School Psychology Review,
2009, Volume 38, No. 2, pp. 249–265

General and Specific Effects on Cattell–Horn–Carroll Broad Ability Composites: Analysis of the Woodcock–Johnson III Normative Update Cattell–Horn–Carroll Factor Clusters Across Development

Randy G. Floyd
The University of Memphis

Kevin S. McGrew
Woodcock–Munoz Foundation, University of Minnesota

Amberly Barry, Fawziya Rafael, and Joshua Rogers
The University of Memphis

Abstract. Many school psychologists focus their interpretation on composite scores from intelligence test batteries designed to measure the broad abilities from the Cattell–Horn–Carroll theory. The purpose of this study was to investigate the general factor loadings and specificity of the broad ability composite scores from one such intelligence test battery, the Woodcock–Johnson III Tests of Cognitive Abilities Normative Update (Woodcock, McGrew, Schrank, & Mather, 2007). Results from samples beginning at age 4 and continuing through age 60 indicate that Comprehension–Knowledge, Long-Term Retrieval, and Fluid Reasoning appear to be primarily measures of the general factor at many ages. In contrast, Visual-Spatial Thinking, Auditory Processing, and Processing Speed appear to be primarily measures of specific abilities at most ages. We offer suggestions for considering both the general factor and specific abilities when interpreting Cattell–Horn–Carroll broad ability composite scores.

School psychologists have been inundated this decade with intelligence test batteries that provide a variety of composites representing specific cognitive abilities. The majority of the specific cognitive abilities targeted by these contemporary test batteries are

This study was based on the third author's master's thesis proposal submitted to The University of Memphis Department of Psychology. The second author, Kevin S. McGrew, has a financial interest in the Woodcock–Johnson III because he is a coauthor of that test battery. We thank Richard Woodcock and Woodcock–Muñoz Foundation for making data from the WJ-III Normative Update sample available for this research. We benefited greatly from the reviews and comments of John Garruto, Ruben Lopez, John Kranzler, Tim Keith, and Matt Reynolds. Katie Cooley and Danielle Murphy are appreciated for providing computer, editing, and library support.

Correspondence regarding this article should be addressed to Randy G. Floyd, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN, 38120; E-mail: rgfloyd@memphis.edu

Copyright 2009 by the National Association of School Psychologists, ISSN 0279-6015

grounded in design blueprints based on the Cattell–Horn–Carroll (CHC) theory[1] (see Alfonso, Flanagan, & Radwan, 2005). The composites from these batteries are most often designed to measure the *broad abilities* of CHC theory, such as Crystallized Intelligence and Fluid Intelligence. In addition to intelligence test batteries based on the CHC design blueprint, interpretive approaches guiding users to form composites within and across batteries based on CHC theory have proliferated (e.g., Flanagan, Ortiz, & Alfonso, 2007; McGrew & Flanagan, 1998). Because broad ability composite scores typically possess substantial validity evidence and have overcome reliability limitations inherent in the interpretation of subtest scores (McGrew, 1997; Watkins, Glutting, & Youngstrom, 2005), there has also been increased research focused on their interpretation within score profiles (e.g., Bergeron & Floyd, 2006; Floyd, Bergeron, & Alfonso, 2006; Proctor, Floyd, & Shaver, 2005).

Despite this increased prevalence of test batteries, interpretive approaches, and research employing composite scores based on CHC theory, some important measurement properties of these broad ability composite scores remain unstudied—the effects of general and specific cognitive abilities. The goal of this article is to produce the estimates of these effects on the broad ability composite scores from the first intelligence test battery based on CHC theory, the Woodcock–Johnson III Tests of Cognitive Abilities (WJIII COG; Woodcock, McGrew, & Mather, 2001).

General and Specific Abilities

Mental ability as a general, unitary trait was first postulated by Spearman (1904) based on his observation that all mental test scores were positively intercorrelated. Spearman supported this postulation with research using factor analysis. This method led him to discover the general factor (Spearman, 1927) underlying the positive manifold across mental ability test scores. Since the time of Spearman, hundreds of studies have demonstrated that the general factor accounts for approximately 25% to 50% of the variance shared by such tests—typically the largest percentage of any factor. In addition, another large body of research has demonstrated that scores representing the general factor (e.g., IQs) are strong predictors of representations of personal competence, such as academic success and job performance (see Jensen, 1998; Schmidt, 2002).

Despite extensive evidence for the general factor and its predictive properties, a number of challenges to its primacy has been levied. Some scholars have argued that largely independent, specific cognitive abilities better account for the relations between and among mental test scores. For example, Thurstone (1935) offered seven primary mental ability factors: Comprehension, Fluency, Memory, Number, Reason, Space, and Speed. Cattell (1943) proposed the existence of Fluid Intelligence (Gf) and Crystallized Intelligence (Gc). Later, Cattell and Horn (e.g., Horn & Cattell, 1967, 1982) concluded that multiple factors existed that mirrored those of Thurstone, and they identified other factors representing understudied perceptual abilities, such as Auditory Processing. They labeled the multiple higher order factors *broad abilities* and the lower order group factors *narrow abilities*. Since then, Horn and colleagues extended Gf-Gc theory and produced an impressive body of evidence supporting the validity of these broad abilities (e.g., Horn, 1991; Horn & Blankson, 2005; Horn & McArdle, 2007).

Rapprochement and Contemporary Support

Both arguments supporting the highly predictive and ubiquitous general factor as well as arguments supporting the specific cognitive ability factors are supported by a sizable body of research, and it seems that these perspectives are not mutually exclusive. For example, Thurstone (1947) conceded that there were correlations among his seven primary mental ability factors that could represent the general factor. In a similar manner, Spearman (1939) conceded that the general factor alone could not account for all of the relations between tests of mental abilities and acknowledged the existence of group factors representing more specific cognitive abilities. Holzinger and Swineford (1939) merged Spearman's model focusing on the general factor with Thurstone's model of primary mental ability factors to produce a hybrid model (Jensen, 1998). In this model, the general factor represents the relations among all test scores and specific ability factors represent the remaining reliable, shared variance across subsets of test scores.

In recent years, explanatory models of human cognitive abilities that specify both a general factor and specific factors have become commonplace. For example, Carroll (1993) developed the three-stratum theory of cognitive abilities based on reanalysis of more than 460 data sets using a factor-analytic technique designed to extract the general factor as well as more specific ability factors from the patterns of relations among test scores. His synthesis of results yielded a model espousing abilities at varying levels of generality: Stratum III (the level of general factor), Stratum II (the level of broad abilities), and Stratum I (the level of narrow abilities). Since Carroll's seminal publication, theoretical models specifying both a general factor and more specific ability factors have been prevalent in test manuals and in recently published articles in school psychology and assessment-oriented journals. For example, some researchers have considered both the general and specific factors through use of Carroll's methods of exploratory factor analysis (EFA; e.g., Nelson, Canivez, Lindstrom, & Hatt, 2007; Watkins, 2006; Watkins, Wilson, Kotz, Carbone, & Babula, 2006). Many others have included at least two orders of factors (i.e., Stratum III—general and Stratum II—broad) in their confirmatory factor analyses (CFA; e.g., Keith, Fine, Taub, M. R. Reynolds, & Kranzler, 2006; Oh, Glutting, Watkins, Youngstrom, & McDermott, 2004; Phelps, McGrew, Knopik, & Ford, 2005; Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007; Tulsky & Price, 2003). These publications provide contemporary support for the rejection of most models that do not consider abilities at more than one stratum.

