Psychology in the Schools, Vol. 47(7), 2010
Published online in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/pits.20496

CATTELL–HORN–CARROLL ABILITIES AND COGNITIVE TESTS: WHAT WE'VE LEARNED FROM 20 YEARS OF RESEARCH

TIMOTHY Z. KEITH

The University of Texas at Austin

MATTHEW R. REYNOLDS

The University of Kansas

This article reviews factor-analytic research on individually administered intelligence tests from a Cattell–Horn–Carroll (CHC) perspective. Although most new and revised tests of intelligence are based, at least in part, on CHC theory, earlier versions generally were not. Our review suggests that whether or not they were based on CHC theory, the factors derived from both new and previous versions of most tests are well explained by the theory. Especially useful for understanding the theory and tests are cross-battery analyses using multiple measures from multiple instruments. There are issues that need further explanation, of course, about CHC theory and tests derived from that theory. We address a few of these issues including those related to comprehension–knowledge (*Gc*) and memory factors, as well as issues related to factor retention in factor analysis. © 2010 Wiley Periodicals, Inc.

It has been 20 years since the publication of the Woodcock-Johnson Psychoeducational Battery-Revised (WJ-R; Woodcock & Johnson, 1989), the first individually administered test of intelligence based explicitly on Cattell and Horn's extended *Gf-Gc* theory. In addition, it has been more than 15 years since the publication of Carroll's *Human Cognitive Abilities* (Carroll, 1993), a monumental study that both presented three-stratum theory and supported key aspects of *Gf-Gc* theory. As chronicled by McGrew (2005), many currently refer to the combination of these theories as Cattell–Horn–Carroll, or CHC, theory; we will do so here, although we will occasionally make the distinction between *Gf-Gc,* three-stratum, and CHC theory. This article will generally reference CHC abilities using abbreviations, without elaboration; more information about the CHC abilities can be found in the introduction to this special issue (Newton & McGrew, 2010).

Most new and revised individually administered tests of intelligence are either based on CHC theory or pay allegiance to the theory. Even the latest versions of the traditional, and traditionally atheoretical, Wechsler scales reference CHC theory in their manuals (Wechsler, 2003, 2008). Test users may legitimately wonder whether this adherence to CHC theory is well founded—that is, do these modern tests of intelligence in fact measure constructs consistent with CHC theory? Furthermore, how has the development of CHC-consistent tests informed the further development and refinement of CHC theory? The first purpose of this article will be to review research on current and previous versions of individually administered intelligence tests within the context of CHC theory. Our review will focus on whether factor analyses of these tests support predictions derived from CHC theory—whether those tests were developed based on the theory or based on some other (or no) theoretical orientation (including versions of these tests that were published prior to CHC theory). The second purpose of this article will be to discuss issues that still need to be addressed related to CHC theory and to suggest future directions for cognitive abilities research.

We are responsible for any errors and for the opinions expressed in this article.
Correspondence to: Timothy Z. Keith, 1 University Station, D5800, The University of Texas–Austin, 78712. E-mail: tim.keith@mail.utexas.edu

Validity of Cognitive Tests from a CHC Perspective

Woodcock–Johnson

The WJ-R was the first major, individually administered intelligence test based on *Gf-Gc* theory, and the Woodcock–Johnson III Tests of Cognitive Abilities (WJ-III) was the first major individual cognitive test based on CHC theory (Woodcock, McGrew, & Mather, 2001). It thus makes sense to begin this examination with research focused on the WJ-R and WJ-III.

Evidence supporting the structure of the WJ-R, and consistent with *Gf-Gc* theory, was reported in the WJ-R Technical Manual (McGrew, Werder, & Woodcock, 1990). The 16 WJ-R tests designed to assess *Glr, Gsm, Gs, Ga, Gv, Gc, Gf,* and *Gq* (Quantitative Knowledge) indeed showed substantial loadings on the "correct" factors, as well as a good model fit in an early application of confirmatory factor analysis (CFA). Similar results were reported at an invited conference on intelligence at Memphis State University and in a subsequent special issue of the *Journal of Psychoeducational Assessment* summarizing the conference (Woodcock, 1990).

Factor analysis of tests within a test battery is indeed critical for construct validation, yet an even more rigorous approach that serves to test both the measurement structure of a test and a specific theory is to factor analyze the tests from one battery with tests from other intelligence test batteries, even those developed with a different theoretical orientation. Carroll (among others) especially valued data sets that used more than a single test battery because different instruments drawn from different orientations offer a better opportunity to disconfirm each instrument's structure. For example, if two tests, drawn from distinct theoretical orientations, are jointly factored, and the results clearly mirror factors predicted from one theory but not the other, such findings offer important support for the first theory and evidence against the second theory. CFA of such data (referred to here as cross-battery CFA, or CB-CFA) can provide even stronger evidence by explicitly testing models drawn from one test/theory againts the other. Woodcock reported several such CB-CFAs of the WJ-R tests with tests from other instruments (Wechsler Intelligence scales, Kaufman Assessment Battery for Children (KABC), and Stanford–Binet, Fourth Edition). CHC-like factors emerged from this analysis, providing important evidence of both the validity of the WJ-R and the theory underlying the scale, although sample sizes were small. *Gf-Gc* (and CHC) theory and the WJ-R structure were likewise supported in joint analyses with the Kaufman Adolescent and Adult Intelligence Test (KAIT; Flanagan & McGrew, 1998), and the *Differential* Ability Scales (DAS) and Detroit Tests of Learning Aptitude-3 (McGhee, 1993).

At least two independent researches used the WJ-R standardization data to test key aspects of three-stratum theory. Bickley and colleagues conducted a higher-order multisample CFA of the WJ-R for ages 6 through 79 using a model based on three-stratum theory (Bickley, Keith, & Wolfle, 1995). Results supported the higher-order structure (with *g* at the apex), first-order factors consistent with three-stratum theory, and a consistent structure across the age span. The findings also supported possible intermediate factors between the broad abilities and *g*, suggested, but not tested, by Carroll (1993). Carroll conducted exploratory factor analyses (EFAs) and CFAs using WJ-R data to test competing hypotheses about the nature of a hierarchical general factor of intelligence (*g*), and found support for the (three-stratum) contention that *g* is hierarchical and separable from *Gf* (Carroll, 2003). Although these two studies were designed to test aspects of the three-stratum theory, they also supported the structure of the underlying test and, by extension, CHC theory.

Because the WJ-R was based on *Gf-Gc* theory, the structure was generally consistent with CHC theory. WJ-III was explicitly based on CHC theory (Woodcock et al., 2001). The structure of the WJ-III, and, indirectly, the underlying theory, were supported in CFAs that compared CHC models to models from other theoretical orientations (McGrew & Woodcock, 2001). In particular, the analyses supported a higher-order *g* factor and nine broad abilities, including seven cognitive (*Gf, Gc, Glr, Gv, Ga, Gs,* and *Gsm*) and two achievement (*Grw* and *Gq*) factors. Additional analyses also offered preliminary support for many of the narrow abilities subsumed under these nine broad abilities. Subsequent research using these data have suggested first- and second-order factorial invariance of the CHC-based structure (cognitive factors) across the 6- to 90-year age range (Taub & McGrew, 2004).

