Pergamon

Learning and Individual Differences
13 (2002) 159–188

Learning and
Individual Differences

The contribution of general and specific cognitive abilities to reading achievement

Michael L. Vanderwood[a,*], Kevin S. McGrew[b], Dawn P. Flanagan[c], Timothy Z. Keith[d]

[a]*University of California, Graduate School of Education, Riverside, CA 92521, USA*
[b]*University of Minnesota, Minneapolis, MN, USA*
[c]*St. John's University, Jamaica, NY, USA*
[d]*University of Texas at Austin, Austin, TX, USA*

Received 9 September 2002; accepted 9 September 2002

Abstract

Since the development of the Weschler scales, significant advances have been made in intelligence theory and testing technology that have the potential to provide a more comprehensive understanding of cognitive abilities than currently exists. For this study, the standardization sample of the Woodcock–Johnson Psychoeducational Battery-Revised (WJ-R)—an empirically supported measure of several constructs within the *Cattell–Horn–Carroll (CHC) theory of cognitive abilities*—was used to analyze the contribution of specific cognitive abilities to reading achievement at five developmental levels. Structural equation modeling (SEM), with calibration and cross-validation samples, of four different models of the hypothesized relations among the variables was conducted to determine if specific abilities can provide relevant information regarding the development of reading skills. The results of this study clearly indicate that *Gc* (comprehension knowledge or crystallized intelligence) and *Ga* (auditory processing) play an important role in the development of reading skills.

Keywords: Reading; Weschler scales; Woodcock–Johnson Psychoeducational Battery-Revised

* Corresponding author. Tel.: +1-909-787-0128; fax: +1-909-787-3942.
E-mail address: mike.vanderwood@ucr.edu (M.L. Vanderwood).

1041-6080/02/$ – see front matter
PII: S1041-6080(02)00077-8

1. Introduction

A cursory review of the cognitive abilities research literature reveals that attempts to move "beyond *g*" (i.e., the addition of specific abilities to *g* in the prediction and explanation of educational and occupational outcomes) have not been successful. In his APA presidential address, McNemar (1964, p. 875) concluded, "the worth of the multitest batteries as differential predictors of achievement in school has not been demonstrated." Cronbach and Snow (1977) reached a similar conclusion in their seminal review of the aptitude–treatment interaction (ATI) research, which demonstrated that interventions interact primarily with general level of intelligence, and that few, if any, meaningful specific ability–treatment interactions exist. Jensen (1984, p. 101) also reinforced this conclusion when he stated that "*g* accounts for all of the significantly predicted variance; other testable ability factors, independently of *g*, add practically nothing to the predictive validity." In the area of applied intellectual assessment, the failure to establish the importance of specific abilities has resulted in the warning to "just say no" to the practice of interpreting subtest scores in individual intelligence batteries (McDermott, Fantuzzo, & Glutting, 1990; McDermott & Glutting, 1997). The inability to move beyond *g* has provided little optimism for the development of interventions designed according to an individual's profile of specific ability strengths and weaknesses.

Despite the failure to demonstrate the importance of specific cognitive abilities, some intelligence scholars have suggested that now is not the time to give up hope. In his seminal review of the cognitive abilities factor analytic research literature, Carroll (1993, p. 676) concluded, "there is no reason to cease efforts to search for special abilities that may be relevant for predicting learning." In the school psychology literature, Flanagan (1999), Keith (1999a), and McGrew, Flanagan, Keith, and Vanderwood (1997) have suggested that recent advances in theories of intelligence (see Flanagan & McGrew, 1997), applied theory-driven measurement of intelligence, and research methodology (e.g., structural equation modeling, SEM) argue for continued efforts to investigate the effects of general and specific abilities on general and specific achievements.

Empirical support for this position was recently provided by Gustafsson and Balke (1993) who reported that some specific cognitive abilities may be important in explaining school performance beyond the robust influence of *g* when: (a) both the predictor (i.e., cognitive) and criterion (i.e., school performance) domains are viewed from multidimensional hierarchical frameworks, (b) the intelligence framework used reflects contemporary knowledge on the structure of cognitive abilities (viz., the hierarchical *Cattell–Horn–Carroll* theory of cognitive abilities [CHC theory])[1], and (c) research methods (viz., SEM) particularly suited

[1] Recently, it has come to our attention that Drs. John Horn and John Carroll would like to have "modern *Gf–Gc* theory"—an integration of the Cattell–Horn *Gf–Gc* and three-stratum theories—referred to as the "Cattell–Horn–Carroll Theory of Cognitive Abilities" or "CHC theory" (R. Woodcock, personal communication, July 16, 1999). In this paper, we have adopted this new terminology and have replaced *Gf–Gc* with "CHC" in reference to modern *Gf–Gc* theory, *Gf–Gc* abilities, *Gf–Gc* cross-battery assessment, and the like to be consistent with the request of these theorists.

to *understanding and explaining* (versus simply *predicting*) phenomena are employed. The positive results of Gustafsson and Balke (1993) suggest that research that incorporates advances in intelligence theory, measurement, and research methodology is needed to determine if specific cognitive abilities can be identified that provide explanatory information above and beyond *g*.

1.1. Advances in theory

A variety of intelligence theories grounded in markedly different research traditions have received increased attention in recent years (e.g., CHC theory of cognitive abilities, Gardner's Theory of Multiple Intelligences, the Luria–Das Model of Information Processing, Sternberg's Triarchic Theory of Intelligence; see Flanagan, Genshaft, & Harrison, 1997 for a review). Of these theories, the psychometrically based CHC theory has been viewed by some intelligence scholars as having the greatest potential for examining the importance of general and specific cognitive abilities.

CHC theory represents an integration of Carroll's (1993) three-stratum theory and the Cattell–Horn *Gf–Gc* theory (Horn, 1994) (see Flanagan, McGrew, & Ortiz, 2000; McGrew, 1997; McGrew & Flanagan, 1998). The CHC conception of intelligence is supported extensively by factor analytic (i.e., structural) evidence as well as developmental, neurocognitive, and heritability evidence (see Horn & Noll, 1997 for a summary). In addition, there is a mounting body of research available on the relations between the broad CHC abilities and many academic and occupational achievements (see McGrew & Flanagan, 1998 for a review of this literature). Furthermore, studies have shown that the CHC structure of intelligence is invariant across the lifespan (e.g., Bickley, Keith, & Wolfe, 1995) and across gender, ethnic, and cultural groups (e.g., Carroll, 1993; Gustafsson & Balke, 1993; Keith, 1997, 1999a). In general, the CHC theory is based on a more thorough network of validity evidence than other contemporary multidimensional ability models of intelligence (see McGrew & Flanagan, 1998; Messick, 1992).

According to Daniel (1997, p. 1042–1043), the strength of the CHC model of cognitive abilities is that it was arrived at "by synthesizing hundreds of factor analyses conducted over decades by independent researchers using many different collections of tests. Never before has a psychometric ability model been so firmly grounded in data." The convergence on the CHC model provides a validated framework from which to examine the importance of general and specific cognitive abilities.

In the CHC model, cognitive abilities are classified at three strata that differ in degree of generality (Carroll, 1993). *General* cognitive ability or *g* is located at *stratum III* and subsumes several *broad* cognitive abilities (located at *stratum II*), which, in turn, subsume approximately 70 *narrow* abilities (located at *stratum I*). Several of the broad cognitive abilities in the CHC model are: fluid Intelligence (*Gf*), crystallized intelligence (*Gc*), short-term acquisition and retrieval (*Gsm*), visual intelligence (*Gv*), auditory intelligence (*Ga*), long-term storage and retrieval (*Glr*), cognitive processing speed (*Gs*), correct decision speed (*CDS*), and quantitative knowledge (*Gq*). A brief description of these abilities is provided in Table 1.

Table 1
A description of nine broad CHC cognitive abilities

Fluid reasoning or intelligence (*Gf*) is measured by tests that require inductive, deductive, conjunctive, and disjunctive reasoning to understand relations among stimuli, comprehend implications, and draw inferences.
Acculturation knowledge (*Gc*) is also called comprehension knowledge, it is measured by tests that indicate the breadth and depth of the knowledge of the dominant culture.
Quantitative reasoning (*Gq*) is measured by tests that require understanding and application of the concepts and skills of mathematics.
Short-term apprehension-retention (*Gsm*) is also called short-term memory and is measured with a variety of tests that require maintaining awareness of and recalling elements of immediate stimulation—i.e., events of the last minute or so.
Fluency of retrieval from long-term storage (*Glr*) is also called long-term memory and is measured by tests that indicate consolidation for storage and require retrieval, through association, of information stored minutes, hours, weeks, and years before.
Visual processing (*Gv*) is measured by tests that involve visual closure and constancy and fluency in "image-ing" the way objects appear in space as they are rotated and flip-flopped in various ways.
Auditory processing (*Ga*) is measured by tests that involve the perception of sound patterns under distraction or distortion, maintaining awareness of order and rhythm among sounds, and comprehending groups of sounds, such as chords, and the relationships among such groups.
Processing speed (*Gs*) is part of almost all intellectual tasks and is measured most purely by tests that require rapid scanning and responding to intellectually simple tasks that almost all people would get right if the task were not highly speeded.
Correct decision speed (*CDS*) is measured by tests that require quick answers based on thinking.