In summary, research published throughout the past century and into the present suggests that scores from all tests of cognitive abilities share common variance. Despite arguments and some empirical evidence offered by Horn and others (e.g., Horn, 1991), it appears that this common variance is best conceptualized as the general factor (Carroll, 1993; Jensen, 1998). However, such test scores also share variance with only measures involving similar task demands and requiring similar cognitive processes. These groups of test scores measure more specific cognitive abilities (e.g., Stratum II—broad abilities), and it appears that, because of the similarities in the identification and labeling of these abilities by Carroll (Carroll, 1993) and Horn and colleagues (e.g., Horn & Blankson, 2005), test authors and other scholars have recently focused significant attention on these specific cognitive abilities (see McGrew, 2005; McGrew, 2009).

How Are the Effects of General and Specific Abilities on Test Scores Demonstrated?

When considering the large body of evidence supporting the existence of general and more specific cognitive abilities, psychologists and other professionals engaged in ability measurement should understand the methods used to determine the effects of these abilities on test scores.

g Loadings

Many researchers have applied factor-analytic techniques to data from intelligence test batteries to determine effects attributable to the general factor. Our preliminary review of peer-reviewed journal articles identified 47 articles (presenting the results from 91 analyses) published since 1981 that have presented *g loadings* of subtest scores from published intelligence test batteries.[2] The *g* loadings are standardized coefficients that have a hypothetical range of .00 to 1.00, and they represent the effect, in standard deviation units, of the general factor on the subtests. Across the studies reviewed, authors employed principal factors analysis (a.k.a. common factor analysis) in 24 articles, whereas authors of only 5 articles employed principal components analysis. Authors of 10 articles employed a hierarchical EFA technique (i.e., Wherry hierarchical factor analysis and the Schmid–Leiman procedure). In contrast, authors of 15 articles used maximum-likelihood estimation via CFA; only one study employed generalized least squares estimation via CFA.

Across all studies, *g* loadings ranged from .91 with data from a heterogeneous sample (Keith & Dunbar, 1984) and .95 with data from a low general ability sample (Kane, Oakland, & Brand, 2006) to .10 with data from adults recruited from a college campus (Visser, Ashton, & Vernon, 2006). The *g* loadings have been interpreted using the following standards: .70 and above indicate a *high* measure of the general factor, .69 to .50 indicate a *medium* measure of the general factor, and .50 and below indicate a *low* measure of the general factor (McGrew & Flanagan, 1998; cf. A. S. Kaufman, 1979, 1994). In 49 of the analyses, measures targeting Crystallized Intelligence had the highest *g* loadings. For example, the Vocabulary subtests from child and adult editions of the Wechsler Intelligence Scales (e.g., Wechsler, 2003) and from the Stanford–Binet Intelligence Scale, Fourth Edition (R. L. Thorndike, Hagen, & Sattler, 1986) produced the highest *g* loadings (ranging from .72 to .90) in many analyses (e.g., Gignac, 2006a, 2006b; Watkins, 2006; Watkins et al., 2006). The Riddles subtests from the Kaufman Assessment Battery for Children and the Kaufman Assessment Battery for Children, Second Edition (A. S. Kaufman & N. L. Kaufman, 1983; A. S. Kaufman & N. L. Kaufman, 2004) produced the highest *g* loadings (ranging from .67 to .91) in other analyses (e.g., A. S. Kaufman & McLean, 1987; Keith & Novak, 1987; Naglieri & Jensen, 1987; M. R. Reynolds, Keith, Fine, Fisher, & Low, 2007).

Measures of Fluid Reasoning demonstrated the highest *g* loadings in 22 analyses.[3] For example, the Arithmetic subtests from the Kaufman Assessment Battery for Children, the Wechsler Adult Intelligence Scale, Third Edition (Wechsler, 1997), and the Wechsler Intelligence Scale for Children, Fourth Edition (Wechsler, 2003) produced the highest *g* loadings (ranging from .77 to .86) in some analyses (e.g., Gignac, 2006b; Keith et al., 2006; Keith & Novak, 1987). Subtests with "quantitative reasoning" in their titles from the Differential Ability Scales (Elliott, 1990) and the Stanford–Binet Intelligence Scale, Fifth Edition (Roid, 2003) produced the highest *g* loadings (ranging from .74 to .79) in other analyses (e.g., DiStefano & Dombrowski, 2006; Sanders et al., 2007). Measures of the following broad abilities also demonstrated the highest *g* loadings in some analyses: Visual-Spatial Thinking (11 analyses), Short-Term Memory (5 analyses), reading ability (i.e., Reading and Writing; 3 analyses), and Processing Speed (1 analysis).

In 43 analyses, measures targeting Visual Processing demonstrated the lowest *g* loadings. For example, the Gestalt Closure subtests from the Kaufman Assessment Battery for Children and the Kaufman Assessment Battery for Children, Second Edition (A. S. Kaufman & N. L. Kaufman, 1983; A. S. Kaufman & N. L. Kaufman, 2004) produced the lowest *g* loadings ranging from .40 to .52 (e.g., A. S. Kaufman & McLean, 1987; M. R. Reynolds et al., 2007), and the Mazes subtests from the Wechsler Intelligence Scale for Children—Revised (Wechsler, 1974) and the Wechsler Intelligence Scale for Children, Third Edition (Wechsler, 1991) produced the lowest *g* loadings (ranging from .16 to .39) in some analyses (e.g., Keith & Witta, 1997; Smith & Stanley, 1987). In 30 analyses, measures targeting Processing Speed demonstrated the lowest *g* loadings. For example, the Coding and Digit Symbol-Coding subtests of the Wechsler Intelligence Scale for Children—Revised, the Wechsler Intelligence Scale for Children, Fourth Edition, the Wechsler Adult Intelligence Scale, Revised (Wechsler, 1981), and the Wechsler Adult Intelligence Scale, Third Edition, produced *g* loadings ranging from .25 to .49 in some analyses (e.g., Ashton & Lee, 2006; Gignac, 2006a, 2006b; Watkins, 2006; Watkins et al., 2006). Measures of the following broad abilities also demonstrated the lowest *g* loadings: Short-Term Memory (7 analyses); Long-Term Retrieval (6 analyses); Crystallized Intelligence, Fluid Reasoning, and reading ability (2 analyses); and Auditory Processing and inspection time (1 analysis).

Specificity

Once general factor effects have been considered, reliable, specific variance and error variance remain. Whereas error variance is the portion of a test score that is not reliable and that is not accounted for by the general and specific abilities, reliable, specific variance can be considered *specificity*. Obtaining estimates of specificity for individual measures is important to those who consider broad and narrow abilities because such estimates may establish that these measures can be interpreted as independent of the general factor.