Not only have the CHC broad (and general) abilities embodied in the WJ-III stood up well when factor analyzed in isolation, they have also stood up well in CB factor analyses. In CB-CFAs of the WJ-III with the Wechsler Intelligence Scale for Children, Third Edition (WISC-III), factor models that combined the tests from the two batteries into CHC-like factors fit the data well, and the WJ-III tests generally loaded on both the broad and narrow factors that would be predicted by CHC theory. A CB-CFA of the WJ-III and the Cognitive Assessment System (CAS; Naglieri & Das, 1997) focused primarily on understanding the nature of the constructs measured by the CAS. Nevertheless, the coherence of the factors from the two instruments and the consistency of those factors with predictions from CHC theory (as opposed to the theory underlying the CAS) supported the WJ-III and CHC theory (Keith, Kranzler, & Flanagan, 2001). A CB-CFA with preschool children using the WJ-III and the DAS supported CHC theory, but showed a combination of *Gf* and *Gv* factors for these children and the complex cross-loading of a visual memory (MV) task on *Glr, Gsm,* and perhaps *Gv* (Tusing & Ford, 2004). A *Gf* factor also proved elusive in a recent CB-CFA of the WJ-III with the KABC, Second Edition (KABC-II) for four- and five-year-olds, with apparent *Gf* tests from both batteries loading on *Gv* and *Gc* factors (Hunt, 2007). This research illustrates a common finding in intelligence research for younger children, that is, that fewer factors (a simpler structure) are often found with younger versus older children. The simpler structures often noted for younger children may reflect actual differences in the structure of intelligence for younger children, or may simply be a result of it being more difficult to measure some abilities with younger children (cf. Carroll, 1993).

DAS

The DAS is another popular, individually administered intelligence test. Although the test was based on an eclectic mix of intelligence theory (including *Gf-Gc* theory), research with the DAS quickly supported a *Gf-Gc* (six broad abilities) interpretation of the test, at least for the school-age version (Elliott, 1990; Keith, 1990; Keith, Quirk, Schartzer, & Elliott, 1999). Although much of this research preceded CHC theory, its findings have been consistent with that theory. Core subtests provided strong measures of *Gc, Gf,* and *Gv,* and the diagnostic tests likely provided measures of *Gs, Glr,* and *Gsm*. Like many tests, less is known about the constructs measured by the DAS for preschool students. The age–differentiation conundrum mentioned in the previous section is obvious with the DAS: The test shows a simpler factor structure for younger children (Elliott, 1990; Keith et al., 1999), but it also appears to measure the same constructs for older versus younger children (Keith, 1990).

CB-CFAs have also supported a CHC-like structure of the DAS. CB analysis with the second edition of the WISC (the WISC-R) supported a DAS-type structure rather than a verbal-performance (WISC-R) structure (Stone, 1992, further analyzed in Keith, 2005). Likewise, a CB-CFA with the DAS and WJ-III supported a CHC-based model with subtests from both instruments loading on the *Gc, Glr, Gv, Gf, Gs,* and *Gsm* factors for school-age children (Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007). Several of the factors (e.g., *Gv*) showed lower loadings by some of the WJ-III subtests than would be expected, however, suggesting that further analysis might be fruitful. As noted earlier in text, CB-CFA with the WJ-III for preschoolers also supported a CHC model, although with a less differentiated structure, including a combined *Gf-Gv* factor (Tusing & Ford, 2004). Such analyses support the validity of both the scales involved and CHC theory.

The second edition of the DAS was published in 2007 (Elliott, 2007). Although not explicitly based on CHC theory, the DAS-II manual discussed the consistency of the test with CHC theory in considerable detail. In addition, the manual reported a series of CHC-consistent CFAs of the new instrument as well as a table listing the narrow and broad CHC abilities associated with each subtest (Elliott, 2007). One independent CFA of the DAS-II demonstrated remarkable consistency of the DAS-II with CHC theory, and suggested that the test indeed measures the same (CHC) constructs across the 4- to 17-year age level (Keith, Low, Reynolds, Patel, & Ridley, 2010).

Kaufman Scales

The KABC was designed to assess the constructs of simultaneous and sequential mental processing derived from the Luria–Das theory of mental processing, along with academic achievement (Kaufman & Kaufman, 1983). Analyses of the KABC suggested the possible alternative that the test measured nonverbal reasoning and verbal memory skills, with the achievement scale measuring a mix of verbal reasoning and reading achievement (e.g., Keith & Novak, 1987; Keith, 1986; Keith, 1985; Keith & Dunbar, 1984). CHC terminology would likely refer to these constructs as *Gv, Gsm, Gc,* and reading achievement (cf. Kaufman, 1993). Indeed, several CB EFAs of the KABC with the WISC-R suggested that the KABC Simultaneous tests loaded with the WISC-R performance tests (*Gv,* from a CHC perspective), the Achievement tests (sometimes excluding reading) with the WISC-R Verbal tests (*Gc*), and the KABC Sequential tests with Digit Span (*Gsm*) (Kaufman & McLean, 1986, 1987; Keith, 1997; Keith & Novak, 1987).

Several CB exploratory and confirmatory factor studies referenced *Gf-Gc*-like constructs, but there is no research (to our knowledge) explicitly testing a *Gf-Gc* or CHC model with the KABC. Because the KABC was designed from a different theoretical orientation than CHC theory, such a model should be especially interesting because it would allow the testing of different and competing theoretical models. Therefore, we tested a CHC model of the KABC and WISC-R, fit to the data reported in Keith and Novak (1987); the results are shown in Figure 1. As can be seen in the figure, the model fit the data fairly well using standard criteria. Of more direct interest, the model fit considerably better than did a model specifying separate KABC and WISC-R factors. (That model was presented in Keith, 1997.) The findings thus suggest that the two tests measure overlapping constructs and that those constructs are explainable from a CHC orientation (they may also be explainable from a Luria orientation). Given the lack of enough clear measures, it was not possible to specify a separate *Gf* factor (Matrix Analogies) or a separate *Gs* factor (Coding). Interestingly, when a higher-order *g* factor was added to the model, *Gq,* referenced by the two Arithmetic tests, had the highest loading on that factor.