Adapted from Horn (1991).

The reader is referred to Carroll (1993, 1997), Flanagan et al. (2000), Horn (1991, 1994), Horn and Noll (1997), and McGrew and Flanagan (1998) for a comprehensive description of CHC theory, and for supporting evidence and limitations of the theory.

1.2. Advances in applied measurement

Recent joint factor analyses research (Flanagan & McGrew, 1998; McGhee, 1993; Stone, 1992; Woodcock, 1990), as well as expert consensus task analysis of the major individually administered intelligence batteries (Flanagan et al., 2000; McGrew, 1997; McGrew & Flanagan, 1998), have suggested that none of the current intelligence batteries assess the broad range of cognitive abilities included in CHC theory. Furthermore, no intelligence battery includes enough qualitatively different *CHC* narrow stratum I ability indicators (i.e., subtests) to warrant the generation of hypotheses about all the broad abilities.

Of the current collection of individually administered intelligence batteries, the Woodcock–Johnson Psychoeducational Battery-Revised (WJ-R) comes closest to measuring the complete breadth of broad cognitive abilities included in the CHC theory. The WJ-R was developed deliberately to operationalize contemporary theory, and as such, measures validly eight *CHC* abilities (Keith, 1997; McGrew, Werder, & Woodcock, 1991; Woodcock, 1990; Ysseldyke, 1990). Despite its breadth of coverage, the WJ-R has not often been used in the *g*/specific abilities research. Given the recent development of an individually administered battery of cognitive and achievement tests specifically designed to operationalize CHC theory (viz., WJ-R), we believe that new *g*/specific ability research is warranted. Progress has been made in both theories and measurement of intelligence. This progress requires a continued examination of the general versus specific abilities research question.

1.3. Advances in research methodology

Multiple regression analysis (MR) has been the primary method used to determine whether specific cognitive abilities improve on the prediction of achievements beyond the prediction provided by *g*. Since MR will not allow the prediction of a criterion from both a composite score and the components that comprise the composite score (i.e., the correlation matrix will be singular), creative "tricks" have been used to conduct such analyses. For example, McDermott et al. (1990) subtracted the average Weschler subtest score from each individual subtest (in effect removing *g* from each subtest). They then used these "ipsatized" scores to predict achievement, and compared the variance explained by the ipsatized subtest scores to the variance explained by the original subtest scores in a separate regression. Thorndike, Hagen, and Sattler (1986) compared the variance explained by all subtests to that explained, in a separate regression, by a general, or *g*, factor. These procedures have not allowed for the direct comparison of the effects of general and specific abilities in a single model. A more suitable approach would be to analyze the effects of general and specific abilities *simultaneously* (McGrew et al., 1997).

In addition, most of the prior MR based research has attempted to partition variance into that accounted for by *g* and that accounted for by specific abilities, a practice that is not well suited to determining the relative importance of the effects of different variables on a criterion (Kenny, 1979, Chap. 4; Pedhazur, 1997, Chap. 9). The partitioning of variance provides a "highly stingy. . .measure of the relationship between two variables" (Abelson, 1995, p. 7; see also Rosenthal & Rubin, 1979), and at times can underestimate severely the effects of one variable on another (Pedhazur, 1997). Finally, most of the MR-based *g*/specific ability research has focused on whether specific abilities improve the *prediction* of some criterion beyond that predicted by *g*. Although important, predictive findings do not easily translate into practice. We believe that an *explanatory* approach is more appropriate. It is not enough to know that ability "x" *predicts* "y"; in order to translate research into practice, it is necessary to know whether or not ability "x" *affects* "y."

Latent variable SEM methods (see Hoyle, 1995; Keith, 1999b; Loehlin, 1998 for further information) have a number of advantages over MR procedures. First, SEM allows for the specification and evaluation of complex theories and models, such as those present when a hierarchical structural model of intelligence (*CHC* theory) and a multidimensional hierarchical achievement model are on the predictor and criterion sides, respectively, of a complex causal model. Second, SEM allows for the significance testing of specific effects of individual abilities rather than the blanket, blind prediction of achievements. Third, SEM reduces the confounding effects of measurement error by estimating and removing this source of variation from the consideration of the effects of one variable on another. As a result, SEM gets closer to the *constructs* of primary interest in research. By removing error (unreliability) and unique variance (invalidity) from a structural causal model, SEM provides for more accurate estimates of the effects of one construct on another. Finally, SEM focuses on understanding (analysis of effects) rather than mere prediction (partitioning of variance).

1.4. Purpose of study

We believe that research that is (a) based on contemporary *CHC* theory; (b) utilizes measures specifically designed to operationalize *CHC* constructs (viz., the WJ-R); and (c) uses SEM methods obviates a number of the methodological difficulties that have confounded prior *g*/specific abilities research. In the present study, validated measures of seven *CHC* abilities were used to define a hierarchical model of intelligence that includes *g* and *multiple CHC cognitive* abilities. Thus, the cognitive theory used to model the predictor side of the causal model is up-to-date and empirically supported. SEM procedures were applied to data from large nationally representative grade-based samples from the WJ-R in order to simultaneously estimate the effects of *g* and specific *CHC* abilities on both general (broad reading) and specific (e.g., word attack and comprehension) reading achievement. Based on a review of relevant literature (e.g., Mather, 1991; McGrew, 1993; McGrew & Flanagan, 1998), causal models were specified that included both the direct and indirect effects of *g* and specific *cognitive* abilities on reading achievement. The final calibration models were cross-validated and compared to determine if specific *cognitive* abilities were important in understanding general and specific reading above and beyond the explanation provided by *g*.

2. Method

2.1. Instrument

Secondary analysis of the nationally representative standardization sample of the WJ-R (Woodcock & Johnson, 1989) was used to examine the relations between specific cognitive abilities and reading achievement. The WJ-R is a norm-referenced set of individually administered tests, which consists of 21 measures of cognitive ability and 14 measures of academic achievement. Fourteen of the cognitive and four of the achievement measures were used in the current study (see Woodcock & Mather, 1989 for a more detailed description of each measure). The WJ-R is considered to be a good operationalization of the *CHC* theory (Horn, 1991; McGrew et al., 1997, 1991). The test authors used a rigorous test development procedure which included devising a *CHC* test blueprint, conducting preliminary exploratory and confirmatory factor analyses, test revision, and final confirmatory analyses (McGrew, 1994).

Independent reviewers have been impressed with the comprehensive and theory driven process that was used to develop the WJ-R (Kamphaus, 1993; Kaufman, 1990; Reschly, 1990; Ysseldyke, 1990). Reschly (1990, p. 260) proposed that research with instruments such as the WJ-R that are based upon a comprehensive theory “provide far greater opportunities to generate predictions, test hypotheses, understand basic phenomena, and revise theoretical constructs" than with instruments based on a dichotomous or limited set of abilities. Horn (1991, p. 197) stated: "The revision of the Woodcock–Johnson in 1989 extended the match between new advances in intelligence theory and the measurement of intelligence." Thus, the WJ-R provides researchers and practitioners with a psychometrically sound and empirically and theoretically grounded instrument, which can be used to examine the structure of human intelligence and individual differences from a multidimensional perspective.

2.2. Sample

The WJ-R standardization sample of 6359 subjects ranged in age from 24 months to 95 years and was selected from over 100 communities using a three-stage stratified sampling design based on the 1980 US Census. The 3 stages of sampling were by community, schools, and then by subjects, and included 5 person and 15 community variables to ensure national representativeness of the sample. Some groups with low percentages (e.g., Asian and Pacific Islanders) were oversampled to ensure accurate contributions to the norming sample. Also to ensure representativeness of the sample, individual subjects were weighted to produce a distribution that was an exact match to the variables included in the sampling design. See McGrew et al. (1991) for a more detailed description of the sample and data collection methods.

