Journal of School Psychology 71 (2018) 42–56

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-horn-Carroll theory

Brian A. Zaboski II*, John H. Kranzler, Nicholas A. Gage

School of Special Education, School Psychology, & Early Childhood Studies, University of Florida, PO Box 117050, Norman Hall 1403, Gainesville, FL 32611, United States of America

ARTICLE INFO

Action Editor: Nicholas Benson

Keywords:
Intelligence
Cognitive ability
Academic achievement
Meta-analysis

ABSTRACT

Interpretation of intelligence tests has changed over time, from a focus on the elevation of general ability in the early 1900s, to the shape and/or scatter of subtest and index scores in the mid-1900s to the early 2000s, and back to elevation today. The primary emphasis of interpretation now, however, is widely recommended to be on normative strengths and weaknesses of scores reflecting broad and narrow abilities in the Cattell-Horn-Carroll (CHC) theory (Schneider & McGrew, 2012). Decisions about which abilities are important to assess for the diagnosis of learning difficulties are based largely on literature reviews by Flanagan, Ortiz, Alfonso, and Mascolo (2006) and McGrew and Wendling (2010). These were narrative research syntheses, however, and did not attempt to estimate the magnitude of the relations between CHC abilities and academic achievement. The purpose of this study, therefore, was to conduct a meta-analysis to determine the effect size for these relations across age groups. Results of our analyses found that psychometric *g* and one or more broad cognitive abilities are substantially related to each area of academic achievement. Across all achievement domains and ages, *g* had by far the largest effect, with a mean effect size of $r^2 = 0.540$. In fact, psychometric *g* explained more variance in academic outcomes than all broad abilities combined. Most broad abilities explained less than 10% of the variance in achievement and none explained more than 20%. Some age-related changes in cognitive ability-achievement relations were also observed. In sum, results of our meta-analysis support the interpretation of the overall score on intelligence tests as a measure of psychometric *g* for diagnosing difficulties in reading and mathematics, but only the interpretation of index scores measuring Comprehension-Knowledge (Gc) when diagnosing difficulties in reading. Implications of these results for research and practice are discussed.

1. Introduction

The Stanford-Binet Intelligence Scale (SBIS; Terman, 1916) was the first widely used test of intelligence used in the United States. The SBIS produced only one score, the "intelligence quotient" (IQ), and interpretation of test performance consisted of the assignment of that score to a descriptive category, such as "superior" or "average." In the early 1900s, therefore, the aim of intelligence testing was the quantification of general cognitive ability. The interpretation of intelligence tests changed dramatically in the late 1930s with the publication of the Wechsler-Bellevue Intelligence Scale (WBIS; Wechsler, 1939). The WBIS produced scaled scores for 11 subtests, as well as for verbal and performance (i.e., non-verbal) IQ scores, in addition to an overall IQ score. Wechsler advocated for the intra-

* Corresponding author at: Department of School of Special Education, School Psychology, & Early Childhood Studies, University of Florida, PO Box 117050, Norman Hall 1403, Gainesville, FL 32611, United States of America.

E-mail address: bzaboski@ufl.edu (B.A. Zaboski).

https://doi.org/10.1016/j.jsp.2018.10.001
Received 7 March 2018; Received in revised form 26 July 2018; Accepted 8 October 2018
0022-4405/

individual (or ipsative) analysis of the pattern of an individual's test scores. Since then, psychologists have routinely conducted these analyses to generate hypotheses about the underlying causes of a student's learning difficulties and their remediation (Kamphaus, Winsor, Rowe, & Kim, 2012).

Scores on intelligence tests can be analyzed in terms of their *elevation*, *scatter*, or *shape* (Cronbach & Gleser, 1953). *Elevation* refers to the level of an individual's scores, typically calculated as the mean of all subtest scores for an individual. Elevation is also reflected in the index and overall standard scores for a given person. *Scatter* denotes the amount of variability in an individual's profile, usually defined as the standard deviation of a person's subtest scores. *Shape* refers to the pattern of high and low test scores for an individual. Beginning with the work of Rapaport, Gil, and Schafer (1945–1946), intra-individual analysis of intelligence test scores has emphasized interpretation of the shape of an individual's subtest and composite scores and de-emphasized interpretation of general level of performance.