We identified 15 articles published since 1981 (producing 18 analyses) that presented specificity estimates.[4] In 8 studies, authors obtained specificity estimates through multiple regression analysis in which multiple correlations between the subtests in question and all other subtests in the battery was squared. Authors then subtracted these squared multiple correlations from the reliable variances of the subtests in question. The squared multiple correlations represent general factor variance, and internal consistency reliability coefficients, such as split-half reliability estimates, represent reliable variance. Authors of 6 articles obtained specificity estimates by subtracting communality estimates of the subtests in question from the subtests' reliable variances. Communality estimates stem from factor analysis rather than multiple regression analysis. Like squared multiple correlations, if only a single general factor is extracted from factor analysis, communality estimates represent general factor variance. In only 1 study were CFA techniques used to calculate specificity estimates (i.e., M. R. Reynolds et al., 2007).

Specificity estimates are typically reported as percentages to represent the proportion of variance that is reliable and specific. Specificity estimates have typically been interpreted using the following rules of thumb: (a) for *high* specificity, estimates must be greater than error variance and represent at least 25% of total variance; (b) for *medium* specificity, estimates must be greater than error variance and represent 15% to 24% of total variance; and (c) for *poor* specificity, estimates must be greater than error variance and represent less than 15% of total variance (McGrew & Flanagan, 1998; cf. A. S. Kaufman, 1994). The specificity estimates from our review ranged from 63% (Kamphaus & Platt, 1992) to 3% (M. R. Reynolds et al., 2007). Measures of Short-Term Memory produced the highest specificity estimates in 7 analyses; they ranged from 41% to 63%. For example, the Digit Span subtests from the WAIS-R and the WISC-III; the Memory for Sentences and Memory for Digits subtests from the Stanford–Binet Intelligence Scale, Fourth Edition; and the Word Order subtests from the Kaufman Assessment Battery for Children, Second Edition often demonstrated the highest specificity (e.g., Gutkin, C. R. Reynolds, & Galvin, 1984; M. R. Reynolds et al., 2007). In addition, measures of Processing Speed produced the highest specificity estimates in 3 analyses; they ranged from 43% to 44% and were evident from the Digit Symbol subtests from the Wechsler Adult Intelligence Scale (Wechsler, 1955) and the WAIS-R and from the Animal Pegs subtest from the Wechsler Preschool and Primary Scale of Intelligence—Revised (Wechsler, 1989). Measures of Processing Speed, Visual Processing, Auditory Processing, and Long-Term Retrieval demonstrated the highest specificity in 2 analyses; a measure of Fluid Reasoning demonstrated the highest specificity in another. In 15 analyses, measures of Crystallized Intelligence demonstrated the lowest specificity. Measures of Visual Processing demonstrated the lowest specificity in 2 analyses, and measures of Fluid Reasoning as well as measures of Short-Term Memory demonstrated the lowest specificity in other studies.

Purpose of the Study

Much research has produced g loadings and specificity estimates for subtests from intelligence test batteries, but in recent years interpretive approaches guiding users to form CHC broad ability composites have proliferated. However, no published studies have produced g loadings or specificity estimates for such composite scores from contemporary intelligence test batteries using large, heterogeneous samples (cf. Kane et al., 2006). The results of this study should inform CHC theory by indicating whether the current move toward rapprochement is warranted. If broad ability composite scores are uniformly highly g loaded, redundancy in measurement across them is identified and specific ability effects may be questioned. On the other hand, if broad ability composite scores demonstrate uniformly high specificity, the independence of each of these measures is apparent and the expectation of an expansive and monolithic general factor will be questioned. Results of this study should also inform practice by providing insights about the strength of specificity effects on the broad ability composite scores, which should be expected because they target specific ability constructs, as well as insights about the strength of general factor effects, which may be considered construct irrelevant when targeting specific ability constructs (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

This study examined the g loadings and specificity of the CHC factor cluster scores from the WJIII COG. The WJIII COG produces seven CHC broad ability composite scores, which is greater in number than any other contemporary intelligence test battery. Based on our review of the literature, we hypothesized the following about these scores: (a) Fluid Reasoning and Comprehension–Knowledge should demonstrate high g loadings and low specificity; and (b) Processing Speed, Short-Term Memory, Long-Term Retrieval, Visual Processing, and Auditory Processing should demonstrate high specificity and moderate to low g loadings. The wide age range coverage of the WJIII COG also allowed us to examine developmental differences in these effects from preschool age through late adulthood.

Method

Participants

All participants were drawn from the WJIII United States standardization sample (McGrew & Woodcock, 2001). This nationally representative standardization sample was formed using a stratified sampling plan that controlled for 10 individual and community variables. Participants ages 4 through 90+ who completed all necessary tests to produce scores for this study were included (N = 3,577). These participants were divided into seven age-based subsamples (n = 179 to n = 875). Information about the size, gender distribution, race and ethnicity distribution, and socioeconomic status distribution of each subsample is presented in Table 1. Gender, race, and Hispanic ethnicity were unequally distributed across groups, $\chi^2(6) = 50.67, p < .001$, $\chi^2(18) = 128.80, p < .001$, and $\chi^2(6) = 39.93, p < .001$, respectively. For gender, there were fewer males than expected in the 20–39 age group, based on comparison to the 14–19 age group, $\chi^2(1) = 10.63, p = .001$. For race, there were fewer Black, Asian and Pacific Islander, and American Indian children than expected for the 4–5 age group, based on comparison to the 6–8 age group, $\chi^2(3) = 9.78, p = .021$, as well as fewer Black and Asian and Pacific Islander participants than expected in the 20–39 age group, based on comparison to the 14–19 age group, $\chi^2(3) = 41.06, p < .001$. For Hispanic ethnicity, there were fewer Hispanic participants than expected for the 20–39 age group, based on comparison to the 14–19 age group, $\chi^2(1) = 6.51, p = .011$.

Parent education level, operationalized by the highest education level of a parent or caregiver, differed significantly across the child and adolescent samples, $F(3, 2392) = 8.59, p < .001$. Tukey post hoc tests indicated that parent education level for those

Table 1
Percentages of Participant Sex, Race, and Ethnicity and Descriptive Statistics for Highest Education Level for Seven Age Levels

		Sex		Race and Ethnicity					Education Level
Age Level	*n*	F	M	W	B	AI	API	Hispanic	Parent Education/Years of Education[a] $M(SD)$
4–5	236	50.0	50.0	84.3	11.9	1.3	2.5	9.7	4.1 (1.0)
6–8	471	46.7	53.3	74.5	17.0	3.6	4.9	8.5	4.0 (1.0)
9–13	875	49.1	50.9	75.4	16.0	3.0	5.6	9.3	3.9 (1.0)
14–19	846	50.2	49.8	78.0	15.4	1.4	5.2	7.8	4.1 (0.9)
20–39	640	58.8	41.3	89.5	7.0	1.9	1.6	4.5	14.6 (2.1)
40–59	330	60.9	39.1	89.4	5.2	3.3	2.1	2.1	14.9 (2.7)
60+	179	68.7	31.3	95.5	1.7	2.2	0.6	1.1	13.1 (2.8)

Note. F = female; M = male; W = White; B = Black; AI = American Indian; API = Asian or Pacific Islander.
[a]Parent education level is reported for age groups 4–5 through 14–19. Years of education is reported for age groups 20–39 through 60+. Some data were missing for the parent education variable and years of education variable, so values for these variables are based on slightly smaller sample sizes than reported for all other variables reported in this study.

in the 4–5 age group was significantly higher than those in the 9–13 age group and that parent education level for those in the 14–19 age group was higher than those in the 6–8 and 9–13 age groups, $p < .05$. Years of education across the adult samples was also significantly different, $F(2, 1139) = 36.26$, $p < .001$. Those in the 60+ age group reported significantly fewer years of education than those in the other two groups, $p < .05$.