The kindergarten to twelfth grade sample was used for the present study and included 3425 subjects. (The remaining 3114 subjects were part of the preschool or adult samples.) Subjects were deleted if they did not have data for each of the 18 subtests used in the study, which produced a sample of 2312 subjects for this study. Adjacent grade-based samples were combined and then randomly split to produce calibration and cross-validation samples ($n = 222-255$) at five levels: Grades 1–2, 3–4, 5–6, 7–9, and 10–12. After splitting the sample, correlation matrices and standard deviations for the 18 subtests were produced at each level for both samples. Covariance matrices, recovered from the correlations and standard deviations, were used in the structural equation models.

2.3. Data analysis

Maximum-likelihood estimation using the Amos SEM program was used to examine the effect of general and specific latent *CHC* factors on reading achievement factors. Maximum-likelihood estimation is considered the standard approach for estimating free parameters with SEM and has been found to be very robust (Hoyle & Panter, 1995). Model modification, cross-validation, and model comparison approaches were used as suggested by MacCallum (1995) to reduce the chance that the relations found were a product of chance. Model estimation at each developmental level was conducted in a two-stage process.

During the calibration stage, the initial theoretically and research-based structural models (described later) were specified, evaluated, and when appropriate, modified using the calibration sample. The first step of this stage was estimating the model and then eliminating structural paths with critical values (estimate/standard error) less than 1.96 (P>.05) and structural paths with negative values. Paths were eliminated one at a time and then the model was re-estimated. The second step in model modification consisted of examining modification indices to determine if adding additional paths would create a better model fit. Paths were only added if they made logical and theoretical sense, and they were only kept if they improved overall model fit based on χ^2.

Based on previous research which indicated that the cognitive measurement model and factor structure of the WJ-R operationalized appropriately *CHC* theory (Bickley et al., 1995), the decision was made a priori not to modify the cognitive measurement model or the relations between g and the specific abilities. To ensure that any changes that occurred in model fit were a product of structural model changes and not a product of measurement model changes, modification of the relations between the reading composite and the four reading achievement variables also was not conducted. The structure of the reading measurement model was based on the work of McGrew (1993, 1994). The focus of this study was the relationship between the specific abilities and the reading variables that comprised the structural model.

The second stage of analysis involved estimating the final modified models from the calibration sample in the cross-validation sample without modification. The results of the second-stage analysis were used for answering the research questions of the study by examining the fit indices, and then the structural paths.

2.4. Models

Based on the results of the confirmatory factor analyses that were conducted during the development of the WJ-R and based on a review of the WJ-R technical manual (McGrew et al., 1991), the target model in Fig. 1 was specified. This model was the starting point during the calibration phase at each of the five developmental levels. Each of the seven *CHC* cognitive abilities (*Gq*, quantitative ability, was not used in this study) measured by the WJ-R Cognitive Battery were represented by two subtests. Four subtests (viz., letter–word identification, word attack, reading vocabulary, and passage comprehension) were selected from the WJ-R Achievement Battery to represent reading achievement. The four achievement battery subtests were used to produce latent variables that were subordinates of a general reading factor. Due to previous research that indicated a relationship between the four types of reading achievement measured by the WJ-R and specific abilities (McGrew, 1993, 1994), single indicators for each of the reading variables were used. (The limitations of this decision are presented later.) Letter–word identification and word attack are generally considered measures of basic reading skills, and reading vocabulary and passage comprehension are generally viewed as measures of reading comprehension (McGrew et al., 1991).

The target model included direct relations between specific cognitive abilities and specific reading achievement variables. As indicated in Fig. 1, *Gc*, *Gs*, and *Ga* were hypothesized to influence several reading achievement variables. Based on the research

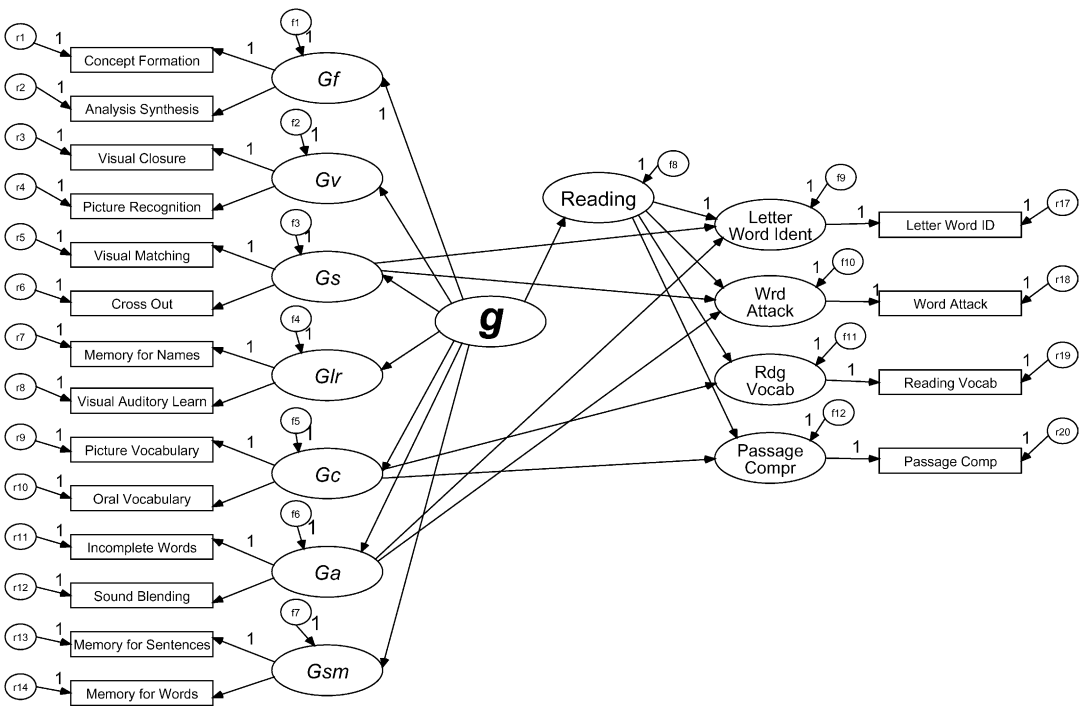

Fig. 1. Initial target model used at all developmental levels.

that indicates auditory processing and phonemic awareness are related to reading achievement, paths between *Ga* (auditory processing) and letter–word identification and word attack were added (Mather, 1991; McGrew, 1993; Torgesen, Wagner, & Rashotte, 1994). *Gc* (comprehension knowledge) is highly related to reading comprehension (Lohman, 1989; Mather, 1991; Snow & Swanson, 1992), and McGrew (1993) reported an increase in the strength of this relation with age. In this model, the two measures of reading comprehension (viz., passage comprehension and reading vocabulary) were hypothesized to be directly affected by *Gc*. The final specific ability, *Gs* (processing speed), is considered an important factor in understanding developmental changes in reading and other forms of achievement (Kail, 1991; Lohman, 1989) and was found by McGrew to have a strong relations with basic reading skills. Therefore, paths between Gs and the two measures of basic reading skills (viz., letter–word identification and word attack) were included in the target model.

To test the contribution of the specific abilities, an alternative model was specified that did not include the relations between the reading achievement variables and *Gc*, *Gs*, and *Ga*. The only path between the achievement and cognitive components of the model was the path between *g* and the reading composite variable. This model was tested at each developmental level during the cross-validation phase. Because paths were not specified between specific cognitive and achievement variables, this model was not tested during the calibration phase.

As recommended by methodologists, multiple indices were used to judge the fit of the models tested in this research. The Root Mean Square Error of Approximation (RMSEA) was used to judge the fit of a single model; RMSEAs at or below .05 suggest a close fit of the model to the data and RMSEAs below .08 suggest a marginal fit (Browne & Cudeck, 1993; Hu & Bentler, 1999). The Goodness-of-Fit Index (GFI; Jöreskog and Sörbom, 1981), Tucker–Lewis Index (TLI; Bentler & Bonett, 1980), and Comparative Fit Index (CFI; Bentler, 1989) were also used to judge the fit of a single model. For each of these indices, values above .95 were taken to suggest a good fit, with values above .90 interpreted as suggesting an adequate fit.

Of more direct interest in this research were comparisons among competing models. To compare competing models, the change in chi-squared ($\Delta\chi^2$) from the target to the alternative model, in relation to the degrees of freedom, was computed at each developmental level. When two models are nested, as these models are, $\Delta\chi^2$ is also distributed as χ^2 and the size of the difference between the two models may be compared to $\chi^2(df)$ to assess the statistical significance of changes to the model. In the present case, a significant increase in χ^2 when going from the target to the alternative models would suggest the superiority of the target model, whereas a nonsignificant increase would suggest the superiority of the more parsimonious alternative (*g* only) model. A comparison of the RMSEAs was also used to assess fit of these competing models. These fit indices are explained in detail in most introductions to SEM (e.g., Hoyle, 1995; Keith, 1999b; Loehlin, 1998). For empirical evaluations of the rules of thumb for judging models, see Hu and Bentler (1999).

