Wechsler-Based CHC Cross-Battery Assessment and Reading Achievement: Strengthening the Validity of Interpretations Drawn from Wechsler Test Scores

Dawn P. Flanagan
St. John's University

The Cattell-Horn-Carroll (CHC) cross-battery approach to assessing cognitive abilities has been introduced to the field recently as an alternative to traditional assessment instruments and techniques (Flanagan, McGrew, & Ortiz, 2000; McGrew & Flanagan, 1998). This theory-based method of assessment was developed to provide a more valid and defensible way of deriving meaning from test scores than that provided by the traditional (and largely atheoretical) Wechsler Scale approach. The present study is the first to investigate the validity of the cross-battery approach using a sample of elementary school children ($N = 166$) who were administered the Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974) and Woodcock-Johnson Psycho-Educational Battery–Revised (WJ-R; Woodcock & Johnson, 1989). Structural Equation Modeling was used to examine whether (a) Wechsler-based CHC cross-battery assessment explained more variance in reading achievement than a traditional Wechsler model, and (b) whether the Wechsler-based CHC cross-battery model provided a more accurate description of the general and specific abilities on general and specific reading achievement than the traditional Wechsler model, as predicted in the extant g/specific abilities literature. Findings showed that the g factor underlying the Wechsler-based CHC cross-battery model accounted for substantially more variance in reading achievement (25%) than the g factor underlying the atheoretical Verbal Comprehension-Perceptual Organization-Freedom From Distractibility (VC-PO-FFD) Wechsler model. In addition, findings demonstrated that when assessments are organized around the strong CHC theoretical model, specific cognitive abilities, including auditory processing (Ga), crystallized intelligence (Gc), and processing speed (Gs) explained a significant portion of variance in reading achievement beyond that accounted for by g. These

findings suggest that a Wechsler-based CHC cross-battery approach is an effective way of ensuring valid representation of multiple cognitive abilities, specifically those that have been found to affect significantly the development of reading skills.

There is little, if any, doubt that the Wechsler Verbal/Nonverbal (Performance) model and Verbal Comprehension-Perceptual Organization-Freedom From Distractibility (VC-PO-FFD) model are among the most widely *recognized* interpretive models of the structure of cognitive abilities to date. Despite the addition of the Processing Speed Index on the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991) and Working Memory Index on the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III; Wechsler, 1997), interpretation of the traditional two- and three-factor Wechsler models remains a consistent practice (for a review, see Flanagan, McGrew, & Ortiz, 2000). Notwithstanding the number of factors underlying the Wechsler Scales (i.e., two, three, or four), it has become increasingly apparent that (all versions of) these instruments are limited and do not reflect current conceptions of intelligence. Thus, some have argued that the interpretations based on any of the underlying interpretive perspectives of the Wechsler Scales are largely outdated and misguided (Flanagan et al., 2000).

Because all versions of the Wechsler Intelligence Scales are not grounded in a strong construct validation research program, it is believed that the application of a strong theory-based construct validity approach to Wechsler-based assessments can strengthen the validity of the interpretations drawn from Wechsler test scores (Flanagan et al., 2000; McGrew & Flanagan, 1998). This article will (a) present an empirically validated model of the structure of intelligence; (b) briefly describe the stages of construct validity that are necessary to establish strong test validity; (c) describe the relation between a commonly applied Wechsler model (viz., VC-PO-FFD) and contemporary, empirically validated models of intelligence; (d) highlight the importance of theory-based research in understanding the effects of specific cognitive abilities on academic achievement and their implications for assessment; and (e) describe how the Wechsler Scales can benefit from advances in applied measurement that operationalize contemporary theory (i.e., Flanagan et al.'s [2000] cross-battery approach).

Next, in response to the calls for more theory-based intelligence test interpretation (Flanagan & McGrew, 1997; Flanagan et al., 2000; Kaufman, 1994, 2000; Kamphaus, Petosky, & Morgan, 1997; Keith, 1997, 1999a; McGrew & Flanagan, 1998), an examination of the effects of general and specific cognitive abilities on reading skills and their implications for deriving score meaning from theoretically based (versus nontheoretically based) cognitive ability assessments will be conducted. According to Keith (1997, 1999a) and McGrew and Flanagan (1998), assessment instruments and approaches that are based on well-validated theoretical models of the structure of intelligence have the potential to contribute more meaningfully to research on the relations between cognitive abilities and different out-

come criteria than instruments and approaches that are not linked to strong theory. Thus, the overriding purpose of the present research is to (a) examine whether a new, theoretically driven applied measurement approach, called "Wechsler-based Cattell-Horn-Carroll (CHC) cross-battery assessment"[1] (Flanagan et al., 2000; McGrew & Flanagan, 1998), explains more variance in reading achievement than an atheoretical, traditional VC-PO-FFD model, and (b) whether the Wechsler-based CHC cross-battery model provides a more accurate description of the general and specific cognitive abilities that contribute to the explanation of reading achievement than a traditional Wechsler model. The major purpose of this research is to compare theoretical and atheoretical models of the structure of intelligence. The VC-PO-FFD model was chosen as the atheoretical model in this research because (a) it has been the prevailing interpretive framework for the Wechsler Scales for the past two decades; (b) it continues to be applied, albeit less frequently, in current practice; and (c) the current research questions of interest (which are largely theoretical in nature) can be addressed appropriately with an available Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974)/Woodcock-Johnson Psycho-Educational Battery–Revised (WJ-R; Woodcock & Johnson, 1989) dataset.

THE CHC THEORY OF COGNITIVE ABILITIES: AN EMPIRICALLY VALIDATED MODEL OF THE STRUCTURE OF INTELLIGENCE

The extant factor-analytic research conducted over the past 60 or more years has converged on a relatively complete multiple-intelligences taxonomy known as the CHC Theory of Cognitive Abilities. As discussed by others (Flanagan et al., 2000; McGrew, 1997; McGrew & Flanagan, 1998; Woodcock & McGrew, 2000; R. W. Woodcock, July 16, 1999, personal communication), CHC theory represents an integration of Carroll's (1993) three-stratum theory and the Cattell-Horn *Gf-Gc* theory (Horn, 1994).

The CHC conception of intelligence is supported extensively by factor-analytic (i.e., structural) evidence and developmental, neurocognitive, and heritability evidence (for a summary, see Horn & Noll, 1997). In addition, there is a mounting body of research available on the relations between the broad CHC abilities and many academic and occupational achievements (for a review of this literature, see McGrew & Flanagan, 1998). Furthermore, studies have shown that the CHC struc-

---

1. Recently, it has come to my 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 article, I 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.

ture of intelligence is invariant across the lifespan (Bickley, Keith, & Wolfe, 1995) and across gender, ethnic, and cultural groups (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 (Kranzler & Keith, 1999; McGrew & Flanagan, 1998; Messick, 1992; Sternberg & Kaufman, 1998).

Given the breadth of empirical support for the CHC structure of intelligence, it provides one of the most useful frameworks for designing and evaluating intelligence tests (Carroll, 1997, 1998; Flanagan & McGrew, 1997; Kaufman, 2000; Keith & Kranzler, 1999; Keith & Witta, 1997; McGrew, 1997; Woodcock, 1990; Ysseldyke, 1990). Moreover, in light of the well-established structural validity of CHC theory (and its corresponding measures, such as the WJ-R), external validity support for the various CHC constructs, derived through sound research methodology, can be used confidently to guide test interpretation (Benson, 1998).

Unlike the WJ-R, all versions of the Wechsler Scales are largely atheoretical. Caution must be used when deriving score meaning from individual Wechsler tests and composites. However, by linking the Wechsler Scales to a strong theory of intelligence (i.e., CHC theory), it is possible to improve our understanding of the constructs underlying the Wechsler Scales and, as a result, strengthen the interpretations that are drawn from Wechsler test scores. Before demonstrating this, it is important to understand the stages that are necessary to establish strong test validity.

STRONG SUBSTANTIVE THEORY IS A PREREQUISITE FOR STRONG TEST VALIDITY

Strong test validity is the result of successful progressions through three broad construct validation stages, namely substantive, structural, and external (Benson, 1998; Cronbach, 1971; Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989). Research conducted at each stage must provide sufficient evidence for "the construct validity interpretation of test scores" (Benson, 1998, p. 15). If, at any given stage, research does not provide sufficient construct validity evidence to support test score interpretations, then a reexamination of the previous stage is warranted (Benson, 1998).

The Substantive Stage of Test Validation

In the substantive stage of test validation, psychological constructs are defined. Underlying each psychological construct is a theoretical domain (i.e., everything that is known about the construct) and an empirical domain (i.e., measures that were developed to operationalize constructs) (Benson, 1998). A strong theory of intelligence is needed to specify, define, and circumscribe both domains of the focal constructs. Because the Wechsler Intelligence Scales are not grounded in a strong theory of intelligence (Carroll, 1993, 1995; Kamphaus, 1993; Shaw, Swerdlik, & Laurent, 1993; Sternberg, 1993), they have a weak substantive foundation. This

conclusion was supported most recently by a comprehensive literature review of validity studies conducted with the Wechsler Intelligence and Memory Scales (Flanagan et al., 2000).

The Substantive Foundation of the Wechsler Scales. Because the CHC taxonomy "provides what is essentially a 'map' of all known cognitive abilities. . .[it] can be used in interpreting scores on the many tests used in individual assessment" (Carroll, 1993, p. 127). The CHC taxonomy can serve as a blueprint from which to evaluate the content validity of intelligence tests. The first systematic attempt to map the major intelligence batteries to the CHC structural framework was presented by Woodcock (1990). Woodcock's (1990) classifications, which were only at the broad (stratum II) level, were extended to the narrow (stratum I) level by McGrew (1997). McGrew and Flanagan (1998) and Flanagan et al. (2000) subsequently refined Woodcock and McGrew's classifications and extended them to more than 200 cognitive ability measures. The Wechsler CHC test classifications of Flanagan et al. demonstrate that the Wechsler Intelligence Scales have weak content validity when evaluated according to the CHC theoretical model. This conclusion is not surprising in light of the fact that the revisions of the Wechsler Scales generally have deviated little from their historical roots. Of the CHC domains specified in contemporary theory (viz., fluid intelligence [*Gf*], crystallized intelligence [*Gc*], short-term memory [*Gsm*], visual processing [*Gv*], auditory processing [*Ga*], long-term retrieval [*Glr*], processing speed [*Gs*]), *Gc* and *Gv* are the only abilities that consistently have strong content or construct representation (defined by the measurement of two or more qualitatively different narrow abilities that represent the broad ability) across all of the Wechsler Intelligence Scales. In addition to *Gc* and *Gv*, the WISC-III adequately represents the construct of *Gs*, whereas the WAIS-III adequately represents *Gs* and *Gsm*.

None of the Wechsler Intelligence Scales contain measures of *Ga* and *Glr*. In the *Gf* domain, the WAIS-III has weak construct representation (i.e., it only measures one narrow ability within this domain), whereas all other Wechsler Scales do not measure this ability. Similarly, *Gsm* is underrepresented on the Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI-R; Wechsler, 1989) and WISC-III (i.e., each scale provides only one measure of this construct) and *Gs* is underrepresented on the WPPSI-R.