Measures

Age-based standard scores ($M = 100$, $SD = 15$) were the metric of analysis. These scores were obtained from a reweighting of the WJIII standardization data based on 2005 United States census data (i.e., the WJIII Normative Update; Woodcock, McGrew, Schrank, & Mather, 2007). The seven WJIII COG CHC factor cluster scores served as measures.[5] Median internal consistency reliability estimates obtained from McGrew, Schrank, and Woodcock (2007) are noted in the following paragraph for each cluster. Cluster reliabilities were calculated based on the obtained reliabilities for component tests. Rasch analysis was used to calculate the reliability of speeded tests (i.e., Visual Matching, Retrieval Fluency, Decision Speed) and tests that employed multiple-point items (i.e., Spatial Relations, Retrieval Fluency, and Picture Recognition). Split-half reliability analyses were used for the remaining tests. Extensive validity evidence supporting cluster scores is reviewed in McGrew and Woodcock (2001), McGrew et al. (2007), and Floyd, Shaver, and McGrew (2003). The development, standardization, and psychometric properties of the WJIII have been evaluated favorably by independent reviewers (e.g., Cizek, 2003; Sandoval, 2003).

The Comprehension–Knowledge cluster measures the breadth and depth of knowledge and the ability to communicate and reason using this knowledge. It is formed from scores on the Verbal Comprehension and General Information tests. Across ages 4–80+, its median reliability coefficient was .95. The Long-Term Retrieval cluster measures the ability to store information and fluently retrieve it later. It is formed from Visual-Auditory Learning and Retrieval Fluency. Its median reliability coefficient was .88. The Visual-Spatial Thinking cluster measures the ability to perceive and manipulate visual stimuli. Its median reliability coefficient was .81. It is formed from Spatial Relations and Picture Recognition. The Auditory Processing cluster measures the ability to analyze, synthesize, and discriminate auditory stimuli. It is formed from Sound Blending and Auditory Attention. Its median reliability coefficient was .91. The Fluid Reasoning cluster measures the ability to reason and form concepts. It is formed from Concept Formation and Analysis-Synthesis. Its median reliability coefficient was .95. The Processing Speed cluster measures the ability to perform simple cognitive tasks quickly and repeatedly. It is formed from Visual Matching and Decision Speed. Its median reliability coefficient was .92. The Short-Term Memory cluster measures the ability to hold language-based information in immediate awareness and use it within a few seconds. It is formed from Numbers Reversed and Memory for Words. Its median reliability coefficient was .88.

Analysis

g Loadings and general factor variance. To obtain the *g* loadings for each factor cluster, standard scores from all seven clusters at each age level were entered into a principal factors analysis in which one factor was extracted.[6] The *g* loadings represent the correlation between the cluster scores and the general factor, and they may be interpreted as effect size estimates (Hunter & Schmidt, 2004; Jensen, 1982). Consistent with previous research, the *g* loadings were squared to obtain estimates of the percentage of variance in the cluster score at each age level attributable to the general factor (i.e., communality estimates).

Specificity and specific variance. Specificity estimates for each factor cluster were obtained at each age level by following three steps. First, communality estimates were obtained from the *g* loadings from principal factors analysis. Second, reliability coefficients for each cluster were obtained from McGrew et al. (2007), and the mean reliability coefficient for each age level was calculated. Third, the communality estimate for each cluster was subtracted from its respective mean reliability coefficient to obtain an estimate of specificity.[7] Because values representing the percentage of "variance accounted for" likely mask the true effects of one variable on another and do not allow for interval-level comparisons to be made across such values, the square roots of each specificity estimate at each age level were calculated to provide effect size estimates mirroring *g* loadings (Hunter & Schmidt, 2004; Jensen, 1982).

Error variance. Estimates of error variance for each cluster were obtained by subtracting reliability coefficients from 1, but specific results are not reported. Error estimates represent the proportion of variance in a cluster attributed to unsystematic sources of score variability.

Results

Preliminary data analysis was conducted with each age-based sample to ensure that the assumptions of absence of outliers, normality, linearity, and factorability of the correlation matrix were not violated. Results revealed 54 univariate outliers (with $p = .001$, two-tailed test), but for no variable were there more than 4 univariate outliers for an age level. Because each age-based subsample was reasonably large, only one case with an extreme univariate outlier ($z = -5.13$) was deleted. Subsequent results revealed 17 multivariate outliers (using Mahalanobis distance with $p = .001$ for the χ^2 value). These cases were also deleted. No variable was notably skewed (all values $< |1.0|$) at any age level. Although two variables demonstrated notable positive kurtosis (i.e., both >1.0 but less than 2.0) for ages 4–5, no kurtosis was severe enough to affect our analyses with such large samples (Waternaux, 1976). Review of scatterplots revealed that the assumption of linearity was not violated. Finally, for each principal factors analysis, the Kaiser–Meyer–Olkin measure of sampling adequacy was .6 or higher ($M = .87$, $SD = .01$, range = .85 to .88), and Bartlett's test of sphericity was statistically significant ($p < .001$).

Table 2
General Factor Effects and Specific Effects for the WJIII CHC Factor Clusters Across Age Levels