The KABC-II (Kaufman & Kaufman, 2004) assesses a considerably broader range of cognitive abilities than did the original; it is still interpretable from a Luria perspective as measuring simultaneous and sequential mental processing, planning, and learning. The KABC-II is also interpretable from CHC theory, however, and from a CHC perspective it is designed to assess *Gc, Gv* (equivalent to Simultaneous), *Gf* (Planning), *Glr* (Learning) and *Gsm* (Sequential). This structure was supported in CFAs reported in the KABC-II technical manual (Kaufman & Kaufman, 2004) and in at least one independent higher-order CFA for school-age children (Reynolds, Keith, Fine, Fisher, & Low, 2007). The analyses by Reynolds and colleagues supported the KABC-II as measuring the broad abilities of *Gc, Gv, Gf, Glr,* and *Gsm* as well as a higher-order *g* for ages 6 through 18. Although the research did not test the factor structure below age 6 (because different tests are administered for younger children), confirmatory comparison of covariance matrices suggested that the test indeed measures

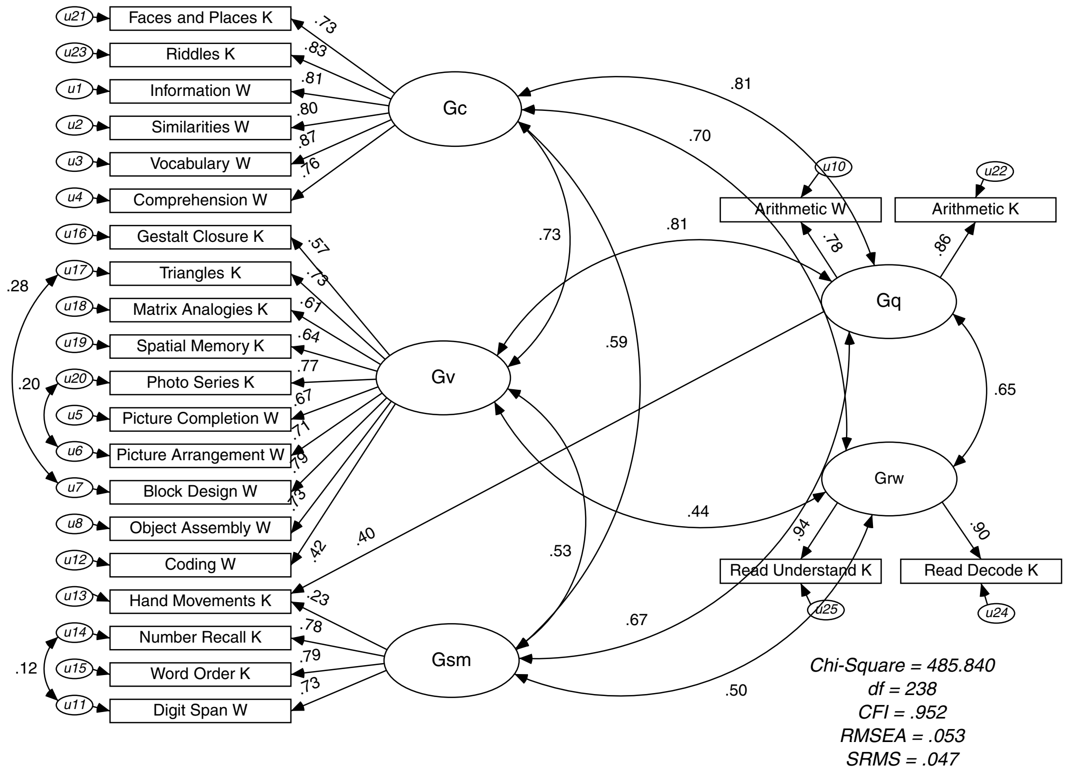

FIGURE 1. CHC CB-CFA of the KABC and the WISC-R. *df:* degrees of freedom; CFI: comparative fit index; RMSEA: root mean square error of approximation; SRMS: standardized root mean square residual.

these same constructs in children as young as 3. A recent CFA of the KABC-II for preschool children (ages 4–5) likewise supported a CHC-type structure, although with *Gv* and *Gf* tests loading on a single factor (a structure that matches the test's actual structure; Morgan, Rothlisberg, McIntosh, & Hunt, 2009).

The KAIT was designed to measure *Gf* and *Gc* (Kaufman & Kaufman, 1993), and also included delayed measures of one *Gc* and one *Gf* test, likely reflecting *Glr*. Factor analyses presented in the manual and elsewhere generally supported the structure of the test as measuring *Gc, Gf,* and delayed recall (Kaufman & Kaufman, 1993; Keith, 2005). CB analyses suggest that the KAIT measures visual memory (a narrow ability under *Gv*) and memory span (under *Gsm*) in addition to *Gf, Gc,* and *Glr* (specifically, associative memory; McGrew & Flanagan, 1998). CB EFA with the KABC suggested factors resembling *Gc, Gv* (Simultaneous), *Gsm* (Sequential), *Gf,* MV, and achievement (Kaufman, 1993). In our opinion, the KAIT was a novel, well-developed instrument; we hope that it is slated for revision incorporating advances in CHC theory.

Stanford–Binet

The Stanford–Binet has likely changed more than any other intelligence test in its recent iterations, going from its classic format as an age scale measuring primarily *g* in Form L-M, to a point scale in its fourth edition based, in part, on *Gf-Gc* theory (Thorndike, Hagen, & Sattler, 1986), back to a modified age scale based largely on CHC theory in the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003; Roid & Pomplun, 2005).

CFAs of the SB4 were generally supportive of the *Gf-Gc*-derived structure of the instrument and suggested that the test (in CHC terminology) measured *Gc, Gf/Gv,* RQ (Quantitative Reasoning), and *Gsm,* although it was difficult to separate the *Gsm* factor from other factors at the youngest age level (ages 2–6; Keith, Cool, Novak, White, & Pottebaum, 1988). Hierarchical analyses did not support the test-derived prediction that the RQ and Verbal factors reflected a higher-order *Gc* factor, but instead suggested a residual relation between the *Gf/Gv* factor and the RQ factor. This residual relation is consistent with the CHC notion that RQ is best conceived as a narrow ability subsumed by *Gf.*

CFAs reported in the technical manual generally supported the five-factor theoretical structure of the SB5, with factors representing *Gf, Gc,* RQ, *Gv,* and *Gsm*/MW (Roid, 2003). Although none of the models tested appeared to fit well, the five-factor CHC model fit better than did other models with fewer factors. In contrast, two independent EFA analyses with the standardization data supported only a two- or one-factor interpretation (Cavinez, 2008; DiStefano & Dombrowski, 2006). DiStefano and Dombrowski also reported CFA results supporting a single factor for all age groups. We are reluctant to accept this interpretation without additional information and analysis, however. For ages 2 through 5, the five-factor solution actually showed the best fit in a CFA using common criteria, although the researchers argued for a one-factor solution based, in part, on high factor correlations. For other age groups, four- and five-factor models "were not able to be evaluated due to estimation problems" (p. 131). Additional analyses and more detailed explanations of results are needed. We would simply note that no one has so far estimated the true theoretical structure of the SB5, a structure that includes both CHC factors and method-like modality factors representing verbal and nonverbal presentation. Such a model would be complex, but could and should be tested via CFA.

The results of a CB-CFA with five subtests from the WJ-III are reported in the SB5 technical manual (Roid, 2003). The five-factor CHC-derived model, with each factor including two SB5 tests and one WJ-III test, fit statistically significantly better than did a one-factor or a two-factor model.