The weak substantive foundation of the Wechsler Scales, from the perspective of CHC theory, is primarily due to the fact that during the development and subsequent revisions of the Wechsler Intelligence Scales, no explicit attempt was made to specify and circumscribe the theoretical domain (and subdomains) of intelligence (Keith & Witta, 1997). As a result, the Wechsler Intelligence Scales have not benefited from a test development process where measures were operationalized according to a theory-based blueprint. The largely atheoretical foundation of the Wechsler Scales constrains and complicates the validity of inferences that can be drawn from many Wechsler individual test and composite scores (past and present) (Flanagan et al., 2000; Keith & Witta, 1997).

Before test validation can advance to the next stage (i.e., structural validity) a strong substantive foundation is necessary. In the absence of a strong substantive foundation, the extant structural validity evidence in support of the Wechsler Scales must be viewed cautiously.

The Structural Stage of Test Validation

The structural or internal validity stage of test validation focuses on whether the tests measure the construct(s) they are purported to measure. This stage is primarily concerned with the internal consistency of a test's variables (Keith & Kranzler, 1999). According to Messick (1995), "the internal structure of the assessment (i.e., interrelations among the scored aspects of tasks and subtask performance) should be consistent with what is known about the internal structure of the domain...[t]his...is called *structural fidelity*" (p. 746). At this stage, positive structural evidence can be inferred when the tests selected or developed from the empirical domain are found to be valid measures of the corresponding constructs in the theoretical domain (Benson, 1998). Construct validation theory postulates that "the absence of structural fidelity precludes construct validity" (Keith & Kranzler, 1999, p. 303). Therefore, in the absence of structural evidence, a theory may need to be refined or the tests purported to measure the theoretical constructs may need to be revised or both (Benson, 1998; Keith & Kranzler, 1999).

Within-Battery Structural Research with the Wechsler Scales. The preponderance of Wechsler Scale structural validity studies (summarized in Flanagan et al., 2000) support the internal validity of these measures. Within-battery factor-analytic studies of the early Wechsler Scales invariably revealed a two-factor solution involving a verbal and nonverbal (or perceptual-organizational) factor (Gutkin & Reynolds, 1981; Silverstein, 1982). Later versions of the Wechsler Scales (WISC-R, WAIS-R) were characterized best by three factors: Verbal Comprehension (VC), Perceptual Organization (PO), and Memory/Freedom From Distractibility (FFD) (Leckliter, Matarazzo, & Silverstein, 1986). More recently, Keith and Witta's (1997) multisample hierarchical confirmatory factor analysis of the WISC-III standardization data found support for similar VC and PO factors and a similar FFD-like factor they labeled, Quantitative Reasoning. In addition, consistent with the WISC-III factor-based index interpretation system, Keith and Witta found support for a separate Processing Speed (PS) factor. Thus, within-battery factor-analytic studies with the Wechsler Scales support the current consensus that a three- or four-factor structure characterizes the various configurations of the common core Wechsler subtests.

Figure 1 provides a summary of the incremental improvement in the internal or structural validity of the factor-based interpretation schemes of the Wechslers. This within-battery structural improvement is represented in Figure 1 by the progression from a two-factor Verbal-Performance model (Wechsler, 1974) to a four-factor Verbal-Performance-FFD/Working Memory-Processing Speed model (Wechsler, 1997).
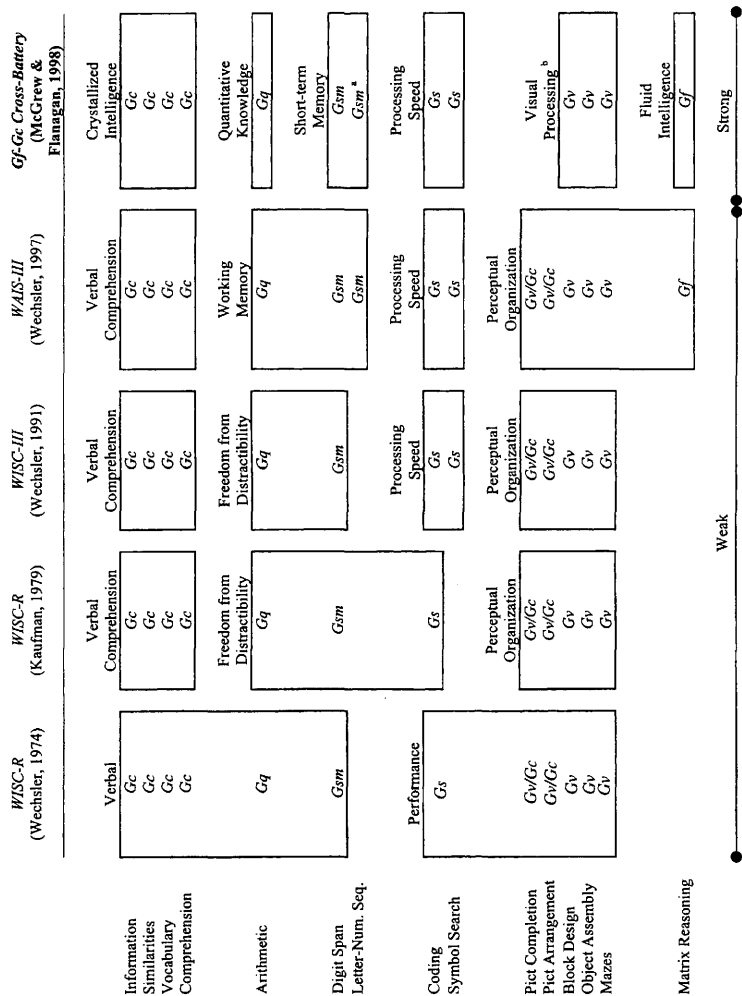
FIGURE 1. Incremental improvement in the structural validity of the factor-based interpretation schemes of the Wechsler Scales

Subtest	WISC-R (Wechsler, 1974)	WISC-R (Kaufman, 1979)	WISC-III (Wechsler, 1991)	WAIS-III (Wechsler, 1997)	Gf-Gc Cross-Battery (McGrew & Flanagan, 1998)
Information	Verbal — Gc	Verbal Comprehension — Gc	Verbal Comprehension — Gc	Verbal Comprehension — Gc	Crystallized Intelligence — Gc
Similarities	Gc	Gc	Gc	Gc	Gc
Vocabulary	Gc	Gc	Gc	Gc	Gc
Comprehension	Gc	Gc	Gc	Gc	Gc
Arithmetic	Gq	Freedom from Distractibility — Gq	Freedom from Distractibility — Gq	Working Memory — Gq	Quantitative Knowledge — Gq
Digit Span	Gsm	Gsm	Gsm	Gsm	Short-term Memory — Gsm
Letter-Num. Seq.				Gsm	Gsm [a]
Coding	Performance — Gs	Gs	Processing Speed — Gs	Processing Speed — Gs	Processing Speed — Gs
Symbol Search			Gs	Gs	Gs
Pict Completion	Gv/Gc	Perceptual Organization — Gv/Gc	Perceptual Organization — Gv/Gc	Perceptual Organization — Gv/Gc	Visual Processing [b]
Pict Arrangement	Gv/Gc	Gv/Gc	Gv/Gc	Gv/Gc	Gv
Block Design	Gv	Gv	Gv	Gv	Gv
Object Assembly	Gv	Gv	Gv	Gv	Gv
Mazes	Gv	Gv	Gv		
Matrix Reasoning				Gf	Fluid Intelligence — Gf

Weak ———————————————————— Strong

301

Although the within-battery factor analysis research of the most recent Wechsler Scales generally supports the internal validity of the batteries, the ever-changing interpretation of the original FFD tests should give cause for concern. Specifically, the ever-changing factorial nature of the Arithmetic, Digit Span, and Coding/Digit Symbol subtests raises concerns about the validity of the factor-based interpretations that have been offered for these subtests (Flanagan et al., 2000; Kaufman, 1994; McGrew & Flanagan, 1998). This variability is the result of the weak substantive or theoretical foundation of the Wechsler Scales.

Cross-Battery Structural Research with the Wechsler Scales. The strong CHC theory of cognitive abilities has been used to guide a series of confirmatory cross-battery factor analyses of all the major intelligence batteries, including the Wechsler Scales. Flanagan et al. (2000) recently synthesized the results of these factor analyses and provided broad and narrow CHC ability classifications for all individual tests from the major intelligence batteries. The results of this research provided a more complete understanding of the structural characteristics and validity of the Wechsler Intelligence Scales (and indeed all intelligence batteries) than that offered by atheoretical and within-battery analyses.

These cross-battery factor analysis studies differ substantially from the within-battery structural validity studies described above in two important ways. First, CHC theory was the framework used to organize each confirmatory factor analysis study. Second, all the data sets included tests from the WJ-R (or WJ-III; Woodcock & McGrew, 2000), a battery comprised of empirically validated indicators of eight broad CHC abilities (McGrew, 1994; McGrew, Werder, & Woodcock, 1991; Reschly, 1990; Woodcock, 1990; Ysseldyke, 1990). As a result, each of the major intelligence batteries was analyzed together with a common set of empirically supported CHC test indicators from the WJ. Most of the broad CHC ability test classifications (Flanagan & McGrew, 1998; Flanagan et al., 2000) were based on empirical research studies. These CHC classifications provide evidence for the construct validity of the tests within individual intelligence batteries. A comparison of within-battery and CHC cross-battery Wechsler Scale analyses, in particular, revealed a number of important findings.

First, when considered within the context of CHC theory, the VC and PO factors are interpreted as measuring mainly *Gc* and *Gv* abilities, respectively. Second, the FFD factor was not consistent with or similar to any broad ability within the CHC taxonomy. In within-battery factor analysis studies, the so-called FFD factor appears to have been an invalid factor that consisted of tests that had no counterparts that measured similar cognitive ability constructs with which they could correlate to form valid factors (for a discussion, see Flanagan et al., 2000). Third, although the PO tests have a long history of being interpreted as measures of *Gf* (or higher-level reasoning) abilities (Kaufman, 1979, 1994), this conclusion is no longer viable. As summarized in Figure 1, cross-battery analyses with the WISC-R and WISC-III show that the PO tests do not load on a *Gf* factor (Woodcock &

McGrew, 2000). These results, as well as other cross-battery factor analyses of the Wechsler's with the Differential Ability Scales (DAS; Elliott, 1990; Stone, 1992) and Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993), call into question the traditional *Gf* interpretation of the Wechsler PO tests (McGrew & Flanagan, 1996).

In conclusion, results of strong programs of theory-based construct validity research show that the Wechsler Intelligence Scales' continued allegiance to largely practical and atheoretical models of intelligence limits the validity of the inferences that can be drawn from some of the Wechsler subtest and composite scores. Figure 1 depicts how the lack of a strong theory can result in conducting structural validity studies with a collection of tests that have significant content validity limitations, which can significantly confound attempts to draw valid inferences from individual test and index scores. Although most of the within-battery structural validity studies reported for the Wechsler Scales have provided evidence for a consistent factor structure (e.g., VC-PO-FFD and VC-PO-FFD-PS), this evidence does not portray accurately the underlying theoretical constructs of these instruments. This reality serves to complicate the interpretation of, or inferences that can be drawn from, the relations between the focal "theoretical constructs" (e.g., PO, FFD) and other constructs (or measures) "outside" of, or external to, the theoretical and empirical domains of primary interest (e.g., academic achievement) (Benson, 1998). In programs of construct validation research, it is futile to attempt to establish external validity before establishing structural validity (Benson, 1998; Keith & Kranzler, 1999).