		CHC Factor Cluster													
		Gc		Glr		Gv		Ga		Gf		Gs		Gsm	
Age Level	Effect	ES	%	ES	%	ES	%	ES	%	ES	%	ES	%	ES	%
4–5	General	.70	.49	**.82***	.68	.45	.20	.59	.35	**.70**	.49	.60	.36	.62	.38
	Specific	.66	.44	.44	.20	**.82***	.68	**.76***	.58	.68	.47	**.75***	.56	**.73**	.54
6–8	General	**.74***	.55	**.73***	.54	.36	.13	.55	.30	.65	.43	.52	.27	.60	.36
	Specific	.61	.37	.58	.34	**.78***	.61	**.75***	.57	**.73**	.53	**.81***	.66	**.72***	.51
9–13	General	**.77***	.60	**.72***	.53	.48	.23	.61	.37	**.78***	.60	.49	.24	.62	.38
	Specific	.59	.34	.59	.34	**.74***	.55	**.71***	.51	.59	.35	**.82***	.67	.69	.48
14–19	General	**.77***	.59	**.74***	.54	.53	.28	.62	.38	**.79***	.63	.52	.27	.63	.39
	Specific	.60	.36	.59	.35	**.73***	.53	**.72***	.52	.56	.32	**.80***	.64	.70	.49
20–39	General	.69	.48	**.74***	.55	.60	.36	.67	.45	**.79***	.62	.51	.26	.61	.37
	Specific	.70	.49	.61	.38	.69	.47	.70	.49	.58	.34	**.82***	.67	**.73***	.53
40–59	General	.70	.49	**.76***	.58	.56	.31	.67	.45	.67	.45	.62	.38	.62	.39
	Specific	.70	.48	.59	.35	**.74***	.55	.70	.49	**.72**	.52	**.75***	.57	**.72**	.52
60+	General	**.75**	.57	**.76***	.58	.48	.23	.63	.39	.66	.43	.61	.37	.61	.37
	Specific	.64	.41	.57	.33	**.79***	.62	**.74**	.55	**.73**	.54	**.76**	.58	**.74**	.55

Note. WJIII CHC = Woodcock–Johnson III Cattell–Horn–Carroll; Gc = Comprehension–Knowledge; Glr = Long-Term Retrieval; Gv = Visual-Spatial Thinking; Ga = Auditory Processing; Gf = Fluid Reasoning; Gs = Processing Speed; Gsm = Short-Term Memory; ES = effect size; % = percentage of variance in CHC factor cluster scores. Bold denotes effect size estimates that are >.70.
*Substantial differences, considering 95% confidence intervals, are evident between general factor effects and specific effects.

g Loadings

When the general factor was extracted from the seven CHC factor cluster scores from each age-based sample, the general factor accounted for approximately 42% of the variance, on average, among scores (SD = 2.50%, range = 36.87% to 44.20%). The coefficient of congruence, the correlation between the extracted general factor and the "true" general factor, was .91, on average (SD = .01, range = .89 to .92).

Table 2 presents *g* loadings and their squared values reported percentage of variance attributable to the general factor. The *g* loadings for Comprehension–Knowledge were typically high (M = .73, Mdn = .74, SD = .04). At only one age level (20–39) was its *g* loading medium (.69). The *g* loadings for Long-Term Retrieval were the highest, on average, of any cluster (M = .76, Mdn = .74, SD = .03). For every age level, its *g* loading was strong. Its highest *g* loading was at ages 4–5 (.82), which was the highest for any cluster. In contrast to Comprehension–Knowledge and Long-Term Retrieval, the *g* loadings for Visual-Spatial Thinking were low or medium. They were the lowest and demonstrated the greatest variability of any cluster (M = .50, Mdn = .55, SD = .08). For the first three age levels, its *g* loading was less than .50, its lowest *g* loading was at ages 6–8 (.36), which was the lowest for any cluster. The *g* loadings for Auditory Processing were also most often medium (M = .62, Mdn = .62, SD = .04). The *g* loadings for Fluid Reasoning were typically high (M = .72, Mdn = .70, SD = .06). For the first five age levels, its g loading was .70 or higher, but its g loadings were lower for the oldest age levels. Its highest g loading was for ages 14–19 and ages 20–39 (.79), whereas its lowest g loading was for ages 60+ (.66). The g loadings for Processing Speed were typically medium (M = .55, Mdn = .52, SD = .05). Only its g loading at ages 9–13 was low (.49). The g loadings for Short-Term Memory were medium, and they varied minimally across age levels (M = .61, Mdn = .62, SD = .01).

Specificity

Table 2 presents for each CHC factor clusters and for each age level both (a) the effect size estimates representing the influence of specificity on test scores and (b) the traditional specificity estimates representing the percentage of variance in scores attributable to specificity. Results indicate that specificity variance exceeded error variance for every cluster across all age levels (M difference = 39.74%, SD = 11.36%, range = 7.05% to 60.62%).

We propose that the effect size estimates for specific variance be evaluated using the exact same standards as g loadings. Specificity effects for Comprehension–Knowledge were from medium to high (M = .64, Mdn = .64, SD = .05). For only ages 20–39 and 40–59 were specificity effects high (.70). Specificity effects for Long-Term Retrieval were typically medium (M = .57, Mdn = .59, SD = .06). For only ages 4–5 was its specificity low (.44). Specificity effects for both Visual-Spatial Thinking and Auditory Processing were typically high (M = .76, Mdn = .74, SD = .05 and M = .73, Mdn = .72, SD = .02, respectively). Specificity effects for Visual-Spatial Thinking and Auditory Processing were highest at ages 4–5. Specificity effects for Fluid Reasoning were medium to high, and they demonstrated the greatest variability of any cluster (M = .66, Mdn = .68, SD = .08). Specificity effects for Processing Speed were high—and the highest of any cluster (M = .79, Mdn = .80, SD = .03). Specificity effects for Short-Term Memory were typically high (M = .72, Mdn = .72, SD = .02). At only one age level (9–13) was its specificity effects medium (.69).

Differences Between General and Specific Effects

Table 2 also presents indications of substantial differences between g loadings and specificity effects for each CHC factor cluster and each age level. For each effect size, 95% confidence intervals were calculated based on sample size. Effect size estimates were transformed to z prime (z') scores, upper and lower limits of the confidence interval for z' scores were calculated, and z' score were transformed back to effect size estimates (Cohen, Cohen, West, & Aiken, 2003). Confidence interval values were compared for each pair of g loadings and specificity effects, and confidence intervals that were not overlapping indicated substantial differences between these values. When substantial differences were identified, the higher of the two effect size estimates was marked with an asterisk in Table 2. (These differences are not statistically significant because of our inability to identify a relevant method to compare interdependent values. The specificity estimate is derived from the g loading, and their sum cannot exceed 1.0.). In addition, to indicate absolute effects, we identified (and made bold in Table 2) g loadings and specificity estimates that exceeded .70.

As evident in Table 2, three CHC factor clusters appear to be primarily measures of the general factor across age levels because they demonstrate both substantial effects of the general factor and substantially greater g loadings than specificity effects. Long-Term Retrieval met both criteria for every age level. Comprehension–Knowledge and Fluid Reasoning met both criteria for three of seven age levels; at other age levels, they were similar in measuring the g factor and specificity. Three CHC factor clusters appear across most age levels to be primarily measures of a more specific factor or factors. Visual-Spatial Thinking, Auditory Processing, and Processing Speed demonstrated both substantial effects of specificity and substantially greater specificity effects than g loadings. Visual-Spatial Thinking met these criteria for every age level, and Auditory Processing and Processing Speed met these criteria for six of seven age levels.

The remaining CHC factor cluster, Short-Term Memory, was most frequently similar in measuring the g factor and specificity. Short-Term Memory demonstrated both substantial effects of specificity and substantially greater specificity effects than g loadings at only two age levels.