Wechsler Scales

As noted earlier, even the recently revised Wechsler scales have become more consistent with CHC theory, although they are not explicitly based on CHC theory. Here we will concentrate primarily on the WISC series, with only a brief review of the WISC-III, and a more detailed examination of the WISC-IV (Wechsler, 2003).

The WISC-III acknowledged in its structure, for the first time, that the test measured something beyond Verbal and Nonverbal (Performance) abilities (Wechsler, 1991). A new test (Symbol Search) was added to improve measurement of the supposed "Freedom from Distractibility" factor, but instead resulted in the bifurcation of that factor. Thus the WISC-III was published with a structure reflecting the constructs of Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed. From a CHC perspective, Verbal Comprehension and Processing Speed represented straightforward *Gc* and *Gs* factors, and Perceptual Organization a likely *Gv* factor, with perhaps a touch of *Gf.* The Freedom from Distractibility factor remained a cipher, as it was in previous versions of the scale. Researchers debated the underlying construct measured by the scale. Working memory and quantitative reasoning were argued as likely candidates; writers were unified only on the argument that this factor did not represent freedom from distractibility.

Keith and Witta (1997) tested the structure of the WISC-III from a three-stratum (CHC) perspective using WISC-III standardization data. They argued that the third factor should be interpreted as a quantitative reasoning factor (RQ, from a three-stratum perspective) because the factor loaded most highly on the second-order *g*, and Arithmetic (rather than Digit Span) had the highest loading on the factor (Keith & Witta, 1997). Also tested was a model, derived from the three-stratum theory, with this factor split into numerical facility (Arithmetic) and memory span (Digit Span) factors. The model did not fit as well as the four-factor model, but then the WISC-III did not include enough subtests to allow an effective test of this model (Carroll, 1997; Keith & Witta, 1997).

Perhaps the most illuminating investigation into the constructs measured by the WISC-III was its CB-CFA with the WJ-III (Phelps, McGrew, Knopik, & Ford, 2005). Most WISC tests performed as expected, and suggested that the Verbal Comprehension, Perceptual Organization, and Processing Speed indices should be considered as measuring the constructs *Gc, Gv,* and *Gs,* respectively. Interestingly, in this research Arithmetic was shown to measure *Gq* (Quantitative Knowledge) and *Gs,* rather than RQ or *Gsm* (MW) (although it is not clear if this final possibility was tested).

The WISC-IV was designed better to incorporate theory (including CHC theory) and research into the classic scale (Wechsler, 2003). In particular, the developers of the WISC-IV sought to add measures of *Gf* to the revised instrument. In addition, for the scale to clearly and cleanly measure *Gsm*/MW and *Gs,* additional measures of MW and *Gs* were included, and the memory demand of Arithmetic was increased. Examination of this latest version of the scale has generally supported its structure with a model mirroring the higher-order structure of the test (four first-order and one second-order factors) showing a good fit to the standardization data (Keith, Fine, Reynolds, Taub, & Kranzler, 2006). Exploratory analyses of the standardization (Watkins, 2006) and other data (Watkins, Wilson, Kotz, Carbone, & Babula, 2006) also supported a four-factor structure mirroring the test's actual structure. From a CHC perspective, this structure can be interpreted as reflecting *Gc* (Verbal Comprehension), *Gsm* or MW (Working Memory), *Gs* (Processing Speed), and a combination of *Gf* and *Gv* (Perceptual Reasoning). A curious finding of the CFA research using a WISC-IV four-factor model, however, was that Arithmetic had the highest loading on the *Gsm* factor, and the *Gsm* factor had the highest loading on *g* (Keith et al., 2006). These findings may suggest that the *Gsm* factor (and the corresponding index) measure something more cognitively complex than *Gsm,* or, alternatively, that this factor is, in fact, an MW, rather than a *Gsm,* factor.

Few researchers have tested the WISC-IV against a stricter CHC model, in part because such a model requires the administration of most of the supplemental tests. Keith and colleagues (2006) compared a WISC-IV-type "scoring" model against a CHC model with the perceptual reasoning subtests split into *Gf* (Matrix Reasoning and Picture Concepts) and *Gv* (Block Design and Picture Completion) factors, with Arithmetic also on the *Gf* factor (as a measure of RQ). This model fit better than did the WISC-IV model, thus supporting the test as a measure of the CHC abilities and also, secondarily, supporting CHC theory. Additional models suggested that the Arithmetic subtest may measure a complex mix of abilities, however, including *Gf* (RQ), *Gsm,* and possibly *Gc*. Recent research with the Taiwanese version of the WISC-IV supported a CHC-derived structure of the test, but also supported a WISC-IV–derived structure (Chen, Keith, Chen, & Chang, 2009). These analyses also suggested that Arithmetic is primarily an excellent measure of *g*. We wonder whether the relative factorial complexity of Arithmetic (and thus its requiring the use of several different abilities) may make it a surrogate for *g*.We also wonder whether different children may use different abilities to solve these problems, meaning that the loadings of Arithmetic would represent averaged rather than fixed values (Wolfle, 1940).

CAS

The CAS (Naglieri & Das, 1997) was designed to assess the constructs of Planning, Attention, and Simultaneous and Successive (PASS) processing derived from the Luria-derived PASS theory of human information processing and cognition. The CAS represented a theoretically derived test that was developed from a theory of human intelligence different than CHC theory, and therefore could provide a crucial test of the validity for CHC theory. Although the test's authors provided evidence in the manual and elsewhere that the CAS in fact measured PASS abilities (Naglieri, 1999, 2005), other researchers provided evidence that the test instead measured constructs that were quite consistent with CHC theory, and not PASS (Carroll, 1995; Keith & Kranzler, 1999; Kranzler & Keith, 1999; Kranzler, Keith, & Flanagan, 2000; Kranzler & Weng, 1995). This research suggested that the CAS Planning and Attention scales measure narrow abilities subsumed under *Gs* (perceptual speed and rate of test taking), that the Simultaneous processing scale measures *Gv* and *Gf,* and that the Successive processing scale measures memory span, a narrow *Gsm* ability. All such CFAs, however, were based on a single instrument, and thus did not allow strong tests of competing hypotheses of the skills measured by the instrument.

A CB-CFA of the CAS with the WJ-III did provide that opportunity, however (Keith et al., 2001). The research compared factors from the CAS with those from the WJ-III for a sample of 155 students to whom both tests were administered. The research showed that the Planning and Attention factors from the CAS were statistically indistinguishable from the WJ-III *Gs* factor and the Successive factor was indistinguishable from the WJ-III *Gsm* factor. The CAS Simultaneous factor showed high correlations with the WJ-III *Gf* (.77) and *Gv* (.68) factors, with subsequent analysis supporting an integrated *Gv* factor including all of the CAS Simultaneous tests on the same factor with the WJ-III *Gv* tests. Last, this research also supported the third stratum of three-stratum/CHC theory. Although PASS theory and the CAS disavow a *g*-type factor, the higher-order *g* factor from the CAS in fact correlated .98 with the WJ-III *g* factor, and the two were statistically indistinguishable (Keith et al., 2001).