The External Stage of Test Validation

Just as a strong substantive foundation is necessary before establishing structural validity, a strong structural foundation is necessary before amassing external validity evidence (Nunnally, 1978). The necessary and sufficient conditions for construct validity are both met when structurally valid measures demonstrate expected convergent and divergent relations with measures of constructs external to the focal measures (Benson, 1998; Benson & Hagtvet, 1996). This "looking outside" of the focal theoretical and empirical domains represents the external stage or component of a strong program of construct validity research.

External Validity Evidence and CHC Theory. Among the major intelligence batteries currently available, the WJ-R appears to be one of the only instruments that has progressed successfully through all three stages of construct validation. Specifically, the theoretical (substantive) "blueprint" for the WJ-R was the empirically supported CHC theory and the structural support for the WJ-R, as an operationalization of several CHC constructs, has been well documented (e.g., McGrew et al., 1991; Keith, 1997). Thus, the extant literature that describes the relations between the CHC constructs underlying the WJ-R and external criteria can

be used confidently to guide interpretations and inferences made from the WJ-R test scores and clusters (McGrew, 1994).

An important and potentially valuable line of external validation research using the CHC theory-based WJ-R is the "*g*/specific abilities" research. Using the WJ-R tests to represent multiple CHC constructs, recent research has demonstrated that specific cognitive abilities are important in understanding specific academic skills (Keith, 1999a; Gustafsson & Balke, 1993; McGrew, Flanagan, Keith, & Vanderwood, 1997; Vanderwood, McGrew, Flanagan, & Keith, 1999). Furthermore, these specific abilities explain a significant portion of the variance in academic criteria (e.g., reading achievement) above and beyond that accounted for by *g* (or general ability) and the importance of these specific abilities in understanding reading and mathematics achievement, in particular, has been found to be indistinguishable across ethnic groups (Keith, 1999a). In general, this research shows that specific cognitive abilities (*Gf*, *Ga*, *Gs*, *Gc*) have significant, cross-validated effects on the development of specific reading and math skills (McGrew et al., 1997). Also, these cognitive abilities are significantly related to academic skills at different developmental levels. This research suggests that certain cognitive abilities, at certain ages, may be important to measure and interpret in addition to *g* (or general intelligence) in reading- and math-related referrals.

According to Keith (1997, 1999a) and McGrew and Flanagan (1998), the findings of the *g*/specific abilities research should apply to other instruments and approaches besides the WJ-R. However, in light of the weak substantive and structural foundations of the Wechsler Scales it is likely that these instruments, despite their widespread use, would only serve to muddy, rather than contribute to, the *g*/specific abilities research. The next section of this article highlights the limitations of the extant Wechsler-based external validity literature and offers a method to strengthen the substantive and structural foundations of the Wechslers, respectively, which is needed to test Keith's and McGrew and Flanagan's prediction and improve Wechsler-based external validation research.

External Validity Evidence and the Wechsler Scales. A review of the Wechsler-based research literature appears to support the external validity of the Wechsler Scales. For example, Zimmerman and Woo-Sam (1997) reported an average correlation (across 55 samples) of .75 between the WISC-III Full Scale score and 11 different ability measures. They also reported correlations ranging from .50 to .65 between the WISC-III Full Scale score and various achievement tests. A recent review of more than a decade of Wechsler-based external validation research reported in Flanagan et al. (2000) produced similar conclusions.

Such positive external validity results reported for the Wechsler Scales appears to raise doubts about the validity of the claim that a strong substantive or theoretical foundation is required for an intelligence test to possess strong construct validity. How can the Wechsler Scales demonstrate significant external validity evidence in the absence of strong substantive and structural foundations? The answer lies in the

fact that although positive external validity evidence has been reported, some of which provides strong evidence for some of the Wechsler tests and indexes, part of this evidence is misleading. For instance, at first glance, the extant external validity research appears to support the validity of the FFD index. Numerous studies have reported significant relations between the Wechsler FFD index and many external variables ranging from reading and math achievement to personality and emotional disturbance (Kaufman, 1994; Wielkiewicz, 1990). However, as Kaufman (1994) reminded clinicians, the mere presence of significant relations between intelligence test scores and external variables is not always indicative of strong validity evidence. In fact, patterns of significant correlations that do not appear consistent with theory may cast doubt on the construct validity of the test scores.

The meaning of the Wechsler PO index score, both from internal and external validity perspectives, also suffers from a poor theoretical-empirical domain match. For example, the current WISC-III PO index score (like its PO WISC-R predecessor), although predominately a measure of Gv abilities, includes two tests (Picture Completion and Picture Arrangement) that the extant factor-analytic research has consistently found to measure both Gv and Gc. Although such factorially complex measures may have important purposes in certain situations, factorial complexity makes it difficult to understand and interpret significant relations between the PO index and external validity criteria.

In contrast with the nebulous interpretations of the FFD and PO indexes, significant relations between the Wechsler VC index and other measures of verbal or Gc abilities and school achievement support the validity of inferences drawn from the tests comprising the Wechsler VC index. This is largely because of the strong Wechsler VC-Gc empirical-theoretical domain match (Fig. 1). To be able to interpret (or reinterpret) the extant external validity evidence for the Wechsler Scales and contribute meaningfully to the Wechsler-based external validity literature, it is necessary to establish a strong empirical-theoretical domain match for all Wechsler tests and composites and not just the VC composite.

THE THREE STAGES OF CONSTRUCT VALIDATION IN PERSPECTIVE: "REPAIRING" THE WECHSLER SCALES THROUGH THE FLANAGAN-MCGREW CHC CROSS-BATTERY APPROACH

In recognition of the theory–practice gap that exists in the field of intelligence testing, Flanagan and McGrew (1997) developed an approach to assessment—the CHC Cross-Battery Approach—that links all major intelligence tests to CHC theory (McGrew & Flanagan, 1998). Specifically, they provided a set of psychometrically and theoretically defensible principles and procedures for supplementing any intelligence battery with tests from other batteries to broaden the range and improve upon the measurement of CHC abilities represented in assessment. In effect, their assessment model serves to strengthen the empirical-theoretical domain match for all cognitive ability tests. Recently, they have

applied their cross-battery approach to the Wechsler Intelligence Scales (Flanagan et al., 2000).

It is beyond the scope of this article to describe the cross-battery approach in detail. Therefore, this discussion will be limited to the manner in which a WISC-R CHC cross-battery was constructed for use in the present research.[2] As discussed above, the individual tests of the Wechsler Scales were classified according to the broad (stratum II) and narrow (stratum I) CHC abilities they measure via the results of theory driven cross-battery factor analyses. These classifications were deemed necessary not only to link the Wechsler Scales to an empirically validated theoretical model but to guard against two major sources of invalidity in assessment: construct irrelevant variance and construct underrepresentation (Comrey, 1988; Flanagan et al., 2000; McGrew & Flanagan, 1998; Messick, 1995).

Furthermore, the classifications of the Wechsler Scales at the broad and narrow CHC ability levels were necessary to provide a stronger substantive foundation for these instruments than that represented by the prevailing two-, three-, and four-factor Wechsler models. An examination of these classifications (Fig. 1) demonstrated that the WISC-R (like the WISC-III) contains several qualitatively different indicators of *Gc* (i.e., Comprehension, Vocabulary, Information, Similarities) and *Gv* (i.e., Object Assembly, Block Design, and Mazes). *Gs*, *Gq*, and *Gsm* are underrepresented and *Glr*, *Ga*, and *Gf* are not measured.[3] Thus, the WISC-R needs to be supplemented to provide adequate measurement of *Gs*, *Gq*, *Gsm*, *Glr*, *Ga*, and *Gf* to more closely approximate a contemporary, empirically supported structure of cognitive abilities, thereby improving its structural foundation.

The WISC-R (or any Wechsler Scale) can be supplemented well with tests from the WJ-R. This is because (a) the WJ-R measures a broader range of CHC abilities than all other intelligence batteries; (b) it contains two strong measures of each of seven broad cognitive abilities; and (c) it was used in all the major cross-battery factor analyses that provided the information necessary to make CHC classifications of cognitive ability tests (McGrew & Flanagan, 1998). After following the recommended procedures for constructing a cross-battery assessment (Flanagan et al., 2000; McGrew & Flanagan, 1998), several WISC-R tests were combined with select WJ-R tests in a psychometrically and theoretically defensible manner. This "Wechsler-based CHC cross-battery" of tests is presented in Table 1.

---

2. The data analyzed in this study were derived from special validity studies conducted during the standardization of the WJ-R (Woodcock & Johnson, 1989). As such, the WISC-R is used in this research rather than the WISC-III. However, because the present research is theoretical in nature and because the same cognitive ability constructs underlie the two latest editions of the WISC (i.e., Verbal, Performance, Freedom from Distractibility, and Processing Speed), the fact that the WISC-R is featured in the present study, rather than the WISC-III, is of little relevance. It should be noted that Processing Speed is underrepresented on the WISC-R (i.e., it contains only one subtest that measures this ability: Coding); and therefore, this construct is not represented in Figure 2.

3. *Gs* is underrepresented on the WISC-R but is well represented on the WISC-III.

TABLE 1. A Wechsler-Based CHC Cross-Battery Example

Gf	Gc	Gv	Gsm	Glr	Ga	Gs
Analysis-Synthesis (RG)	**Information (K0)**	**Block Design (SR)**	**Digit Span (MS)**	Visual-Auditory Learning (MA)	Incomplete Words (PC:A)	**Coding (R9)**
Concept Formation (I)	Vocabulary (VL)	**Object Assembly (CS)**	Memory for Words (MS)	Memory for Names (MA)	Sound Blending (PC:S)	Cross Out (P)

Note. The test names in bold are from the WISC-R (the core battery). Tests names in regular type are from the WJ-R (the supplemental battery). All tests have been found to be strong measures of their respective cognitive abilities as defined empirically (McGrew, Werder, & Woodcock, 1991; Woodcock, 1990). RG = General Sequential Reasoning (Deductive); I = Inductive reasoning; K0 = General Information; VL = Lexical Knowledge; SR = Spatial Relations; CS = Closure Speed; MS = Memory Span; MA = Associative Memory; PC:A = Phonetic Coding (Analysis); PC:S = Phonetic Coding (Synthesis); R9 = Rate of Test Taking; P = Perceptual Speed. See Carroll (1993) or Flanagan et al. (2000) for definitions of these abilities.