Discussion

The goal of this study was to examine the effects of general and specific abilities on the WJIII CHC factor clusters scores across a wide range of age levels. Results revealed that, despite the labels reflecting the CHC broad abilities given to the factor clusters, three clusters appear to be primarily measures of the general factor at many age levels: Long-Term Retrieval, Fluid Reasoning, and Comprehension–Knowledge. That is, across at least three age levels, their g loadings were .70 or greater, and the g loadings were notably higher than specificity effects. These results support our hypothesis about Fluid Reasoning and Comprehension–Knowledge, but Long-Term Retrieval demonstrated more consistently strong g loadings than expected. However, these general patterns are not surprising based on (a) Carroll's (1993) synthesis, (b) recent research using CFA techniques to form broad ability factors from some of the same tests forming these composites (e.g., Floyd, Keith, Taub, & McGrew, 2007; Taub & McGrew, 2004), and (c) research indicating that intelligence test subtests requiring the greatest number of mental processes during completion of items (i.e., cognitive complexity) tend to demonstrate high g loadings (Marshalek, Lohman, & Snow, 1983; McGrew, 2002).

Consistent with the design of the WJIII CHC factor clusters and other broad ability composite scores, specificity effects were sizable across age levels and CHC factor clusters. As hypothesized, Visual-Spatial Thinking, Auditory Processing, and Processing Speed appear to be primarily measures of specific abilities, whereas Short-Term Memory demonstrated sizable specificity effects at only two age levels (cf. Kamphaus & Platt, 1992; R. M. Thorndike, 1990). It is perhaps most notable that, when both g loadings and specificity effects were put on the same scale, on average, specificity effects (M = .69, SD = .08, range .44 to .82) were similar to (and in fact exceeded) general factor effects (M = .64, SD = .10, range .36 to .82). Indeed, all CHC factor clusters except Long-Term Retrieval demonstrated high specificity effects for at least two age levels. Furthermore, only Long-Term Retrieval at one age level (1 of 48) demonstrated low specificity effects. These collective results are most likely from the greater reliability of composite scores than subtest scores.

Age Differences

There was not great variability in g loadings and specificity effects of the CHC factor clusters across age levels. Standard deviations of g loadings across the age levels ranged from .01 for Short-Term Memory to .08 for Visual-Spatial Thinking (Mdn = .04), and standard deviations of specificity estimates ranged from .02 for Auditory Processing and Short-Term Memory to .08 for Fluid Reasoning (Mdn = .05). Correlations revealed negligible to weak relations between the rank ordering of (a) the age levels and (b) the g loadings or specificity effects ($\rho < .40$) for most clusters. Only g loadings for Visual-Spatial Thinking and Auditory Processing appear to increase with age, $\rho = .65$ and $\rho = .85$, respectively, and only specificity effects for Auditory Processing appeared to decrease with age, $\rho = -.60$. Because of significant differences in composition of our age-based samples (e.g., gender, race, ethnicity, and socioeconomic status), we cannot eliminate the possibility that some of the differences in g loadings and specificity effects attributable to age-related effects may instead be effects associated with other personal variables and their interactions (see Sirin, 2005 and Keith, M. R. Reynolds, Patel, & Ridley, 2008).

Other Limitations and Alternate Explanations

Because our results stemmed from an analysis of normative samples from the standardization of the WJIII, it is likely that our results will not generalize perfectly to children and adults near the tails of the general factor ability distribution. Spearman's law of diminishing returns conveys that the general factor effects on test scores tend to decrease as levels of the general factor increase (Jensen, 2003; M. R. Reynolds & Keith, 2007). As a result, it is possible that the *g* loadings we present underestimate general factor effects and overestimate specificity for very low-functioning individuals and that the *g* loadings we present overestimate general factor effects and underestimate specificity for very high-functioning individuals. Some researchers have suggested that the *g* loadings (and specificity effects) vary to a substantial degree based on the variety of tests included in the factor analysis (e.g., McGrew & Flanagan, 1998; Woodcock, 1990). For example, some may argue that the Long-Term Retrieval cluster's *g* loadings were spuriously high because of speed-related ability influencing one of its tests, Retrieval Fluency. We tested the influence of including both (a) Long-Term Retrieval scores and (b) Processing Speed scores in the same factor analysis. Our results revealed the influence of psychometric sampling error to be negligible at most age levels. At four of seven age levels, the difference between the *g* loadings for Long-Term Retrieval obtained with and without Processing Speed was .01 or less. The median difference was .01, and the average difference was .04 (SD = .06). At only ages 9–13 and ages 60+ did these differences exceed .05 (i.e., .08 and .16, respectively). Consistent effects of psychometric sampling error are also improbable because we selected composite scores stemming from tests that have been demonstrated to well represent different broad ability factors (McGrew & Woodcock, 2001).

Some may argue that our use of EFA is inappropriate when more theory-driven and controlled analyses, such as CFA, can be used. We agree that theory and control are needed for extracting general factors, and in fact, we believe that we used EFA methods in a confirmatory manner—as have most all other authors who have used EFA to extract *g* loadings. Others may argue that we failed to consider first-order factors stemming from two or more factor clusters from which a higher order general factor can be specified (e.g., Bickley, Keith, & Wolfle, 1995). However, analysis of scree plots, consideration of eigenvalues greater than or equal to 1.0, and results from Horn's parallel analysis (Horn, 1965) suggested only a single factor should be extracted during analysis at each age level. Multiple viable first-order factors are improbable.

Implications

Theory. The results of this study support the rapprochement that appears in much contemporary literature devoted to CHC theory. A review of items from the 14 tests yielding the CHC factor cluster scores will reveal that a variety of stimuli and output mechanisms and a plethora of cognitive processes must be employed during test performance; thus, these tests appear to measure different processes and specific abilities. However, on average, the resultant cluster scores are greatly affected by the general factor. Well more than one-third of the variance (i.e., ~42%) across these scores can be considered shared. If the specific abilities that effect CHC factor cluster scores were independent (or even largely independent), this percentage would be much lower. On the other hand, these results reveal that the general factor is neither expansive nor monolithic in its effects. Clearly, several CHC factor cluster scores demonstrate substantial effects from specific abilities; thus, some scores are more independent than others. In addition, when the effects of specific abilities were put on a level playing field with general factor effects (putting aside consideration of only the percentage of variance attributable to specificity), these effects were relatively equal influences on the CHC factor cluster scores.

School psychologists embedded in cognitive ability assessment, measurement and ability researchers, and wonks steeped in CHC theory and psychometric interpretive frameworks may rejoice in the variety of abilities apparently measured by the CHC factor clusters. These results reveal great diversity in the cognitive abilities measured by these clusters. They indicate that at least eight abilities—the general factor and at least one independent specific cognitive ability per CHC factor cluster—can be reliability measured in children and adults. In fact, one can lavish in the array of variation across the cluster scores and across age levels. Clearly, these results are noteworthy from a theoretical perspective, and CHC theory should steadfastly continue to embrace a model specifying both a general ability and more specific abilities.

