievement, and appear to explain substantially more of the achievement measures than the fluid ability index. The results from the Western Reserve Twin Project sample using TBCP with more rigorous experimental paradigms were similar, suggesting that it may be practically feasible to adopt TBCP with experimental paradigms into the psychometric testing tradition. Results based on the latent factors in structural equation models largely confirmed the findings based on the observed aggregates and composites.

Keywords: Criterion validity; Crystallized intelligence; Fluid intelligence; General intelligence (*g*); Processing speed; Scholastic performance; Tasks of basic cognitive processes, Aggregates; Working memory, Structural equation models

Tasks of basic cognitive processes (TBCP), also known as elementary cognitive tasks, are measurement instruments designed to unravel the most fundamental mechanisms of human cognitive processing, such as those of sensation, perception, and memory. The cognitive processes underlying TBCP share two important characteristics that render them basic or elemental to human cognition. First, unlike conventional tests of crystallized intelligence, TBCP require only minimal formal instruction.

☆ This work was supported by NICHD Grant HD21947 and NICHD Training Grant HD07176.

* Corresponding author. Tel.: +1 724 357 4518; fax: +1 724 357 2214.
E-mail address: dluo@iup.edu (D. Luo).

0160-2896/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2004.12.003

Secondly, in contrast to tests purported to measure the ability to educe correlates and relations (general intelligence and fluid intelligence; Cattell, 1963; Horn & Cattell, 1965; Jensen, 1998, 2001; Spearman, 1904), TBCP apparently engage only primitive processes of sequencing, classification, etc., but rarely, if at all, higher-order processes theorized for reasoning, such as rule abstraction, rule mapping, etc. Because the demands for formal knowledge and skills or for complex inductive and deductive reasoning are usually incompatible with the intention to unequivocally measure the basic processes, most TBCP are deliberately designed to minimize these demands. The elemental features of TBCP are naturally appealing to those who intend to provide a reductionist account of intelligence (Detterman, 1987; Hunt, 1978; Jensen, 1982a, 1982b, 1985, 1987; Larson & Saccuzzo, 1989; Nettlebeck & Lally, 1976; see also Brody, 1992; Deary, 2000; Jensen, 1998; Vernon, 1987 for comprehensive reviews).

TCBP have still another important feature. Conventional intelligence tests cannot be readily subjected to experimental analyses, and are thus elusive in their cognitive underpinnings. TBCP, on the other hand, are traditionally used in experimental settings. Some of these tasks, for example, reaction time and stimulus discrimination, are classic tasks of psychophysics. Others, such as those of working memory processing, also directly originate from cognitive-experimental psychology (Baddeley, 1986; Baddeley & Hitch, 1974). These experimental tasks can be systematically manipulated and the mechanisms that control the task performance can be explicitly analyzed through the experimental manipulations. TBCP are also sometimes adapted to psychological testing and assessment, but the adapted TBCP still maintain the most essential features of their experimental prototypes.

Despite the advantage of TBCP in their basic and explicable features, TBCP have not been used intensively in the practical testing of intelligence (Carroll, 1992, 1993), primarily because they seem to have relatively poor psychometric properties. Individual TBCP typically have only low to moderate correlations with indexes of general intelligence (*g* loadings), and do not manifest strong criterion validity. By comparison, conventional intellectual ability tests have high *g* loadings and strong criterion validity, and are thus considered more suitable testing instruments of intelligence. Although recent years have seen an increase in the employment of TBCP in batteries of intelligence (Wechsler, 1991, 1997; Woodcock, McGrew, & Mather, 2001a, 2001c), TBCP are still far outweighed by conventional ability tests in most batteries. The contrasts between experimental tasks of basic processes and conventional ability tests reflect to a degree a fundamental division that runs deep between the experimental and the psychometric traditions in psychology. The division dates back to the early years of modern psychology, and is considered unfortunate for an integrated science of psychology (Cronbach, 1957).

Recent studies of intelligence have accumulated a considerable body of evidence suggesting that intelligence, particularly general intelligence, or *g*, can be largely reduced to a few basic cognitive components. Two of these components, processing speed and working memory, have attracted special attention among students of human intelligence. The two basic cognitive components, determined statistically either as linear composites of TBCP via multiple regression and principal component analyses, or as latent factors of TBCP, have been found to have substantial correlations with indexes of *g* (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway & Engle, 1996; Deary, 1986; Engle, Cantor, & Carullo, 1992; Engle, Kane, & Tuholski, 1999; Engle, Tuholski, Laughlin, & Conway, 1999; Jensen, 1982a, 1982b, 1985, 1987; Kyllonen & Christal, 1990; Kyllonen & Stephens, 1990; Luo & Petrill, 1999; Luo, Thompson, & Detterman, 2003a, 2003b; Miller & Vernon, 1992; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Nettlebeck & Lally, 1976; Vernon, 1983, 1988; Vernon, Nador, & Kantor, 1985; see also Brody, 1992; Jensen, 1998; Deary, 1980; Vernon, 1988 for comprehensive reviews). They also seem to transcend modalities, whether visual or auditory, and are adaptable to content domains of verbal, spatial, and numeric stimuli, suggesting they are general to different intellectual activities. The strong relations between the TBCP-based components and *g*, along with the more explicit cognitive mechanisms of TBCP and their generality, make TBCP in theory a plausible set of indicators for intelligence, although the seemingly weak *g* loadings and criterion validity of individual TBCP remain to be a concern for the practical testing of intelligence.

The seemingly poor psychometric properties of individual TBCP may be overcome by properly aggregating TBCP according to their known mechanisms (Detterman, 1982, 1984, 1986, 1992, 1994). Individual TBCP are predominantly controlled by specific cognitive processes, and are associated with *g* and other complex omnibus abilities only through these specific processes. The process-specific variances of TBCP, however, can contribute to *g* and other complex abilities in an additive manner when tasks measuring different fundamental processes are aggregated. TBCP also typically consist of substantial method-specific variances, which tend to arise from tasks measuring "purer" processes as task performance reflecting such processes can be easily influenced by other factors such as the format of stimulus presentation, the modality of stimulus presentation, response requirement, the content of stimuli, etc. The method-specific variances depress not only the correlations of TBCP with *g* and other complex abilities, but also the correlations of TBCP with each other. The method-specific variances of TBCP can also be effectively reduced by aggregating measures of TBCP. TBCP that involve a common cognitive mechanism but have different stimulus presentation formats, stimulus modalities, response requirements, or stimulus contents, can be aggregated to strengthen the variability of the shared mechanism. In other words, TBCP could have both clearly defined cognitive mechanisms and strong *g* loadings when they are properly aggregated.

TBCP that follow closely a rigorous experimental paradigm may have still other merits. These tasks use readily manipulated stimuli and task demands, which may facilitate the design and application of test items. For example, one can vary stimulus and demand features of these tasks, while still maintaining the same cognitive mechanisms underlying the tasks, to easily generate item banks. Experimental TBCP may also provide richer testing information than their less experimental, adapted counterparts. For example, information processing speed tasks presented in a rigorous experimental paradigm, e.g., choice reaction time, stimulus discrimination time, allow a dissociation between decision time and movement time, which is not possible for less experimental processing speed tests, such as Digit Symbol and Symbol Search in conventional test batteries. Experimental manipulations may also permit a separation of the accuracy component from the speed component in information processing speed tasks (Dosher, 1979; Smith, Kounios, & Osterhout, 1997). For working memory tasks, experimental manipulations are sometimes used to differentiate between the subcomponent of short-term storage and the subcomponent of central executive, and the two subcomponents seem to have disparate relations with intellectual abilities (Engle et al., 1999). Such task parameter information from more experimental TBCP may be of practical value in diagnoses and assessments. Moreover, using experimental tasks for the practical testing of intellectual abilities may to a certain degree bring closer the psychometric tradition and the experimental tradition, and help integrate the two main branches of scientific psychology. These advantages notwithstanding, tasks with a rigorous experimental paradigm may have properties that are deemed undesirable in practical testing of intelligence. One property of experimental tasks is that they tend to appear more "impersonal" than individually administered tests, and may thus give rise to confounding psychological affects, e.g., anxiety or weakened motivation. Another likely concern is that experimental tasks typically have predetermined, rigid procedures, oftentimes controlled by a computer. Once started, these procedures may not be stopped at will as the interruptions may mitigate the testing results. The need for interruptions may nonetheless arise in practical testing when testees, particularly testees young in age or low in ability, need to clarify task demands, causing a conflict between the rigidly designed experimental procedures and the required flexibility in individual treatment. Given that most TBCP have unmistakably simple and non-intimidating task demands, these concerns about experimental TBCP may not be actually relevant. Still, empirical evidence is needed to demonstrate that aggregated TBCP in a more rigorous experimental paradigm have acceptable psychometric properties.

The objectives of the present study were: (1) to demonstrate that processing speed and working memory aggregates could have criterion validity comparable to that of conventional ability test scores; (2) to examine whether these aggregates of TBCP can provide supplemental explanatory power to conventional measures of intelligence; and (3) to assess specifically the criterion validity of aggregated TBCP that incorporate rigorous experimental tasks. We chose to examine the criterion validity of TBCP measuring information processing speed and those measuring working memory, because both cognitive components have been found to play a very salient role in *g* and other complex intellectual abilities. The two components, while sharing a sizable amount of variability, also appear to represent distinct psychological mechanisms. Processing speed has been found to play a dominant mediating role in developmental changes in intelligence (Fry & Hale, 1996; Salthouse, 1996, 2000), whereas working memory seems to outperform processing speed in predicting the within-age individual differences, customarily termed as age-normed differences, in intellectual abilities (Conway et al., 2002; Kyllonen & Christal, 1990). It is therefore prudent to place both components under scrutiny for their criterion validity, and evaluate the joint as well as the unique contributions of the two components to criterion measures. We adopted measures of scholastic achievement as the criterion measures in the present study, because scholastic achievement is widely accepted as the primary criterion for intellectual ability tests, especially for school age children and adolescents.

Previous studies have shown that TBCP measuring processing speed and working memory have significant correlations with scholastic achievement measures (Benton, Kraft, Glover, & Flake, 1984; Carlson & Jensen, 1982; Daneman & Carpenter, 1980, 1983; Daneman & Green, 1986; King & Just, 1991; Luo et al., 2003a, 2003b; MacDonald, Just, & Carpenter, 1992; Ormrod & Cochran, 1988). These correlations, however, were typically observed on a limited scale, both in terms of age range and in terms of sample size, and were not evaluated in comparison to conventional tests of intelligence. Despite the theoretical merits of TBCP, a wider adoption of these tasks in practical testing of intelligence still needs to be implemented with caution, as these tasks have not been used as major assessment instruments of intelligence in clinical and educational settings, and may have properties not fully known. Large-scale evaluations of TBCP for their criterion validity, with conventional tests of intelligence serving as comparison standards, are needed to demonstrate that TBCP are qualified instruments to be used widely in future testing of intelligence.

In the present study, the conventional ability measures to which the aggregates of TBCP were compared were indexes of crystallized intelligence, fluid intelligence, and general intelligence. We intended to compare the aggregates of TBCP to these indexes of intelligence because these indexes represent the most pervasive latent constructs underlying intellectual ability tests, and the variances they each share with the criterion of scholastic achievement reflect distinct sources of influence on the criterion. Of particular interest in our comparison is the evaluation of TBCP in conjunction with fluid ability measures. The two families of measures are similar in that both putatively tap abilities less affected by cultural and educational backgrounds, and are thus more likely to reflect an individual's "potential" rather than his/her acquired knowledge and skills. We attempted to compare TBCP to two specific kinds of fluid ability measures, namely, measures of more strictly defined reasoning abilities, including those of induction and deduction, and measures of more broadly defined problem-solving skills, such as those indicated by the Performance subscales employed in the Wechsler Scales of Intelligence. The two kinds of fluid ability measures represent the most commonly adopted means of testing fluid abilities, and the comparison between TBCP and fluid ability measures needs to be conducted with reference to both kinds of fluid ability measures. The extent to which the TBCP aggregates supercede and surpass the conventional indexes of intelligence in the explanation for the criterion will reveal a great deal not only the practical value of the TBCP aggregates in the testing of intelligence, but also the cognitive nature of the conventional indexes.

The measures of TBCP in the present study were taken from two data sources, Woodcock-Johnson III Cognitive Abilities and Achievement Tests (W-J III; Woodcock et al., 2001a, 2001c; Woodcock, McGrew, & Mather, 2001b) normative data and the Western Reserve Twin Project (WRTP) data. The W-J III battery includes two TBCP aggregates, the Processing Speed cluster and the Working Memory cluster. The TBCP in the WRTP study included mostly experimental tasks, which measured information processing speed and working memory/short-term memory processing. The fluid reasoning cluster of W-J III includes an inductive reasoning test and a deductive reasoning test, whereas the fluid ability tests adopted by WRTP are the Wechsler Intelligence Scale for Children-Revised (WISC-R) Performance subscales. The aggregated TBCP of the WRTP data could be used to evaluate the criterion validity of TBCP that follow a rigorous experimental paradigm, and the WISC-R Performance subscales used by the WRTP provide an additional dimension of the fluid ability in our comparison between TBCP and fluid ability measures. For both data sets, the criterion validity of aggregated TBCP was examined in comparison to composite scores of crystallized intelligence and fluid intelligence, as well as indexes of general intelligence.

The aggregates of TBCP were formed on different composition levels in the present study. Aggregates of TBCP can be made separately according to their respective underlying mechanisms, e.g., the processing speed aggregate, the working memory aggregate, and aggregates of other cognitive processes. They can also be incorporated into a single TBCP composite, or together with certain conventional test composites (crystallized ability composite, fluid ability composite, etc.). The overall criterion validity of aggregated TBCP may vary with the level of composition. When separate aggregates of TBCP are used as individual predictors in a multiple regression model, these aggregates are combined according to a set of optimal weights. When these aggregates are used to form a single total score, the explanatory power of such a single total score is generally not the same as, and typically lower than, that of an optimally weighted sum. Since optimal weights obtained from multiple regression models are susceptible to extraneous sampling features, simple sums instead of optimally weighted sums are more often used in practice. In other words, if TBCP are to be posited as potent contenders against conventional ability tests, they are more likely to be presented in a form of simple composites rather than optimally weighted composites, and the criterion validity of such simple sums of aggregated TBCP needs to be evaluated and compared to that of conventional tests. Similarly, a simple sum combining both aggregates of TBCP and certain conventional test composites, particularly those of crystallized abilities, may be of practical value because such a hybrid of TBCP and conventional measures is a likely form of test battery in practical settings, covering most comprehensively the cognitive foundation of criterion performance. The criterion validity of such a hybrid with a concentration in aggregates of TBCP therefore needs to be examined.

In the present study, we examined the criterion validity of aggregated TBCP on three levels of composition. On Level 1, aggregates of TBCP were formed according to their known cognitive mechanisms (e.g., processing speed, short-term memory/working memory, etc.), and these aggregates were used as separate predictors in multiple regression models to examine their predictive power (R^2s) for scholastic performance. Moreover, the separate TBCP aggregates were compared to counterparts of fluid abilities on the first level of composition to see whether they provide incremental validity over fluid ability tests. On the second level, TBCP aggregates were further combined into a simple sum, and the explanatory power of the sum, which can be viewed as the criterion validity of a test battery consisting exclusively of TBCP, was evaluated and compared to that of conventional ability tests. On the third level, a total score was obtained from both aggregates of TBCP and certain composites of conventional ability tests, and the criterion validity of such a hybrid is indicative of the power of a possible new kind of battery made of both an extensive collection of aggregated TBCP and a selected type of conventional ability measures.