Rapaport et al.'s (1945–1946) method consisted of the visual analysis of the pattern of strengths and weaknesses in an individual's profile. Although initially popular, research supporting the diagnostic utility of this intuitive approach was lacking (Kamphaus et al., 2012). To move test interpretation toward a more solid empirical footing, Kaufman (1979, 1994) proposed the *intelligent testing* approach, which combined clinical acumen with measurement science. In this approach, statistical tests are conducted to examine whether composites differ from each other and whether subtests differ from their overall mean. By doing so, one can differentiate between scatter that is likely random and scatter that is likely not due to chance. For scores that are both statistically significantly different and have enough unique variance to warrant interpretation (e.g., 25%), base-rate data are examined to determine the commonness of the size of score differences in the normative sample. Scores that are both statistically significantly different and uncommon are interpreted as clinically important when supported by other data, such as test behavior, background information, and other assessment results.

Although the intelligent testing approach is an improvement over Rapaport et al.'s (1945–1946) "arm-chair" intra-individual analysis, serious concerns have been raised about its usefulness as an evidence-based practice (for a review, see Watkins, Glutting, & Youngstrom, 2005). Many of these concerns are related to reliance on difference scores for clinical decision-making. Difference scores are notoriously unreliable (e.g., Cattell, 1983). Because the reliability of difference scores is related to the internal consistency of the scores being compared and the correlation between those scores, their use is especially problematic for the ipsative analysis of subtests, given their weaker psychometric properties and substantial inter-correlations, but it is also problematic for composite scores. Over the last 20 years, research has shown that diagnoses based on the intelligence testing approach have poor stability and do not reliably classify groups of children and adolescents with and without disabilities (e.g., Smith & Watkins, 2004; Watkins, Kush, & Glutting, 1997). Moreover, results of ipsative analyses of subtest and composite scores have not been found to validly inform the design of interventions for academic difficulties (e.g., Braden & Shaw, 2009; Burns et al., 2016; Cronbach & Snow, 1977; Elliott & Resing, 2015; Kearns & Fuchs, 2013; Stuebing et al., 2015). In sum, despite the longstanding practice of the intra-individual analysis of intelligence test scores, at the current time very little empirical research supports its clinical utility (e.g., Glutting, Watkins, Konold, & McDermott, 2006; Macmann & Barnett, 1997; McDermott, Fantuzzo, & Glutting, 1990; Watkins, 2000).

1.1. Contemporary intelligence test interpretation

According to Kamphaus et al. (2012), the field of psychology is currently experiencing a "new wave" of interpretation wherein research, theory, and intelligence testing are merging. Since their inception until only recently, standardized tests of intelligence had been criticized on the grounds that they were largely atheoretical (e.g., Brody, 1992). Over the past decade, however, test developers have increasingly used the Cattell-Horn-Carroll (CHC) theory as a framework for designing tests (CHC; see Schneider & McGrew, 2012). CHC theory is an integration of the Horn-Cattell fluid and crystallized (Gf-Gc) theory (e.g., Horn & Noll, 1997) and Carroll's (1993) three-stratum theory. It is essentially a taxonomy of human cognitive abilities that specifies the number of different abilities and their arrangement. In CHC theory, "intelligence" is multidimensional and consists of three hierarchically arranged strata of cognitive abilities with varying degrees of *referent generality*. Referent generality refers to "the variety of behaviors or mental activities to which [a construct] relates and the degree to which it relates to them" (Coan, 1964, p. 138). At the apex of the hierarchy at Stratum III is psychometric *g*, the most general factor. Every cognitive ability measures *g* to some extent (Jensen, 1998). Stratum II consists of eight or more broad cognitive abilities that reflect various group factors (e.g., Fluid Reasoning [Gf], Comprehension-Knowledge [Gc], Short-Term Memory [Gsm]). Group factors are common only to tests that require the same kind of item content (e.g., verbal, numerical, or spatial) or cognitive processes (e.g., oral fluency, short-term memory, and perceptual speed). At Stratum I, there are over 80 narrow cognitive abilities (e.g., spatial scanning, phonetic coding, and memory for sound patterns). Today, the constructs measured by the most widely used tests of intelligence are often interpreted within the framework of CHC theory (e.g., Flanagan, Ortiz, & Alfonso, 2013).