Reynolds Intellectual Assessment Scales

The Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003) is a relatively new intelligence test, and CHC theory was used as a theoretical guide in its development. Four subtests make up the core scale, which was developed to provide a measure of *g* along with verbal and nonverbal intelligence (corresponding to crystallized and fluid intelligence, respectively). A memory composite index was also included. This index includes measures of verbal and nonverbal memory, which may measure aspects of *Gsm, Glr,* and MV (*Gv*). A CHC interpretation would suggest that the RIAS measures *g*, *Gf, Gv, Gc, Gsm,* and *Glr*.

Because the RIAS is a relatively new test, independent research on it is limited. EFA was used to examine the structure of the RIAS in a referred sample of children (Nelson, Cavinez, Lindstron, & Hatt, 2007). The extraction criteria used in that study suggested only one factor. In contrast, CFA results of the RIAS norming sample, a sample of referred children, and a reanalysis of the Nelson and colleagues (2007) data suggested that the RIAS likely measures Nonverbal (*Gf*) and Verbal (*Gc*) factors (Beaujean, McGlaughlin, & Margulies, 2009). The use of CFA procedures allowed for direct comparison of a one-factor to a two-factor model in the three samples. The two-factor model provided the best fit in each. An attempt at a three-factor model did not converge, but a two-factor solution in which the nonverbal memory test loaded on the Nonverbal factor and the verbal memory test loaded on the Verbal factor fit quite well.

CB-CFAs with other intelligence batteries would help to clarify the nature of the constructs measured by the RIAS. Future research should include these subtests in CB-CFAs to clarify what is measured by these memory subtests as well as by the subtests included in the Nonverbal and Verbal composites within a CHC framework.

SUMMARY: CHC THEORY AND RESEARCH

We believe that CHC theory offers the best current description of the structure of human intelligence. It is by no means perfect or settled (and we will discuss some of the unsettled issues below), but it functions well as a working theory, as a "Rosetta Stone" (McGrew, 2005, p. 147) or "periodic table" (Horn, 1998, p. 58) for understanding and classifying cognitive abilities, and as a guide for new test development. We conclude that it is quite clear that research on recent, individually administered tests of intelligence supports a CHC-like set of constructs as underlying those tests, and this conclusion holds whether the test was based on CHC theory, an alternative theory, or an eclectic mix of theories. There are caveats to this conclusion, however, and additional issues that need to be considered. A few of these will be briefly addressed now.

A Potpourri of Remaining Issues

CHC Theory

Much progress has been made in understanding human cognitive abilities, yet there is much work to be done. Next we discuss a few of the unanswered questions related to CHC theory.

Gc. Questions related to the nature of *Gc* were recently pointed out by the late John Horn (Horn & Blankson, 2005; Horn & McArdle, 2007). We discuss some of his concerns and our own observations. *Gc* remains an elusive construct, and researchers often talk past each other when discussing *Gc,* with it being referred to as crystallized intelligence, academic achievement, verbal ability, or comprehension/knowledge, to name a few. Meanwhile, the measurement of *Gc* on individual intelligence tests has seemingly become more narrow. *Gc* is often referred to as crystallized intelligence, derived from Cattell's investment theory according to which crystallized intelligence is determined by the investment of historical *Gf* into learning via experience and schooling (Cattell, 1987; Cattell & Bernard, 1971). In school-age children, this simple view of investment theory, where *Gf* is invested in *Gc,* is yet to be supported empirically (Ferrer & McArdle, 2004). In the psychological literature, *Gc* is often indicated by achievement tests, or the term *Gc* is used interchangeably with achievement. It been hypothesized that *Gc* mediates the influence of historical *Gf* on academic achievement, but *Gc* is not considered to be equivalent to academic achievement (Cattell, 1987). Although *Gc* is often indicated by tests requiring verbal answers, this verbal component has been reduced greatly in recent versions of intelligence tests (e.g., the KABC-II), and many other non-*Gc* tests also require verbal responses, verbal directions, and verbal mediation. Clarification about the nature of *Gc* versus verbal ability and achievement would be useful.

In current CHC theory, *Gc* is referred to as comprehension/knowledge. Measurements of *Gc,* however, rarely measure a "depth" of knowledge or an understanding in a domain. Yet, the term comprehension seems to imply this type of understanding. Observers have suggested that someone who scores well on current measures of *Gc* is likely to be "a person flitting over many areas of knowledge" rather than someone who has a true understanding of a domain (Horn & McArdle, 2007, p. 239). Someone with high *Gc* thus could be described as someone who knows a lot, but nothing well, meaning that current indicators of *Gc* cannot differentiate someone who does know something well from someone who does not. Depth of knowledge is rarely measured on current tests of intelligence, even though this "expertise" would seem to be crucial to someone who is considered to have a great deal of comprehension/knowledge (see Horn & Blankson, 2005). Horn and McArdle (2007) suggested that *Gc* represents a *capacity* to retain information in accessible storage. Perhaps this capacity is what current *Gc* measures are tapping. Much is to be learned about *Gc,* but consumers of intelligence tests need to be aware that they are not assessing a depth of knowledge.

Memory. Related questions remain about the various memory abilities (Carroll, 1993). Horn defined memory abilities as short-term apprehension and retrieval (SAR) and tertiary storage and retrieval (TSR) (Horn & Blankson, 2005). Carroll (1993) defined them as general memory and learning (*Gy*) and broad retrieval ability (*Gr*). CHC theory uses the terms *Gsm* and *Glr*. The indicators of these abilities depend on the definition. For example, paired associate tasks are indicators of *Glr* in CHC theory, but they are indicators of short-term memory according to Horn and Blankson (2005). Some may argue that *Glr* factors are indexed by measures of rote learning ability, and thus the relation between rote learning and *Glr* (or memory and learning) needs to be defined more clearly. An additional problem in the measurement of *Glr* arises when rate or fluency of retrieval is used as an indicator. Although rate and fluency of retrieval may be aspects of *Glr,* these measures may also be confounded with cognitive speediness (*Gs*).

Other questions remain about memory constructs. For example, why do verbal memory tasks load on *Gsm* factors, but visual memory tasks on *Gv* factors? Future research should investigate the nature of visual and verbal presentation and responses with memory. CB-CFAs with the WAIS-III and Wechsler Memory Scales, Third Edition, have provided some insight (Tulsky & Price, 2003). Tulsky & Price (2003) found that, in general, the visual memory tasks were cognitively complex, with some visual memory tasks loading on both Visual Memory and Perceptual Organization factors. Spatial Span, however, loaded on Working Memory and Perceptual Organization factors, but did not load on the Visual Memory Factor.

Memory factors are common in factor analytic studies, but questions about how to satisfactorily identify these factors have a long history (Wolfle, 1940). Additional memory-related questions have been explicated in more detail elsewhere (e.g., McGrew, 2005). CB-CFAs should be able to continue to provide more insight. A priori predictions about which memory tests should load on which factors could be explicitly tested, a methodology advocated long ago by Thurstone (Wolfle, 1940).