To further validate the results obtained using the aggregated tests and composite scores, the explanatory power of TBCP for scholastic achievement was also evaluated and compared to that of the crystallized and fluid abilities using the structural equation modeling (SEM) method (Jöreskog, 1973). In the structural equation models, latent factors of processing speed, working memory, crystallized ability, fluid ability, and scholastic achievement were specified, and the explanatory power of the processing speed and working memory factors for the achievement factor was evaluated in a way analogous to the evaluation of observed TBCP aggregates and composite scores. The results based on the manifest composites and those from the SEM analyses each are important in their own right, and the consistency between the two sets of results can provide cross-validation as well. Assessment of criterion validity on the basis of the observed aggregates and composite scores is necessary, as in practical testing only the observed measures, not the latent factors, are used. The SEM analyses, on the other hand, were intended to provide a more theoretical account for the explanatory power of processing speed and working memory, because the latent factors are theoretically not interwoven with measurement errors of the manifest variables. In other words, a latent factor can be viewed as an ideally aggregated composite that only pertains to the shared variability among the aggregated manifest variables, but not the specific variances of these variables. The merit of latent factors as ideal aggregates is of particular relevance in the present study, because, as discussed above, the strength of TBCP typically is not fully displayed until they are sufficiently aggregated. The results from the SEM analyses, if commensurate with those based on the observed aggregates and composite scores, will theoretically confirm the conclusions drawn from the observed variables.

1. Method

1.1. The W-J III normative data

The W-J III was normed in a nationally representative sample and consists of a wide range of cognitive and scholastic performance variables. It has a number of characteristics that are important for the present study: (a) it includes tests measuring basic cognitive processes such as those of processing speed and working memory, as well as complex intellectual abilities such as crystallized intelligence, fluid intelligence, and scholastic achievement; (b) tests are aggregated into clusters according to their hypothetical latent constructs; (c) the Tests of Cognitive Ability battery of W-J III was co-normed with the Tests of Achievement battery, permitting an evaluation of the criterion validity of the cognitive ability tests; (d) the normative sample of W-J III was large (total size: 8818) and was obtained using a quota sampling scheme to represent the characteristics of region, sex, race, age, school grade, and occupation of the US population.

1.2. Selected variables of the W-J III

The W-J III clusters selected for the present study were the clusters of Processing Speed (WJ Processing Speed Cluster), Working Memory (WJ Working Memory Cluster), Knowledge-Comprehension (WJ Comprehension Knowledge Cluster), Fluid Reasoning (WJ Fluid Reasoning Cluster), and some clusters of the Achievement Tests battery. In addition, a general index of W-J III General Intellectual Ability-Standard (WJ General Intellectual Ability) was included in the analyses with the cluster scores.

1.2.1. Aggregates of TBCP

The WJ Processing Speed and WJ Working Memory clusters are aggregates of cognitive tasks measuring basic cognitive processes. WJ Processing Speed Cluster is a composite of two tests, Visual Matching and Decision Speed. Both are easy tasks of symbol matching, requiring largely the speed of perception and stimulus discrimination. The WJ Working Memory Cluster consists of Numbers Reversed and Auditory Working Memory, which require actively operating on the information held in awareness. In Numbers Reversed, the testee is asked to listen to a list of numbers recorded on a tape, and recall these numbers in a reversed order after the completion of the list presentation. The list length varies from 2 digits to as many as 8 digits. In Auditory Working Memory, the testee is required to listen to a tape recording that reads numbers and objects alternately and to recall the read items after the termination of each list by grouping the numbers and objects in two categories and arranging them in their respective presentation orders (see Woodcock et al., 2001a, 2001b, 2001c, for a detailed description of these tasks).

The W-J III working memory tasks, such as the memory span task of Numbers Reserved, may be suspected to be mainly short-term storage processing measures but not central executive processing indicators (Engle et al., 1999), and may thus not adequately represent working memory processing, but such a suspicion may only be valid for the college student and adult populations. With younger age testees, such as those included in the present study, whose short-term storage capacity is not fully developed, an extra attention management effort may be needed in these tasks, and the tasks may still tap effectively central executive processing of the testees. Furthermore, since working memory tasks with more intensive central executive processing typically have higher correlations with complex intellectual abilities (Engle et al., 1999), memory span tasks such as Numbers Reversed may provide a conservative estimate of the criterion validity of working memory tasks. In other words, it is possible to further strengthen the criterion validity of TBCP demonstrated by the working memory tasks in the present study by employing working memory tasks with a more intensive central executive component.

1.2.2. Conventional tests of intelligence

The W-J III Comprehension Knowledge and Fluid Reasoning clusters (WJ Comprehension Knowledge Cluster and WJ Fluid Reasoning Cluster) were designed to mainly represent the two major categories of intellectual abilities, crystallized intelligence and fluid intelligence. WJ Comprehension Knowledge Cluster is a composite of Verbal Comprehension, a test of synonyms and antonyms, and General Information, a test assessing general knowledge, similar to the WISC-R Information. WJ Fluid Reasoning Cluster is based on Concept Formation, a test of inductive reasoning, and Analysis Synthesis, a test of deductive reasoning. Questions in Concept Formation consist of objects that can be classified according to their colors or shapes, and the testee is asked to identify in each question one object that does not quite fit the classification. In Analysis Synthesis, a number of colored objects are first presented to the testee, and the rules governing the relations between the objects are described. The testee is then shown objects with a missing piece and asked to identify the missing object according to the rules described before. The tests of WJ Comprehension Knowledge Cluster and WJ Fluid Reasoning Cluster have high *g* loadings, with a median *g* loading of 0.77 in the 6 to adult age range. The median *g* loading of the WJ Processing Speed Cluster and WJ Working Memory Cluster tests is 0.51. The disparity in *g* loading between the complex intellectual ability tests and the TBCP is nonetheless expected to diminish with the aggregation of TBCP, as the aggregation would decrease the method-specific variances of the TBCP variables on the one hand, and increase the breadth of the cognitive processes represented by the TBCP on the other.

In addition to WJ Comprehension Knowledge Cluster and WJ Fluid Reasoning Cluster scores, a general score, WJ General Intellectual Ability, was also used for the analysis. The WJ General Intellectual Ability index is the first principal component score from a collection of tests. It is a weighted sum of seven W-J III standard tests, including one test each from the WJ Comprehension Knowledge, WJ Fluid Reasoning, WJ Processing Speed, and WJ Working Memory clusters as well as tests from three other clusters. The general score can be considered as a proxy of general intelligence, or *g*, and could serve as a reference for the evaluation of the criterion validity of aggregated TBCP.

1.2.3. Achievement measures

The W-J III achievement measure used in the present study as the criterion measure is Total Achievement (WJ Total Achievement), which is a summary score for a number of achievement test clusters, including Broad Reading, Broad Math, and Broad Written Language.[1]

1.2.4. Manifest variables for the SEM analyses

The same individual tests that made up the clusters, i.e., Visual Matching and Decision Speed for WJ Processing Speed, Numbers Reversed and Auditory Working Memory for WJ Working Memory, Verbal Comprehension and General Information for WJ Comprehension Knowledge, Concept Formation and Analysis Synthesis for WJ Fluid Reasoning, and Broad Reading, Broad Written Language, and Broad Math for WJ Total Achievement, were used as manifest variables in the structural equation models to define their respective latent constructs. Broad Reading, Broad Written Language, and Broad Math are cluster scores each made of a number of subtest scores, and the correlations/covariances between these achievement cluster scores and the test scores of the other factors were determined analytically (Johnson & Wichern, 1992, pp. 117–118) on the basis of the reported correlations in the W-J III technical manual among all individual subtest scores.

1.3. Subgroups of W-J III normative sample

The total W-J III normative sample contain all age ranges from the primary school years to senior years. We analyzed data from three school age groups, 6–8 (N=1095), 9–13 (N=2241), and 14–19 (N=1641). The selection of these age groups is based on the consideration that scholastic achievement constitutes the best criterion performance for these age groups, but less so for older age groups.[2] Results from these

---

[1] Results based on the subcomposites of the W-J III Total Achievement, Broad Reading, Broad Math, and Broad Written Language composites, were largely the same as the results based on the W-J III Total Achievement, and were omitted in the paper. These results can be obtained upon request.

[2] Similar results were found from the older age groups of the W-J III normative sample (20–39 and 40+), but are not presented based on the consideration that scholastic achievement is less relevant as criterion performance for the older age groups.

groups could be used to demonstrate the strength of TBCP in predicting criterion measures throughout the pre-college school age span.

1.4. Analysis

1.4.1. Analyses of observed aggregates and composites

Based on the reported correlations among the adopted measures (McGrew & Woodcock, 2001), the multiple regression analysis method was used to evaluate the criterion validity of WJ Processing Speed and WJ Working Memory clusters on three composition levels. On the first level, WJ Processing Speed Cluster, WJ Working Memory Cluster, and WJ Fluid Reasoning Cluster each were treated as individual predictors, and their zero-order and higher-order contributions to the variability of WJ Total Achievement were assessed in terms of R^2 and R^2 changes.

On Level 2, the TBCP clusters were combined into a more agglomerated composite, the WJ Processing Speed and WJ Working Memory composite (WJ Processing Speed+Working Memory Composite). The criterion validity of this TBCP composite was compared to that of WJ Comprehension Knowledge Cluster, WJ Fluid Reasoning Cluster, and of the WJ Comprehension Knowledge and WJ Fluid Reasoning combination (WJ Comprehension Knowledge+Fluid Reasoning Composite). These composites were made of equally weighted sums of member clusters, and their intercorrelations and correlations with the other measures could be analytically determined on the basis of Level 1 cluster correlations. Multiple regression analyses relating the Level 2 TBCP composite to WJ Total Achievement were conducted based on the Level 2 correlations. R^2s and R^2 changes from these regression models could be evaluated in relation to WJ Comprehension Knowledge Cluster, WJ Fluid Reasoning Cluster, and WJ Comprehension Knowledge+Fluid Reasoning Composite with WJ Total Achievement as the criterion measure.

On Level 3, the aggregates of TBCP were combined with some of the complex ability composites. Three Level 3 composites were formed, the WJ Processing Speed, WJ Working Memory, and WJ Comprehension Knowledge composite (WJ Processing Speed+Working Memory+Comprehension Knowledge Composite), the WJ Processing Speed, WJ Working Memory, and WJ Fluid Reasoning composite (WJ Processing Speed+Working Memory+Fluid Reasoning Composite), and the WJ Processing Speed, WJ Working Memory, WJ Comprehension Knowledge, and WJ Fluid Reasoning composite (WJ Processing Speed+Working Memory+Comprehension Knowledge+Fluid Reasoning Composite). Member clusters were equally weighted in each of the composites. The composites were created to examine whether batteries using mainly aggregates of TBCP would have criterion validity comparable to that of conventional batteries, which typically use more omnibus ability measures than TBCP measures. Comparisons could also be made on Level 3 to see whether conventional fluid intelligence measures are expandable when sufficient TBCP measures are employed. R^2s resulted from the simple regression models using each of these third level composites and the WJ General Intellectual Ability index as predictors for achievement measures were compared.[3]

---

[3] To ensure that the results of the present study were not seriously affected by the reliability of the predictors, analyses were also conducted on correlations corrected for test reliability. These results were very similar to those based on the uncorrected correlations. In fact, the corrections had a more positive impact on the criterion validity of the aggregated TBCP than on that of the conventional test composites. Since the estimation of the theoretical correlations between perfectly reliable measures could be achieved more effectively on the basis of the SEM analyses than on the basis of corrected correlations, these results were not presented. Interested readers can request the results based on the corrected correlations from the authors.

1.4.2. The SEM analyses

Five latent factors were specified for the SEM analyses, a processing speed factor (WJ Speed Factor) indicated by Visual Matching and Decision Time, a working memory factor (WJ Working Memory Factor) manifested by Numbers Reversed and Auditory Working Memory, a crystallized intelligence factor (Gc) defined by Verbal Comprehension and General Information, a fluid intelligence factor (Gf) with loadings on Concept Formation and Analysis Synthesis, and an achievement factor (WJ Achievement) underlying Reading, Written Language, and Math.

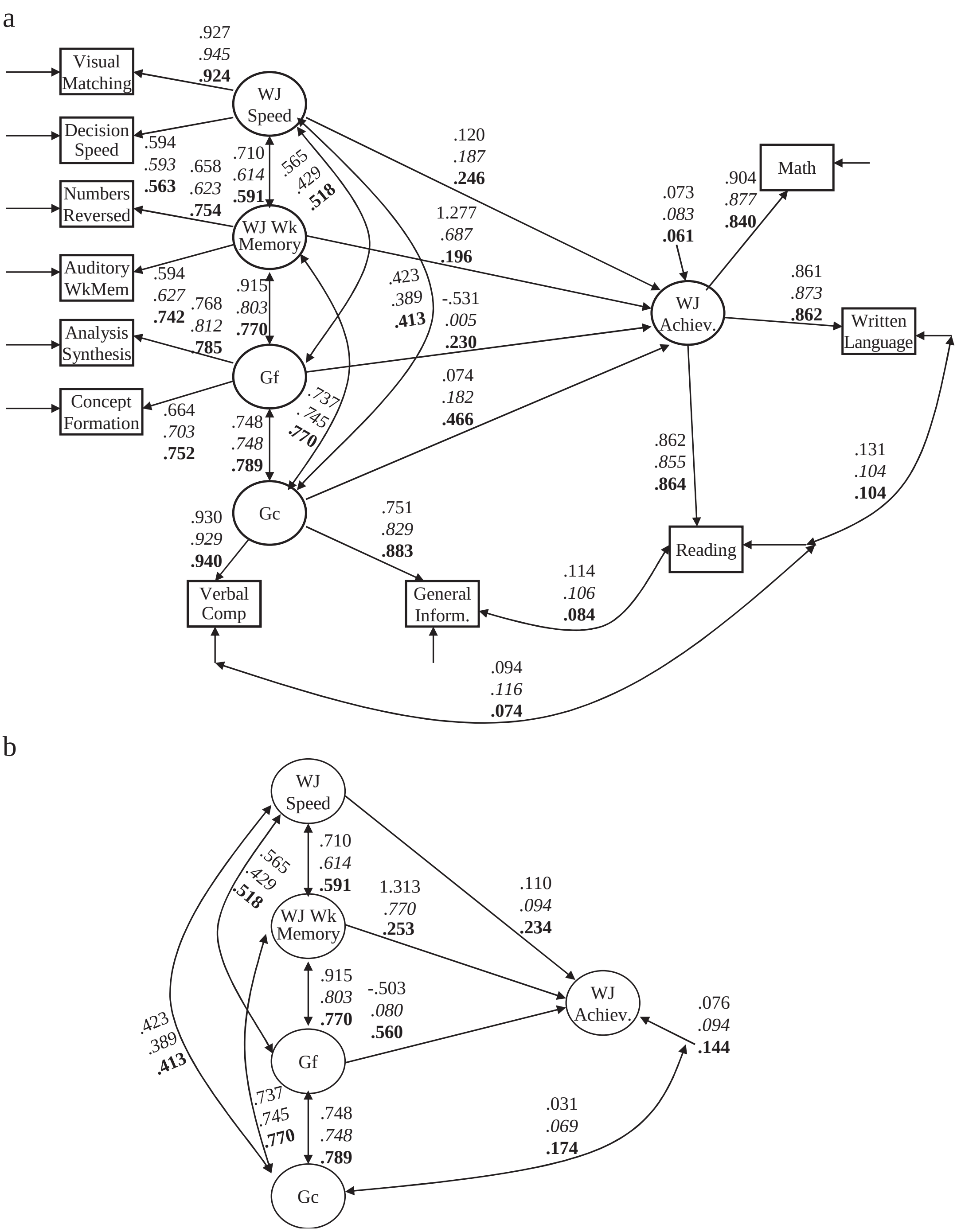
Based on these latent factors, four mathematically equivalent structural equation models (Models 1A, 1B, 1C, and 1D) were specified. In Model 1A, WJ Speed Factor, WJ Working Memory Factor, Gf, and Gc are all treated as correlated explanatory variables for the WJ Achievement Factor, and thereby have one-way paths from these factors to the Achievement Factor and two-way paths between each other. The model is analogous to a multiple regression model with four predictors for the dependent measure, and the standardized residual variance in Achievement, or alternatively, one minus the residual (R^2), indicates the amount of total variability in Achievement explained by these factors (see Fig. 1a for a graphic description).