3. Results

The results of the calibration phase will be presented first, followed by an examination of the cross-validation model comparison and fit statistics. The standardized path coefficients for the target models will also be presented.

3.1. Calibration phase

Several modifications were made to the target model at each developmental level during the calibration phase. The model in Fig. 1 was used at each developmental level as a starting point for model modification. At Grades 1–2, both paths from *Gs* were eliminated. The path from *Ga* to letter–word identification and both *Gs* paths were removed at Grades 3–4. At the next developmental level, Grades 5–6, *Ga* to letter–word Identification, and both original *Gs* paths were eliminated, and paths from *Gs* to reading vocabulary and passage comprehension were added. The only change that occurred at Grades 7–9 was the elimination of the *Gs* paths. The only specific ability paths remaining at Grades 10–12 were from *Gc* to reading vocabulary and passage comprehension. The modifications made at each developmental level were tested with the cross-validation sample, which was used to produce the fit statistics described in

Section 3.2. For all the target models, diagnostic fit information (e.g., modification indices and standardized residuals) suggested that changes to the measurement portions of the CHC and reading models to improve the fit of the models. To provide consistency in models across grades, however, an a priori decision was made not to modify the measurement portions of the models.

3.2. Cross-validation phase

As shown in Table 2, most of the target (specific abilities) models showed an adequate fit to the data as judged by the RMSEA, GFI, TLI, and CFI, although a few such models showed a good fit. Of more direct interest are the comparisons between the target (specific abilities) and the alternative (*g* only) models. At each grade level, the *g* only model resulted in a statistically significant increase in χ^2. This means that, for each grade level, a model, which excludes paths from specific abilities to reading skills, fit worse than did models that included

Table 2
Comparison of cross-validation fit indices across models and developmental levels

Models	$\chi^2(df)$	P	GFI	TLI	CFI	RMSEA
Grades 1–2						
Target	206.80(123)	<.001	.91	.95	.96	.04–.07
Alternative	243.96(127)	<.001	.89	.94	.95	.05–.08
	$\Delta\chi^2$ = 37.17(4), P < .001					
Grades 3–4						
Target	228.84(125)	<.001	.90	.93	.94	.05–.07
Alternative	295.83(127)	<.001	.87	.88	.90	.06–.09
	$\Delta\chi^2$ = 134.422(3), P < .001					
Grades 5–6						
Target	234.99(123)	<.001	.90	.90	.92	.05–.08
Alternative	292.09(127)	<.001	.87	.85	.88	.07–.09
	$\Delta\chi^2$ = 143.29(5), P < .001					
Grades 7–9						
Target	236.20(124)	<.001	.90	.92	.94	.05–.07
Alternative	328.55(127)	<.001	.87	.86	.89	.07–.09
	$\Delta\chi^2$ = 231.18(4), P < .001					
Grades 10–12						
Target	243.11(126)	<.001	.91	.93	.94	.05–.07
Alternative	295.90(129)	<.001	.89	.90	.91	.06–.08
	$\Delta\chi^2$ = 149.76(2), P < .001					

GFI = Goodness-of-Fit Index, TLI = Tucker-Lewis Index, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation.

such effects. In other words, the inclusion of paths from specific abilities to reading skills provides a more complete explanation of those skills than does a model that assumes that only general intelligence affects reading skills.

Additional support for the better fit of the target versus *g* only models was provided by the results of the RMSEA statistic. Using a 90% confidence interval of the RMSEA fit index, a traditional hypothesis test was conducted to determine if the models produced a close fit. RMSEA confidence interval values that were at or below 0.05 were considered a close or good fit (Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996). The null hypothesis was not rejected for every grade of the target model, which indicates these models fit or replicated the true relations of the variables closely. For the alternative model, the null hypothesis was rejected at every grade except Grades 1–2.

3.3. Path coefficients

As shown in Table 3, the standardized path coefficients from *g* to reading indicated a strong effect at all grade levels. (When interpreting the strength of the standardized path coefficients, the following guidelines suggested by McGrew (1994) were used: (a) path values below .10 are considered weak or not important, (b) values from .10 to .29 suggest a

Table 3
Direct effects on reading variables for target and alternative models

Paths	Grades				
	1–2	3–4	5–6	7–8	10–12
Target model					
To reading					
From *g*	.63	.77	.68	.57	.88
To letter word identification					
From *Ga*	.33	–	–	.32	–
To word attack					
From *Ga*	.49	.24	.27	.50	–
To reading vocabulary					
From *Gc*	.56	.76	.71	.84	.98
From *Gs*	–	–	.09 *	–	–
To passage comprehension					
From *Gc*	.47	.53	.49	.69	.90
From *Gs*	–	–	.21	–	–
Alternative model					
To reading					
From *g*	.88	.93	.92	.99	1.00

All values are standardized direct effects and are significant at $P < .05$ unless otherwise noted.

* Not significant at $P < .05$.

moderate relationship, and (c) path coefficients above .30 indicate a strong relationship.) *Gc* maintained a strong effect on reading vocabulary and passage comprehension and had the largest value of any specific ability across grade levels. The contribution of *Ga* fluctuated from a strong influence on letter–word identification and word attack at Grades 1–2 and 7–9, to a moderate effect on word attack at Grades 3–4 and 5–6. At Grades 5–6, the only level that *Gs* was included in the cross-validation sample, the path from *Gs* to passage comprehension was of moderate strength and the path to reading vocabulary was nonsignificant ($P > .05$).

4. Discussion

Theories of intelligence have developed significantly in the last two decades and moved far beyond the atheoretical approach that was used to design the Weschler Scales (Weschler, 1939). *CHC* theory is considered a strong alternative to either an atheoretical or general ability approach to intelligence testing (Carroll, 1993; Flanagan et al., 2000; Horn, 1994; McGrew, 1997; McGrew & Flanagan, 1998; Reschly, 1990) and has been incorporated into the WJ-R (Woodcock & Johnson, 1989). An underlying question of this study was given recent advances in intelligence theory, applied measurement, and research methodology, to what extent do constructs of a contemporary theory of intelligence (*CHC* theory) explain reading achievement.

Three general conclusions are supported by the current research. First, models, which contain specific cognitive ability/reading achievement relations, fit better than models without those relations. Second, the relations between the *CHC* specific cognitive abilities and reading achievement changed at each developmental level studied in this investigation. Third, *Gc* and *Ga* were related strongly to basic reading and reading comprehension skills.

The data from this study support the belief that *CHC* specific cognitive abilities can be used to explain and better understand academic achievement, above and beyond the effects of *g*. First, the specific cognitive ability/achievement models outperformed the general ability/achievement models as indicated by the significant ($P < .001$) decrease in chi-square for the specific ability models in every case when they were compared to the general ability/achievement models. Additional support was provided by the results of the RMSEA fit statistic. Although only 20% of the general ability/achievement models produced a close fit, 100% of the specific ability/achievement models met the standards for a close fit.

The better fit of the specific ability/achievement models was expected and is consistent with the theories proposed by Carroll (1993) and Horn (1991, 1994) and research conducted by Gustafsson and Balke (1993). The addition of the three specific abilities that prior research indicated influence reading achievement, did, in fact, account for a significant increase in fit over a model that only contained *g*. These results are also in direct opposition to the belief that specific abilities do not add anything useful to the explanation of school achievement (McDermott et al., 1990; McNemar, 1964). Although Jensen (1984) argued that his review of the extant literature indicated that *g* accounts for almost all of the predicted variance, and other factors add little or nothing, the studies he cited did not use tests that were based on a theoretically supported multidimensional approach to intelligence. This research clearly supports the idea that specific abilities do add relevant information above and beyond that provided by *g*.

4.1. Contribution of Gc

The significant paths between *Gc* and reading comprehension are consistent with previous research that indicated prior achievement affects future achievement and that comprehension-knowledge influences reading achievement (Aaron, 1995; Felton & Pepper, 1995; McBride-Chang, 1995; McGrew, 1994). It also seems logical that prior exposure to the culture and language of the test should be of benefit to the test taker when attempting to garner understanding from a reading passage or define vocabulary words. And, the stronger relation of *Gc* to reading vocabulary versus *Gc* to passage comprehension in every case suggests that previous exposure to language plays a bigger role in providing a synonym or antonym than it does in providing a missing word in a sentence. Another interesting point is that the aforementioned relationship has a general positive linear trend from Grades 1–2 to 10–12, which indicates that, as acquired knowledge (*Gc*) accumulates, it plays a bigger role in identifying words and comprehending sentences.