307

Table 1 includes the strongest and most theoretically pure tests from the WISC-R (based on cross-battery analyses; viz., McGrew & Flanagan, 1998; Woodcock, 1990) and a set of strong measures from the WJ-R. The augmentation of the WISC-R with tests from the WJ-R serves to expand and strengthen the structural foundation of the WISC-R (in the same way that it would serve to strengthen all versions of the Wechsler Scales). To the extent that the WISC-R-based CHC cross-battery (depicted in Table 1) measures the constructs it was designed to measure (e.g., *Gf*, *Gc*, *Gv*, etc.), like the WJ-R complete battery, it should prove useful in measuring the specific cognitive abilities highlighted in the emerging *g*/specific abilities research (as predicted by Keith, 1997, 1999a; McGrew & Flanagan, 1998; and described above). To this end, the validity of inferences that can be drawn from Wechsler scores, within the context of the theory-driven CHC cross-battery approach, may improve substantially and the utility of the extant CHC external validity research may serve to inform and enrich interpretation.

RESEARCH QUESTIONS

The first part of this research was conducted to determine how much variance in reading achievement is explained by general and specific abilities underlying the atheoretical Wechsler VC-PO-FFD model, and the Wechsler-based CHC cross-battery model. An investigation of these research questions will assist in determining whether an assessment approach designed to improve the substantive and structural foundations of the Wechsler Scales (i.e., the Wechlser-based CHC cross-battery approach) explains more variance in reading achievement than an atheoretical VC-PO-FFD Wechsler model (i.e., WISC-R) and whether it provides a clearer understanding of the relationship between cognitive abilities and reading abilities.

The second part of this research was conducted to answer the following questions: (a) Does a specific intelligence battery that is based on the current CHC model (i.e., WJ-R) explain more variance in reading achievement than a traditional Wechsler battery (i.e., WISC-R)?; and (b) Is there a difference in variance explained in reading achievement between CHC designed assessment models (i.e., CHC cross-battery assessments, WJ-R) and a traditional model (Wechsler)? An investigation of these research questions will also assist in determining whether theory-based approaches (cross-battery) and batteries (WJ-R) explain more variance in reading achievement than atheoretical approaches (Wechsler Scales). It is important to recognize that the focus of this research is on the CHC theoretical constructs—constructs that are measured to varying degrees across all major intelligence batteries—and not on an individual intelligence battery. This research is an extension of Keith's (1999a) and McGrew et al.'s (1997) *g*/specific abilities research with the WJ-R norm data and represents the first attempt to provide validity evidence for Flanagan and McGrew's theory-driven CHC cross-battery approach. This research is theory-focused, rather than test-kit-focused.

METHODS

Participants

Participants were 166 students (72 males and 94 females) from two special validity studies included in the WJ-R Technical Manual (McGrew et al., 1991). Ninety-two participants were third-grade students from two schools in rural, north central Texas (Scarr, 1991) and the remaining 72 participants were third- and fourth-grade students from approximately 20 rural and urban school districts in Texas (McCullough & Wiebe, 1991). Of the total sample, 160 participants were Caucasian, 3 were African American, and 3 were Asian (Pacific Islander). All participants were administered the same select WJ-R Tests of Cognitive Ability (Extended Battery) and Achievement in addition to the complete WISC-R. An examination of the means and standard deviations presented in Table 2 demonstrates that the present sample performed within the average range of ability on all tasks.

Instruments

The WJ-R COG is appropriate for use with individuals ranging in age from 24 months to 95 or more years. It contains 21 tests of cognitive ability that are divided into standard and supplemental batteries. The standard battery contains seven tests, one measure for each of seven CHC factors (*Gf, Gc, Gv, Ga, Gs, Gsm, Glr*) assessed by the WJ-R COG. The supplemental battery contains 14 tests. Of these, the first seven (i.e., tests 8–14) provided complimentary measures of the previously mentioned cognitive factors and constitute the additional tests that are necessary to calculate the seven Broad Cognitive Ability Clusters. The remaining seven tests on the WJ-R COG supplemental battery (i.e., tests 15–21) provide mixed measures of CHC abilities and may be administered to derive additional information about an individual's cognitive strengths and weaknesses. Tests 1–14 were used in the present study. The WJ-R COG has strong psychometric properties including standardization, reliability, and validity (see WJ-R *Technical Manual*; McGrew et al., 1991) and good support for the use of its tests as markers for the eight CHC factors (e.g., McGrew et al., 1997; Woodcock, 1990; Ysseldyke, 1990).

The WISC-R is an individually administered test of intellectual functioning used with individuals ranging from 6 years, 0 months, 0 days to 16 years, 11 months, 30 days. It contains 12 subtests, two of which are optional (i.e., Digit Span and Mazes) and not included in the calculation of the IQ if the 10 major subtests have been administered. Overall, the WISC-R provides (a) three Deviation IQs—Verbal, Performance, and Full Scale—obtained by comparing the examinees' scores with those earned by a representative sample of age peers ($M = 100; SD = 15$); (b) standard scores for each of the 12 subtests ($M = 10; SD = 3$); and (c) three factor scores—Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility. The WISC-R's psychometric strengths include its excellent standardization, outstanding reliability, and adequate validity (Sattler, 1992). The WISC-R was succeeded by the WISC-III in 1991.

TABLE 2. Descriptive Statistics for WJ-R and WISC-R tests (*N* = 166)

Battery Test	Mean (# of Cases)	Range	*SD*
WJ-R Cognitive			
Memory for Names	100.02 (166)	47–143	19.93
Memory for Sentences	101.16 (166)	59–152	16.72
Visual Matching	101.56 (166)	67–153	15.48
Incomplete Words	100.06 (166)	64–143	15.10
Visual Closure	103.23 (166)	71–138	14.51
Picture Vocabulary	100.52 (166)	51–142	13.81
Analysis-Synthesis	103.43 (166)	74–141	13.56
BCA Standard Scale	100.84 (166)	52–139	14.46
Visual Auditory Learning	102.34 (166)	54–148	16.12
Memory for Words	101.14 (164)	65–139	14.30
Cross-Out	101.73 (164)	61–141	15.45
Sound Blending	105.79 (164)	64–150	15.59
Picture Recognition	106.46 (164)	66–148	18.04
Oral Vocabulary	103.32 (165)	63–138	14.14
Concept Formation	101.76 (165)	55–133	14.20
Long-Term Memory (*Glr*)	101.57 (164)	64–143	17.05
Short-Term Memory (*Gsm*)	101.26 (164)	60–150	15.37
Processing Speed (*Gs*)	101.74 (164)	61–146	15.52
Auditory Processing (*Ga*)	103.48 (164)	62–150	14.74
Visual Processing (*Gv*)	104.83 (164)	73–142	14.66
Comprehension Knowledge (*Gc*)	101.95 (165)	54–131	13.07
Fluid Reasoning (*Gf*)	102.06 (165)	63–133	12.96
Broad Cognitive Ability Extended Scale	102.09 (164)	48–139	14.52
WJ-R Achievement			
Letter-Word Identification	103.42 (166)	65–135	13.21
Passage Comprehension	107.57 (164)	60–145	14.21
Word Attack	103.17 (149)	60–151	16.53
WISC–R Subtests			
Information	10.70 (166)	1–17	2.69
Similarities	12.93 (166)	1–19	3.60
Arithmetic	9.80 (166)	1–18	2.99
Vocabulary	10.34 (166)	1–19	3.62
Comprehension	11.09 (166)	1–19	3.53
Digit Span	9.84 (155)	1–19	3.20
Picture Completion	11.12 (166)	4–18	2.98
Picture Arrangement	11.93 (166)	1–19	2.74
Block Design	10.61 (166)	1–19	3.32
Object Assembly	10.45 (166)	1–19	3.02
Coding	10.38 (166)	1–19	3.30
Mazes	11.04 (142)	2–18	2.94
WISC–R Scales			
Verbal IQ	105.84 (166)	47–147	16.87
Performance IQ	106.01 (166)	63–139	14.58
Full Scale IQ	106.64 (166)	50–144	15.85

Note. WJ–R = Woodcock–Johnson Tests of Cognitive Ability and Achievement–Revised; WISC–R = Wechsler Intelligence Scale for Children – Revised. The WJ–R and WISC–R scale test scores are on a scale with a mean of 100 and a *SD* of 15. The WISC–R subtest test scores are on a scale with a mean of 10 and a *SD* of 3.

Models

Traditional Wechsler/g Causal Reading Model (Wechsler/g Model). To determine how much variance in reading achievement is explained by a general ability factor underlying the atheoretical Wechsler VC-PO-FFD model, a WISC-R and Reading Achievement causal model was proposed (Fig. 2). This model will be referred to as the Wechsler model. A hierarchical confirmatory factor model of the WISC-R subtests is illustrated in the left half of Figure 2. The WISC-R tests are represented in rectangles and are called measured variables. These variables are assumed to measure the latent variables underlying the WISC-R, namely Verbal Comprehension (VC), Perceptual Organization (PO), and Freedom From Distractibility (FFD). The latent variables are represented in ellipses in Figure 2. The arrows pointing from the latent variables to the measured variables indicate that an individual's abilities (as represented by the latent variable) partially cause their scores on the measured variables (subtests). Measurement error and uniqueness, represented by the small circles labeled r1 through r11, constitute additional sources of variance that also influence test scores. Likewise, the latent variables, or first-order factors (VC, PO, FFD), are assumed to be partially caused by g, or general intelligence (a second-order factor). Other influences on the first-order factors are represented by the circles (labeled f1 through f3) pointing to each of these factors. From a measurement perspective, these circles represent error variance and that which is unique to the first-order factors (Keith, 1997, 1999a). The validity of the hierarchical framework presented in the left side of Figure 2, although supported in the within-battery Wechsler-based factor analysis literature, does not approximate the comprehensive multiple cognitive abilities structure of intelligence specified in contemporary theory and research. This model, therefore, represents a traditional, atheoretical structure of intelligence that is confined to the core Wechsler tests (past and present).

The right side of Figure 2 shows the structure of key WJ-R Reading Achievement tests. Similar to the research conducted by Keith (1999a) and McGrew et al. (1997), it is assumed that each test measured a corresponding achievement (e.g., Word Attack, etc.), and that each of these specific achievements was affected, in part, by general Reading Achievment.[4] The arrow pointing from g to Reading in Figure 2 demonstrates that g (or general ability) affects general reading achievement directly and affects narrower reading skills indirectly through its affect on general reading.