In Model 1B, the Speed, Working Memory, and Gf factors have one-way paths to Achievement, but not Gc, which has two-way paths with all the other factors (Fig. 1b). Such a model is comparable to a multiple regression model with three instead of four predictors, and the resultant R^2 can be compared to that of the model with 4 explanatory variables described above to attain a R^2 change indicating the unique contribution from Gc to Achievement over and above WJ Speed Factor, WJ Working Memory Factor, and Gf.

Similarly, the R^2 change caused by Gf over WJ Speed Factor and WJ Working Memory Factor for WJ Achievement can be obtained by comparing the model with only two explanatory factors, WJ Speed and Working Memory, in Model 1C (see Fig. 1c) to that in Model 1B.

In Model 1D, WJ Speed Factor, WJ Working Memory Factor, Gf, and Gc are all covariates of the WJ Achievement Factor, with two-way paths among all five latent factors (Fig. 1d). In the model, the estimated correlation between the Achievement Factor and each of the other four factors can be squared to provide the related R^2, which may be compared to the R^2 obtained from Model 1A, Model 1B, or Model 1C to evaluate the R^2 change of interest. For example, the unique contribution from WJ Speed Factor and WJ Working Memory Factor to WJ Achievement Factor after controlling the influence of Gf can be evaluated by obtaining the change in R^2 from Model 1B with WJ Speed Factor, WJ Working Memory Factor, and Gf as the explanatory variables to the R^2 ascribed to Gf only in Model 1D. By the same token, the incremental explanatory power of the Working Memory factor beyond the Speed factor for WJ Achievement can be gauged by obtaining the counterpart R^2 change from Model 1C to Model 1D.

Although the R^2s and R^2 changes aforementioned can be analytically determined from model parameter estimates from any one of the four models, for instance, from the zero-order factorial correlation estimates of Model 1D, specifying these models has an advantage of testing statistically unique contributions made by certain predictors controlling the others in these models. In Models 1A, 1B, and 1C, one or more predictor-to-criterion one-way paths (beta weights) can be fixed to zero to evaluate the chi-square changes caused by the zero constraints and the magnitude of the changes can indicate the impact of one or more predictors over that of the others for the explanation for the Achievement Factor. For example, fixing a beta weight associating Gc with the Achievement Factor in Model 1A leads to the chi-square change indicative of the unique impact of Gc over that of the other


a
Visual Matching
Decision Speed
Numbers Reversed
Auditory WkMem
Analysis Synthesis
Concept Formation
WJ Speed
WJ Wk Memory
Gf
Gc
Verbal Comp
General Inform.
WJ Achiev.
Math
Written Language
Reading
.927 .945 .924
.594 .593 .563
.658 .623 .754
.710 .614 .591
.565 .429 .518
.594 .627 .742
.768 .812 .785
.915 .803 .770
.423 .389 .413
.664 .703 .752
.748 .748 .789
.737 .745 .770
.930 .929 .940
.751 .829 .883
.120 .187 .246
1.277 .687 .196
-.531 .005 .230
.074 .182 .466
.073 .083 .061
.904 .877 .840
.861 .873 .862
.862 .855 .864
.131 .104 .104
.114 .106 .084
.094 .116 .074
b
WJ Speed
WJ Wk Memory
Gf
Gc
WJ Achiev.
.710 .614 .591
.565 .429 .518
1.313 .770 .253
.110 .094 .234
.915 .803 .770
-.503 .080 .560
.423 .389 .413
.737 .745 .770
.748 .748 .789
.076 .094 .144
.031 .069 .174

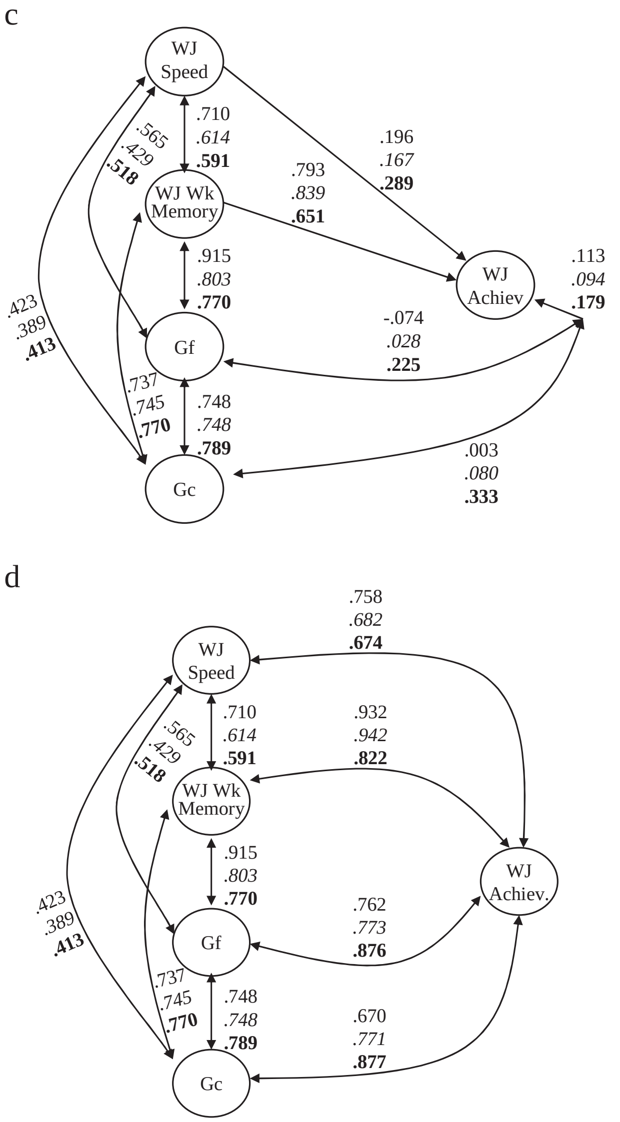


Fig. 1. (a) The five-factor model with WJ Speed Factor, WJ Working Memory Factor, Gf, and Gc as explanatory variables for WJ Achievement Factor. Note: Parameter estimates shown in the diagram are all standardized values. Variables symbolized by ovals are latent factors, and those in rectangles are manifest variables. The open-ended one-way arrow to a manifest or latent variable represents the residual variance to the variable. Variable acronyms: Auditory WkMem—Auditory Working Memory; Verbal Comp—Verbal Comprehension; General Inform.—General Information, Reading—Broad Reading; Written Language—Broad Written Language; Math—Broad Math; WJ Wk Memory—WJ Working Memory Factor; WJ Achiev.—WJ Achievement Factor. Estimates in regular font: 6–8 age group; italics: 9–13 age group; boldface: 14–19 age group. (b) The five-factor model with WJ Speed Factor, WJ Working Memory Factor, and Gf as explanatory variables for WJ Achievement Factor. (c) The five-factor model with WJ Speed Factor and WJ Working Memory Factor as explanatory variables for WJ Achievement Factor. (d) The five-factor model with correlated factors but no explanatory variable for WJ Achievement Factor. Note: The models described in b–d are mathematically equivalent to that depicted in a, and parameter estimates associated with the manifest variables are thus identical in all these models. The diagrams only present model parameters related to the latent factors.

three predictors, whereas the chi-square change caused by the zero constraints placed on the beta weights related to the Speed and Working Memory factors in the same model indicates the unique impact of the Speed and Working Memory factors beyond that of Gc and Gf. Beta weights in Models 1B and 1C can also be fixed to zero to evaluate the related unique impacts of the predictors.

Models 1A through 1D are mathematically equivalent in that they yield the same expected covariance/correlation matrix with identical model fit indexes. The four predictors in Model 1A are progressively converted to covariates of the Achievement Factor in the other three models without changing the model fit (Stelzl, 1986; Rule 3). The R^2 values estimated from these models are all based on the "reduced form" (Jöreskog & Sörbom, 1996), and can be interpreted directly as proportions of variability explained without any adjustment. The four mathematically equivalent models constitute a model sequence with a theorized decrease in the cognitive complexity of the predictors from Model A to Model C. In Model 1A, one of the predictors, Gc, presumably reflects a hybrid of acquired knowledge/skills and "culture-fair" fluid abilities and is thus on the highest level of complexity in underpinning, with Gf on the next complexity level representing only "culture-fair", although psychologically complex, traits, and with the basic processing factors on the lowest complexity level for their apparent simplicity in cognitive mechanisms. The sequence therefore decreases in the predictor complexity level from Model A to Model C (no predictor in Model D) because the most complex predictor in a preceding model is converted to a covariate in the next model. The sequence permits one to evaluate the relative importance of complex processing and basic processing for the Achievement Factor according to the magnitude of the R^2 changes and the chi-square changes obtained from the model sequence.

All these mathematically equivalent models were fitted to the covariance matrices from the 6–8, 9–13, and 14–19 age groups using the multi-group maximum likelihood estimation method. To facilitate the evaluation of the R^2s and R^2 changes described above, all factors in these models were constrained to have unit variances. For the explained Achievement Factor, the unit variance constraint was set according to the rationale described by McDonald, Parker, and Ishizuka (1993) so that the estimated residual variance can be understood as $1-R^2$. The SEM analyses were conducted using the MX program (Neale, 1999).

2. The WRTP data

The Western Reserve Twin Project was a study of cognitive abilities and scholastic achievement in elementary school twins. It included variables from TBCP, conventional intellectual ability tests, and measures of scholastic performance. One important characteristic of the TBCP employed in the WRTP was that most of the TBCP were standard experimental tasks. The characteristic is of particular relevance to the present study because we intended to evaluate the criterion validity of TBCP that follow closely an experimental paradigm. Another notable feature of the WRTP is that it employed the Performance subscales of WISC-R. These subscales presumably measure, among other things, the ability of spatial relations, and are also commonly considered tests of fluid intelligence, albeit more loosely defined. It would be of interest to compare the criterion validity of TBCP to that of these fluid ability tests. Two other features of the WRTP are further worth noting. The WRTP included a wider selection of processing speed variables than W-J III, and used mostly visuospatial memory span tasks instead of verbal and numerical working memory tasks.

2.1. Participants

Participants in the WRTP were twins recruited through public, private, and parochial schools within a six-county area surrounding the city of Cleveland, Ohio. The 568 participants were in the first through sixth grade at the time of participation, with an age rage from 6 to 13. The sample represented the full range of individual differences in general intelligence with a WISC-R Full Scale IQ mean of 104.5 and a standard deviation of 15.8.

To obtain covariances and correlations for the present study, all twin pairs were split and randomly assigned into two nonoverlapping subsamples. Correlations and covariances were then pooled from the two subsamples so that subjects were treated as though they were unrelated individuals. Correlations/covariances between all variables were based on 512 observations after listwise deletion of cases with missing values. Multiple regression and SEM analyses were conducted using these correlations and covariances.

2.2. Measures

Each child in the study was tested a total of 8.5 h over three sessions, the psychometric ability test session, the elementary cognitive task session, and the achievement test session.

2.2.1. Aggregates of TBCP

Six tasks from the Cognitive Abilities Test (CAT; Detterman, 1988) were administered to all participants. They were: Learning, Probe Recall, Reaction Time, Stimulus Discrimination, Self-Paced Probe Recall, and Tachistoscopic Threshold (TT).

Stimulus presentation was similar in Probe Recall, Learning, and Self-Paced Probe Recall. In each trial, a participant was presented with a row of blank windows slightly below the center of the computer screen, and a probe window centered above this row. Stimuli comprised of matrix diagrams were shown one by one in each blank window. Each diagram appeared and disappeared before the presentation of the next diagram. After the last diagram disappeared on the screen, one of the previously presented diagrams reappeared in the probe window, and participants were asked to indicate which bottom window contained the probed diagram.

In Probe Recall, only one of the six previously presented diagrams appeared after the presentation, and the participant was required to identify which of the six windows originally presented the target diagram. In Learning, the presented diagrams varied in number (3, 5, 7, or 9) during each trial, and each of the presented diagrams was probed after the presentation. The participants were asked to respond to all of the probed diagrams. Both Probe Recall and Learning are presumed to tap largely visuospatial short-term memory, or visuospatial sketchpad, to use the epithet in the literature of working memory studies (Baddeley & Logie, 1999). Self-Paced Probe Recall was the same as Learning, except that participants could control the presentation time of each stimulus diagram. The control over the stimulus presentation time allows the participant to perceive and remember each new stimulus while rehearsing in short-term memory the items already presented. In other words, there is putatively an additional central executive component involved in Self-Paced Probe Recall, although, as indicated by a previous study (Miyake et al., 2001) and also by the highly similar observed correlations of Learning and Self-Paced Probe Recall with other intellectual ability tests in the WRTP data, visual–spatial working memory tasks tapping the additional central executive component may not be highly distinguishable from those only engaging the visuospatial sketchpad component.

In the Reaction Time task, an array of 1, 2, 4, 6, or 8 empty windows appeared on the screen and one of the windows would light up in each trial. The participant was required to touch the lit window as quickly as possible.

Participants in Stimulus Discrimination were presented with six blank windows in the bottom portion of the screen, and a probe window in the upper portion of the screen. The six windows each displayed a different diagram, and the probe window presented a diagram identical to one of the six diagrams below it. The participant's task was to find the match to the probe, and touch it as quickly as possible.

In both Reaction Time and Stimulus Discrimination, the participant was instructed to keep a finger on the home button in the lower part of the touch screen before the diagrams appeared. The diagrams then came on after a randomly varied waiting period. The total response time was the time from the onset of the diagram presentation to the touching of the target diagram on the screen. In both tasks, the total response time was decomposed into two disjointed time components, decision time and movement time. The time from the onset of the diagram presentation to the movement of the finger from the home button was the decision time. The time from the release of the finger on the home button to the target diagram was the movement time, recorded separately from the decision time. The decision time is known to have a more pronounced correlation with intelligence than the movement time.

In Tachistoscopic Threshold, two diagrams were presented simultaneously for a very brief duration and were then masked, and participants were asked to determine whether they were the same. The presentation time varied until a threshold duration for correct identification, presumably measuring inspection time, was determined.

All participants during the CAT sessions used a touch screen device to make their responses to the stimuli, and both their response times and response choices were automatically recorded by a computer.

Six measures of elemental cognitive processes were obtained from the six CAT tasks. These measures were: percent correct from Probe Recall, percentage correct from Learning, percent correct from Self-Paced Probe Learning, decision time from Reaction Time, decision time from Stimulus Discrimination, and threshold time from Tachistoscopic Threshold. Threshold Time of the Tachistoscopic Threshold task indicates the minimum stimulus presentation time needed for a participant to respond correctly 75% of the times for the task. The six variables selected were the most informative indicators of their respective tasks. All six CAT variables were age and gender adjusted standardized multiple regression residuals.