Although the previously mentioned findings provide support for the *Gc* construct, its high loading on *g* in both models across all developmental levels could be interpreted as meaning that *Gc* is not differentiated from *g*. The relation between *Gc* and *g* was very strong, ranging from .89 to .96, which if interpreted without looking at previous research suggests that *Gc* and *g* may be the same factor. But, when using the same WJ-R norm data and factor structure, Bickley et al. (1995) tested the hypothesis that *Gc* was indistinguishable from *g* by determining the change in chi-square between models that had the *Gc* to *g* loading set at 1.0 and the model used in this study. These authors found that the fit of the model with *Gc* set to 1.0 was significantly worse than with the two factors separate, which provides support for the use of the model in this study and the interpretation that *Gc* is distinct from *g*.

4.2. Contribution of Ga

The initial specifications between *Ga* and the basic reading skills variables were based on a growing consensus that awareness of and ability to manipulate sounds (which is typically defined as phonological awareness) is the best predictor of reading achievement in young children (Morris et al., 1998; Stanovich, Cunningham, & Cramer, 1984; Torgesen et al., 1994) and that fluent reading is made up of at least two independent components: reading comprehension and word recognition (Aaron, 1995; Joshi, 1995). Therefore, the expected result of the *Ga* to basic reading skills specifications was that *Ga* would have a strong relationship with the basic reading skills variables (which measure word recognition) in the early elementary grades and that by the secondary grades the relationship would be nonsignificant.

At Grades 1–2, *Ga* had a strong relationship to both basic reading skills variables and, at Grades 3–4, the path to letter–word identification was eliminated and the path to word attack was of moderate strength. Grades 5–6 maintained a moderate relationship between *Ga* and word attack. The surprising result occurred at Grades 7–9, which had strong relationships between *Ga* and both basic reading skills variables. This result is not supported by previous research and does not fit the expected trend of the relationships between *Ga* and the reading variables. The most plausible explanation for this result may be that for the upper grades the reading measurement model was not adequate and spurious relationships were created. In summary, the data support the belief that phonological awareness is related to reading achievement for early elementary students and that the WJ-R is able to measure this relationship.

4.3. Contribution of Gs

In contrast to *Gc* and *Ga*, the initial a priori specifications of *Gs* were not supported in any case across the five models. During the calibration phase, the paths from *Gs* to basic reading skills (letter–word identification and word attack) were eliminated in every case and paths between *Gs* and the reading comprehension variables (passage comprehension and reading vocabulary) were added at Grades 5–6. The paths between *Gs* and reading vocabulary were either weak or not significant in the cross-validation sample.

Therefore, the only significant *Gs* path in the cross-validation sample was to passage comprehension at Grades 5–6 (moderate strength), and this path was not initially specified to have an effect. Although these results suggest that processing speed does not play a major role in the development of reading skills, previous research clearly indicates that speed is important for reading achievement (Aaron, 1995; Carroll, 1993; McGrew, 1993; Snow & Swanson, 1992). A possible cause for the difference between previous research and this study was the multivariate approach used here that included *g* may have eliminated the variance that is usually accounted for by *Gs*.

4.4. Limitations

Two types of limitations are presented in this section. The first type is termed technical and focuses on the design and data analysis of the study. The second area of study limitations addresses the theoretical problems associated with the *CHC* theory and SEM.

4.4.1. Technical

To ensure replicable results, a priori model development guidelines were used which dictated that the *CHC* and reading measurement models could not be modified. By not allowing changes in the reading construct, which were indicated by the modification indices during the calibration phase, the specification of ability/achievement relationships may have been prevented and other spurious relationships may have been created. For example, for the target model at Grades 10–12, the paths from the reading comprehension variables to reading were nonsignificant and, at Grades 7–9, the path from reading to passage comprehension was also nonsignificant, which indicates that the reading model is inadequately specified and should be changed. Therefore, the decision to maintain the same reading model across grades may have created spurious relationships or prevented genuine relationships from becoming significant.

Another technical limitation of this study was the use of single indicators for the reading achievement variables. Although reliability estimates were used for each subtest to counter the effects of single indicators, Bollen (1989) suggests the use of at least three indicators for each latent variable. Although the intent of only using one indicator was to examine different aspects of reading achievement, the study would have been stronger, and, possibly, clearer relationships between specific abilities and the subtypes of reading achievement would have been found if more than one subtest was used to create the latent variables.

4.4.2. Theoretical

One interpretation of these results is that they provide substantial support for the *Gc* construct and its relationship to reading achievement. Carroll (1993) included reading comprehension as a stratum level I factor of crystallized intelligence (*Gc*) and argued that where the line is drawn between what is considered reading achievement and what is considered ability is arbitrary. Therefore, the *Gc* to reading comprehension associations supported by this study would be interpreted by Carroll as relationships between a factor and its subfactor and not as a relationship between ability and achievement.

Humphreys (1973) provided a similar argument and suggested one alternative explanation for the relationships between the cognitive abilities and reading achievement is that, in fact, the tests that make up the cognitive variable and the achievement variable are actually measures of the same construct, but at a different level. Humphreys argued that intelligence and achievement tests are not distinct and are actually measuring the same skills at different levels. When applied to this study, Humphreys' argument suggests that in essence, when relationships are found between cognitive abilities and achievement, the true relationship is between a broad and narrow operationalization of a skill. This argument can be reworded to say the tests are both measuring achievement, which is what several authors proposed about the original Woodcock–Johnson Battery (Sattler, 1988; Shinn, Algozzine, Marston, & Ysseldyke, 1982) but was not supported by further analyses (McGrew, 1993; Woodcock, 1990). Furthermore, joint confirmatory factor analyses indicated that the WJ-R is only 14% achievement like, while other intelligence tests have more than 50% of their content that is achievement like (Woodcock, 1990).

4.5. Implications

One of the most critical implications of this research for applied psychologists is the use of intelligence tests for the diagnosis and treatment of reading disabilities. Authors have cautioned against using the traditional intelligence tests (e.g., Weschler scales) to diagnose reading disabilities because they are unable to differentiate types of reading problems and the tests do not provide instructionally relevant information (Aaron, 1995; Joshi, 1995). Although some authors promote attempting to gain instructionally relevant information from IQ tests using subtest (i.e., ipsative) and profile analysis (e.g., Kaufman, 1994), the practice is not supported by empirical research (McDermott et al., 1990). To address this problem, the current study was designed to determine if a theoretically based and multidimensional IQ test could provide information that is relevant for instructional planning and intervention development. As a first step, a review of the literature indicated that the WJ-R measures at least three specific abilities (viz., *Gs*, *Ga*, and *Gc*) that research indicated influences reading achievement (Aaron, 1995).

The results of this study indicate that *Ga*, which is considered an essential component of reading, was clearly related to basic reading skills, especially word attack skills in early elementary grades. And, even when *g* was included as a construct in the analyses, relationships between *Ga* and reading were significant. The results from this study also make it clear that there are developmental changes in the role of some *CHC* specific abilities. While *Ga* is clearly important for understanding reading achievement in the early elementary grades, the relationship between *Gc* and reading comprehension became stronger while the effect of *Ga* was reduced with age.

4.6. Future directions

An exciting aspect of this study is that additional support was provided for a theory that has the potential of breaking some of the mirrors in Cronbach's (1975) "hall of mirrors" that is entered when conducting ATI research. Using *CHC* theory, it may be possible to determine, for instance, that students with low *Ga* and high *Gc* scores benefit from instruction that focuses on phonological awareness, while students with high *Ga* and low *Gc* scores benefit from instruction in comprehension strategies.

Given the nonsignificant paths for the reading measurement model at the upper grade levels, use of a more comprehensive reading assessment that has a well replicated factor structure should help clarify the results obtained here. It is also important to further examine the changing nature of the structure of *CHC* abilities by conducting a longitudinal study versus the cross-sectional approach used here.

Clearly, the results of this study provide more support for the idea that the contribution of specific abilities to achievement should be examined further. Although the results needs to be replicated with other measures of reading achievement before more definitive statements can be made, the following statement made by Gustafsson and Balke (1993, p. 432) is supported by this study:

> The conclusion that little is to be gained by differentiation of different factors of ability may thus be challenged, and it seems differentiation among at least a limited number of broad abilities would be worthwhile.