Traditional Wechsler/g + Specific Causal Reading Model (Wechsler/g + Specific Model). To determine how much variance in reading achievement is explained

---

4. The single indicator reading first-order factors were estimated by "fixing" the residual parameters to be equal to the reliable variance of each respective indicator. This was accomplished by multiplying the variance for each indicator by the value of 1 - r (where r equals the median reliability reported for the same ages in the WJ-R technical manual).
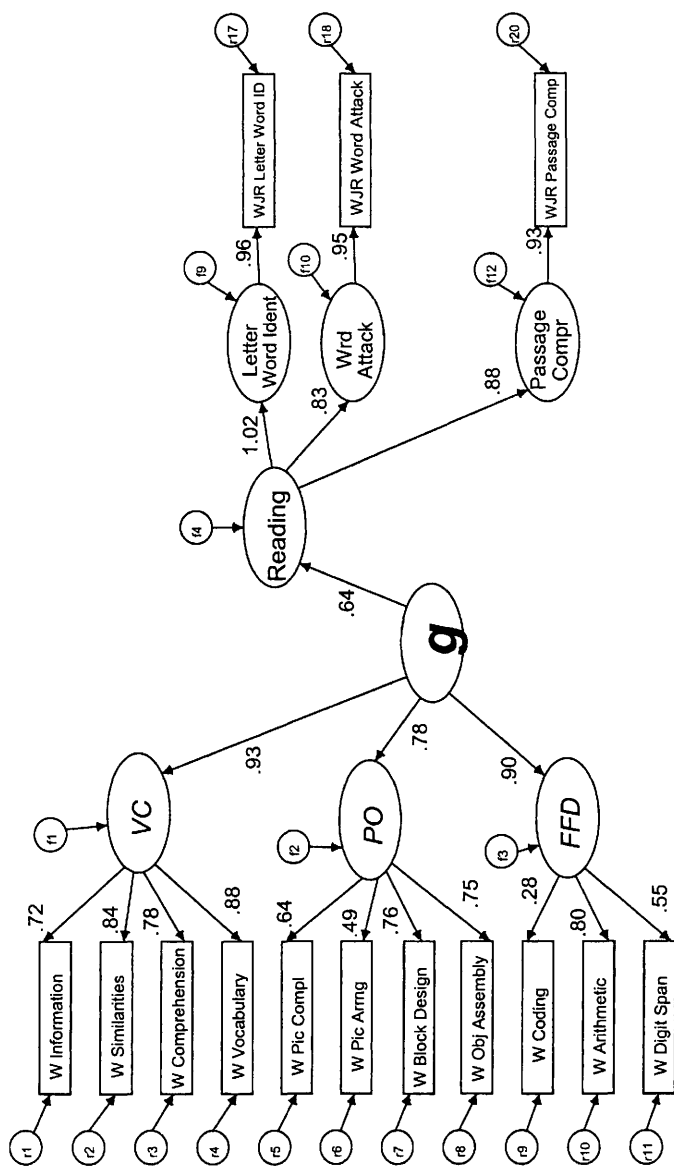
FIGURE 2. Traditional Wechsler/g Causal Reading Model (Wechsler/g model)

Note. $\chi^2 = 153.195$; df = 73; $p = .000$; AGFI = .821; CFI = .925; GFI = .876; PCFI = .742; PGFI = .609; RMSEA = .086.

by the specific ability factors underlying the atheoretical *Wechsler/g* model, the same causal model as presented in Figure 2 was proposed. However, this model allowed for the examination of paths between specific cognitive abilities and specific achievements. This model will be referred to as the *Wechsler/g + specific* model (Fig. 3). The purpose of this model is to test whether the popular specific Wechsler abilities (VC, PO, FFD) affect specific reading achievements in addition to the effects of *g* on Reading.

Wechsler-Based CHC Cross-Battery/g Causal Reading Model (Wechsler CB/g model). To determine how much variance in reading achievement is explained by the general ability factor underlying the Wechsler-based CHC cross-battery (CB) model, a WISC-R cross-battery and reading achievement causal model was proposed. This model will be referred to as the *Wechsler CB/g* model (Fig. 4). This model is similar to the model presented in Figure 2 except the WISC-R has been augmented via cross-battery principles and procedures in the manner described earlier (Flanagan et al., 2000). Whereas the *Wechsler/g* model (Fig. 2) included the more traditional, atheoretical specific abilities of VC, PO, and FFD, the model depicted in Figure 4 shows the *Wechsler-based CHC cross-battery*, which includes seven specific broad cognitive ability constructs (*Gf, Gv, Glr, Gs,* etc.)[5] that are consistent with strong theory.

Wechsler-Based CHC Cross-Battery/g + specific Causal Reading Model (Wechsler CB/g + Specific Model). To determine how much variance in reading achievement is explained by the specific ability factors underlying the Wechsler-based CHC cross-battery model, the same causal model as depicted in Figure 4 was proposed. However, this model (Fig. 5) allowed for the examination of paths between specific CHC abilities and specific achievements. This model will be referred to as the *Wechsler CB/g + specific* model. The purpose of the paths specified in this model is to test whether the empirically validated specific CHC ability constructs (i.e., *Gf, Gv, Glr, Gsm, Gv, Ga, Gs*) represented by Wechsler-based cross-battery data, affect specific reading achievements in addition to the effects of *g* on reading in a manner consistent with previous research.

WJ-R/g and WJ-R CB/g Models. To determine how much variance in reading achievement is explained by the general ability factor underlying the WJ-R CHC model, a WJ-R and reading achievement causal model was proposed (Fig. 6). This *WJ-R/g model* is similar to the model presented in Figure 2 in that it is confined to the specific tests within a single battery. Whereas the measurement model in Figure 2 represents the within-battery VC-PO-FFD structure of the WISC-R, the *WJ-R/g model* represents the within-battery CHC structure of the WJ-R. The *WJ-R/g model*

---

5. It is important to note that if the WISC-III battery was used in this study, rather than the WISC-R, *Gs* would be well represented in this model. That is, like *Gc* and *Gv, Gs* would not need to be augmented. However, the abilities of *Gf, Glr, Ga,* and *Gsm* are equally underrepresented or not measured on both the WISC-R and WISC-III.
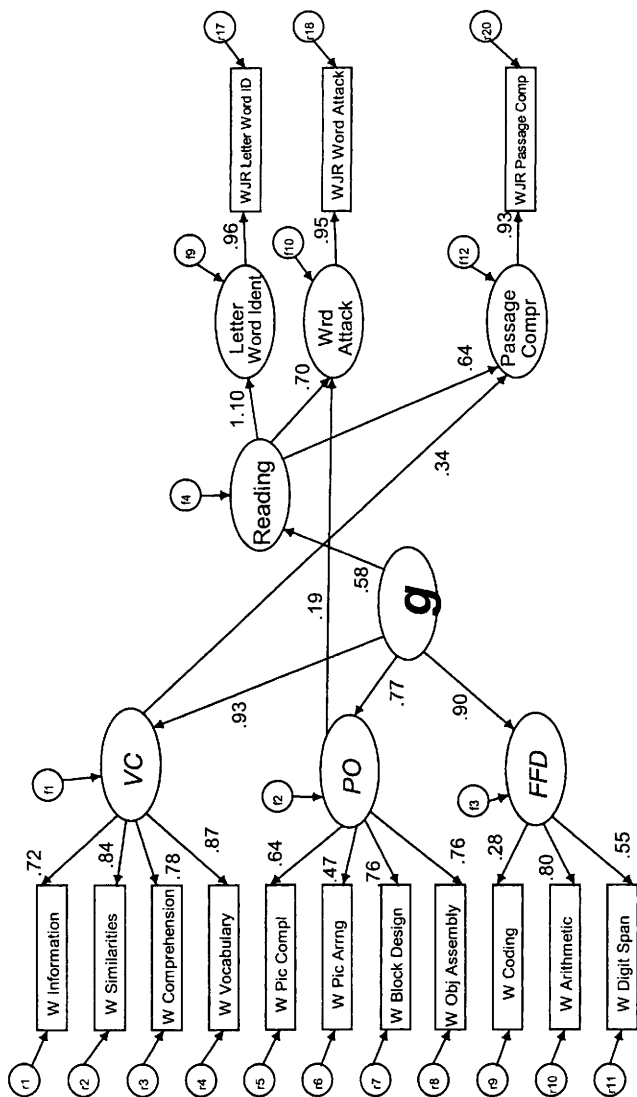
FIGURE 3. Traditional Wechsler/g + specific Causal Reading Model (Wechlser/g + specific model)

Note. $\chi^2 = 128.929$; df = 71; *p* = .000; AGFI = .840; CFI = .840; GFI = .946; GFI = .892; PCFI = .738; PGFI = .603; RMSEA = .074.
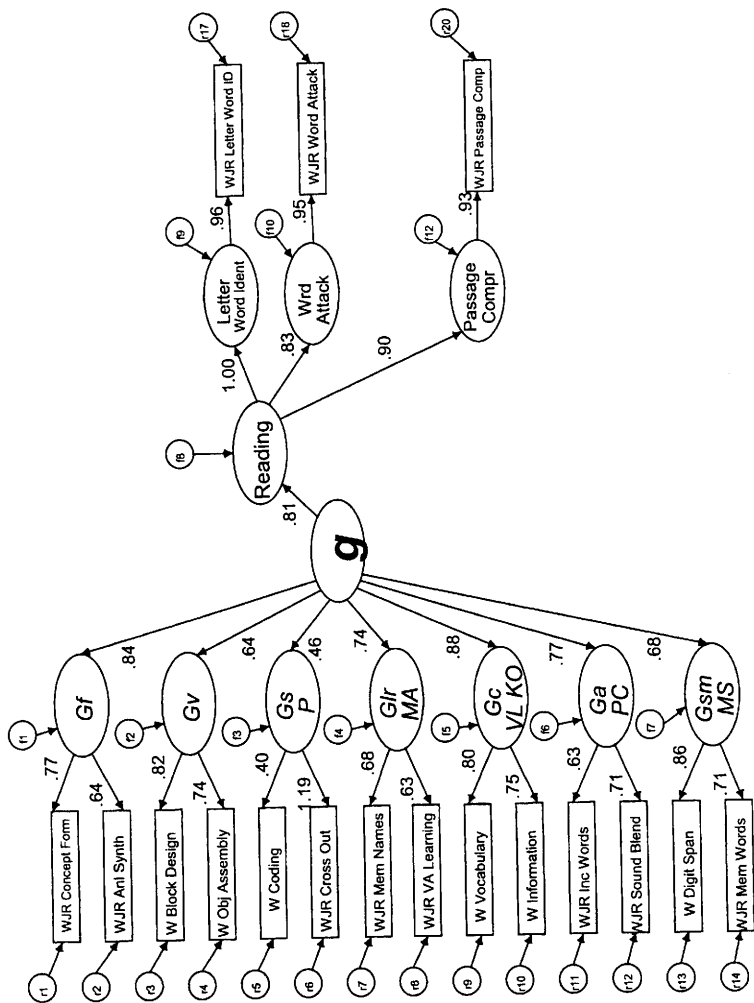
FIGURE 4. Wechsler-based CHC CB/g Causal Reading Model (Wechsler CB/g model)

Note. χ² = 164.657; df = 111; *p* = .001; AGFI = .844; CFI = .948; GFI = .887; PCFI = .774; PGFI = .643; RMSEA = .057.