Which CHC factors do intelligence tests measure? Due to the practical limitations of psychological assessment, the number of cognitive abilities measured on standardized tests of intelligence tests tends to be fairly small. In addition to psychometric *g*, virtually all intelligence tests measure no more than 3 to 5 broad abilities at Stratum II of CHC theory (e.g., see Flanagan et al., 2013, p. 85). For example, the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014) is the most widely used intelligence test with children and youth (Benson, Floyd, Kranzler, Eckert, & Fefer, 2018). According to Reynolds and Keith (2017), the WISC-V measures *g* and the following five broad abilities: Fluid Reasoning, Verbal Comprehension, Visual Spatial Ability, Working Memory, and Processing Speed (cf. Canivez, Dombrowski, & Watkins, 2018; Canivez, Watkins, & Dombrowski, 2016, 2017; Dombrowski, Canivez, & Watkins, 2018a). A thorough assessment of an individual's broad cognitive abilities, therefore, requires the administration of one or more additional intelligence tests measuring different abilities at Stratum II of CHC theory.

With regard to contemporary practices in intelligence test interpretation, researchers and test developers now widely recommend that the primary level of clinical interpretation be done at the index score level, which reflects the broad cognitive abilities at Stratum II of CHC theory, especially when there is considerable scatter within an individual's test score profile (e.g., Wechsler, 2014). The interpretation of subtest scores is discouraged due to their relatively poor psychometric properties. Moreover, given the paucity of evidence supporting interpretation of the shape and scatter of intelligence test profiles, the focus of interpretation is now on the elevation of the index and overall scores.

In *Essentials of WISC-V Assessment*, for example, Flanagan and Alfonso (2017) provided a logical, step-by-step approach for interpretation of the WISC-V. Although noting that the overall score is the most reliable and valid score, Flanagan and Alfonso asserted that "analysis of the primary index scores is a critical, and arguably necessary, component of WISC-V interpretation" (p. 209). In their approach, emphasis is placed on the interpretation of population-relative (or normative) strengths and weaknesses in an individual's index scores. The identification of intra-individual strengths and weaknesses is optional. As they stated, "truth be told, we find limited value in conducting this level of analysis, particularly as it applies to personal strengths and weaknesses, which is why we include it as an optional step" (p. 254).

Thus, the interpretation of intelligence test profiles has shifted over time from a focus on the elevation of general ability in the early 1900s, to an emphasis on shape, scatter, or both in the mid-1900s to the early 2000s, and back to a focus on elevation today. In contrast to the infancy of intelligence testing, however, the primary emphasis of test interpretation at present is on normative strengths and weaknesses of an individual's index scores and not general ability, although some scholars have argued for a return to an emphasis on interpretation of the overall score (e.g., Dombrowski et al., 2018a, b; Kranzler & Floyd, 2013).