Appendix A

Table A1. Correlations and standard deviations for Grades 1 and 2, cross-validation sample

Variable												
1	2	3	4	5	6	7	8	9	10	11	12	13
1.000												
.380	1.000											
.282	.325	1.000										
.299	.365	.304	1.000									
.166	.196	.253	.251	1.000								
.397	.540	.246	.479	.304	1.000							
.270	.369	.313	.444	.108	.397	1.000						
.473	.432	.350	.378	.209	.516	.456	1.000					
.289	.521	.266	.259	.042	.261	.333	.275	1.000				
.193	.212	.689	.345	.265	.331	.341	.356	.150	1.000			
.320	.349	.313	.474	.102	.358	.398	.358	.382	.320	1.000		
.140	.317	.328	.279	.322	.323	.244	.372	.187	.303	.220	1.000	
.365	.602	.405	.519	.296	.638	.491	.495	.381	.386	.466	.403	1.000
.424	.432	.378	.446	.174	.445	.525	.418	.379	.353	.349	.339	.540
.232	.294	.477	.344	.120	.300	.416	.314	.291	.351	.374	.325	.403
.353	.417	.464	.346	.174	.417	.503	.440	.345	.381	.419	.417	.568
.245	.408	.497	.447	.168	.471	.515	.470	.327	.454	.468	.412	.628
.348	.500	.460	.432	.222	.429	.493	.479	.429	.387	.576	.351	.600
.346	.466	.421	.418	.199	.400	.467	.469	.410	.353	.626	.343	.577
.345	.579	.445	.459	.221	.493	.529	.542	.456	.372	.541	.408	.683
.368	.538	.462	.423	.268	.502	.458	.466	.355	.365	.502	.412	.638
.369	.448	.487	.449	.205	.413	.514	.504	.416	.381	.555	.344	.605
.365	.413	.446	.363	.194	.294	.505	.414	.446	.310	.468	.331	.503
.322	.364	.406	.412	.164	.306	.415	.462	.335	.378	.485	.310	.449
.258	.325	.447	.308	.115	.211	.436	.395	.382	.315	.418	.258	.431
S.D.												
16.458	15.771	17.380	17.414	14.914	16.054	15.719	17.064	15.300	15.814	17.610	14.641	16.612

Variables: (1) Memory_for_Names, (2) Memory_for_Sentences, (3) Visual_Matching, (4) Incomplete_Words, (5) Visual_Closure, (6) Picture_Vocabulary, (7) Analysis_Synthesis, (8) Visual_Auditory_Learn, (9) Memory_for_Words, (10) Cross_Out, (11) Sound_Blending, (12) Picture_Recognition, (13) Oral_Vocabulary, (14) Concept_Formation, (15) Calculation, (16) Applied_Problems, (17) Quant_Concepts, (18) Letter_Word_ID, (19) Word_Attack, (20) Reading_Vocab, (21) Passage_Comp, (22) Dictation, (23) Proofing, (24) Writing_Samples, (25) Writing_Fluency.

14	15	16	17	18	19	20	21	22	23	24	25
1.000											
.427	1.000										
.500	.621	1.000									
.470	.557	.627	1.000								
.458	.517	.558	.659	1.000							
.429	.456	.479	.581	.803	1.000						
.525	.462	.522	.609	.773	.743	1.000					
.495	.518	.544	.628	.783	.647	.754	1.000				
.469	.575	.623	.664	.760	.714	.689	.672	1.000			
.434	.487	.515	.568	.699	.669	.697	.585	.696	1.000		
.410	.473	.467	.506	.628	.592	.613	.566	.679	.566	1.000	
.400	.411	.486	.543	.585	.603	.597	.540	.622	.617	.592	1.000
15.442	16.035	16.747	16.148	16.524	15.574	16.682	15.155	17.320	16.000	15.052	16.589

Table A2. Correlations and standard deviations for Grades 3 and 4, cross-validation sample

Variable												
1	2	3	4	5	6	7	8	9	10	11	12	13
1.000												
.305	1.000											
.144	.143	1.000										
.232	.209	.128	1.000									
− .007	.022	.191	.085	1.000								
.366	.398	.079	.295	.149	1.000							
.127	.177	.253	.099	.195	.212	1.000						
.486	.286	.181	.202	.193	.281	.313	1.000					
.248	.616	.087	.337	.013	.194	.076	.212	1.000				
.027	.166	.578	.108	.275	.093	.303	.178	.124	1.000			
.213	.461	.159	.375	.169	.332	.178	.270	.362	.204	1.000		
.222	.270	.224	.129	.254	.162	.192	.326	.259	.287	.226	1.000	
.315	.544	.246	.341	.224	.597	.321	.334	.359	.232	.453	.340	1.000
.293	.316	.307	.224	.182	.290	.468	.399	.202	.306	.306	.287	.437
.202	.224	.361	.179	.084	.243	.396	.188	.137	.276	.221	.159	.403
.368	.426	.402	.306	.212	.394	.425	.376	.309	.372	.431	.355	.618
.350	.357	.367	.289	.123	.414	.429	.374	.228	.275	.351	.221	.570
.411	.382	.322	.426	.155	.433	.303	.315	.309	.259	.448	.256	.616
.361	.343	.249	.358	.124	.349	.283	.378	.299	.190	.468	.240	.454
.281	.501	.252	.396	.154	.515	.317	.261	.334	.249	.464	.301	.747
.340	.453	.245	.333	.119	.447	.308	.258	.283	.243	.396	.210	.629
.379	.268	.393	.353	.160	.403	.312	.276	.221	.285	.412	.278	.477
.346	.349	.407	.288	.179	.309	.397	.343	.295	.298	.405	.317	.517
.293	.312	.344	.279	− .002	.241	.275	.227	.277	.216	.385	.196	.415
.269	.268	.440	.226	.182	.248	.268	.237	.123	.409	.344	.262	.411
S.D.												
17.373	15.703	15.558	15.557	15.904	15.072	16.484	16.602	16.133	15.603	16.555	16.510	15.233

Variables: (1) Memory_for_Names, (2) Memory_for_Sentences, (3) Visual_Matching, (4) Incomplete_Words, (5) Visual_Closure, (6) Picture_Vocabulary, (7) Analysis_Synthesis, (8) Visual_Auditory_Learn, (9) Memory_for_Words, (10) Cross_Out, (11) Sound_Blending, (12) Picture_Recognition, (13) Oral_Vocabulary, (14) Concept_Formation, (15) Calculation, (16) Applied_Problems, (17) Quant_Concepts, (18) Letter_Word_ID, (19) Word_Attack, (20) Reading_Vocab, (21) Passage_Comp, (22) Dictation, (23) Proofing, (24) Writing_Samples, (25) Writing_Fluency.

14	15	16	17	18	19	20	21	22	23	24	25
1.000											
.341	1.000										
.487	.527	1.000									
.393	.523	.629	1.000								
.379	.422	.551	.583	1.000							
.357	.405	.510	.493	.724	1.000						
.447	.470	.609	.614	.659	.487	1.000					
.356	.464	.574	.593	.668	.532	.721	1.000				
.384	.430	.592	.622	.693	.617	.535	.568	1.000			
.467	.518	.643	.614	.621	.565	.558	.589	.694	1.000		
.404	.417	.458	.492	.509	.492	.497	.547	.589	.658	1.000	
.422	.255	.482	.501	.557	.458	.440	.472	.620	.524	.473	1.000
15.622	15.271	17.396	15.559	14.736	15.399	14.879	15.494	15.253	15.254	13.607	15.778

Table A3. Correlations and standard deviations for Grades 5 and 6, cross-validation sample