In addition to the CAT variables, three more variables, the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) Coding and Digit Span scale scores and Perceptual Speed of the Colorado Specific Cognitive Ability Test (SCA) were also used as variables of TBCP. Digit Span requires the recall of number lists both in the forward fashion and in the backward fashion, and is thus very similar to the W-J III working memory test of Numbers Reversed. Coding measured the speed of pattern perception/recognition. Perceptual Speed measured the subject's total number of correct and incorrect responses within a given time when they searched for a set of target stimuli from a list of symbols. Of the three variables not taken from the CAT, Digit Span could be easily adapted to an experimental paradigm. Coding and Perceptual Speed were similar to Visual Matching in the W-J III, and may also be rigorously analyzed in experimental studies, particularly when they are used in conjunction with some other experimental tasks, such as Reaction Time and Stimulus Discrimination. The participant in Reaction Time and Stimulus Discrimination responded to only a single target stimulus per trial, but in Coding and Perceptual Speed the participant needed to attend to multiple target stimuli simultaneously, thereby initiating additional cognitive processing. The processing components in Coding and Perceptual Speed not shared by Reaction Time and Stimulus Discrimination may be isolated if Coding and Perceptual Speed are adapted to an experimental paradigm similar to those of Reaction Time and Stimulus Discrimination, and used along with Reaction Time and Stimulus Discrimination.

Two aggregated variables were created according to the known mechanisms of the nine TBCP variables. One of them, a processing speed variable (WRTP Processing Speed Aggregate), was an aggregate of Stimulus Discrimination, Reaction Time, Tachistoscopic Threshold, Coding, and Perceptual Speed, and presumably represented the speed of information processing. The aggregate was constructed in two steps, with Stimulus Discrimination, Reaction Time, Tachistoscopic Threshold, Coding added to form an initial aggregate in the first step, and then with the initial aggregate and Perceptual Speed both standardized and summed in the second step. The rationale for the stepwise aggregation was that Perceptual Speed has a noticeably higher *g* loading than those of the other WRTP Speed variables, and aggregating variables with disparate *g* loadings in a stepwise fashion ensures that the final aggregate has a *g* loading higher than any of its ingredients'.[4] The other TBCP aggregate was made of four memory tasks, Probe Recall percent correct, Learning percent correct, Self-Paced Probe Recall percent correct, and Digit Span scaled score. All four tasks involved mostly short-term memory and working memory processing, so the aggregate could be construed as mainly a short-term memory/working memory aggregate (WRTP STM/Working Memory Aggregate). The weights of the aggregation were determined by the inverse of the variable standard deviations.

2.2.2. Conventional tests of intelligence

The WISC-R IQ scores (WISC Verbal IQ, WISC Performance IQ, and WISC Full IQ) were used to represent traditional measures of intelligence to which the aggregated TBCP were compared. WISC Verbal IQ and WISC Performance IQ were often treated as estimates of crystallized and fluid intelligence, and WISC Full IQ was usually used as a proxy of general intelligence. The WISC-R Performance subscales were originally designed as measures of non-verbal, problem-solving abilities, and loosely fit the definition of "culture-fair" and "education-lite" fluid intelligence (Horn & McArdle, 1992). They have been used in clinical settings to provide an additional dimension of assessment beyond the WISC-R Verbal subscales, which closely fit the definition of crystallized intelligence. WISC-R Full IQ as a general index is different from WJ General Intellectual Ability in that it is made of a different type of fluid abilities. Moreover, WISC-R Full IQ is more heavily weighed by conventional measures of intelligence than WJ General Intellectual Ability. It is therefore possible that the WRTP Processing Speed and STM/Working Memory Aggregates may add sizable additional explanatory power to WISC Full IQ for the criterion measures of scholastic performance, as the aggregates of TBCP may reflect certain cognitive aspects of the criterion measures not fully measured by the composite of crystallized intelligence and fluid intelligence measures.

---

[4] The rationale for the increment in *g* loading through a sufficient aggregation of initially moderate or low *g* loaded TBCP can be construed in light of the classic Spearman–Brown formula (McDonald, 1999), which ensures the *g* loading to be an increasing function of aggregation. The formula stipulates, however, that the to-be-aggregated items have the same true score variance, or in the context of *g* variability, the same *g* loading. If this assumption is not met, the aggregation may not increase or may even decrease the resultant *g* loading. On the other hand, given a large enough item (task) pool, one may always start with items with lower but similar *g* loadings to form preliminary aggregates that have *g* loadings higher than any of their ingredients', and progressively include additional items with higher *g* loading into the aggregate. The process will finally create an aggregate with a *g* loading higher than those of all its ingredients.

2.2.3. Measures of scholastic achievement

The total score (MAT Total Achievement) summing over the scores of the Language, Math, and Reading subscales of the Metropolitan Achievement Test 6 battery was used as the criterion measure for the analyses.

The MAT scores were grade-normed, and thus had a substantial within-grade age variability, i.e., younger pupils scoring consistently lower than their older grade peers. Although grade-normed scholastic measures have their merits, correlations between intelligence and scholastic achievement tend to be suppressed by the within-grade age variability inherent in the grade-normed scholastic measures. To remove this suppressing age effect, we used the age-adjusted multiple regression residuals of the MAT Total Achievement, Language, Math, and Reading scores in our analyses.[5]

2.2.4. Manifest variables for the SEM analyses

The nine TBCP variables that were aggregated to form the WRTP Processing Speed and STM/Working Memory Aggregates were also used as manifest variables for their respective latent constructs (WRTP Speed and STM/Working Memory Factors). Four WISC-R Verbal subscales, Information, Similarity, Vocabulary, and Comprehension were used to indicate a verbal factor (Verbal Factor), and four Performance subscales, Picture Completion, Picture Arrangement, Block Design, and Object Assembly, were treated as manifest variables for a performance factor (Performance Factor). The MAT Language, Math, and Reading scores were treated as observed indicators of an achievement factor (MAT Achievement Factor).

2.3. Analysis

2.3.1. Analyses of observed aggregates and composites

Correlations, multiple R^2s, and R^2 changes were computed from the pooled subsamples to evaluate the criterion validity of the aggregated TBCP. Similar to the analysis of the W-J III data, the analysis of the WRTP data was also conducted on three levels. On Level 1, the WRTP Processing Speed and STM/Working Memory Aggregates were treated as two separate predictors for the scholastic criterion measures, and the R^2 changes controlling the WISC-R Performance IQ score reflect the higher-order variability shared between the aggregates of TBCP and the scholastic measure. Alternatively, the higher-order variability shared between the conventional Performance IQ measure and the scholastic criterion index controlling for the TBCP predictors was also assessed according to the R^2 changes. Results of the analysis on this level could indicate, in addition to the explanatory power of the respective TBCP aggregates, whether the aggregates of TBCP provide incremental validity over the fluid ability measure of WISC-R Performance IQ.

On Level 2, the two aggregates were simply added to form a single WRTP Processing Speed+STM/Working Memory Composite, and on the third level, the TBCP composite was added to the WISC-R Verbal, Performance, and Full IQ scores, respectively, to create more agglomerated composites (WRTP Processing Speed+STM/Working Memory+WISC Verbal IQ Composite, WRTP Processing Speed+-

---

[5] Results of the analyses using the age-adjusted subtest scores of MAT Language, MAT Math, and MAT Reading were very similar to those using the age-adjusted total score of the MAT. Similar results were also obtained from the grade-normed MAT scores, and from the Wide Range Achievement Test Reading, Spelling, and Math scores. These results are not presented, but can be attained upon request.

Working Memory+WISC Performance IQ Composite, and WRTP Processing Speed+Working Memory+WISC Full IQ Composite) as hybrids of TBCP and selected conventional tests. The criterion validity of the Level 2 and Level 3 composites could be compared to that of their W-J III counterparts to examine the influence of a rigorous experimental paradigm on the tasks' criterion validity.

2.3.2. SEM analyses

The SEM approach adopted for the W-J III data was also adopted for the WRTP data. Five factors, WRTP Speed Factor, STM/Working Memory Factor, Verbal Factor, Performance Factor, and MAT Achievement Factor, were specified by their respective observed indicators. Four mathematically equivalent models (Models 2A, 2B, 2C, and 2D) that are analogous to Models 1A through 1D were fitted to the observed covariances using the maximum likelihood estimation procedure. R^2s and R^2 changes indicating the proportion of variability in MAT Achievement Factor explained by some or all of the other factors, as well as the unique contribution made by one of the explanatory factors beyond any of the others, were evaluated through these models. The unique impacts of predictors were also evaluated by gauging the chi-square changes caused by the zero constraints placed on the related beta weights in these models. Fig. 2a–d provide figural descriptions of these models.

3. Results

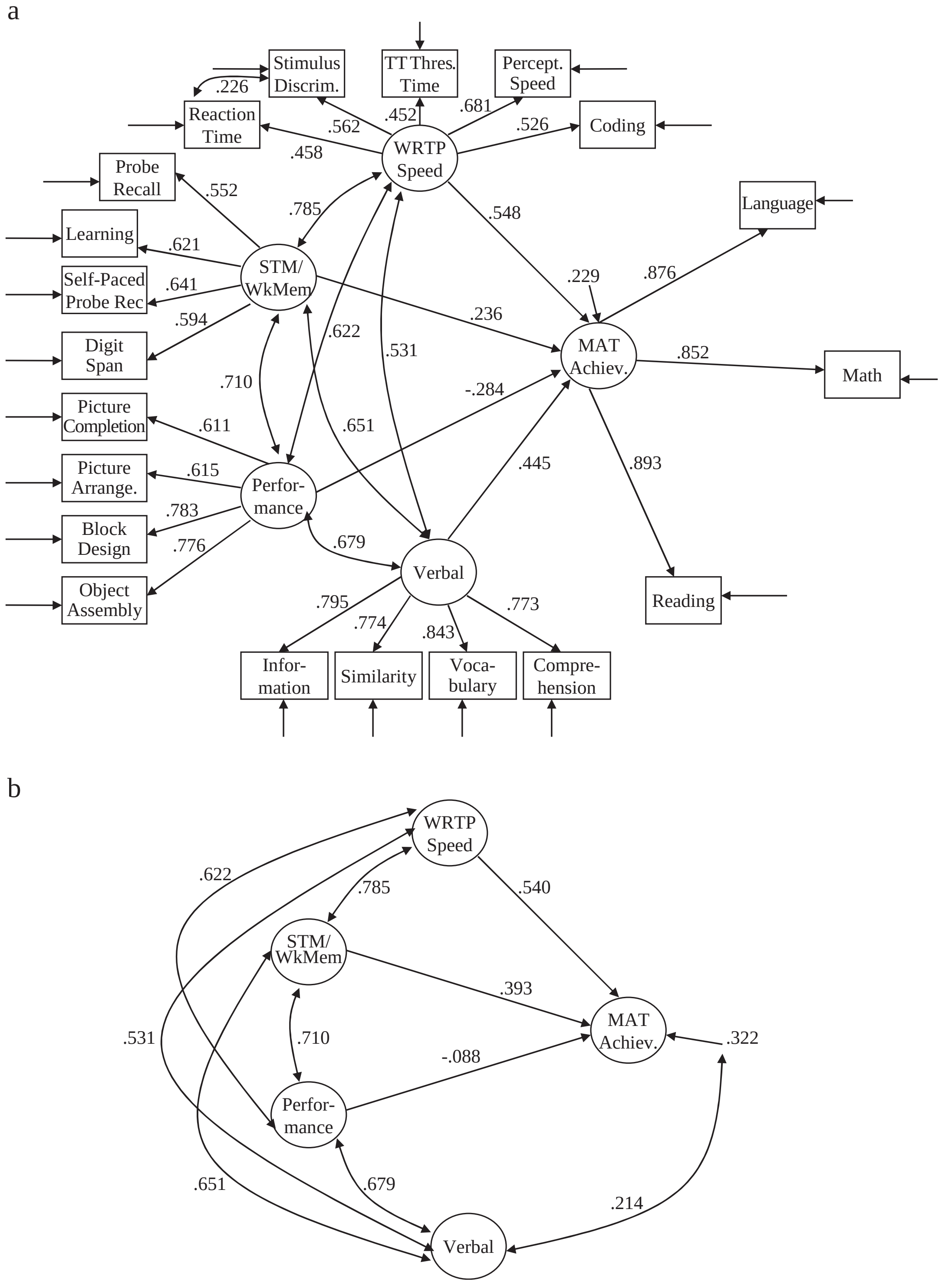
3.1. The Woodcock-Johnson III normative data

3.1.1. Results based on observed aggregates

On Composition Level 1, where the two aggregates of TBCP, WJ Processing Speed Cluster and WJ Working Memory Cluster, were treated as separate predictors, the explanatory power of the two separate TBCP predictors for the general scholastic performance score, WJ Total Achievement, was as strong as, or greater than, that of the fluid intelligence index. This is indicated by the multiple correlations (R^2s) in Table 1.

In Table 1, the R^2 for models using WJ Processing Speed Cluster and WJ Working Memory Cluster as predictors and WJ Total Achievement as the predicted variable ranges from 0.336 to 0.397. The counterpart R^2 based on WJ Fluid Reasoning Cluster is between 0.281 and 0.410. Moreover, WJ Processing Speed Cluster and WJ Working Memory Cluster both provide substantial incremental validity over the fluid ability index with the R^2 increments varying between 0.110 and 0.164 beyond WJ Fluid Reasoning Cluster. WJ Processing Speed Cluster and Working Memory Cluster add substantial explanatory power to one another: average increments in proportion of explained variability are, respectively, 0.125 and 0.149.