315

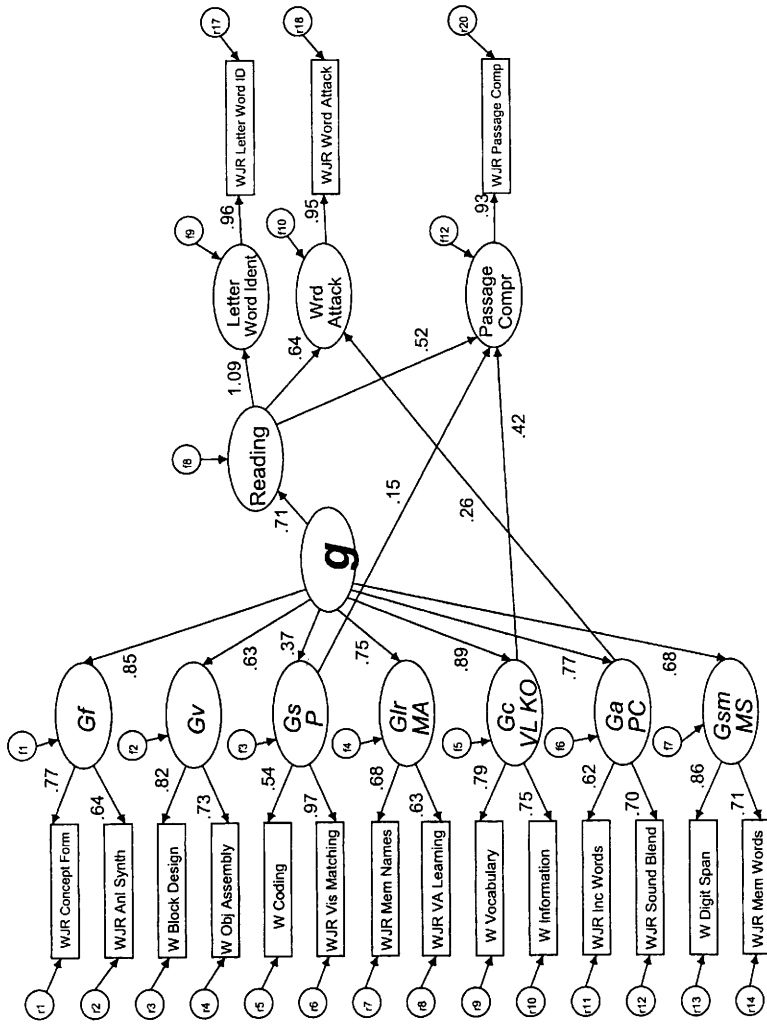FIGURE 5. Wechsler-based CHC CB/g + specific Causal Reading Model (Wechsler CB/g + specific model)

Note. χ^2 = 143.718; df = 108; p = .012; AGFI = .854; CFI = .965; GFI = .897; PCFI = .766; PGFI = .633; RMSEA = .047.
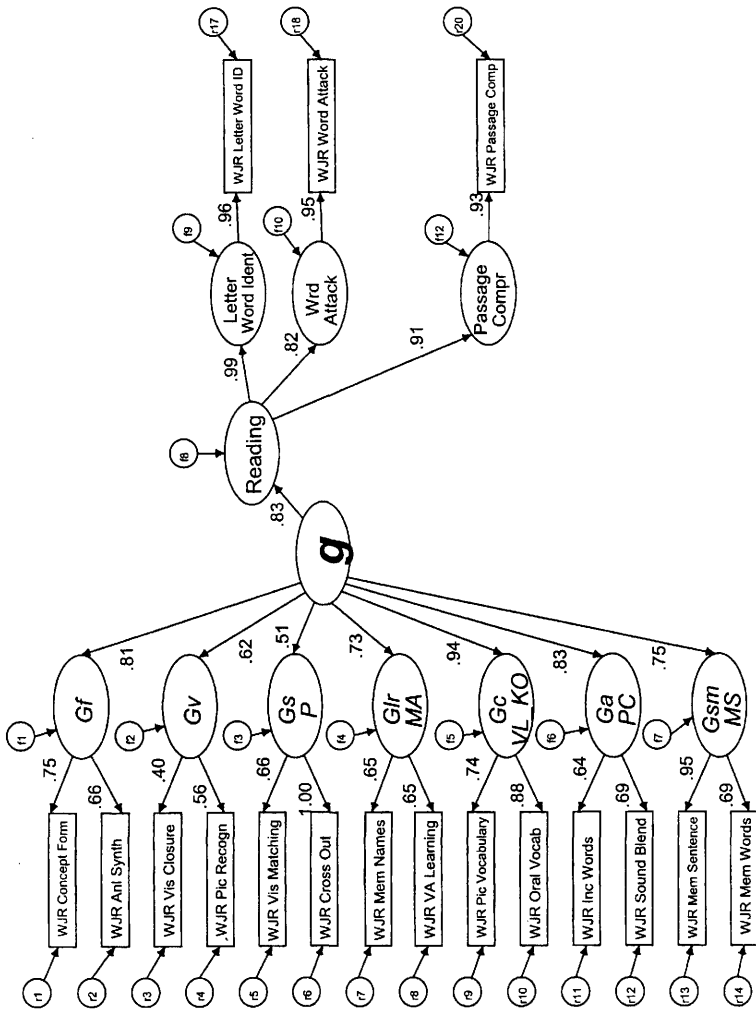
316

FIGURE 6. WJ-R/g model

Note. $\chi^2 = 175.912$; df = 111; $p = .000$; AGFI = .832; CFI = .941; GFI = .832; PCFI = .878; PCFI = .768; PGFI = .637; RMSEA = .063.
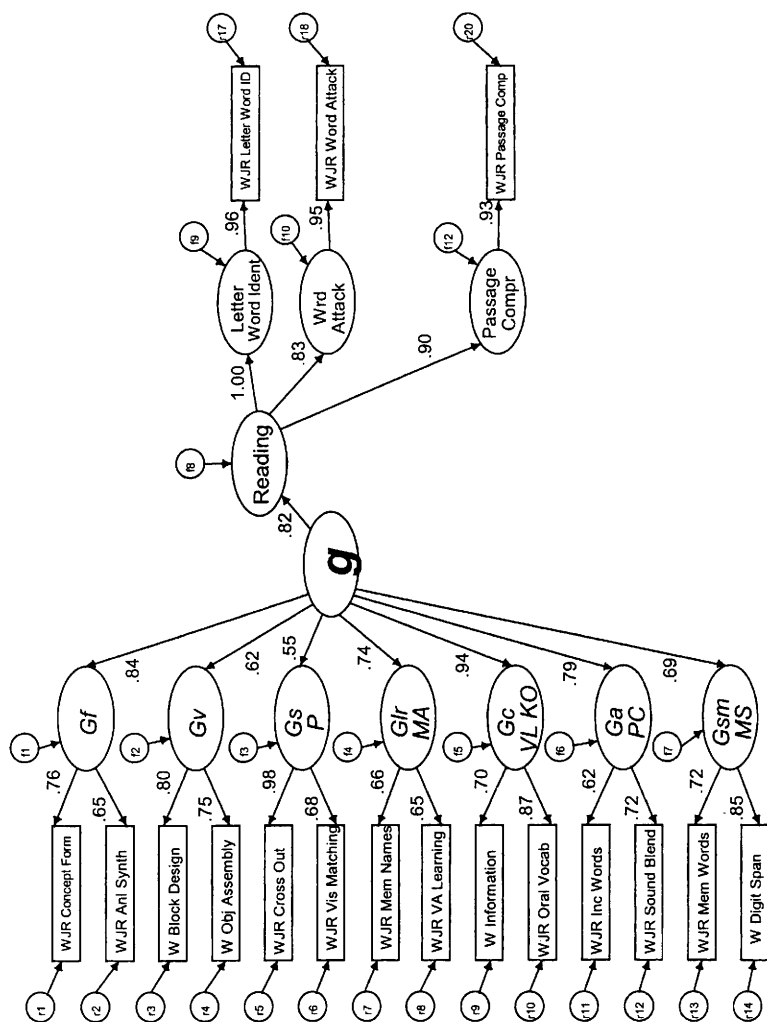
317

FIGURE 7. WJ-R CB/*g* model

Note. $\chi^2 = 163.808$; df = 111; $p = .000$; AGFI = .846; CFI = .953; GFI = .888; PCFI = .778; PGFI = .644; RMSEA = .057.

was proposed to investigate whether the empirically validated CHC theoretical model underlying the WJ-R explains more variance in reading achievement than the atheoretical *Wechsler/g* model.

The *WJ-R CB/g* model depicted in Figure 7 is similar to the model presented in Figure 4 because it represents CHC constructs using cross-battery data. Whereas the model in Figure 4 used the WJ-R to supplement the WISC-R in a manner consistent with contemporary CHC theory and research, the *WJ-R CB/g* model used the WISC-R to supplement the WJ-R to provide better and stronger construct representation of certain CHC constructs that underlie the WJ-R. Specifically, the Wechsler Block Design and Object Assembly tests provide a more robust *Gv* factor than the WJ-R Visual Closure and Picture Recognition tests (Flanagan & McGrew, 1998). Therefore, the Wechsler tests define the *Gv* factor in this model. Also, because the *Gc* construct underlying the WJ-R is primarily defined by tests of lexical knowledge, it is underrepresented on the WJ-R (McGrew & Flanagan, 1998). As such, the *WJ-R CB/g* model uses the Wechsler Information subtest with the WJ-R Oral Vocabulary test to provide better representation (through qualitatively different indicators) of the *Gc* construct. Finally, the WJ-R *Gsm* cluster includes a mixed measure of memory span (*Gsm*) and crystallized (*Gc*) abilities (i.e., Memory for Sentences). As such, the *Gsm* construct underlying the WJ-R includes variance (*Gc*) that is irrelevant to the interpreted construct (i.e., *Gsm*). A purer *Gsm* construct is represented in the *WJ-R CB/g* model by combining the WJ-R Memory for Words subtest with the Wechsler Digit Span subtest. This model was proposed to investigate whether a *WJ-R CB/g* model explains more variance in reading achievement than the previously proposed models.

Data Analysis

Maximum-likelihood estimation using the AMOS (Analysis of Moment Structures; Arbuckle, 1997) structural equation modeling program was used to examine the relationship between manifest cognitive variables used to define the VC, PO, FFD, *g*, CHC, and reading achievement latent factors and the causal paths between the VC, PO, FFD, *g*, CHC, and reading achievement latent factors. Maximum-likelihood estimation is considered the standard approach for estimating free parameters with structural equation modeling (SEM) and has been found to be very robust.

RESULTS

A review of the results in Figure 2 shows the typical pattern of factor loadings for the WISC-R tests on the respective VC, PO, and FFD latent factors (Kaufman, 1979) and the latent factors on the higher-order *g* factor (Keith & Witta, 1997). As expected, given the manner in which the single-indicator first-order reading factors were specified, the three respective WJ-R reading tests consistently loaded above

.90 on their respective factors. All three first-order reading factors displayed signif-icant ($p < .05$) [6] loadings on the general Reading factor, with the Word Attack (.83) and Passage Comprehension (.88) factors loading at similar levels. As reflected by the 1.02 loading (a "Heywood" case) for the Letter-Word Identification factor, per-formance on this test is isomorphic with the latent factor.