Variable												
1	2	3	4	5	6	7	8	9	10	11	12	13
1.000												
.285	1.000											
.176	.043	1.000										
.148	.222	.121	1.000									
.134	.003	.107	.133	1.000								
.408	.443	.136	.255	.124	1.000							
.304	.271	.146	.223	.129	.400	1.000						
.549	.262	.150	.211	.129	.391	.303	1.000					
.193	.515	.173	.346	.083	.222	.089	.178	1.000				
.091	.111	.539	.130	.157	.165	.215	.210	.198	1.000			
.165	.231	.151	.341	.121	.296	.147	.168	.387	.185	1.000		
.228	.165	.186	.123	.383	.343	.277	.285	.093	.209	.178	1.000	
.342	.542	.067	.334	.158	.634	.370	.312	.307	.141	.286	.286	1.000
.381	.182	.238	.166	.198	.296	.495	.340	.103	.265	.273	.241	.359
.114	.157	.327	.017	.083	.105	.311	.188	.100	.297	.090	.118	.237
.311	.359	.359	.228	.160	.378	.391	.301	.191	.281	.267	.255	.522
.199	.363	.289	.188	.010	.357	.361	.236	.256	.226	.215	.151	.469
.292	.462	.196	.242	.007	.414	.336	.198	.316	.193	.251	.215	.537
.267	.307	.182	.273	−.029	.179	.276	.202	.241	.110	.349	.123	.292
.302	.496	.203	.309	.083	.572	.405	.328	.267	.252	.255	.241	.727
.274	.410	.291	.323	.080	.474	.389	.333	.228	.242	.331	.229	.515
.264	.218	.360	.134	.083	.248	.324	.187	.169	.270	.241	.205	.343
.215	.183	.465	.170	.100	.211	.345	.152	.126	.331	.155	.137	.265
.192	.217	.250	.104	.030	.296	.333	.156	.094	.164	.131	.195	.309
.138	.203	.450	.223	.127	.185	.283	.187	.220	.391	.217	.106	.209
S.D.												
15.763	14.738	15.430	16.891	14.719	14.005	16.147	16.617	14.559	15.461	16.067	14.622	14.787

Variables: (1) Memory_for_Names, (2) Memory_for_Sentences, (3) Visual_Matching, (4) Incomplete_Words, (5) Visual_Closure, (6) Picture_Vocabulary, (7) Analysis_Synthesis, (8) Visual_Auditory_Learn, (9) Memory_for_Words, (10) Cross_Out, (11) Sound_Blending, (12) Picture_Recognition, (13) Oral_Vocabulary, (14) Concept_Formation, (15) Calculation, (16) Applied_Problems, (17) Quant_Concepts, (18) Letter_Word_ID, (19) Word_Attack, (20) Reading_Vocab, (21) Passage_Comp, (22) Dictation, (23) Proofing, (24) Writing_Samples, (25) Writing_Fluency.

14	15	16	17	18	19	20	21	22	23	24	25
1.000											
.202	1.000										
.399	.531	1.000									
.297	.522	.636	1.000								
.221	.306	.460	.472	1.000							
.252	.260	.407	.391	.597	1.000						
.285	.272	.547	.477	.596	.414	1.000					
.406	.358	.552	.485	.532	.442	.607	1.000				
.249	.385	.514	.435	.618	.477	.418	.464	1.000			
.317	.396	.477	.413	.478	.429	.422	.475	.682	1.000		
.244	.345	.388	.404	.434	.327	.389	.419	.519	.534	1.000	
.310	.272	.314	.298	.339	.314	.256	.371	.371	.490	.264	1.000
14.815	17.465	14.754	14.837	17.565	15.412	16.390	14.984	14.794	14.162	14.973	15.904

Table A4. Correlations and standard deviations for Grades 7 and 9, cross-validation sample

Variable

1	2	3	4	5	6	7	8	9	10	11	12	13
1.000												
.351	1.000											
.253	.162	1.000										
.144	.228	.157	1.000									
.245	.183	.240	.254	1.000								
.458	.355	.087	.067	.304	1.000							
.283	.308	.208	.136	.169	.317	1.000						
.541	.329	.286	.207	.199	.311	.400	1.000					
.267	.541	.208	.266	.092	.143	.168	.246	1.000				
.284	.177	.621	.188	.319	.193	.266	.297	.156	1.000			
.307	.373	.115	.383	.222	.310	.239	.280	.285	.175	1.000		
.416	.262	.266	.087	.337	.335	.212	.255	.211	.318	.187	1.000	
.443	.530	.154	.185	.334	.681	.389	.326	.280	.194	.401	.352	1.000
.335	.438	.201	.144	.215	.248	.443	.447	.226	.250	.228	.222	.358
.314	.360	.311	.146	.112	.244	.400	.383	.204	.225	.210	.117	.456
.286	.387	.237	.082	.236	.431	.457	.309	.256	.241	.255	.221	.524
.384	.401	.236	.038	.160	.431	.401	.395	.252	.210	.234	.176	.542
.413	.405	.265	.289	.190	.479	.296	.289	.314	.235	.490	.280	.564
.358	.391	.283	.299	.149	.279	.287	.295	.411	.205	.479	.272	.385
.390	.524	.196	.234	.290	.586	.387	.334	.277	.221	.404	.348	.777
.389	.460	.258	.184	.249	.515	.322	.344	.238	.275	.324	.316	.604
.413	.372	.377	.151	.200	.463	.252	.356	.268	.238	.349	.311	.564
.440	.402	.408	.203	.240	.392	.316	.428	.304	.307	.363	.379	.550
.344	.356	.371	.226	.248	.400	.320	.401	.213	.374	.377	.331	.566
.266	.277	.412	.170	.241	.295	.204	.233	.194	.348	.327	.232	.336
S.D.												
14.891	14.861	16.247	15.438	15.275	14.931	15.312	16.661	15.482	15.412	14.900	14.957	14.461

Variables: (1) Memory_for_Names, (2) Memory_for_Sentences, (3) Visual_Matching, (4) Incomplete_Words, (5) Visual_Closure, (6) Picture_Vocabulary, (7) Analysis_Synthesis, (8) Visual_Auditory_Learn, (9) Memory_for_Words, (10) Cross_Out, (11) Sound_Blending, (12) Picture_Recognition, (13) Oral_Vocabulary, (14) Concept_Formation, (15) Calculation, (16) Applied_Problems, (17) Quant_Concepts, (18) Letter_Word_ID, (19) Word_Attack, (20) Reading_Vocab, (21) Passage_Comp, (22) Dictation, (23) Proofing, (24) Writing_Samples, (25) Writing_Fluency.

14	15	16	17	18	19	20	21	22	23	24	25
1.000											
.423	1.000										
.422	.578	1.000									
.437	.595	.664	1.000								
.249	.376	.423	.414	1.000							
.298	.349	.362	.366	.622	1.000						
.460	.500	.560	.566	.578	.442	1.000					
.338	.329	.481	.434	.483	.394	.619	1.000				
.196	.480	.449	.496	.615	.530	.541	.469	1.000			
.329	.459	.379	.502	.561	.559	.593	.520	.660	1.000		
.365	.469	.488	.457	.529	.433	.609	.524	.542	.584	1.000	
.186	.234	.240	.269	.394	.408	.376	.362	.389	.432	.361	1.000
16.915	16.631	16.039	16.061	15.836	17.510	15.665	14.455	17.282	15.329	15.367	16.226

Table A5. Correlations and standard deviations for Grade 12, cross validation-sample

Variable												
1	2	3	4	5	6	7	8	9	10	11	12	13
1.000												
.215	1.000											
.160	.152	1.000										
.084	.206	.162	1.000									
.208	.130	.189	.234	1.000								
.268	.461	.078	.341	.324	1.000							
.313	.348	.229	.142	.211	.435	1.000						
.575	.215	.197	.188	.309	.260	.429	1.000					
.221	.513	.212	.194	.128	.300	.337	.317	1.000				
.132	.197	.616	.113	.200	.128	.220	.120	.194	1.000			
.226	.269	.065	.288	.170	.276	.279	.255	.287	.102	1.000		
.473	.224	.217	.104	.392	.380	.295	.403	.208	.299	.249	1.000	
.277	.441	.192	.304	.298	.689	.472	.331	.395	.227	.418	.463	1.000
.291	.443	.273	.239	.277	.438	.484	.460	.270	.302	.323	.357	.461
.293	.220	.479	.191	.168	.270	.436	.373	.252	.384	.259	.261	.459
.263	.352	.452	.208	.219	.435	.438	.319	.268	.344	.282	.298	.584
.359	.369	.381	.274	.213	.532	.520	.421	.356	.297	.303	.415	.674
.321	.466	.229	.379	.346	.613	.457	.386	.459	.233	.463	.460	.722
.224	.329	.234	.334	.192	.385	.263	.339	.369	.160	.348	.270	.474
.333	.497	.214	.382	.359	.670	.521	.400	.447	.201	.450	.489	.819
.286	.458	.221	.312	.301	.523	.375	.404	.299	.169	.288	.385	.611
.315	.359	.360	.307	.157	.453	.388	.342	.418	.253	.310	.297	.586
.355	.395	.437	.238	.205	.373	.420	.422	.344	.339	.375	.329	.555
.294	.372	.332	.234	.172	.356	.327	.325	.331	.292	.381	.331	.573
.137	.242	.330	.270	.151	.165	.228	.218	.137	.266	.283	.205	.359
S.D.												
15.067	15.797	14.874	15.319	15.916	18.078	17.494	16.354	16.332	15.465	14.982	16.170	16.587

Variables: (1) Memory_for_Names, (2) Memory_for_Sentences, (3) Visual_Matching, (4) Incomplete_Words, (5) Visual_Closure, (6) Picture_Vocabulary, (7) Analysis_Synthesis, (8) Visual_Auditory_Learn, (9) Memory_for_Words, (10) Cross_Out, (11) Sound_Blending, (12) Picture_Recognition, (13) Oral_Vocabulary, (14) Concept_Formation, (15) Calculation, (16) Applied_Problems, (17) Quant_Concepts, (18) Letter_Word_ID, (19) Word_Attack, (20) Reading_Vocab, (21) Passage_Comp, (22) Dictation, (23) Proofing, (24) Writing_Samples, (25) Writing_Fluency.