More noteworthy are the results related to WJ Fluid Reasoning Cluster on Composition Level 2. WJ Fluid Reasoning Cluster as a predictor for WJ Total Achievement not only results in a lower R^2 than the predictor of the TBCP composite, but also makes a rather meager exceptional contribution to WJ Total Achievement after the TBCP composite is statistically controlled. On the other hand, the TBCP Composite, WJ Processing Speed+Working Memory Composite, substantially surpass the contribution made by WJ Fluid Reasoning Cluster to WJ Total Achievement. The average R^2 increment induced by WJ Fluid Reasoning Cluster beyond WJ Processing Speed+Working Memory Composite is 0.049,


a
Stimulus Discrim.
TT Thres. Time
Percept. Speed
Coding
Reaction Time
.226
.562
.452
.681
.526
.458
WRTP Speed
Probe Recall
.552
.785
.548
Language
Learning
.621
Self-Paced Probe Rec
.641
STM/ WkMem
.229
.876
.594
.236
Digit Span
.622
.531
MAT Achiev.
.852
Math
.710
-.284
Picture Completion
.611
.651
.893
Picture Arrange.
.615
.445
Perfor- mance
Block Design
.783
.679
Object Assembly
.776
Verbal
Reading
.795
.773
.774
.843
Infor- mation
Similarity
Voca- bulary
Compre- hension
b
WRTP Speed
.622
.785
.540
STM/ WkMem
.393
.531
.710
MAT Achiev.
.322
-.088
Perfor- mance
.651
.679
.214
Verbal

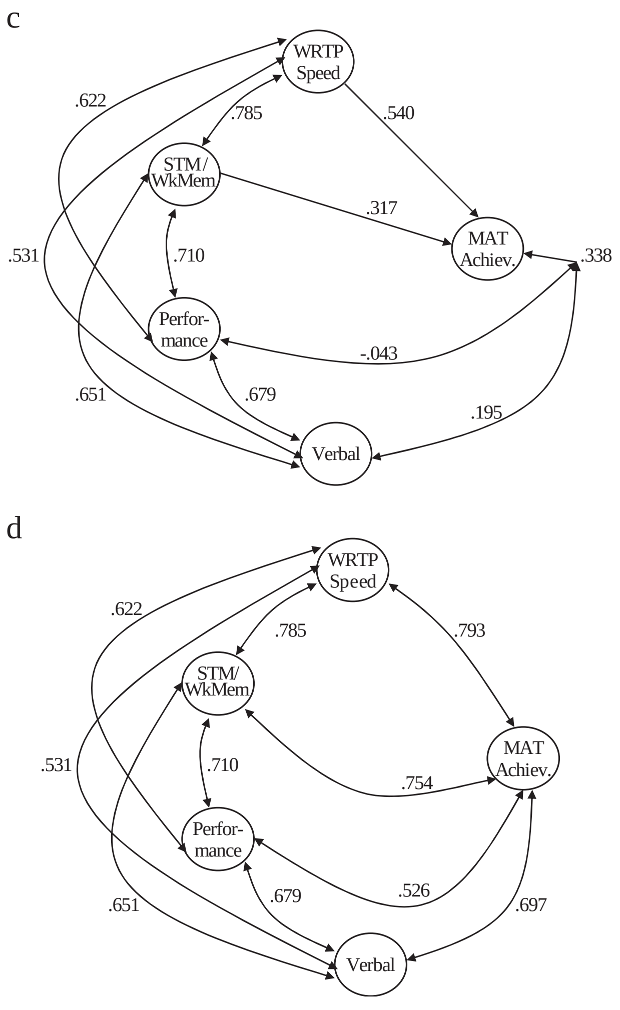


Fig. 2. (a) The five-factor model with WRTP Speed Factor, WRTP STM/Working Memory Factor, Performance, and Verbal as explanatory variables for MAT Achievement Factor. Note: Parameter estimates shown in the diagram are all standardized values. Variables symbolized by ovals are latent factors, and those in rectangles are manifest variables. The open-ended one-way arrow to a manifest or latent variable represents the residual variance to the variable. Variable acronyms: Stimulus Discrim.—Stimulus Discrimination; TT Threshold—Tachistoscopic Threshold Threshold Time; Percept. Speed—Perceptual Speed, Self-Paced Probe Rec—Self-Paced Probe Recall; Language—MAT Language; Math—MAT Math; Reading—MAT Reading; STM/WkMem—WRTP STM/Working Memory Factor; MAT Achiev—MAT Achievement Factor. (b) The five-factor model with WRTP Speed Factor, WRTP STM/Working Memory Factor, and Performance as explanatory variables for MAT Achievement Factor. (c) The five-factor model with WRTP Speed Factor and WRTP STM/Working Memory Factor as explanatory variables for MAT Achievement Factor. (d) The five-factor model with correlated factors but no explanatory variable for MAT Achievement Factor. Note: The models described in b–d are mathematically equivalent to that depicted in a, and parameter estimates associated with the manifest variables are thus identical in these models. The diagrams only present model parameters related to the latent factors. Values in parentheses are estimated 95% confidence intervals for the related parameters.

Table 1
R^2s and R^2 increments on composition levels 1, 2, and 3 of the W-J III data analysis

	Criterion measure: WJ Total Achievement					
	Age: 6–8, N=1095		Age: 9–13, N=2241		Age: 14–19, N=1641	
	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2
Composition Level 1						
Working Memory Cluster	0.360		0.360		0.397	
Processing Speed and Working Memory Clusters	0.475	0.115	0.494	0.134	0.523	0.126
Processing Speed Cluster	0.348		0.336		0.360	
Processing Speed and Working Memory Clusters	0.475	0.127	0.494	0.158	0.523	0.163
Processing Speed Cluster	0.348		0.336		0.360	
Processing Speed and Fluid Reasoning Clusters	0.436	0.088	0.467	0.131	0.539	0.179
Fluid Reasoning Cluster	0.281		0.303		0.410	
Processing Speed and Fluid Reasoning Clusters	0.436	0.155	0.467	0.164	0.539	0.129
Fluid Reasoning Cluster	0.281		0.303		0.410	
Working Memory and Fluid Reasoning Clusters	0.423	0.142	0.438	0.135	0.520	0.110
Working Memory Cluster	0.360		0.360		0.397	
Working Memory and Fluid Reasoning Clusters	0.426	0.063	0.438	0.078	0.520	0.113
Composition Level 2						
Fluid Reasoning Cluster	0.281		0.303		0.410	
Processing Speed+Working Memory Composite and Fluid R Cluster	0.504	0.223	0.538	0.235	0.597	0.187
Processing Speed+Working Memory Composite	0.475		0.494		0.522	
Processing Speed+Working Memory Composite and Fluid R Cluster	0.504	0.029	0.538	0.044	0.597	0.075
Comprehension Knowledge Cluster	0.348		0.462		0.563	
Processing Speed+Working Memory Composite and Comp K Cluster	0.554	0.179	0.645	0.183	0.715	0.152
Processing Speed+Working Memory Composite	0.475		0.494		0.522	
Processing Speed+Working Memory Composite and Comp K Cluster	0.554	0.079	0.645	0.151	0.715	0.193
Comprehension Knowledge+Fluid Reasoning Composite	0.410		0.488		0.596	
Processing Speed+Working Memory Composite and Comp K+Fluid R Composite	0.551	0.141	0.623	0.135	0.698	0.102
WJ Processing Speed+Working Memory Composite	0.475		0.494		0.522	
Processing Speed+Working Memory Composite and Comp K+Fluid R Composite	0.551	0.076	0.623	0.129	0.698	0.176
Composition Level 3						
General Intellectual Ability Index	0.518	0.563	0.656			
Processing Speed+Working Memory+Fluid R Composite	0.498	0.535	0.597			
Processing Speed+Working Memory+Comp K+Fluid R Composite	0.548	0.623	0.696			

Note: The R^2 values listed in two adjacent rows are obtained from two multiple regression models employing the indicated predictors. The model in the bottom row employs an additional predictor not included in the model above, and the ΔR^2 value stands for the R^2 increment caused by the additional predictor.

amounting to 1/11 of the mean total explained variability in WJ Total Achievement (0.049 versus mean total R^2 of 0.546). In contrast, the R^2 increment provided by WJ Processing Speed+Working Memory Composite over WJ Fluid Reasoning Cluster averages 0.215, equivalent to more than 1/3 of the total explained variability in WJ Total Achievement. It seems that the relation-educing fluid intelligence tests provide little more to scholastic performance than the composite of TBCP, whereas the composite of TBCP far surpasses the explanatory power of the fluid intelligence index.

It can also be seen on Composition Level 2 that the TBCP composite and WJ Comprehension Knowledge Cluster are complementary to each other for the prediction of WJ Total Achievement. WJ Processing Speed+WJ Working Memory Composite on average adds 0.170 of explained variability above WJ Comprehension Knowledge Cluster, while WJ Comprehension Knowledge Cluster provides a mean additional explanatory power of 0.141 in variability over WJ Processing Speed+WJ Working Memory Composite for the same criterion measure.

On Composition Level 3, the aggregates of TBCP were combined with the conventional indexes of crystallized intelligence and fluid intelligence to seek for an optimal combination of aggregated TBCP and conventional ability tests. WJ Processing Speed+Working Memory+Comprehension Knowledge Composite, WJ Processing Speed+Working Memory+Fluid Reasoning Composite, and WJ Processing Speed+Working Memory+Comprehension Knowledge+ Fluid Reasoning Composite were formed on this level, and were used as single predictors for the criterion measures.

Values in Table 1 indicate that the composites consisting of the aggregated TBCP and the crystallized intelligence index, WJ Processing Speed+Working Memory+Comprehension Knowledge Composite, and WJ Processing Speed+Working Memory+Comprehension Knowledge+Fluid Reasoning Composite lead to very similar R^2s for WJ Total Achievement, averaging 0.628 and 0.622, respectively. The lack of influence from the fluid intelligence index, WJ Fluid Reasoning Cluster, is again evident when comparing the R^2s between composites without WJ Fluid Reasoning Cluster and those employing WJ Fluid Reasoning Cluster, i.e., WJ Processing Speed+Working Memory+Comprehension Knowledge Composite versus WJ Processing Speed+Working Memory+Comprehension Knowledge+Fluid Reasoning Composite. Adding the Fluid Reasoning cluster to the composite comprising the aggregates of TBCP and the crystallized intelligence index does not lead to any R^2 increments. In fact, including WJ Fluid Reasoning into the composite slightly depresses the total R^2 in two of the age groups. The role of the conventional fluid intelligence index seems to be both insufficient in its own right and redundant when used along with aggregates of TBCP and crystallized intelligence indexes.

It is also of interest to compare these composites to WJ General Intellectual Ability as these composites may provide alternatives to the latter. In Table 1, the mean R^2 relating WJ General Intellectual Ability to WJ Total Achievement is 0.579, which is not higher than any of the R^2s based on the Level 3 composites. It appears that test composites containing both aggregates of TBCP and crystallized intelligence tests, with similar weights on both classes of measures, such as the WJ Processing Speed+Working Memory+Comprehension Knowledge Composite, could do equally well in predicting general scholastic performance as conventional general intelligence indexes, which typically are more heavily weighted by conventional complex ability tests than by TBCP.

3.1.2. Results based on SEM analyses

Three alternative multi-group models were specified and tested, (1) a model with equivalent factorial relations (factor loadings, factorial correlations, and latent structural relations) across the three age groups, (2) a model with invariant residual variances and residual covariances for the observed indicators across age groups, and (3) a model without any equality constraints on the factorial relations or the residuals of the observed indicators (congeneric). The chi-square statistic and several derived fit indexes from the three models are listed in Table 2. The derived indexes were used to offset the bias of the chi-square statistic that tends to inflate with the sample size for models that are "true" approximately, but not exactly. The derived indexes listed include the Tucker–Lewis Index (TLI; Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988; Tucker & Lewis, 1973), the Comparative Fit Index (CFI; Bentler, 1990), and the Root Mean Square Error of Approximation (RMSEA, Steiger & Lind, 1980). The TLI and CFI values for the multi-group models were derived on the basis of a null model described by Widaman and Thompson (2003). It can be seen from the derived indexes that the congeneric five-factor model described in Method, with three pairs of correlated residuals (see Fig. 1a), fits the data reasonably well, with both TLI and CFI in a well-fitting range (0.973 and 0.987), and an adequate RMSEA (0.058). The two crystallized ability variables, Verbal Comprehension and General Information, have residual correlations with Reading, and Reading and Written Language also have a residual correlation. These correlated residuals probably represent the specific academic skills shared between these tests. The derived fit indexes of the equivalent-factorial-relations model and the equivalent-residuals model are also in the well-fitting and adequate ranges, but the chi-square values of the two more constrained models are noticeably worse than that of the congeneric model. Moreover, model estimates from the congeneric model seem to indicate certain notable between-age differences in relations among the predictor factors and the Achievement Factor. To more fully describe the outcomes from the SEM analyses of the W-J III data, results based on the congeneric model were reported.

Table 2 also presents the R^2s and R^2 increments indicating the explanatory power of WJ Speed Factor, WJ Working Memory Factor, WJ Gc and WJ Gf for the criterion measure of WJ Achievement Factor, and the unique contributions made by some of the explanatory factors controlling for some of the other factors. These results confirm the general conclusion drawn from the observed aggregates and composites that the basic processing components of processing speed and working memory provide the explanatory power for scholastic achievement comparable to, or even stronger than, that offered by the conventionally measured crystallized and fluid abilities.

In particular, the two basic components jointly explain substantially more of scholastic achievement than fluid reasoning in the primary school age range. When both WJ Speed Factor and WJ Working Memory Factor are explanatory variables, they together account for about 90% of the variability in the WJ Achievement Factor in the 6–8 and 9–13 age groups, and over 70% in the 14–19 age group. On the other hand, when Gf is the only explanatory variable, it accounts for lower than 60% of the variability in the Achievement Factor in the primary school age range, and over 70% in the 14–19 age group. Such direct comparisons suggest that the basic processing factors outperform the fluid reasoning factor by 30% in predictive power in the primary school age range, and are comparable to the latter as predictors in the 14–19 age range. Values of the R^2 changes provide a similar picture. In the primary school age range, when Gf is controlled, the basic processing factors still explain more than 30% of the variability in the criterion factor, whereas Gf offers little when the factors of TBCP are controlled. Table 2 (the third section) also shows that the zero constraints placed on β_{Speed} and $\beta_{\text{Working Memory}}$ in the 6–8 and the 9–13 age groups in Model 1B resulted in highly significant chi-square changes whereas such constrains on β_{Gf} do not cause significant changes at $p<0.01$.

Table 2
Results from SEM analyses of W-J III Data (Age: 6–19, N=4979)

	Multi-group model fit indexes of the five-factor model				
	χ^2	df	TLI	CFI	RMSEA (95% confidence interval)
Null	35126.626	198			
Inv. Factorial Rel.	1026.455	138	0.964	0.975	0.060 (0.056, 0.064)
Inv. Observed Resid.	1009.364	121	0.958	0.975	0.064 (0.059, 0.068)
Congeneric	741.926	94	0.973	0.987	0.058 (0.052, 0.063)

Contributions of the WJ Speed, WJ Working Memory, Gf, and Gc Factors to the WJ Achievement Factor

Predictors in model	R^2 changes in achievement factor					
	6–8		9–13		14–19	
	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2
Speed, Working Memory, Gf (Model 1B)	0.924		0.906		0.856	
Speed, Working Memory, Gf, Gc (Model 1A)	0.927	0.003	0.917	0.011	0.939	0.083
Speed, Working Memory (Model 1C)	0.887		0.904		0.730	
Speed, Working Memory, Gf (Model 1B)	0.924	0.037	0.906	0.002	0.856	0.126
Gf (Model 1D)	0.580		0.598		0.767	
Speed, Working Memory, Gf (Model 1B)	0.924	0.344	0.906	0.308	0.856	0.089
Working Memory (Model 1D)	0.868		0.887		0.676	
Speed, Working Memory (Model 1C)	0.887	0.019	0.904	0.017	0.730	0.054
Speed (Model 1D)	0.575		0.465		0.454	
Speed, Working Memory (Model 1C)	0.887	0.312	0.904	0.439	0.730	0.276

Note: R^2 is based on 1 minus the estimated residual variance of the Achievement Factor in the related model. The R^2 values listed in two adjacent rows are obtained from two models employing the indicated predictors. The model in the bottom row employs one or more predictors not included in the model above, and the ΔR^2 value stands for the R^2 increment caused by the additional predictor(s).

	Changes in model fit		
	6–8	9–13	14–19
	$\Delta\chi^2/\Delta df$	$\Delta\chi^2/\Delta df$	$\Delta\chi^2/\Delta df$
(1) Equivalent Model A: four predictors (Speed, Working Memory, Gf, and Gc)			
Impact of βGc	0.342†/1	8.031/1	76.224/1
Impact of β_{Speed} and $\beta_{\text{Working Memory}}$	192.840/2	240.168/2	116.473/2
(2) Equivalent Model B: three predictors (Speed, Working Memory, and Gf)			
Impact of β_{Gf}	5.207†/1	0.872†/1	89.396/1
Impact of β_{Speed} and $\beta_{\text{Working Memory}}$	95.689/2	309.044/2	104.191/2
(3) Equivalent Model C: two predictors (Speed and Working Memory)			
Impact of $\beta_{\text{Working Memory}}$	117.444/1	385.097/1	202.498/1
Impact of β_{Speed}	4.862†/1	12.503/1	66.321/1

Note: The impact of β coefficient(s) was evaluated by fixing the relevant β parameter(s) to zero, and calculating the chi-square change induced by the zero constraint(s) placed on the full model. Symbol † indicates the related chi-square change is insignificant at p=0.01.