Speed and Measurement

The role and nature of speed in performing tasks leads to many measurement problems. In general, the importance of speed on performance has been reduced with each revision of the major intelligence tests. This change is important because using timed scores for bonus points might differentiate some individuals who are at higher levels of ability, but factor models using timed versions of tests often fit worse than those using untimed versions (Keith et al., 2010; Reynolds et al., 2007). When bonus points for timed tasks are used, the scores may represent differences in both the ability and some other construct (e.g., speed). Speed may be important, but it may introduce construct-irrelevant variance.

Differentiation of Abilities

An issue already addressed several times in this article is the reduced complexity of intelligence tests, and perhaps intelligence itself, for younger children (or, said differently, the increasing complexity of intelligence with development; Carroll, 1993, chap. 17). According to Carroll, the issue of differentiation of intelligence has been debated since the early 20th century, yet it is still unclear whether intelligence is less complex for younger children or simply more difficult to measure. *Gf* factors seem especially fragile for younger children. Carroll suggested that Piagetian Reasoning is a narrow ability subsumed by *Gf* (Carroll, 1993; Newton & McGrew, 2010), so it seems reasonable that such differences could be due to changes in the structure of intelligence as children develop. Recent studies have found contradictory evidence for differentiation and the closely related phenomenon that *g* appears more important for younger children than for older children and adults (cf., Bickley et al., 1995; Hartmann & Reuter, 2006; Juan-Espinosa, Garcia, Colom, & Abad, 2000; Kane & Brand, 2006; Tideman & Gustafsson, 2004). In addition, it is not uncommon to find that a test measures the same constructs when analyzed across ages, while paradoxically showing different factor structures when different ages are analyzed separately (Keith, 1990; Keith et al., 2010;

Reynolds et al., 2007). CB-CFA methods, along with tests of measurement invariance, have promise to help settle this issue (Reynolds & Keith, 2007).

Future Research

There is plenty of theorizing and empirical work needed to understand even some of the most commonly measured and well-researched broad abilities. Improvement in understanding will come from future refinements and improvements in theory and the collection of empirical evidence. The following suggestions may aid in the development of future studies.

CB-CFA

Many of the questions posed in this section and earlier in this review could be answered, at least in part, using CFA approaches analyzing subtests from multiple batteries. An analysis with measures of multiple narrow *Glr* and *Gc* abilities could help improve the understanding of the nature of these abilities, for example, yet no one battery includes more than a few measures of each, and thus CB analysis is especially important. The problem with CB analysis, however, is the time commitment required for participants, and such studies rarely include more than two batteries and rarely include large samples. One efficient way of dealing with these limitations is the use of reference variable methodology to take advantage of planned missingness (McArdle, 1994). This methodology can be used to combine data from different studies, given an overlapping series of tests, the reference variables (e.g., Keith et al., 2010).

Number of Factors

Much of the research discussed in this article has used CFA, and has explicitly tested for constructs derived from CHC theory (or *Gf-Gc* or three-stratum theory). Others have argued, however, that modern tests are likely overfactored, that is, that analysts and test publishers have begun retaining too many factors in their analyses (including CFAs; Frazier & Youngstrom, 2007). If tests are being overfactored, that would mean that such tests really measure fewer abilities than their developers think they are measuring. As a solution, the authors suggested the use of EFA or principal components analysis (PCA; a method that is distinct from factor analysis) and the use of the parallel analysis or minimum average partial (MAP) analysis to determine the number of components to retain. Obviously, if the "true" numbers of factors underlying recent tests were lower than those predicted by CHC theory, that finding would be a blow to CHC theory. If, for example, the WJ measured four factors rather than seven, those factors would not be consistent with CHC theory. This critique, then, should be addressed.

Frazier and Youngstrom (2007) showed that whereas the number of constructs that tests are supposed to measure has increased over time, the number of components identified using objective criteria has not. There are several possible reasons for this discrepancy; Frazier and Youngstrom suggested that publishers and researchers are likely overfactoring newer tests. Another possibility, of course, is that the criteria used to decide the number of factors did not accurately capture the true factors present, that is, this research underfactored current tests. We illustrate this second possibility with a small simulation.

The model shown in Figure 2 was used to generate an implied matrix that we then analyzed using a variety of approaches. The higher-order model shown is consistent with CHC theory and with modern approaches (such as that used with the WJ-III) in which each broad ability is referenced by at least two measures. The magnitude of first- and second-order factor loadings is also consistent with CHC theory and research. Because the factor model was known (it was used to create the data), a corresponding CFA model fit the data perfectly (i.e., $\chi^2 = 0$, root mean square error of