14	15	16	17	18	19	20	21	22	23	24	25
1.000											
.459	1.000										
.531	.726	1.000									
.531	.742	.760	1.000								
.468	.456	.483	.603	1.000							
.334	.426	.414	.479	.633	1.000						
.503	.429	.565	.655	.720	.497	1.000					
.434	.392	.521	.564	.496	.371	.629	1.000				
.352	.491	.471	.560	.642	.592	.583	.442	1.000			
.458	.579	.548	.612	.584	.556	.567	.461	.638	1.000		
.393	.525	.529	.530	.569	.429	.505	.531	.575	.574	1.000	
.305	.355	.348	.319	.369	.345	.377	.384	.381	.438	.424	1.000
16.511	16.978	17.503	17.444	16.328	15.847	17.524	16.491	15.299	16.417	15.596	16.922

References

Aaron, P. G. (1995). Differential diagnosis of reading disabilities. *School Psychology Review*, *24*(3), 345–360.

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles: GMDP Statistical Software.

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

Bickley, P. A., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: test of the structure of intelligence across the life span. *Intelligence*, *20*, 309–328.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

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

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 122–130). New York: Guilford.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116–127.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.

Daniel, M. H. (1997). Intelligence testing: status and trends. *American Psychologist*, *52*, 1038–1045.

Felton, R. H., & Pepper, P. P. (1995). Early identification and intervention of phonological deficits in kindergarten and early elementary children at risk for reading disability. *School Psychology Review*, *24*, 405–414.

Flanagan, D. P. (1999). Wechsler-based CHC cross-battery assessment and reading achievement: strengthening the validity of interpretations drawn from Wechsler test scores. Manuscript submitted for publication.

Flanagan, D. P., Genshaft, J. L., Harrison, P. L. (Eds.) (1997). *Contemporary intellectual assessment: theories, tests, and issues*. New York: Guilford Press.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 314–325). New York: Guilford.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf–Gc theory: joint confirmatory factor analysis of the WJ-R and KAIT in a non-white sample. *Journal of School Psychology*, *36*, 151–182.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf–Gc theory: a contemporary approach to interpretation*. Boston: Allyn and Bacon.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*(4), 407–434.

Horn, J. L. (1991). Measurement of intellectual capabilities: a review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual*. Chicago: Riverside.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York: Macmillan.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf–Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 53–91). New York: Guilford.

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Humphreys, L. G. (1973). The misleading distinction between aptitude and achievement tests. In D. R. Green (Ed.), *The aptitude–achievement distinction* (pp. 262–278). Carmel, CA: CTB/McGraw-Hill.

Jensen, R. A. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structure*, *7*, 93–118.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.

Joshi, R. M. (1995). Assessing reading and spelling skills. *School Psychology Review*, *24*(3), 361–375.

Kail, R. (1991). Development of processing speed in childhood and adolescence. In H. W. Reese (Ed.), *Advances in child development and behavior, vol. 23* (pp. 151–185). San Diego, CA: Academic Press.

Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence*. Needham Heights, MA: Allyn and Bacon.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn and Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 373–402). New York: Guilford.

Keith, T. Z. (1999a). Effects of general and specific abilities on student achievement: similarities and differences across ethnic groups. *School Psychology Quarterly*, *14*, 239–262.

Keith, T. Z. (1999b). Structural equation modeling in school psychology. In C. R. Reynolds, & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed.) (pp. 78–107). New York: Wiley.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Loehlin, J. C. (1998). *Latent variable models: an introduction to factor, path, and structural analysis*. (3rd ed.). Hillsdale, NJ: Erlbaum.

Lohman, D. F. (1989). Human intelligence: an introduction to advances in theory and research. *Review of Educational Research*, *59*(4), 333–373.

MacCallum, R. C. (1995). Model specification: procedures, strategies and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 16–36). Thousand Oaks, CA: Sage.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *2*, 130–149.

Mather, N. (1991). *An instructional guide to the Woodcock–Johnson Psycho-Educational Battery-Revised*. Brandon, VT: Clinical Psychology Publishing.

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, *87*(2), 179–192.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: a critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, *8*, 290–302.

McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic leaning behavior, disposition, and achievement through ability subtests-Or, more illusions of meaning? *School Psychology Review*, *26*, 120–142.

McGhee, R. L. (1993). Fluid and crystallized intelligence: confirmatory factor analysis of the Differential Abilities Scale, Detroit Tests of Learning Aptitude-3, and Woodcock–Johnson Psycho-Educational Assessment Battery-Revised. *Journal of Psychoeducational Assessment, Monograph Series: Woodcock–Johnson Psycho-Educational*, 20–38.

McGrew, K. S. (1993). The relationship between the WJ-R Gf–Gc cognitive clusters and reading achievement across the life-span. *Journal of Psychoeducational Assessment, Monograph Series: WJ-R*, 39–53.

McGrew, K. S. (1994). *Clinical interpretation of the Woodcock–Johnson Tests of Cognitive Ability-Revised*. Needham Heights, MA: Allyn and Bacon.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf–Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests and issues* (pp. 151–179). New York: Guilford.

McGrew, K. S., & Flanagan, D. P. (1998). *The Intelligence Test Desk Reference (ITDR): Gf–Gc cross-battery assessment*. Boston, MA: Allyn and Bacon.

McGrew, K. S., Flanagan, D. P., Keith, T., & Vanderwood, M. L. (1997). Beyond g: the impact of Gf–Gc specific cognitive abilities research on the future use and interpretation of intelligence test batteries in the schools. *School Psychology Review*, *26*(2), 189–210.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ-R technical manual*. Chicago: Riverside.

McNemar, Q. (1964). Lost: our intelligence? Why? *American Psychologist*, *19*, 871–882.

Messick, S. (1992). Multiple intelligences or multilevel intelligence? Selective emphasis on distinctive properties of hierarchy: on Gardner's Frames of Mind and Sternberg's Beyond IQ in the context of theory and research on the structure of human abilities. *Psychological Inquiry*, *3*(4), 365–384.

Morris, R. D., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Steubing, K. K., Fletcher, J. M., Lyon, G. R., Francis, D. J., & Shaywitz, B. A. (1998). Subtypes of reading disability: variability around a phonological core. *Journal of Educational Psychology*, *90*(3), 347–373.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: prediction and explanation* (3rd ed.). New York: Holt, Rinehart and Winston.

Reschly, D. J. (1990). Found: our intelligences: what do they mean? *Journal of Psychoeducational Assessment*, *8*, 259–267.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology*, *9*, 395–396.

Sattler, J. (1988). *Assessment of children*. (3rd ed.). San Diego, CA: Author.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. (1982). A theoretical analysis of the performance of learning disabled students on the Woodcock–Johnson Psycho-Educational Battery. *Journal of Learning Disabilities*, *15*, 221–226.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: aptitude, adaptation, and assessment. *Annual Review of Psychology*, *43*, 583–626.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: issues of task compatibility. *Journal of Experimental Child Psychology*, *38*, 175–190.

Stone, B. J. (1992). Joint confirmatory factor analyses of the DAS and WISC-R. *Journal of School Psychology*, *30*, 185–195.

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

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*(5), 276–286.

Weschler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams and Wilkins.

Woodcock, R.W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, *8*, 231–258.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery-Revised*. Chicago: Riverside.

Woodcock, R. W., & Mather, N. (1989). WJ-R tests of cognitive ability-standard and supplemental batteries: examiner's manual. In R. W. Woodcock, & M. B. Johnson (Eds.), *Woodcock–Johnson Psycho-Educational Battery-Revised*. Chicago: Riverside.

Ysseldyke, J. E. (1990). Goodness of fit of the Woodcock–Johnson Psycho-Educational Battery-Revised to the Horn–Cattell Gf–Gc theory. *Journal of Psychological Assessment*, *8*, 268–275.