The causal path from g to Reading (.64) in Figure 2 is significant and indicates that the g factor, defined by the WISC-R VC, PO, and FFD factors, accounts for ap-proximately 41% of the general Reading variance. Finally, although the fit statis-tics reported in Figure 2 (and subsequently for Figs. 3–7) suggest a "good" model fit (see Keith, 1997, 1999b for discussion of fit indices), a comparison of the *rela-tive fit* of each of the models presented herein is *not* the important result of this re-search and will not be discussed in detail. These model fit statistics suggest that the Wechsler and WJ-R within-battery and cross-battery models are equally plausible representations of the underlying structure of the data. Instead, the presence or ab-sence of significant causal paths and the amount of reading achievement variance explained by the atheoretical (VC-PO-FFD) and theoretical (CHC) models are the relevant data for the research questions posed in this investigation.

The *Wechsler/g+specific* model (see Fig. 3) was derived from a review of the modification indices reported for the *Wechsler/g* model. Significant paths from the three specific WISC-R factors (VC, PO, and FFD) to the specific reading factors were added to the model in an iterative process until no other significant paths were suggested. It is important to note that this model "tweaking" process can capitalize on chance errors and the results need to be cross-validated in independent samples. The final model included (in addition to the significant .58 g/Reading causal path) significant VC/Passage Comprehension (.34) and PO/Word Attack (.19) paths.

Two Wechsler CHC cross-battery (i.e., *Wechsler CB/g* and *Wechsler CB/g + specific*) models (see Figs. 4 and 5, respectively) were specified in a similar man-ner. The parameter values for the reading measurement models for both models mirrored those previously discussed for the WISC-R atheoretical models (Figs. 2 and 3). Also, the magnitude and pattern of loadings for the *Wechsler CB/g* model and the CHC factor loadings on the higher-order g factor, were generally consistent with previous research reported by Bickley et al. (1995), McGrew et al. (1991), and McGrew et al. (1997).

As described previously, the *Wechsler CB/g* model (Fig. 4) specified seven first-order CHC factors that were subsumed by a second-order g factor. The g factor in this model was specified to have a causal path to general Reading, a path that was found to be high (.81) and significant. This .81 g/Reading causal path indicates that the g factor defined by seven first-order Wechsler-based CHC cross-battery latent factors accounted for approximately 66% of the Reading factor variance. This value reflects a 25% increase in the reading variance explained by the *Wechsler/g*

---

6. Significant structural paths in all models were those that had critical values (estimate/standard error) greater than 1.96 ($p < .05$).

model. Similar to the *Wechsler/g+specific* model (Fig. 3), the *Wechsler CB/g+specific* model (Fig. 5) was an extension of the *Wechsler CB/g* model (Fig. 4) in that the AMOS modification indices guided the addition of paths between the Wechsler-based cross-battery CHC factors and the specific reading factors (see Fig. 5). As seen in Figure 5, additional specific cognitive and achievement causal paths were found for *Ga*/Word Attack (.26), *Gc*/Passage Comprehension (.42), and *Gs*/Passage Comprehension (.15). With one exception (viz., *Gc*/Passage Comprehension), the specific cognitive/achievement causal paths retained in this model differed from those found for the *Wechsler/g+specific* model (Fig. 3). A significant (.19) PO/Word Attack path found in the atheoretical *Wechsler/g+specific* model was absent in the theory-based *Wechsler CB/g+specific* model. In contrast, significant *Ga*/Word Attack (.26) and *Gs*/Passage Comprehension (.15) paths were found in the theory-based *Wechsler CB/g+specific* model but not in the atheoretical *Wechsler/g+specific* model.

Finally, the CHC theoretical *WJ-R/g* model (Fig. 6) and *WJ-R CB/g* model (Fig. 7) showed nearly identical (.83 and .82, respectively) *g* to general Reading paths. These parameters indicate that the *g* factor defined by either the CHC factor structure underlying the WJ-R or the WJ-R CHC cross-battery factor structure accounted for essentially the same amount of Reading factor variance (i.e., 68% and 67%, respectively). The percent of variance accounted for by the *WJ-R/g* and *WJ-R CB/g* models was nearly identical to that reported for the *Wechsler CB/g* model (i.e., 66%). These percent of variance values (.66–.68) for the models that define *g* based on the seven-factor CHC model are consistently higher (by at least 25%) than that reported for a *g* factor defined by the Wechsler VC, PO, and FFD factors. These percentages of variance values are summarized in Figure 8.

DISCUSSION

In the traditional, atheoretical *Wechsler/g* model (Fig. 2), *g* accounted for approximately 41% of the variance in Reading achievement. In the *Wechsler/g + specific* model (Fig. 3), *g* also played a significant role in Reading, as demonstrated by its significant direct effect (.58) on this variable. In addition, VC and PO had large and moderate direct effects on Passage Comprehension (.34) and Word Attack (.19), respectively, above and beyond the variance accounted for by *g* (Keith, 1999b).[7] These findings suggest that Verbal and Perceptual Organization abilities are important to consider in addition to a Wechsler Full Scale score in the explanation of reading achievement.

Overall, these findings demonstrate that although the Wechsler factor scores are based on a weak substantive foundation and, thus, a weak structural validity model,

---

7. According to Keith (1999b), rules of thumb for interpreting effect sizes for manipulable influences on learning or achievement are as follows: less than .05 = not meaningful; .05 and above = small effect; above .10 or .15 = moderate effect; above .25 = large effect.

	Wechsler	WJ-R Cog	Difference
Published Battery	41%	68%	+ 27 %
CHC Cross-Battery	66%	67%	+ 01 %
Difference	+ 25%	- 01%	

FIGURE 8. Percent of reading variance explained by atheoretical (Wechsler) and theoretical (CHC) cognitive ability structural models.

positive external validity evidence exists for the WISC-R Full Scale, VC, and PO scores. Initial inspection of these results appears to argue against the central tenet of this article, namely, intelligence batteries must be grounded in a strong program of validation research.

However, the substantial improvement in external validity demonstrated by the theory-based *Wechsler CB/g* model (Fig. 4), provides strong support for the author's central tenet. Specifically, when the WISC-R was supplemented with strong CHC test indicators from the WJ-R, the structural path from g to Reading increased from .64 to .81, reflecting a 25% increase in prediction and explanation of the total reading variance. This increase in percent of variance explained is not trivial (McGrew et al., 1997) and is not merely a mathematical inevitability (e.g., as more indicators and paths are added to the structural model, the structural coefficient associated with the path from g to Reading will increase).[8] Rather, the WISC-R was supplemented in a systematic manner, following psychometrically and theoretically defensible CHC cross-battery procedures. The resultant CHC structural

8. When latent CHC factors (that were found not to predict reading achievement significantly) were dropped from the *Wechsler CB/g* model following the g/specific abilities research, the variance in reading achievement increased. For example, when *Glr, Gsm, Gv,* and *Gf* were dropped from the *Wechsler CB/g* model, the variance in reading achievement accounted for by a *Ga-Gc-Gs/g* model (comprised of three WISC-R and three WJ-R subtests) was 76%. These analyses can be obtained by contacting the author.

TABLE 3. Direct and Indirect Effects of Cognitive Abilities on Reading Skills

Paths	Wechsler/g+specific model		Wechsler CB/g+specific model	
	Direct Effect	Indirect Effect	Direct Effect	Indirect Effect
To Reading	—	—	—	—
From g	.58	—	.71	—
To Word Attack	—	—	—	—
From g	—	.41	—	.45
From Reading	.70	—	.64	—
From PO	.19	—	—	—
From Ga	—	—	.26	—
To Passage Comprehension	—	—	—	—
From g	—	.37	—	.37
From Reading	.64	—	.52	—
From VC	.34	—	—	—
From Gc	—	—	.42	—
From Gs	—	—	.15	—
To Letter or Word Identification	—	—	—	—
From g	—	.64	—	.77
From Reading	1.1	—	1.1	—

Note. According to Keith (1999b), the following are rules of thumb for interpreting effect sizes for manipulable influences on learning or achievement: less than .05 = not meaningful; .05 and above = small effect; above .10 or .15 = moderate effect; above .25 = large effect.

model underlying the Wechsler-based cross-battery in the present research has strong support in the extant factor-analytic cognitive abilities research (Flanagan et al., 2000). Moreover, the CHC test indicators in the cross-battery models presented here show significant paths to the first-order CHC factors, which show significant paths to the second-order g factor. The theory-based CB models in this research are justified from both theoretical and statistical standpoints.

The theory-based *Wechsler CB/g + specific* model (Fig. 5), like the *Wechsler/g + specific* model (Fig. 3), shows significant paths from cognitive constructs to specific reading abilities. However, the number and nature of the paths in the *Wechsler CB/g + specific* model differ from those found in the traditional *Wechsler/g + specific* model. Furthermore, the significant specific paths in the *Wechsler CB/g + specific* model are more consistent with the extant CHC reading research literature than those found in the traditional *Wechsler/g + specific* model. The direct and indirect effects of specific cognitive abilities on specific reading skills (from Figs. 3 and 5) are summarized in Table 3.

In both the atheoretical Wechsler causal model (Fig. 3) and theory-based Wechsler causal model (Fig. 5) the direct effects of Crystallized Intelligence (i.e., VC and Gc) on Passage Comprehension were large (i.e., .34 and .42, respectively). A review of Table 3 shows that the direct effect of Gc on Passage Comprehension in the *Wechsler CB/g + specific* model is *larger* than the indirect effect of g on Passage Comprehension, as mediated through its strong direct effect on general reading. In particular, this result suggests that reading comprehension could improve by almost one half of a *SD* for each *SD* increase in Gc ability, after the effect of g is al-

ready taken into account. This finding is consistent with the *Gc*/reading achievement research (McGrew et al., 1997; Vanderwood et al., 1999).

In contrast, the direct effect of PO on Word Attack (.19) in the atheoretical *Wechsler/g + specific* model is difficult to interpret in light of the extant research literature that provides little, if any, support for a relationship between Word Attack skills and a cognitive construct (PO) that is primarily *Gv* in nature. The lack of a significant relation between the valid and more theoretically pure *Gv* factor and reading achievement in the *Wechsler CB/g + specific* model (Fig. 5) is more consistent with the research literature. It is possible that the significant PO/Word Attack path found in the *Wechsler/g + specific* model (Fig. 3) was largely because of irrelevant *Gc* variance in the PO index associated with the Picture Completion and Picture Arrangement subtests (McGrew & Flanagan, 1998; Woodcock, 1990). This explanation highlights the fact that a poor theoretical-empirical domain match (or the lack of a substantive foundation to guide the construction of measures) appears to have confounded the Wechsler Scales.

The significant direct effects of *Ga* on Word Attack (.26) and *Gs* on Passage Comprehension (.15) in the *Wechsler CB/g + specific* model (Fig. 5) are consistent with existing research on the relations between specific cognitive abilities and reading achievement. The large effect size for *Ga*/Word Attack is particularly important because, although it is approximately half that of the indirect effect of *g* on Word Attack (see Table 3), it demonstrates that reading decoding skills can improve substantially with a one *SD* increase in *Ga* (or phonological awareness). Admittedly, phonological awareness is only one important factor that contributes to our understanding of reading acquisition. However, the general consensus of the literature is that more is known about the relations between phonological processing and reading achievement than relations between reading and nearly anything else (Wagner & Torgeson, 1987). Indeed, phonological awareness appears to be the core processing deficit in individuals with reading problems (see Morris et al., 1998). Because the research shows that specific interventions lead to significant improvement in phonological awareness *(Ga)*, this construct *(Ga)* seems important to include in assessments of individuals with reading difficulties. The *Ga*/Word Attack path was not found in the traditional *Wechsler/g + specific* model (Fig. 3), simply because none of the Wechsler Scales include direct measures of *Ga* or phonological awareness.