Results in Fig. 1d and Table 2 seem to indicate that on the level of latent factors, the two omnibus ability factors, Gc and Gf, have an increased impact on the Achievement Factor, and the two basic processing factors have a decreased impact in the 14–19 age group. Gc and Gf both have correlations lower than 0.80 with the Achievement Factor in the two primary school age groups, but the correlations increase to above 0.85 in the 14–19 group. Working Memory has correlations over 0.90 with Achievement in the younger groups, but a correlation of 0.82 with the Achievement Factor in the 14–19 age group. One possible explanation is that scholastic performance may be more related to the complex skills and abilities of Gc and Gf in the high school age range. Meanwhile, as was posited by Engle and Tuholski et al., tasks of working memory such as those employed in the W-J III may tap more of memory span than central executive control in older age groups, leading to a decrease in the predictive power of the WJ Working Memory factor. It should be noted, however, that the basic processing factors still have explanatory power close to that of the omnibus abilities in this age range, and account for over 70% of the variability in the Achievement Factor. Furthermore, the increased impact of the omnibus factors and the decreased impact of the basic processing factors are largely theoretical, as the changes are notable mostly on the level of latent factors. On the level of observed aggregates and composites, the TBCP composite still substantially exceeds the fluid reasoning index in explanatory power in this age range (R^2 change: 0.187), whereas the fluid reasoning index only adds a moderate amount of explained variance (0.075) above the TBCP composite.

Of the two basic processing factors, the Working Memory factor is apparently more powerful than the Speed factor, although the latter also accounts for over 45% of the Achievement variability. The Speed factor does not seem to provide a lot of unique contribution to the criterion factor over Working Memory. The related R^2 changes are all less than 0.10, and the chi-square changes caused by the zero constraints on β_{Speed} in Model 1C are moderate or insignificant in the primary school age range.

3.2. The WRTP data

Table 3 lists the pooled zero-order correlations among the WRTP Processing Speed and the STM/Working Memory Aggregates, the WISC-R IQ scores, and the MAT Total Achievement score. The two aggregates of TBCP have correlations with both the WISC-R IQ scores and the MAT Total Achievement score that are comparable to correlations typically observed among conventional ability tests and scholastic measures, indicating that the aggregates of mainly experimental TBCP have appropriate psychometric properties for practical testing of intelligence. The age-adjusted MAT Total Achievement

Table 3
Zero-order correlations between aggregates of TBCP, WISC-R IQ scores, and MAT total scores based on WRTP data (Age: 6–13, N=512)

	WISC-R Verb. IQ	WISC-R Perf. IQ	WISC-R Full IQ	Age-adjusted MAT Total	WRTP processing speed aggregate
WISC-R Perf. IQ	0.661				
WISC-R Full IQ	0.820	0.878			
Age-adjusted MAT Total	0.677	0.505	0.678		
WRTP Processing Speed Aggregate	0.464	0.575	0.571	0.635	
WRTP STM/Working Memory Aggregate	0.563	0.539	0.616	0.600	0.506

score has correlations with the WISC-R crystallized and fluid intelligence indexes, WISC Verbal IQ and WISC Performance IQ, (0.677 and 0.505, respectively) close to those of WJ Total Achievement with WJ Comprehension Knowledge Cluster and WJ Fluid Reasoning Cluster (averaged 0.636 and 0.540, respectively, in the same age range of 6 to 13), indicating MAT Total Achievement is a similar scholastic measure to WJ Total Achievement.

3.2.1. Results based on observed aggregates and composites

The results of analyses on Composition Levels 1, 2, and 3 are displayed in Table 4, which lists the R^2 and R^2 changes using the MAT Total Achievement score as the explained variable and the WISC-R IQ scores and the WRTP Processing Speed and STM/Working Memory Aggregates as the explanatory variables.

Table 4
R^2s and R^2 increments on Composition Levels 1, 2, and 3 of WRTP data analysis (Age: 6–13, N=512)

Criterion measure: age-adjusted MAT Total Achievement

	R^2	ΔR^2
Composition Level 1		
STM/Working Memory Aggregate	0.360	
Processing Speed and STM/Working Memory Aggregates	0.505	0.135
Processing Speed Aggregate	0.404	
Processing Speed and STM/Working Memory Aggregates	0.505	0.101
Performance IQ	0.255	
Processing Speed Aggregate and Perf. IQ	0.433	0.178
Processing Speed Aggregate	0.404	
Processing Speed Aggregate and Perf. IQ	0.433	0.029[a]
Performance IQ	0.255	
STM/Working Memory Aggregate and Perf. IQ	0.406	0.151
STM/Working Memory Aggregate	0.360	
STM/Working Memory Aggregate and Perf. IQ	0.406	0.046
Composition Level 2		
Performance IQ	0.255	
Processing Speed+STM/Working Memory Composite and Perf. IQ	0.507	0.252
Processing Speed+STM/Working Memory Composite	0.503	
Processing Speed+STM/Working Memory Composite and Perf. IQ	0.507	0.004[a]
Verbal IQ	0.458	
Processing Speed+STM/Working Memory Composite and Verb IQ	0.605	0.147
Processing Speed+STM/Working Memory Composite	0.503	
Processing Speed+STM/Working Memory Composite and Verb IQ	0.605	0.102
Composition Level 3		
Processing Speed+STM/Working Memory+Verb IQ Composite	0.604	
Processing Speed+STM/Working Memory+Full IQ Composite	0.561	
Processing Speed+STM/Working Memory+Perf. IQ Composite	0.449	

Note: The R^2 values listed in two adjacent rows are obtained from two multiple regression models employing the indicated predictors. The model in the bottom row employs an additional predictor not included in the model above, and the ΔR^2 value stands for the R^2 increment caused by the additional predictor.

[a] Refers to non-significance at p=0.01.

Table 4 shows that, on Composition Level 1 when WRTP Processing Speed Aggregate and STM/Working Memory Aggregate are treated as separate predictors, the proportions of variability in MAT Total Achievement explained are 0.404 and 0.360, respectively, higher in magnitude than the variability in MAT Total Achievement explained by the fluid intelligence index of the WISC-R, WISC Performance IQ (R^2=0.255). The incremental explanatory power that the TBCP aggregates provide above WISC Performance IQ for MAT Total Achievement is equivalent to 0.178 for WRTP Processing Speed Aggregate and 0.151 for WRTP STM/Working Memory Aggregate in proportion. The two WRTP TBCP aggregates each add over 0.100 (0.135 and 0.101) over the other in predicting the MAT Total Achievement score.

On Composition Level 2, where WRTP Processing Speed Aggregate and STM/Working Memory Aggregate were summed to form a composite, WRTP Processing Speed+STM/Working Memory Composite added 0.252 of explained variability above WISC Performance IQ for MAT Total Achievement, which was almost one half of the total explained variability of 0.507, whereas WISC Performance IQ's unique contribution over the TBCP composite is negligible (0.004).

Also on Composite Level 2, when WRTP Processing Speed+STM/Working Memory Composite was used together with WISC-R Verbal IQ as predictors, the TBCP composite added 0.147 over the crystallized ability index to the explained variability in the criterion measure, or close to 1/4 of the total explained variability of R^2=0.605. Such a pattern of R^2s and R^2 increments are quite similar to that observed for WJ Total Achievement in the W-J III normative sample data where the processing speed and working memory composite and the crystallized intelligence index each add about 0.150 or 1/5 of total explained variability beyond the other for the total achievement score—the TBCP composite and the crystallized intelligence index are supplemental to each other in predicting scholastic performance in both data sets.

On Composition Level 3, the TBCP composite, WRTP Processing Speed+STM/Working Memory Composite, was further combined with either WISC-R Full IQ, WISC-R Verbal IQ, or WISC-R Performance IQ, to explore the best form of combination consisting of TBCP and selected types of conventional tests. Once again, the composite of processing speed, short-term memory/working memory, and crystallized intelligence variables, WRTP Processing Speed+STM/Working Memory+WISC-R Verbal IQ Composite has the highest correlations with the criterion variable, followed by the TBCP and WISC Full IQ composite, WRTP Processing Speed+STM/Working Memory+-WISC Full IQ Composite. The TBCP and fluid intelligence composite, WRTP Processing Speed+STM/Working Memory+WISC Performance IQ Composite, has the weakest correlations with the scholastic measures. All Level 3 composites have higher correlations with the criterion measure than their conventional member indexes, namely, WISC-R Verbal IQ, WISC-R Full IQ, and WISC-R Performance IQ. For example, WRTP Processing Speed+STM/Working Memory+WISC Verbal IQ has a R^2 of 0.604 with regard to the criterion measure of MAT Total Achievement, which is a 0.146 increase in explanatory power compared to the R^2 based on WISC Verbal IQ alone.

WISC Full IQ has a slightly lower correlation with MAT Total Achievement (r=0.665 or R^2=0.442) than the TBCP composite (R^2=0.503). Unlike the W-J III general index, WJ General Intellectual Ability, which is an optimally weighed composite of both TBCP and conventional crystallized and fluid intelligence indexes, WISC-R Full IQ is a sum of mostly conventional crystallized and fluid intelligence tests. As a result, its explanatory power is not as strong as WJ General Intellectual Ability. Adding the WRTP TBCP composite to WISC-R Full IQ noticeably increases the explanatory power from R^2=0.442

to R^2=0.561, although the R^2 is still not as high as that associated with the best hybrid, the TBCP and crystallized ability composite.

The findings on all three composition levels of the WRTP data were analogous to those based on the W-J III data, demonstrating that experimental TBCP, while being readily analyzable to explicate their underpinnings, have desirable criterion validity for the practical testing of intelligence.

Table 5
Results from SEM analyses of WRTP data (Age: 6–13, N=512)

Model fit indexes of the five-factor model

χ^2	*df*	TLI	CFI	RMSEA (95% confidence interval)
367.893	159	0.945	0.958	0.051 (0.043, 0.059)

Contributions of the WRTP Speed, WRTP STM/Working Memory, WISC Performance, and WISC Verbal Factors to the MAT Achievement Factor

R^2 changes in Achievement Factor

Predictors in model	R^2	ΔR^2
Speed, STM/Working Memory, Performance (Model 2B)	0.678	
Speed, STM/Working Memory, Performance, Verbal (Model 2A)	0.773	0.095
Speed, STM/Working Memory (Model 2C)	0.662	
Speed, STM/Working Memory, Performance (Model 2B)	0.678	0.016
Performance (Model 2D)	0.277	
Speed, STM/Working Memory, Performance (Model 2B)	0.678	0.401
STM/Working Memory (Model 2D)	0.564	
Speed, STM/Working Memory (Model 2C)	0.662	0.098
Speed (Model 2D)	0.629	
Speed, STM/Working Memory (Model 2C)	0.662	0.039

Note: R^2 is based on 1 minus the estimated residual variance of the Achievement Factor in the related model. The R^2 values listed in two adjacent rows are obtained from two models employing the indicated predictors. The model in the bottom row employs one or more predictors not included in the model above, and the ΔR^2 value stands for the R^2 increment caused by the additional predictor(s).

Changes in model fit

	$\Delta\chi^2/\Delta df$
(1) Equivalent Model A: four predictors (Speed, STM/Working Memory, Performance, and Verbal)	
Impact of β_{Verbal}	29.162/1
Impact of β_{Speed} and $\beta_{\text{STM/Working Memory}}$	147.703/2
(2) Equivalent Model B: three predictors (Speed, STM/Working Memory, and Performance)	
Impact of $\beta_{\text{Performance}}$	1.589†/1
Impact of β_{Speed} and $\beta_{\text{STM/Working Memory}}$	156.020./2
(3) Equivalent Model C: two predictors (Speed and STM/Working Memory)	
Impact of $\beta_{\text{STM/Working Memory}}$	7.695/1
Impact of β_{Speed}	18.849/1

Note: The impact of β coefficients was evaluated by fixing the relevant β parameter(s) to zero, and calculating the chi-square change induced by the zero constraint(s) placed on the full model. Symbol † indicates the related chi-square change is insignificant at p=0.01.

3.2.2. Results based on SEM analyses

The sequence of mathematically equivalent five-factor models, with correlated residuals between Reaction Time and Stimulus Discrimination (Fig. 2a), fit the data adequately, as indicated by the fit indexes in Table 5. TLI and CFI are close to or above 0.95, and RMSEA is 0.051.

The evaluation of R^2s and R^2 increments shown in Table 5 validates the basic findings from the WRTP observed aggregates and composites. The cognitive components of processing speed and short-term memory/working memory seem to relate to the criterion performance of scholastic achievement at least as strongly as, and probably even more so than, the conventionally determined intellectual abilities, particularly the fluid ability defined by the WISC-R Performance tests. For example, WRTP Speed Factor and WRTP STM/Working Memory Factor jointly account for 0.662 of the MAT Achievement Factor variance in Model 2C, and including Performance Factor as an additional explanatory variable in Model 2B adds little (R^2 change: 0.016) to it. Performance Factor explains 0.277 of variability in MAT Achievement Factor (Model 2D, Fig. 2d), which is considerably weaker in strength than the variance explained by the WRTP Speed and STM/Working Memory factors together (R^2=0.662).

An inspection of individual roles of the WJ Speed and STM/Working Memory factors reveals that the two factors are similar in strength (correlations 0.793 and 0.754 as shown in Fig. 2d), with neither substantially exceeding the other in explaining for MAT Achievement Factor. However, both appear to do substantially better than Performance Factor (correlation with MAT Achievement: 0.526), and the chi-square increment after fixing β_{Speed} and $\beta_{\text{STM/Working Memory}}$ to zero in Model 2B is 156.020 (*df*=2) whereas the counterpart chi-square change when $\beta_{\text{Performance}}$ is fixed to zero is insignificant.

The explained variance in MAT Achievement Factor is moderately increased, from R^2=0.678 to R^2=0.773 when Verbal Factor joins force with the WRTP Speed Factor, WRTP STM/Working Memory Factor and the Performance factor as explanatory variables in Model 2A. The chi-square change caused by the zero constraint on β_{Verbal} in Model 2A is also moderate (29.162 with *df*=1). On the other hand, the basic processing factors seem to exercise an substantial unique impact above Verbal Factor on scholastic achievement, as the chi-square increases by 147.703 (*df*=2) after β_{Speed} and $\beta_{\text{STM/Working Memory}}$ are fixed to zero in Model 2A.

4. Discussion

Despite recent evidence indicating a dominant role of basic cognitive components in *g*, and in the *g*-scholastic performance correlation, TBCP designed to measure these components are still under-represented in current test batteries of intelligence. A major concern over the practical use of TBCP is that individual TBCP are predisposed to be both process-specific and method-specific, and thus do not have high *g* loadings or strong criterion validity. High *g* loadings and strong criterion validity can nonetheless be achieved from TBCP if they are properly aggregated. In the present study, we demonstrated that aggregates of the processing speed and working memory tasks could have criterion validity comparable to that provided by traditional IQ measures. In the present study, we used the W-J III processing speed and working memory aggregates as predictors for the criterion measure of scholastic achievement, and obtained evidence for a desirable criterion validity of aggregated TBCP in all pre-college school age groups. We used TBCP aggregates tapping similar mechanisms but with more rigorous experimental paradigms in the WRTP sample, and observed comparable results regarding the criterion validity of the TBCP aggregates. These TBCP aggregates, whether treated as separate predictors or as simple sums, have substantial correlations with scholastic performance, and these correlations are similar to correlations between conventional test composites and scholastic performance.