Practice. Theory and research findings should guide the interpretation of test scores. One goal of this study was to inform psychologists about the measurement properties of composite scores that target CHC broad abilities, and it used an established method central to the history of psychology to identify these properties. Despite our enthusiasm from a theoretical perspective, it is possible that considering the general factor and more specific abilities (as well as error) muddies the waters of clinical interpretation of composites targeting CHC broad abilities. These composite scores are not pure measures of CHC cognitive abilities—even when random error is removed (Oh et al., 2004)—and it is highly improbable that any other such composite across intelligence tests would yield a pure measure of any cognitive ability. Consequentially, psychologists interpreting CHC broad ability composite scores should consider the following: (a) composite scores, like subtest scores, are influenced by the general factor, which contributes construct-irrelevant effects when targeting CHC broad abilities, per se (Watkins, 2006; Watkins et al., 2006); (b) composite scores are influenced by specific abilities (sometimes substantially), which contributes construct-irrelevant effects when targeting the general factor; and (c) there is great variation across broad ability composites in their measurement of the general factor and in their specificity.

With these considerations in mind, some implications may follow. Some WJIII CHC factor clusters are predominantly measures of the general factor across many age levels. Hypotheses regarding environmental effects can be offered to explain variation among such scores representing the general factor. For example, a score on one high *g*-loading cluster that is notably higher than scores on other high *g*-loading clusters may be attributed to environmental enhancements, such as explicit training, whereas a score on a high *g*-loading cluster that is notably lower than other *g*-loading clusters may be attributed to environmental inadequacies, such as being from an impoverished background (A. S. Kaufman, 1979). These hypotheses may be important when basic assessment techniques, such as informant interviews, cannot be used to obtain more accurate information about potential environmental effects, and such hypotheses should be tested when possible (see Hale & Fiorello, 2004). Although the high *g*-loading clusters may be used in place of the IQ in the diagnosis of mental retardation when there is reason to doubt the validity of the IQ (National Research Council Committee on Disability Determination for Mental Retardation, 2002), we would not recommend routinely using them to represent the general factor because of sizable specificity effects. The practice of aggregating multiple measures of mental ability to operationalize the general factor and weighting these measures according to their *g* loadings appears to be ideal (Floyd, Clark, & Shadish, 2008; McGrew & Woodcock, 2001).

Specificity, which may be attributable in large part to broad and narrow abilities, appears to be much more consistently sizable for composite scores than subtest scores and more sizable for some composites than others. It is possible that interpretation of item-level performance on tests contributing to high-specificity clusters will allow clinicians to develop hypotheses relevant to intervention development, but this task is daunting because item-level performance is besieged by random error. For example, random error may be evident in failure to retrieve known answers, use of detrimental and facilitative strategies, guessing, and serendipitous exposure to item content. It is also possible that some composites with high specificity may offer incremental value (beyond the general factor) in predicting some specific outcomes and that profiles of high specificity composites may provide diagnostic information or inform intervention development, but previous research has indicated that the actuality of these possibilities is unlikely (see McDermott, Fantuzzo, & Glutting, 1990; Schmidt, 2002).

Conclusion

A large body of research evidence and the model espoused by the CHC theory informs test users that most every cognitive ability test score—be it a composite or a subtest score—is reflective of some part general factor, some part broad ability or abilities, and some part narrow ability or abilities (in addition to random error). It is an asset to recognize this simple psychometric reality because assessment instruments are only as good as the professionals using them (Meyer et al., 2001). Knowing which composite or test scores are more saturated with general factor variance, which are more saturated with variance from a broad ability, and so on should lead to more appropriate inferences in practice and research settings. Without (a) the availability of scores targeting one cognitive ability freed from the effects of other abilities, an elusive psychometric goal (Carroll, 1993); or (b) sophisticated, statistically based interpretive methods designed to overcome construct-irrelevant influences in ability measurement (e.g., use of suppressor variables and differential weighting; Woodcock & Johnson, 1977; Woodcock et al., 2001), professionals may fall prey to misguided decisions about cognitive ability strengths and weaknesses and fall short in developing appropriate interventions.

Footnotes

[1]CHC theory is only one model describing human cognitive abilities, and its foundation is factor-analytic evidence. Other models focus more on cognitive processes and are formed from other sources of evidence (see Flanagan & Harrison, 2005; Sattler, 2008; Sternberg, 2000).

[2]We realize that *g* loadings have been published in books (e.g., Sattler, 2008) and in test manuals, but we included only those from peer-reviewed journals published during 1982 and afterward. A table summarizing these results can be obtained from the first author or online (http://www.memphis.edu/psychology/people/faculty/floyd.php).

[3]Measures of Quantitative Knowledge and Quantitative Reasoning were considered measures of Fluid Reasoning (Carroll, 1993).

[4]A table summarizing these results can be obtained from the first author or his website.

[5]A table presenting the means and standard deviations of standard scores for each cluster and for each sample can be obtained from the first author or his website.

[6]Analysis was also completed using principal components analysis. Results indicated that across CHC factor clusters, the principal components analysis yielded *g* loadings that were approximately .01 higher on average than those from principal factors analysis (SD = .07).

[7]Use of communality estimates allows the sum of squared *g* loadings (communality), specificity, and error to equal 100% or 1.0. We compared communalities to squared multiple correlations and found that results were minimally different (mean difference across clusters and age levels = 5.13%, SD= 3.07%). The communality estimates were larger in magnitude, on average, for every CHC factor cluster (range of difference = 9.21% for Long-Term Retrieval to 2.21% for Processing Speed).

References

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell–Horn–Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp., 185–202). New York: Guilford Press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Ashton, M. C., & Lee, K. (2006). "Minimally biased" g-loadings of crystallized and non-crystallized abilities. *Intelligence, 34,* 469–477.

Bergeron, R., & Floyd, R. G. (2006). Broad cognitive abilities of children with mental retardation: An analysis of group and individual profiles. *American Journal of Mental Retardation, 111,* 417–432.

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.

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

Cizek, G. J. (2003). [Review of the Woodcock-Johnson III.] In B. S. Plake & J. C. Impara (Eds.), *The fifteenth mental measurements yearbook* (pp. 1020–1024). Lincoln, NE: Buros Institute of Mental Measurements.

Cohen, J. Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

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.* San Antonio, TX: Psychological Corporation.

Flanagan, D. P., & Harrison, P. (Eds.). (2005). *Contemporary intellectual assessment* (2nd ed.). New York: Guilford Press.

Flanagan, D. P., Ortiz, S., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). New York: Wiley.

Floyd, R. G., Bergeron, R., & Alfonso, V. C. (2006). Cattell–Horn–Carroll cognitive ability profiles of poor comprehenders. *Reading and Writing: An Interdisciplinary Journal, 19,* 427–456.

Floyd, R. G., Clark, M. H., & Shadish, W. R. (2008). The exchangeability of intelligent quotients: Implications for professional psychology. *Professional Psychology: Research and Practice, 39,* 414–423.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22,* 200–233.

Floyd, R. G., Shaver, R. B., & McGrew, K. S. (2003). Interpretation of the Woodcock–Johnson III Tests of Cognitive Abilities: Acting on evidence. In F. A. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation* (pp. 1–46). New York: Academic Press.