Unlike the situation with the *Ga* construct, the absence of a significant *Gs*/Passage Comprehension path in the atheoretical *Wechsler/g + specific* model, compared with the *Wechsler CB/g + specific* model, is likely because of the underrepresentation of the *Gs* construct on the WISC-R. *Gs* is represented by only one indicator on the WISC-R (i.e., Coding). Two or more qualitatively different indicators of each construct are necessary for appropriate construct representation in assessment, evaluation, and research (for an explanation, see McGrew & Flanagan, 1998; Messick, 1995). Due to the atheoretical combination of the Coding subtests with the Arithmetic and Digit Span subtests on the WISC-R, the potential contribu-

tion of *Gs* (as operationalized by the Coding subtest) to the prediction and explanation of reading was not realized in the traditional *Wechsler/g + specific* model (Fig. 3), despite the consistent and significant *Gs*/reading achievement finding in the literature. It is important to note that more current versions of the Wechsler Scales include adequate representation of the *Gs* construct (i.e., WISC-III, WAIS-III).

In summary, the present research found that the *g* factor underlying the *Wechlser CB/g* model accounted for substantially more variance in reading achievement (25%) than the *g* factor underlying the atheoretical *Wechsler/g* model. Similarly, the *g* factor underlying the *WJ-R/g* model and the *WJ-R CB/g* model also accounted for substantially more variance in reading achievement (approximately 26%–27%) than the traditional *Wechsler/g* model. More importantly, there was no significant difference in reading variance accounted for by *g* across the three CHC models (i.e., *Wechsler CB/g, WJ-R/g,* and *WJ-R CB/g*). This finding reflects the indifference of the indicator (Jensen, 1998) and implies that the important aspect of cognitive assessment is the organization of tests according to a well-validated theoretical model, or the selection of a battery that operationalizes the broad range of abilities specified in contemporary theory. The present research also demonstrated (as predicted by Keith [1997, 1999a] and McGrew and Flanagan [1998]) that when assessments are organized around the strong CHC theoretical model, specific cognitive abilities, including *Ga, Gc,* and *Gs* explained a significant portion of variance in reading achievement beyond that accounted for by *g*. This finding is consistent with the extant literature on the relations between cognitive abilities and reading achievement and the emerging *g*/specific abilities research. It is important to note that the size of the direct effects of *Gc, Ga,* and *Gs* on specific reading skills (summarized in Table 3) are consistent with (and many are larger than) those that have been judged to have enormous practical and scientific significance (i.e., effect sizes of .10–.20) in the medical and behavioral sciences (Gage, 1996). The present findings suggest that *Gc, Ga,* and *Gs* abilities may be particularly important to assess in addition to *g* or general intellectual functioning in young children referred for reading difficulties.

CONCLUSIONS

The extant factor-analytic cognitive abilities research shows that the preponderant two- (e.g., Verbal-Performance), three- (e.g., VC-PO-FFD), and four-factor (e.g., VC-PO-FFD-PS) Wechsler models do not provide a good explanation of the structure of intelligence when a wide range of cognitive abilities is factor analyzed (McGrew & Flanagan, 1998). Although the three- and four-factor models that underlie the most recent editions of the Wechslers (i.e., WPPSI-R and WISC-III/WAIS-III, respectively), have been found to provide a good fit of their respective standardization data (e.g., Keith & Witta, 1997; Wechsler, 1997), it is argued here that the WISC-R (and all editions of the Wechsler Scales) have weak substantive and structural foundations that significantly limit the validity of the infer-

ences that can be drawn from many of the Wechsler subtests and composite scores. The lack of a strong theory (or absence of a substantive foundation) to drive the construction of the Wechsler Scales questions the nature and meaning (i.e., the structural validity) of the constructs underlying the Wechsler Scales, despite the robustness of their respective within-battery factor structures and the invariance of these factor structures across a large age range. Likewise, a weak structural foundation questions the meaningfulness of the relations between test scores and external criteria (i.e., external validity). These latter two points were demonstrated through the evaluation of the nature and meaning of the Wechsler FFD and PO factors, in particular.

In an attempt to strengthen the validity of interpretations drawn from Wechsler test scores, the present study "repaired" the substantive and structural foundations of the Wechsler Scales (i.e., WISC-R) via Flanagan and McGrew's (1997) CHC cross-battery approach. This substantive and structural repair led to a clearer and more robust explanation of the relations between general and specific cognitive abilities and reading achievement. These findings were consistent with the emerging g/specific abilities research and provide preliminary validity support for the CHC cross-battery approach. The present findings demonstrated that the Wechsler-based CHC cross-battery principles and procedures were effective in operationalizing an assessment model that predicted and explained reading achievement in a manner consistent with the extant CHC external validity research. The application of the cross-battery approach to the WISC-R resulted in structurally valid CHC measures because these measures demonstrated the expected relations with the external measures included in this study (i.e., general and specific reading abilities). The results of this research provide preliminary validity support for the construction of cognitive assessments (specifically, augmenting the Wechslers with tests from the WJ-R) following Flanagan and McGrew's CHC cross-battery approach.

Study limitations suggest fruitful avenues for future research. First, although the restricted composition and size of the present sample may limit generalizability to other populations, knowledge of the relations between general and specific cognitive abilities and reading achievement in the current sample, based on the CHC cross-battery approach is a contribution to the relatively limited g/specific abilities and CHC cross-battery research. Because the current sample is limited demographically, caution in generalizing the results to non-Caucasian samples or similar samples at different developmental levels should be exercised until appropriate cross-validation studies are conducted. Second, the current study evaluated a Wechsler-based CHC cross-battery using the WJ-R as the supplemental instrument. To demonstrate fully the utility of the cross-battery approach, it is necessary to conduct similar research with different intelligence batteries using the cross-battery principles and procedures.

Overall, the CHC cross-battery approach appears to be a viable method for repairing the substantive and structural foundations of the Wechsler tests. The result is better-informed cognitive ability interpretations that are grounded in strong theory and research. Future research will determine the robustness of the present findings and their generalizability to other core intelligence batteries and populations.

REFERENCES

Arbuckle, J. (1997). *Amos user's guide 3.6.* Chicago: Small Waters.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10–22.

Benson, J., & Hagtver, K. (1996). The interplay between design, data analysis and theory in the measurement of coping. In M. Zeidner & N. Endler (Eds.), *Handbook of coping: Theory, research, applications* (pp. 83–106). New York: Wiley.

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

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

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429–452.

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.

Carroll, J. B. (1998). Foreword. In K. S. McGrew & D. P. Flanagan (Eds.), *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment* (pp. xi–xii). Boston: Allyn & Bacon.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56,* 754–761.

Cronbach, L. J. & Meehl, P. (1955). Construct validity of psychological tests. *Psychological Bulletin, 52,* 281–302.

Cronbach, L. (1971). Construct validation after thirty years. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy. Proceedings of a Symposium in Honor of Lloyd Humphreys* (pp. 147–167). Urbana, IL: University of Chicago Press.

Elliott, C. D. (1990). *Differential ability scales.* San Antonio, TX: The Psychological Corporation.

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 & Bacon.

Gage, N. L. (1996). Confronting counsels of despair for the behavioral sciences. *Educational Researcher, 25*(3), 5–22.

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

Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for white and black children from the standardization sample. *Journal of Educational Psychology, 73,* 227–231.

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.

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

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

Kamphaus, R. W., Petoskey, M. D., & Morgan, A. W. (1997). A history of test intelligence interpretation. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 32–51). New York: Guilford.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley.

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

Kaufman, A. S. (2000). Foreword. In D. P. Flanagan, K. S. McGrew, & S. O. Ortiz (Eds.), *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Boston: Allyn & Bacon.

Kaufman, A. S., & Kaufman, N. L. (1993). *The Kaufman Adolescent and Adult Intelligence Test.* Circle Pines, MN: American Guidance Service.

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.

Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the Cognitive Assessment System does and does not measure. *School Psychology Review, 28,* 303–321.

Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly, 12,* 89–107.

Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System (CAS): What does the CAS measure? *School Psychology Review, 28,* 117–144.

Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology, 42,* 332–242.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694.

McCullough, S., & Wiebe, M. (1991). Standardization data collected at Texas Woman's University. In R. W. Woodcock & N. Mather (Eds.), *Woodcock-Johnson tests of cognitive ability: Standard and supplemental batteries Examiner's Manual.* Allen, TX: DLM Teaching Resources.

McGrew, K. S. (1994). *Clinical interpretation of the Woodcock-Johnson Tests of Cognitive Ability–Revised.* Boston: 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. 151–180). New York: Guilford.

McGrew, K. S., & Flanagan, D. P. (1996). The Wechsler Performance Scale debate: Fluid intelligence (*Gf*) or visual processing (*Gv*)? *Communiqué, 24* (6), 14–16.

McGrew, K. S. & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

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

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock-Johnson Psycho-Educational Battery-Revised technical manual.* Chicago: Riverside.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education.

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,* 365–384.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

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

Nunnally, J. S. (1978). *Psychometric theories*. New York: McGraw Hill.

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

Sattler, J. (1992). *Assessment of children* (3rd ed.). San Diego, CA: Sattler.

Scarr (1991). Standardization data collected in north central Texas. In R. W. Woodcock & N. Mather (Eds.), *Woodcock-Johnson tests of cognitive ability: Standard and supplemental batteries Examiner's Manual.* Allen, TX: DLM Teaching Resources.

Shaw, S. E., Swerdlik, M. E., & Laurent, J. (1993). Review of the WISC-III [WISC-III Monograph]. *Journal of Psychoeducational Assessment,* 151–160.

Silverstein, A. B. (1982). Pattern analysis as simultaneous statistical inference. *Journal of Consulting and Clinical Psychology, 50,* 234–240.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49,* 479–502.

Sternberg, R. J. (1993). Rocky's back again: A review of the WISC-III. [WISC-III Monograph]. *Journal of Psychoeducational Assessment,* 161–164.

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

Vanderwood, M., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (1999). *Examination of the contribution of general and specific cognitive abilities to reading achievement.* Manuscript submitted for publication.

Wagner, R. K., & Torgeson, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised.* New York: The Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence —Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-Third Edition.* San Antonio, TX: The Psychological Corporation.

Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R third factor: It's more than distractibility. *Psychological Assessment, 2,* 91–97.

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

Woodcock, R. W. (1993). An information processing view of *Gf-Gc* theory. [Monograph Series: WJ-R Monograph]. *Journal of Psychoeducational Assessment, ,* 80–102.

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

Woodcock, R. W. & McGrew, K. S. (2000). *Woodcock-Johnson Psychoeducational Battery—Third Edition (WJ-3).* Chicago: Riverside.

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

Zimmerman, I. L., & Woo-Sam, J. M. (1985). Clinical applications. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 873–898). New York: Wiley.