We compared these aggregates of TBCP to the crystallized intelligence indexes in both the W-J III normative sample and the WRTP sample, and found that the two families of measures were complementary to each other for the explanation of scholastic performance. The incremental predictive power of processing speed and working memory beyond crystallized intelligence was also confirmed on the level of latent traits in both samples. Since some of the basic processes measured by TBCP are likely to be the cognitive determinants for both crystallized intelligence and scholastic performance, the primary emphasis probably should be placed on TBCP rather than on tests of crystallized intelligence. Furthermore, conventional tests of crystallized intelligence as measures of accumulated knowledge and skill through formal instruction are good predictors mainly for one particular criterion, scholastic achievement. TBCP, on the other hand, tap basic processes more general to various intellectual activities, and are likely to be more versatile predictors for other criteria.

The moderate to substantial supplemental explanatory power of the processing speed/working memory variables and of the crystallized intelligence measures to one another for scholastic performance are revealing theoretically in that they shed light on the cognitive underpinnings of scholastic achievement. The mutual supplementing power from the two classes of measures seems to implicate a partitioning of four variability components in scholastic performance, i.e., variability attributable to both basic processes and crystallized intelligence, variability contributed exclusively by the basic processes, variability only ascribable to crystallized intelligence, and variability explained by neither basic processes nor crystallized intelligence. It should be noted that the basic processes are likely to play a causal role in the accumulation of knowledge and skills, thereby also causally determining to a degree the variability they share with crystallized intelligence. Considering that the basic processes and crystallized intelligence could jointly account for about 60% of the variability in scholastic achievement on the level of observed aggregates and composites, and up to 70% or more on the level of latent traits in the present study, the total amount of variability in scholastic performance attributable to the basic processes is remarkable.

Our analyses of both the W-J III normative data and the WRTP data also suggest that the optimal criterion validity for scholastic performance could be achieved by combining the aggregates of TBCP with the index of crystallized intelligence. Such a hybrid is at least as good as the general indexes of the W-J III and the WISC-R, and slightly outperforms the composite of crystallized intelligence and fluid intelligence indexes. This brings about the possibility for a new class of intelligence batteries. These batteries will include an extensive collection of TBCP, and a selection of crystallized intelligence tests, whose content may vary with the cultural, educational and occupational background of the targeted testees, and will have both explicit underlying cognitive mechanisms and desirable criterion validity.

The outcome from our analyses of the observed WRTP TBCP aggregates and composites largely paralleled that of the W-J III data, suggesting that aggregates of TBCP following more closely an experimental paradigm also have desirable criterion validity. Since these TBCP have the advantage of being more readily subjected to experimental analysis, thereby having better defined processing mechanisms and providing richer task parameter information, and with the advance of computer technology they also become increasingly easier to design and administer, their use in the practical testing of intelligence should be encouraged. From a broader perspective, the use of experimental TBCP in the practical testing of intelligence may help bridge the two long divided psychological traditions, the experimental tradition and the psychometric tradition, leading to an integrated science of psychology.

One of the most notable findings of the present study is that the value of fluid ability tests in predicting scholastic achievement appears to be doubtful. Fluid intelligence tests are thought to reflect a person's relatively "culture-fair", intellectual potential, and seem to measure a dimension of intelligence more general than that of crystallized intelligence. Their cognitive underpinnings, however, are not as readily explicable as those of TBCP, and are thus theoretically vague. In the present study, we found them both redundant and insufficient, particularly in the primary school age range, for explaining scholastic achievement when compared to the aggregated TBCP that have more basic and explicable processing mechanisms. The results from the SEM analyses further add to the suspicion that conventional fluid ability measures may not add much more to scholastic achievement beyond the TBCP of processing speed and working memory. Moreover, the questionable role of fluid abilities was manifested both by more strictly constructed reasoning ability tests, as those in the W-J III battery, and by tests measuring more broadly defined problem solving abilities, such as those of the WISC-R Performance subscales employed in the WRTP. It seems that the practical value and the theoretical significance of fluid intelligence tests need to be critically reexamined.

It is possible that the apparent weakness of the fluid ability measures arises from certain measurement properties of these measures. For example, in younger age groups, questions in conventional fluid ability tests may largely measure children's comprehension of task requirements, but not truly their problem solving abilities. It also appears plausible, however, that the majority of general variability in human intelligence is determined by certain basic cognitive processing components, and tests of fluid abilities are to a considerable degree indirect, and therefore oftentimes inept, measures of these basic components. The fluid ability tests are mostly designed to indicate a person's ability to educe correlates and relations, and are arguably better measures of *g* than other tests. The TBCP analyzed in the present study apparently did not engage the higher-order processes hypothesized for reasoning, such as rule abstraction, rule mapping, etc., yet they almost exhaustively mediate the observed correlations between fluid reasoning and scholastic achievement, both of which are analytical-skills laden. Individual differences in reasoning may be mainly ascribable to the basic components of processing speed and working memory, and higher-order processes such as those of rule abstraction and mapping do not constitute the main source of individual differences. This possibility has been contemplated by those who observed a dominant influence of working memory on the reasoning ability (Carpenter, Just, & Shell, 1990; Kyllonen & Christal, 1990), and the results of the present study seem to empirically support the conjecture. Moreover, the TBCP of processing speed and working memory appear to uncover sources of individual differences not tapped effectively by tests of fluid abilities. The seeming superiority of TBCP in criterion validity may stem from a wider range of processing mechanisms reflected by these tasks, including mechanisms that control stimulus encoding, memory span, and central executive, whereas the range for the fluid ability tests may be more restrictive, putatively focusing on only the central executive component engaged in reasoning. A wider range of cognitive processing may nonetheless be crucial for the success in scholastic achievement, making the TBCP more effective predictors than tests of fluid abilities. Such a theoretical account, however, is still largely speculative, as studies supporting a bottom-up notion of fluid abilities, the present one included, have been mostly correlational in nature, and the correlations are typically observed from concurrent measures. Evidence for concurrent criterion validity has its limitations in that the concurrent correlations may to a degree stem from the similarity in the concurrent test settings. More compelling evidence for predictive validity needs to be obtained from longitudinal studies in which TBCP administered at an earlier date are shown to effectively predict criterion measures at a later data. Ultimately, however, the causal account that explains how and why TBCP predict criterion measures can only be established through experimental analyses of fluid abilities and scholastic achievement, and with systematic manipulations of mechanisms controlling processing speed and working memory tasks.

The results of the study may have important implications for educators and practitioners. Scholastic achievement represents acquired knowledge and skills, and the acquisition of the knowledge and skills is the ultimate goal of formal instruction. When there are difficulties in the acquisition, instructional resources are more customarily devoted to higher-order mental processing, whereas the difficulties may actually come from processing mechanisms that are rather basic. The results of the present study suggest that instructional resources may be better spent if they are directed to where the difficulties really exist, namely, in basic information processing, and are dedicated to those who need support and compensation in this regard.

The results of the present study also give rise to a few unanswered questions. They are: (1) how processing speed and working memory relate to one another; (2) whether visuospatial working memory differs from verbal–numerical working memory in criterion validity; and (3) to what extent the explanatory power found in TBPC of processing speed and working memory is applicable to criterion performance other than scholastic achievement.

The possible relationship between processing speed and working memory has been an issue of discussion in recently years (Conway et al., 2002; Kyllonen & Christal, 1990). The processing speed component extracted from the W-J III normative data seems to have a weaker explanatory power than the working memory component, but the explanatory power of processing speed may be strengthened with a wider selection of processing speed variables to reduce the method variances arising from the procedures of specific tasks. The larger criterion related R^2s contributed by the WRTP Processing Speed Aggregate and their latent factor seem to result from such a wider variable selection. Still, processing speed does not appear to far exceed working memory in accounting for the criterion variance, implicating that processing speed may represent cognitive underpinnings encompassed by working memory, while working memory may pertain to additional mechanisms. The W-J III working memory factor can explain as much as 90% of the variability in the achievement factor, and such strong explanatory power, although obtained on the level of latent factors and thus only theoretical, is unlikely to be observed from a latent processing speed factor even with the widest possible selection of processing speed variables. On the other hand, the strength of processing speed variables in practical testing of intellectual abilities needs to be reckoned, as they are not only relevant to criterion performance in their own right, but appear to supplement the explanatory power of working memory measures on the level of observed aggregates as well. Processing speed also seems to be a distinct theoretical component, with tasks measuring the component typically involving neither intensive memory span nor central executive required by prototypical working memory tasks. Some of the processing speed tasks, stimulus discrimination for instance, are found to evoke neurological activities highly similar to those of analogical reasoning (Duncan et al., 2000), and, as noted earlier, they also appear to mediate age changes in intellectual abilities more effectively than working memory tasks. It is possible that processing speed represents a set of highly age-related cognitive mechanisms that partly constitute working memory. These mechanisms mediate most of the age changes in working memory, and also account for a sizable proportion of the within-age variability in working memory. Working memory, however, seem to comprise mechanisms additional to those of processing speed. These additional mechanisms, while adding a considerable amount of variability to the within-age individual differences above processing speed mechanisms, are relatively stable across age groups, and thus reduce the total between-age variability in working memory. A thorough explication of the roles of the two basic components needs to be conducted experimentally, preferably with the aid of neurological techniques, and the explication may shed important light on the foundation of human intelligence.

The possible differential relations of visuospatial working memory and verbal–numerical working memory with other intellectual abilities are another issue of discussion in recent research of working memory and intelligence (Miyake et al., 2001). The results of the present study seem to indicate stronger criterion validity of verbal–numeric working memory tasks than that of visuospatial working memory tasks. The visuospatial short-tem memory and working memory tasks adopted by WRTP, Probe Recall, Learning, and Self-Pace Probe Recall have lower correlations with the achievement tests than the WISC-R Digit Span, a numeric working memory task. Their correlations with the scholastic achievement measures also seem to be weaker than those of the W-J III verbal–numerical working memory tests, whose criterion validity is comparable in magnitude to that of the WISC-R Digit Span. The stronger explanatory power of WJ Working Memory Factor than that of WRTP STM/Working Memory Factor may also reflect the difference between the two types of working memory tasks. Because a direct comparison between the two types of working memory tasks was not possible on the level of aggregates or factors in the present study, and the differences noted between the two working memory types on the level of individual tests are confined to the primary school age group only (6–13), it is still unclear as how they actually differ in criterion validity in general. A direct comparison of this kind will be both theoretically and practically informative. It should also be noted that the predictive strength of verbal–numeric working memory may be underestimated in the present study, as even the strongest verbal–numeric working memory predictors in the present study, WJ Numbers Reversed and WISC-R Digit Span, may be argued as representing mainly the short-term memory subcomponent in older age groups. With a strengthened representation of the central executive subcomponent in working memory tasks, the tasks' criterion validity is likely to further increase (Engle et al., 1999). On the other hand, working memory tasks emphasizing the central executive component may sometimes be contended as engaging complex thinking processes not easily distinguishable from those of reasoning tests, rendering the implication of the criterion validity provided by TBCP ambiguous, whereas the short-term memory and working memory tasks employed in the present study stand clearly apart from reasoning tests, demonstrating more unequivocally the power of basic processes in criterion performance.

The present study only assessed the criterion validity of aggregated TBCP with regard to scholastic performance. There are other criterion measures of intelligence, e.g., skills in various occupations and social settings, and their relations with aggregates of TBCP should also be examined. TBCP may have differential criterion validity depending on the nature of the criterion adopted, and for different criteria there may be different aggregates of TBCP that are more predictive. Investigations in the possible relationship between aggregated TBCP and various criteria may provide insights into the cognitive underpinnings of a wide range of intellectual aptitudes.

5. Uncited references

Cattell & Farand, 1890
Jastak & Wilkinson, 1984

Meeker & Escobar, 1995
Neale & Miller, 1997

Appendix A. Correlations between W-J III observed variables

W-J III clusters

Age: 6–8 (*N*=1095)

	WJ Processing Speed Cluster	WJ Working Memory Cluster	WJ Comprehension Knowledge Cluster	WJ Fluid Reasoning Cluster	WJ General Intellectual Ability Index
WJ Working Memory Cluster	0.49				
WJ Comprehension Knowledge Cluster	0.38	0.49			
WJ Fluid Reasoning Cluster	0.45	0.53	0.53		
WJ General Intellectual Ability Index	0.64	0.77	0.74	0.79	
WJ Total Achievement	0.59	0.60	0.59	0.53	0.72

Age: 9–13 (*N*=2241)

	WJ Processing Speed Cluster	WJ Working Memory Cluster	WJ Comprehension Knowledge Cluster	WJ Fluid Reasoning Cluster	WJ General Intellectual Ability Index
WJ Working Memory Cluster	0.41				
WJ Comprehension Knowledge Cluster	0.34	0.47			
WJ Fluid Reasoning Cluster	0.37	0.52	0.55		
WJ General Intellectual Ability Index	0.57	0.74	0.76	0.79	
WJ Total Achievement	0.58	0.60	0.68	0.55	0.75

Age: 14–19 (*N*=1641)

	WJ Processing Speed Cluster	WJ Working Memory Cluster	WJ Comprehension Knowledge Cluster	WJ Fluid Reasoning Cluster	WJ General Intellectual Ability Index
WJ Working Memory Cluster	0.45				
WJ Comprehension Knowledge Cluster	0.37	0.51			
WJ Fluid Reasoning Cluster	0.43	0.55	0.62		
WJ General Intellectual Ability Index	0.61	0.77	0.78	0.81	
WJ Total Achievement	0.60	0.63	0.75	0.64	0.81

W-J III tests (Age: 6–8, N=1095)

	Visual Matching	Decision Speed	Numbers Reversed	Auditory Working Memory	Verbal Comprehension	General Information	Analysis Synthesis	Concept Formation	Broad Reading	Broad Written Language
Decision Speed	0.55									
Numbers Reversed	0.46	0.30								
Auditory Working Memory	0.35	0.26	0.39							
Verbal Comprehension	0.34	0.29	0.44	0.43						
General Information	0.27	0.28	0.36	0.33	0.70					
Analysis Synthesis	0.38	0.28	0.45	0.44	0.55	0.47				
Concept Formation	0.37	0.28	0.39	0.36	0.45	0.37	0.51			
Broad Reading	0.58	0.39	0.53	0.47	0.65	0.55	0.50	0.49		
Broad Written Language	0.62	0.39	0.50	0.45	0.55	0.44	0.47	0.45	0.88	
Broad Math	0.64	0.42	0.57	0.50	0.54	0.43	0.50	0.49	0.77	0.78

W-J III tests (Age: 913, N=2241)

	Visual Matching	Decision Speed	Numbers Reversed	Auditory Working Memory	Verbal Comprehension	General Information	Analysis Synthesis	Concept Formation	Broad Reading	Broad Written Language
Decision Speed	0.56									
Numbers Reversed	0.38	0.25								
Auditory Working Memory	0.34	0.25	0.39							
Verbal Comprehension	0.30	0.27	0.40	0.47						
General Information	0.28	0.27	0.33	0.43	0.77					
Analysis Synthesis	0.31	0.27	0.40	0.41	0.58	0.49				
Concept Formation	0.30	0.24	0.35	0.36	0.49	0.42	0.57			
Broad Reading	0.48	0.34	0.50	0.50	0.74	0.68	0.54	0.44		
Broad Written Language	0.55	0.38	0.40	0.50	0.64	0.61	0.52	0.41	0.86	
Broad Math	0.50	0.37	0.53	0.51	0.59	0.52	0.56	0.53	0.73	0.75

W-J III tests (Age: 1419, *N*=1641)