Gignac, G. E. (2006a). Evaluating subtest "g" saturation levels via the single trait-correlated uniqueness (STCU) SEM approach: Evidence in favor of crystallized subtests as the best indicators of "g." *Intelligence, 34,* 29–46.

Gignac, G. E. (2006b). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences, 27,* 73–86.

Gutkin, T. B., Reynolds, C. R., & Galvin, G. A. (1984). Factor analysis of the Wechsler Adult Intelligence Scale—Revised: An examination of the standardization sample. *Journal of School Psychology, 22,* 83–93.

Hale, B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York: Guilford Press.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (Supplementary Education Monographs, No. 48). Chicago: Department of Education, University of Chicago.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual* (pp. 197–232). Itasca, IL: Riverside Publishing.

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* (2nd ed., pp. 41–68). New York, NY: Guilford Press.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized general intelligences. *Acta Psychologica, 26,* 107–129.

Horn, J. L., & Cattell, R. B. (1982). Whimsy and misunderstandings of Gf-Gc theory: A comment on Guilford. *Psychological Bulletin, 91,* 623–633.

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. 205–249). Mahwah, NJ: Erlbaum.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Jensen, A. R. (1982). *Bias in mental testing.* New York: Free Press.

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

Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence, 31,* 95–105.

Kamphaus, R. W., & Platt, L. O. (1992). Subtest specificities for the WISC-III. *Psychological Reports, 70,* 899–902.

Kane, H. D., Oakland, T. D., & Brand, C. R. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *The Journal of Genetic Psychology, 167,* 327–341.

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

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

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

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & McLean, J. E. (1987). Joint factor analysis of the K-ABC and WISC-R with normal children. *Journal of School Psychology, 25,* 105–118.

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

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, 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., & Novak, C. G. (1987). Joint factor structure of the WISC-R and K-ABC for referred school children. *Journal of Psychoeducational Assessment, 5,* 370–386.

Keith, T. Z., Reynolds, M. R., Patel, P., & Ridley, K. R. (2008). Sex differences in latent cognitive abilities ages 6 to 59. Evidence from the Woodcock–Johnson III Tests of Cognitive Abilities. *Intelligence, 36,* 502–525.

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.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–128.

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

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. 131–150). New York: Guilford Press.

McGrew, K. S. (2002). *McGrew interpretive multidimensional ability pattern schematic.* Unpublished manuscript.

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

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10.

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

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual. Woodcock–Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

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

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128–165.

Naglieri, J. A., & Jensen, A. R. (1987). Comparison of Black–White differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence, 11,* 21–43.

National Research Council Committee on Disability Determination for Mental Retardation. (2002). *Mental retardation: Determining eligibility for Social Security benefits.* Washington, DC: National Academy Press.

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

Oh, H.-J., Glutting, J. J., Watkins, M. W., Youngstrom, E. A., & McDermott, P. A. (2004). Correct interpretation of latent versus observed abilities: Implications from structural equation modeling applied to the WISC-III and WIAT linking sample. *The Journal of Special Education, 38,* 159–173.

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

Proctor, B., Floyd, R. G., & Shaver, R. B. (2005). CHC broad cognitive ability profiles of low math achievers. *Psychology in the Schools, 42,* 1–12.

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. E., & Low, J. (2007). Confirmatory factor structure 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 Scale, Fifth Edition.* Itasca, IL: Riverside Publishing.

Sanders, S., McIntosh, D. A., 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—Third Edition. *Psychology in the Schools, 44,* 119–138.

Sandoval, J. (2003). [Review of the Woodcock–Johnson III]. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook* (pp. 1024–1028). Lincoln, NE: Buros Institute of Mental Measurements.

Sattler, J. M. (2008). *Assessment of children: Cognitive applications* (5th. ed.). San Diego: Author.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15,* 187–211.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417–453.

Smith, G. A., & Stanley, G. (1987). Comparing subtest profiles of g loadings and correlations with RT measures. *Intelligence, 11,* 291–298.

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

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Spearman, C. (1939). Thurstone's work reworked. *Journal of Educational Psychology, 30,* 1–16.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence.* Cambridge, UK: Cambridge University Press.

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

Thorndike, R. M. (1990). Would the real factors of the Stanford–Binet Fourth Edition come forward? *Journal of Psychoeducational Assessment, 8,* 412–435.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet Intelligence Scale: Guide for administering and scoring the Fourth Edition.* Chicago: Riverside Publishing.

Thurstone, L. L. (1935). *Vectors of mind: Multiple-factor analysis of the isolation of primary traits.* Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Methods of factor analysis: A development and expansion of vectors of mind.* Chicago: University of Chicago Press.

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.

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence, 34,* 487–502.
Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrica, 63,* 639–645.
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., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 251–268). New York: Guilford Press.
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. (1955). *Wechsler Adult Intelligence Scale.* San Antonio, TX: Psychological Corporation.
Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised.* San Antonio, TX: Psychological Corporation.
Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* San Antonio, TX: Psychological Corporation.
Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised.* San Antonio, TX: Psychological Corporation.
Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition.* Antonio, TX: Psychological Corporation.
Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: Psychological Corporation.
Wechsler, D. (2003). *The Wechsler Intelligence Scale for Children—Fourth Edition.* Antonio, TX: Psychological Corporation.
Woodcock, R. W. (1990). The theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.
Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock–Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III.* Itasca, IL: Riverside Publishing.
Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2007). *Woodcock–Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

Date Received: October 1, 2007
Date Accepted: February 5, 2009
Action Editor: James DiPerna ■

Randy G. Floyd is an associate professor of psychology at The University of Memphis. He received his doctoral degree in school psychology from Indiana State University in 1999. His research interests include the structure, measurement, and correlates of cognitive abilities; the technical properties of early numeracy measures; and the process of professional publication.

Kevin S. McGrew is research director for the Woodcock–Muñoz Foundation, director of the Institute for Applied Psychometrics, associate director of Measurement Learning Consultants, and visiting professor in educational psychology at the University of Minnesota. He received his doctoral degree in educational psychology from the University of Minnesota in 1989. His research interests focus on theories and assessment of intelligence, applied psychometrics, models of personal competence, and neurotechnology.

Amberly Barry is currently employed as the fourth-grade teacher at The Redmont School in the Waldorf Tradition in Birmingham, Alabama. She received her master's degree in school psychology at The University of Memphis in 2008. Her interests focus on the practical applications of assessment.

Fawziya Rafael received her master's degree in school psychology at The University of Memphis in 2006. Her interests focus on the identification and development of behavioral interventions for children with pervasive developmental disorders and working with multicultural communities.

Joshua Rogers received his BA in psychology at The University of Memphis in 2007.

COPYRIGHT INFORMATION

TITLE: General and Specific Effects on Cattell-Hom-Carroll Broad Ability Composites: Analysis of the Woodcock-Johnson III Normative Update Cattell-Hom-Carroll Factor Clusters Across Development
SOURCE: Sch Psychol Rev 38 no2 Je 2009

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited. To contact the publisher: http://www.nasponline.org/index2.html