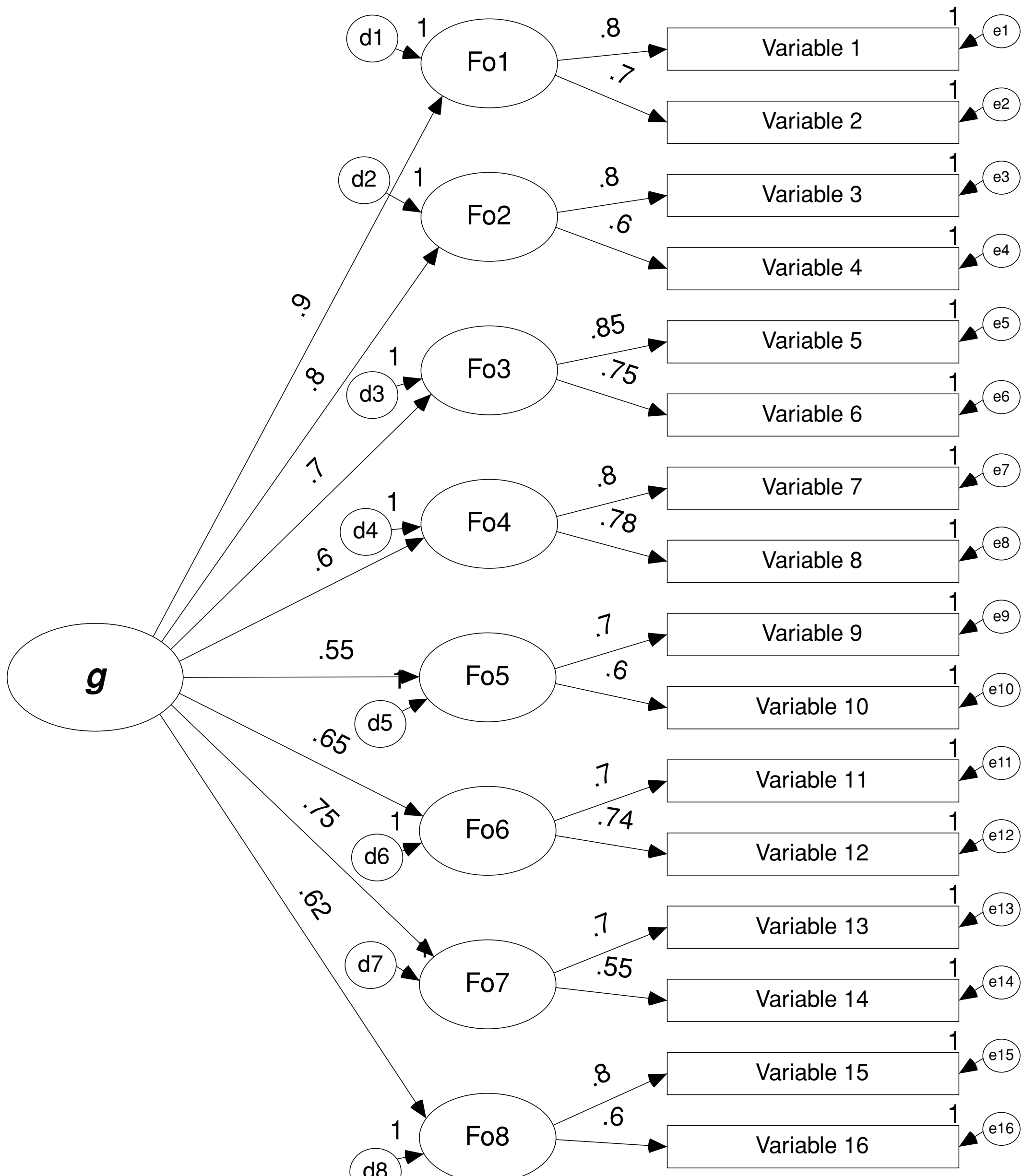

FIGURE 2. Higher-order model used to generate simulated data.

approximation [RMSEA] = 0, comparative fit index [CFI] = 1.0) when it was estimated using the simulated data. Other (incorrect) models, with fewer factors, did not fit the data as well, although some still showed an adequate or good fit using conventional criteria. Nevertheless, the best-fitting CFA model successfully and perfectly uncovered the correct factor structure.

The implied matrix was also analyzed using the techniques recommended by Frazier and Youngstrom (2007). Rather than the correct eight-factor solution, parallel analysis using principal components suggested either four (95% confidence interval) or five (average random data eigenvalues) components, and the MAP method suggested only one factor! The most common method used in the literature (eigenvalues greater than 1) also underfactored the implied matrix, suggesting five components. A scree plot of factors suggested either one or (correctly) eight components, depending how the straight line was drawn.

These findings are not definitive. A more complete simulation would add error and would manipulate first- and second-order factor loadings. The findings certainly suggest, however, that EFA or PCA procedures for uncovering the number of factors (or components) are not always accurate and are not always superior to theory-guided CFA in determining true factor structure. Furthermore, the recommendations ignore the importance of theory in test development and factor analysis. Increasingly, developers of intelligence tests have a theory in mind when they develop their tests, so they develop subtests and write items to reflect that theory. True science makes predictions based on theory and then tests those predictions against data. We believe that when a test is designed to measure constructs from a theory it is important to evaluate its conformance to predictions based on that theory, a process that is generally better captured through confirmatory as opposed to exploratory approaches. Even stronger evidence supporting or refuting a theory (and a test developed from that theory) is obtained when the researcher's or test developer's theory is tested against predictions from alternative, competing theories. This goal is also well implemented in CFA, with its ability to test specific models and compare competing models.

What We've Learned

Here we discussed what we have learned over the past 20 years of research on the nature and measurement of intelligence from a CHC perspective. Before moving forward, it is important to take a step back. It is likely inconceivable for new students in school psychology to imagine that a psychological test would be developed *without* an underlying theoretical underpinning. This reliance on underlying theory is a relatively recent advancement in the measurement of human cognitive abilities, however. The original KABC, for example, was groundbreaking for its reliance on theory in the mid-1980s. The reliance of most intelligence tests on theory is thus a major advance in itself. This advance can, in turn, improve research on intelligence and intelligence tests by moving research from focusing on exploratory analysis to focusing on tests of theory.

The increased reliance on CHC theory as a primary driver of new tests is also opportune. CHC theory appears to be a valid foundation on which to build the current and next generation of intelligence tests. There are certainly questions that remain about the theory, however, and test developers will also no doubt find new and better methods for translating theory into practice. Intelligence and its measurement remain fascinating topics, and we look forward to participating in the next 20 years of research.

References

Beaujean, A. A., McGlaughlin, S. M., & Margulies, A. S. (2009). Factorial validity of the Reynolds Intellectual Scales for referred students. Psychology in the Schools, 46, 932–950.

Bickley, P. G., 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.

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

Carroll, J. B. (1995). Book review – Assessment of cognitive processes: The PASS theory of intelligence. Journal of Psychoeducational Assessment, 13, 397–409.

Carroll, J. B. (1997). Commentary on Keith and Witta's hierarchical and cross-age confirmatory factor analysis of the WISC-III. School Psychology Quarterly, 12, 108–109.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), The scientific study of general intelligence: Tribute to Arthur R. Jensen (pp. 5–21). Boston: Pergamon.

Cattell, R. B. (1987). Intelligence: Its structure, growth, and action. New York: North-Holland.

Cattell, R. B., & Bernard, R. (1971). Abilities: Their structure, growth, and action. Boston: Houghton Mifflin.

Cavinez, G. L. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales–Fifth Edition for children and adolescents. School Psychology Quarterly, 23, 533–541.

Chen, H.-Y., Keith, T. Z., Chen, Y.-H., & Chang, B.-S. (2009). What does the WISC-IV measure for Chinese students? Validation of the scoring and CHC-based interpretive approaches in Taiwan. Journal of Research in Education Sciences, 54(3), 85–108.

DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet–Fifth Edition. Journal of Psychoeducational Assessment, 24, 123–136.

Elliott, C. D. (1990). Differential Ability Scales: Introductory and technical manual. San Antonio, TX: Psychological Corporation.

Elliott, C. D. (2007). Differential Ability Scales (2nd ed.): Introductory and technical manual. San Antonio, TX: Harcourt Assessment.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. Developmental Psychology, 40, 935–952.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from modern Gf-Gc theory: Joint confirmatory factor analysis of the WJ-R and Kaufman Adolescent and Adult Intelligence Test (KAIT). Journal of School Psychology, 36, 151–182.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? Intelligence, 35, 169–182.

Hartmann, P. H., & Reuter, M. (2006). Spearman's "Law of Diminishing Returns" tested with two methods. Intelligence, 34, 47–62.

Horn, J. L. (1998). A basis for research on age differences in cognitive abilities. In J. J. McArdle & R. W. Woodcock (Eds.), Human cognitive abilities in theory and practice (pp. 57–92). Mahwah, NJ: Erlbaum.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 41–68). New York: Guilford.

Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. C. MacCallum (Eds.), Factor analysis at 100: Historical developments and future directions (pp. 205–247). Mahwah, NJ: Erlbaum.

Hunt, M. S. (2007). A joint confirmatory factor analysis of the Kaufman Assessment Battery for Children, Second Edition, and the Woodcock-Johnson Tests of Cognitive Abilities, Third Edition, with preschool children. Unpublished Dissertation, Ball State University, Muncie, Indiana.