	Visual Matching	Decision Speed	Numbers Reversed	Auditory Working Memory	Verbal Comprehension	General Information
Decision Speed	0.52					
Numbers Reversed	0.42	0.25				
Auditory Working Memory	0.39	0.29	0.56			
Verbal Comprehension	0.34	0.29	0.45	0.51		
General Information	0.31	0.28	0.36	0.47	0.83	
Analysis Synthesis	0.36	0.29	0.44	0.47	0.61	0.53
Concept Formation	0.36	0.28	0.43	0.43	0.55	0.50
Broad Reading	0.51	0.36	0.56	0.54	0.79	0.77
Broad Written Language	0.54	0.37	0.53	0.52	0.72	0.70
Broad Math	0.54	0.32	0.52	0.50	0.67	0.63

	Analysis Synthesis	Concept Formation	Broad Reading	Broad Written Language
Concept Formation	0.59			
Broad Reading	0.60	0.56		
Broad Written Language	0.58	0.54	0.85	
Broad Math	0.57	0.61	0.72	0.74

Appendix B. Correlations between WRTP observed tests

	Reaction Time	Stimulus Discrimination	Tachistoscopic Threshold	Perceptual Speed	Coding	Probe Recall
Stimulus Discrim.	0.483					
Tachis. Threshold	0.178	0.318				
Perceptual Speed	0.264	0.387	0.279			
Coding	0.261	0.287	0.179	0.405		
Probe Recall	0.244	0.333	0.233	0.324	0.234	
Learning	0.260	0.316	0.317	0.302	0.235	0.340
Self-Paced Probe Rec.	0.194	0.215	0.274	0.300	0.236	0.405
Digit Span	0.218	0.198	0.208	0.348	0.191	0.286
Information	0.252	0.341	0.230	0.317	0.197	0.266
Similarity	0.207	0.252	0.232	0.269	0.158	0.233
Vocabulary	0.169	0.270	0.245	0.264	0.165	0.228
Comprehension	0.202	0.214	0.237	0.282	0.149	0.264
Picture Completion	0.152	0.245	0.294	0.182	0.152	0.162
Picture Arrangement	0.147	0.249	0.321	0.207	0.155	0.200
Block Design	0.177	0.287	0.313	0.389	0.273	0.316
Object Assembly	0.171	0.277	0.310	0.235	0.215	0.299
MAT Language	0.354	0.355	0.256	0.471	0.354	0.306
MAT Math	0.355	0.365	0.294	0.502	0.382	0.366
MAT Reading	0.386	0.394	0.271	0.493	0.356	0.380

	Learning	Self-Paced Probe Recall	Digit Span	Information	Similarity	Vocabulary	Comprehension
Self-Paced Probe Recall	0.465						
Digit Span	0.291	0.367					
Information	0.337	0.299	0.425				
Similarity	0.305	0.297	0.375	0.626			
Vocabulary	0.297	0.304	0.400	0.660	0.646		
Comprehension	0.294	0.305	0.403	0.574	0.597	0.690	
Picture Completion	0.340	0.249	0.278	0.371	0.349	0.391	0.337
Picture Arrangement	0.252	0.204	0.260	0.400	0.394	0.408	0.396
Block Design	0.417	0.405	0.338	0.444	0.428	0.407	0.410
Object Assembly	0.383	0.334	0.224	0.385	0.374	0.388	0.355
MAT Language	0.353	0.356	0.468	0.523	0.426	0.471	0.420
MAT Math	0.361	0.415	0.494	0.512	0.463	0.454	0.474
MAT Reading	0.370	0.393	0.531	0.582	0.483	0.543	0.460

	Picture Completion	Picture Arrangement	Block Design	Object Assembly	MAT Language	MAT Math
Picture Arrangement	0.378					
Block Design	0.440	0.440				
Object Assembly	0.497	0.495	0.627			
MAT Language	0.259	0.235	0.368	0.275		
MAT Math	0.256	0.241	0.433	0.297	0.758	
MAT Reading	0.313	0.305	0.446	0.352	0.791	0.740

Note: Correlations related to Reaction Time, Stimulus Discrimination, Tachistoscopic Threshold, Perceptual Speed, Probe Recall, Learning, and Self-Paced Probe Reccall, are based on the age- and sex-adjusted regression residuals of these variables.

Appendix C. Reliability estimates

Reliability estimates of the W-J III composites

Composite/age	68	913	1419
WJ Processing Speed Cluster	0.937	0.915	0.919
WJ Working Memory Cluster	0.905	0.898	0.911
WJ Comprehension Knowledge Cluster	0.924	0.939	0.954
WJ Fluid Reasoning Cluster	0.956	0.947	0.947
WJ General Intellectual Ability	0.963	0.962	0.968
WJ Total Achievement	0.973	0.972	0.972

Reliability estimates of the WRTP composites	
WRTP Processing Speed Aggregate	0.90
WRTP STM/Working Memory Aggregate	0.86
WISC Verbal IQ	0.94
WISC Performance IQ	0.90
WISC Full IQ	0.96
Age-adjusted MAT Total Achievement	0.91

Note: The reliability estimates of the W-J III composites were weighted age group averages on the basis of the W-J III reported cluster reliability estimates (McGrew & Woodcock, 2001). The reliability estimates of the WISC Verbal IQ, WISC Performance IQ, and WISC Full IQ were reported in the manuals of the WISC-R (Wechsler, 1974). The reliability estimates of WRTP Processing Speed Aggregate were based on the split-half reliability estimates of the decision time of Reaction Time (0.97), the decision time of Stimulus Discrimination (0.85), the threshold time of Tachistoscopic Threshold (0.62), the testretest reliability of Coding (0.72), and the testretest reliability estimate of Perceptual Speed (0.81). The reliability estimate of WRTP STM/Working Memory Aggregate was obtained on the basis of the four member tests, the percent correct of Probe Recall (split-half reliability: 0.60), the percent correct of Learning (internal consistency reliability: 0.68), the percent correct of Self-Paced Probe Recall (split-half reliability: 0.79), and Digit Span (testretest reliability: 0.78). The composite reliability estimates of WRTP Processing Speed Aggregate and STM/Working Memory Aggregate were computed according to the following formula:

$$r_{cc} = 1 - \frac{\sum s_j^2 - \sum s_j^2 r_{jj}}{\sum s_j^2 + 2\sum s_j s_k r_{jk}}$$

where r_{cc} is the reliability of a composite; s_j is the standard deviation of test j; s_k is the standard deviation of test k; r_{jj} is the reliability of test j; and r_{jk} is the correlation between tests j and k. The reliability estimates for the age-adjusted MAT Total Achievement is the Crobach alpha coefficient obtained on the basis of the correlations among the age-adjusted MAT Language, Math, and Reading subscales.

Appendix D. Descriptive Statistics

Descriptive statistics of the W-J III variables (*N*=4979)

Test name	Mean	Standard deviation
Visual Matching	98.66	14.03
Decision Time	98.80	14.31
Numbers Reversed	98.86	14.68
Auditory Working Memory	98.74	14.26
Verbal Comprehension	98.85	14.44
General Information	97.99	13.88
Concept Formation	98.77	14.99
Analysis Synthesis	99.64	14.16
Broad Reading	99.36	13.77
Broad Written Language	99.62	13.77
Broad Math	99.40	14.12

Note: The descriptive statistics of the W-J III tests are based on the age-normed standardized scores. For Broad Reading, Broad Writing, and Math, the statistics are determined as the equally weighted-composites of the means and variances of their ingredient tests (5 subtests for Broad Reading, 4 for Broad Writing, and 4 for Math). To reduce the disparity in scaling between the manifest variables (standard deviations close to 15) and their latent factors (standard deviations constrained to be 1.0), the manifest variables are rescaled to have 1/15 of their original standard deviations (close to 1.0) in the SEM analyses.

Descriptive statistics of the WRTP variables (N=512)

Variable name	Mean	Standard deviation
Standardized Residual of Decision Time for Reaction Time	0.00	1.01
Standardized Residual of Tachistoscopic Threshold Threshold Time	0.01	1.00
Standardized Residual of Decision Time for Stimulus Discrimination	0.02	0.97
WISC-R Coding Scale Score	10.24	3.50
Standardized Residual of Perceptual Speed	0.01	0.99
Standardized Residual of Percent Correct of Probe Recall	0.02	0.99
Standardized Residual of Percent Correct of Learning	0.01	0.99
Standardized Residual of Percent Correct of Self-Paced Probe Recall	0.02	0.99
WISC-R Digit Span Scale Score	9.96	3.14
WISC-R Information Scale Score	10.11	3.13
WISC-R Vocabulary Scale Score	11.74	3.26
WISC-R Similarity Scale Score	10.59	3.18
WISC-R Comprehension Scale Score	10.67	3.18
WISC-R Picture Completion Scale Score	10.92	2.74
WISC-R Picture Arrangement Scale Score	11.63	3.09
WISC-R Block Design Scale Score	10.54	3.35
WISC-R Object Assembly Scale Score	10.25	3.24
Standardized Residual of MAT Language Scale Score	0.01	0.99
Standardized Residual of MAT Math Scale Score	0.02	0.99
Standardized Residual of MAT Reading Scale Score	0.03	0.97

Note: The standardized residuals of the WRTP variables are age-adjusted regression residuals that partial out the age effect from the original variables.

References

Baddeley, A. D. (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*, *vol. 8*. (pp. 47–89) New York: Academic Press.

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

Benton, S. L., Kraft, R. G., Glover, J. A., & Flake, B. S. (1984). Cognitive capacity differences among writers. *Journal of Educational Psychology*, *76*, 820–834.

Carlson, J. S., & Jensen, C. M. (1982). Reaction time, movement time, and intelligence: A replication and extension. *Intelligence*, *6*, 265–274.

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*(3), 404–431.

Carroll, J. B. (1992). Cognitive abilities: The state of the art. *Psychological Science*, *3*(5), 266–270.

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

Cattell, J. M., & Farand, L. (1890). Physical and mental measurements of the students of Columbia University. *Psychological Review*, *3*, 618–648.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22.

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*(2), 163–183.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, *4*, 577–590.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684.

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

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *9*, 561–584.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.

Deary, I. J. (1980). *How general is the mental speed factor in "general" intelligence: An attempt to extend inspection time to the auditory modality.* B.S. honors thesis, University of Edinburgh, Scotland.

Deary, I.J. (1986). Inspection time: discovery or rediscovery? *Personality and Individual Differences*, *7*, 625–632.

Deary, I. J. (2000). *Looking down on human intelligence*. New York: Oxford University Press, Inc..

Detterman, D. K. (1982). Does "*g*" exist. *Intelligence*, *6*, 99–108.

Detterman, D. K. (1984). Understand cognitive components before postulating metacomponents, etc.: Part II. *Behavioral and Brain Sciences*, *7*, 289–290.

Detterman, D. K. (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg, & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 57–61). Norwood, NJ: Ablex Publishing Corporation.

Detterman, D. K. (1987). What does reaction time tell us about intelligence? In P. Vernon (Ed.), *Speed and information processing and intelligence*. Norwood, NJ: Ablex.

Detterman, D. K. (1988). *Manual for the cognitive assessment tasks battery.* (Unpublished Technical Manual).

Detterman, D. K. (1992). Mopping up: The relation between cognitive abilities and intelligence. *American Journal of Mental Retardation*, *97*, 295–301.

Detterman, D. K. (1994). A system theory of intelligence. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 85–115). Norwood, NJ: Ablex Publishing Corporation.

Dosher, B. A. (1979). Empirical approaches to information processing: Speed–accuracy tradeoff functions or reaction time—A reply. *Acta Psychologica*, *43*, 347–359.

Duncan, J., Seltz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., et al. (2000). A neural basis for general intelligence. *Science*, *289*, 457–460.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*, 972–992.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge University Press.

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

Fry, A. G., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241.

Horn, J., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117–144.

Horn, J. L., & Cattell, R. B. (1965). Refinement and tests of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253–270.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review*, *85*, 109–130.

Jastak, S., & Wilkinson, G. S. (1984). *Wide range achievement test administration manual*. Wilmington, DE: Jastak Association, Inc.

Jensen, A. R. (1982a). Reaction time and psychometric *g*. In H. J. Eysenck (Ed.), *A model for intelligence*. Berlin: Springer-Verlag.

Jensen, A. R. (1982b). The chronometry of intelligence. R. J. Sternberg *Advances in research in intelligence*, *vol. 1*. Hillside, NJ: Erlbaum.

Jensen, A. R. (1985). Methodological and statistical techniques for the chronometric study of mental abilities. In C. R. Reynolds, & V. L. Willson (Eds.), *Methodological and statistical advances in the study of individual differences*. New York: Plenum.

Jensen, A. R. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information processing and intelligence*. Norwood, NJ: Ablex.

Jensen, A. R. (1998). *The g factor*. 88 Post Road West, Westport, CT 06881: Praeger Publishers.

Jensen, A. R. (2001). *Psycholoquy Reviews of The g Factor and Replies by Arthur Jensen.* Psycholoquy.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis*. (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Golderger, & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's references guide*. Chicago: Scientific Software International.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

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

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, *2*(2), 129–160.

Larson, G. E., & Saccuzzo, D. P. (1989). Cognitive correlates of general intelligence: Toward a process theory of *g*. *Intelligence*, *13*, 3–5.

Luo, D., & Petrill, S. (1999). Elementary cognitive tasks and their roles in psychometric *g*. *Intelligence*, *27*(2), 157–174.

Luo, D., Thompson, L. A., & Detterman, D. K. (2003a). The causal factor underlying the correlation between psychometric *g* and scholastic performance. *Intelligence*, *31*, 67–83.

Luo, D., Thompson, L. A., & Detterman, D. K. (2003b). The phenotypic and behavioral genetic covariation between elemental cognitive components and scholastic performance. *Behavior Genetics*, *33*, 221–246.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391–410.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: L. Erlbaum Associates.

McDonald, R. P., Parker, P. M., & Ishizuka, T. (1993). A scale-invariant treatment for recursive path models. *Psychometrika*, *58*(3), 431–443.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Riverside Publishing.

Meeker, W. Q., & Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *American Statistics*, *49*, 48–53.

Miller, L. T., & Vernon, P. A. (1992). The general factor in short term memory, intelligence, and reaction time. *Intelligence*, *16*, 5–30.

Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.

Neale, M. C. (1999). MX: Statistical modeling. Box MCV, Richmond, VA 23298: Department of Human Genetics.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, *18*, 347–356.

Nettlebeck, T., & Lally, M. (1976). Inspection time and measured intelligence. *British Journal of Psychology*, *67*, 17–22.

Ormrod, J. E., & Cochran, K. F. (1988). Relationship of verbal ability and working memory to spelling achievement and learning to spell. *Reading Research and Instruction*, *28*(1), 33–43.

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

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological psychology, Vol. 54 (1–3), Special issue: Behavior, event-related potentials, and neuroimaging: A lifespan perspective* (pp. 35–54).

Smith, R. W., Kounios, J., & Osterhout, L. (1997). The robustness and applicability of speed–accuracy decomposition, a technique for measuring partial information. *Psychological Methods*, *2*(1), 95–120.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Article presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stelzl, I. (1986). Change a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, *21*, 309–331.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, *7*, 53–70.

Vernon, P. A. (1987). *Speed of information processing and intelligence*. Norwood, NJ: Ablex.

Vernon, P. A. (1988). Intelligence and speed of information processing. *Human Intelligence Newsletter*, *9*, 8–9.

Vernon, P. A., Nador, S., & Kantor, L. (1985). Reaction times and speed-of-processing: Their relationship to timed and untimed measures of intelligence. *Intelligence*, *9*, 357–374.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Cooperation.

Wechsler, D. (1991). *Wechsler intelligence scale for children*. (Third ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler adult intelligence scale*. (Third ed.). San Antonio, TX: Psychological Corporation.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*, 16–37.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001a). Woodcock-Johnson III. Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001b). Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001c). Woodcock-Johnson III tests of cognitive abilities. Itasca, IL: Riverside Publishing.