Juan-Espinosa, M., Garcia, L. F., Colom, R., & Abad, F. J. (2000). Testing the age related differentiation hypothesis through the Wechsler's scales. Personality and Individual Differences, 29, 1069–1075.

Kane, H. D., & Brand, C. R. (2006). The variable importance of general intelligence (g) in the cognitive abilities of children and adolescents. Educational Psychology, 26, 751–767.

Kaufman, A. S. (1993). Joint exploratory factor analysis of the Kaufman Assessment Battery for Children and the Kaufman Adolescent and Adult Intelligence Test for 11- and 12-year-olds. Journal of Clinical Child & Adolescent Psychology, 22, 355–364.

Kaufman, A. S., & Kaufman, N. L. (1983). Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1993). Manual: Kaufman Adolescent and Adult Intelligence Test. Circle Pines, MN: AGS.

Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman Assessment Battery for Children (2nd ed.): Technical manual. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & McLean, J. E. (1986). Joint factor analysis of the K-ABC and WISC-R with normal children. Journal of School Psychology, 25, 105–118.

Kaufman, A. S., & McLean, J. E. (1987). K-ABC/WISC-R factor analysis for a learning disabled population. Journal of Learning Disabilities, 19, 145–153.

Keith, T. Z. (1985). Questioning the K-ABC: What does it measure? School Psychology Review, 14, 9–20.

Keith, T. Z. (1986). Factor structure of the K-ABC for referred school children. Psychology in the Schools, 23, 241–246.

Keith, T. Z. (1990). Confirmatory and hierarchical confirmatory analysis of the Differential Ability Scales. Journal of Psychoeducational Assessment, 8, 391–405.

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. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (2nd ed., pp. 581–614). New York: Guilford Press.

Keith, T. Z., Cool, V. A., Novak, C. G., White, L. J., & Pottebaum, S. M. (1988). Confirmatory factor analysis of the Stanford–Binet Fourth Edition: Testing the theory-test match. Journal of School Psychology, 26, 253–274.

Keith, T. Z., & Dunbar, S. B. (1984). Hierarchical factor analysis of the K-ABC: Testing alternate models. Journal of Special Education, 18(3), 367–375.

Keith, T. Z., Fine, J. G., Reynolds, M. R., Taub, G. E., & Kranzler, J. H. (2006). Higher-order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children–Fourth Edition: What does it measure? School Psychology Review, 35, 108–127.

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., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock–Johnson Tests of Cognitive Ability (3rd ed.). School Psychology Review, 30, 89–119.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales–II: Consistency across ages 4 to 17. Psychology in the Schools, 47(7), 676–697.

Keith, T. Z., & Novak, C. G. (1987). Joint factor structure of the WISC-R and K-ABC for referred school children. Journal of Psychoeducational Assessment, 5(4), 370–386.

Keith, T. Z., Quirk, K. J., Schartzer, C., & Elliott, C. D. (1999). Construct bias in the Differential Ability Scales? Confirmatory and hierarchical factor structure across three ethnic groups. Journal of Psychoeducational Assessment, 17, 249–268.

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.

Kranzler, J. H., Keith, T. Z., & Flanagan, D. P. (2000). Independent examination of the factor structure of the Cognitive Assessment System (CAS): Further evidence challenging the construct validity of the CAS. Journal of Psychoeducational Assessment, 18, 143–159.

Kranzler, J. H., & Weng, L.-J. (1995). Factor structure of the PASS cognitive tasks: A reexamination of Naglieri et al. (1991). Journal of School Psychology, 33, 143–157.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. Multivariate Behavioral Research, 29, 409–454.

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

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (2nd ed., pp. 136–181). New York: Guilford.

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

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1990). WJ-R technical manual. Allen, TX: DLM Teaching Resources.

McGrew, K. S., & Woodcock, R. W. (2001). Technical manual. Woodcock–Johnson III. Itasca, IL: Riverside.

Morgan, K. E., Rothlisberg, B. A., McIntosh, D. E., & Hunt, M. S. (2009). Confirmatory factor analysis of the KABC-II in preschool children. Psychology in the Schools, 46, 515–525.

Naglieri, J. A. (1999). How valid is the PASS theory and CAS? School Psychology Review, 28, 146–162.

Naglieri, J. A. (2005). The Cognitive Assessment system. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (2nd ed., pp. 441–460). New York: Guilford.

Naglieri, J. A., & Das, J. P. (1997). Cognitive assessment system: Interpretive handbook. Itasca, IL: Riverside.

Nelson, J. M., Cavinez, G. L., Lindstron, W., & Hatt, C. V. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. Journal of School Psychology, 45, 439–456.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell–Horn–Carroll–based assessment. Psychology in the Schools, 47(7), 621–634.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery investigation. School Psychology Quarterly, 20, 66–88.

Reynolds, C. R., & Kamphaus, R. W. (2003). Reynolds Intellectual Assessment Scales. Lutz, FL: Psychological Assessment Resources.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's Law of Diminishing Returns in hierarchical models of intelligence for children and adolescents. Intelligence, 35, 267–281.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. F., & Low, J. A. (2007). Confirmatory factor analysis of the Kaufman Assessment Battery for Children–Second edition: Consistency with Cattell–Horn–Carroll theory. School Psychology Quarterly, 22, 511–539.

Roid, G. H. (2003). Stanford Binet Intelligence Scales (5th ed.): Technical manual. Itasca, IL: Riverside.

Roid, G. H., & Pomplun, M. (2005). Interpreting the Stanford-Binet Intelligence Scales (5th ed.). In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (2nd ed., pp. 325–343). New York: Guilford.

Sanders, S., McIntosh, D. E., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the Differential Ability Scales and the Woodcock–Johnson Tests of Cognitive Abilities (3rd ed.). Psychology in the Schools, 44, 119–138.

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

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell–Horn–Carroll theory and cross-age invariance of the Woodcock–Johnson Tests of Cognitive Abilities III. School Psychology Quarterly, 19, 72–87.

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

Tideman, E., & Gustafsson, J.-E. (2004). Age-related differentiation of cognitive abilities in ages 3–7. Personality and Individual Differences, 36, 1965–1974.

Tulsky, D. S., & Price, L. R. (2003). The joint WAIS-III and WMS-III factor structure: Development and cross-validation of a six-factor model of cognitive functioning. Psychological Assessment, 15, 149–162.

Tusing, M. E., & Ford, L. (2004). Examining preschool cognitive abilities using a CHC framework. International Journal of Testing, 4(2), 91–114.

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children–Fourth edition. Psychological Assessment, 18, 123–125.

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children–Fourth edition among referred students. Educational and Psychological Measurement, 66, 975–983.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children (3rd ed.): Manual. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). Wechsler Intelligence Scale for Children (4th ed.): Technical and interpretive manual. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). Wechsler Adult Intelligence Scale (4th ed.): Technical and interpretive manual. San Antonio, TX: Pearson.

Wolfle, D. (1940). Factor analysis to 1940. Psychometric Monographs, 3.

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 Psychoeducational Battery–Revised: Tests of cognitive ability. Chicago: Riverside.

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