In the late 1990s, Flanagan and colleagues introduced a new approach to intelligence test use and interpretation called Cross-Battery Assessment (XBA; Flanagan & McGrew, 1997; Flanagan, McGrew, & Ortiz, 2000). In the XBA approach, the administration of a single intelligence test is augmented by additional tests measuring somewhat different abilities to gain a more thorough understanding of an individual's normative strengths and weaknesses at Stratum II of CHC theory. Although they initially encouraged practitioners to assess seven broad abilities, Flanagan and Alfonso (2017) recently stated that, given results of research on the relationships between broad and narrow abilities and academic achievement,

> it no longer seems critically important that *all* the major cognitive domains assessed by intelligence and cognitive batteries (i.e., Gc, Gf, Gv, Gsm, Glr, Gs, Ga) be represented *broadly* via at least two subtests that measure qualitatively different aspects of the cognitive domain. Rather, it seems more important to understand (1) whether composites represent a broad or a narrow ability such that appropriate inferences may be drawn from test scores; and (2) whether the battery provides adequate measurement of the abilities and processes that research shows are important for understanding referrals, particularly those related to suspected learning disability [SLD]. (p. 221, emphases in the original)

Results of a recent survey by Sotelo-Dynega and Dixon (2014) indicated that roughly 25% of school psychologists currently use the XBA approach for intelligence test interpretation. In addition, the interpretation of index scores on intelligence tests is also central to the dual discrepancy/consistency (DD/C) pattern of strengths and weaknesses (PSW) approach to SLD identification (Flanagan et al., 2013). In the DD/C PSW approach, SLD is defined as unexpected underachievement and corresponding weakness in one or more specific cognitive abilities that are "empirically related" to that area of academic deficit. Moreover, proponents of the DD/C PSW method asserted that "this consistency is a necessary marker for SLD because SLD is *presumably* caused by cognitive processing weaknesses or deficits" (Flanagan & Alfonso, 2017, p. 436, emphasis in the original), which are reflected in the index scores on intelligence tests. Methods based on the PSW approaches are currently used in at least 14 states for the identification of SLD (Maki, Floyd, & Roberson, 2015).

1.1.1. Research on CHC abilities and academic achievement relations

At present, decisions about which cognitive abilities are the most germane to measure for the diagnosis of academic learning difficulties are based largely on two reviews of the literature by Flanagan, Ortiz, Alfonso, and Mascolo (2006) and by McGrew and Wendling (2010). According to McGrew and Wendling, research on the relations between cognitive abilities and academic achievement indicates that there is a Cognitive Ability by Academic Achievement Area by Age three-way interaction effect. Table 1 summarizes the broad cognitive abilities that they determined are "importantly" related to basic reading skills (BR), reading comprehension (RC), basic mathematics skills (BM), and mathematics reasoning (MR) across age. For example, according to their review the broad abilities that are importantly related to BR are Comprehension-Knowledge (Gc), Long-Term Storage and Retrieval (Glr), Processing Speed (Gs), and Short-Term Memory (Gsm), but not Auditory Processing (Ga), Fluid Intelligence (Gf), and Visual Processing (Gv). Moreover, they concluded that the pattern of these relations varies across age and academic domain. For instance, they determined that RC is related to Gc, Ga, and Gsm at ages 6–8 years, but only to Gc at ages 14–19 years. Based on the results of their review, McGrew and Wendling asserted that they "believe that less emphasis should be placed on the overall full-scale IQ and that cognitive assessment should be more selective and focused. For example, there should be selective testing of key markers for screening at-risk children" (p. 651).

It is important to note, however, that the literature reviews conducted by Flanagan et al. (2006) and by McGrew and Wendling (2010) were narrative in nature. Traditional narrative reviews, which tend to rely on the "box score" of the outcomes of significance tests across studies, have been criticized as "imprecise" and lacking in "explicit standards of proof" (Cooper, 2010, p. 7). In particular, as McGrew and Wendling noted, Flanagan et al. did not provide the following necessary information in their review: (a) specification of literature search terms and date ranges; (b) criteria for determining significant CHC ability-academic achievement relations; (c) criteria for including or excluding studies in their review; and (d) method for control of publication bias. Consequently, Flanagan

Table 1
Broad cognitive abilities in the Cattell-Horn-Carroll theory importantly related to academic achievement by age group in McGrew and Wendling (2010).

Achievement area	Age group		
	6–8 years	9–13 years	14–19 years
Basic Reading Skills	Gc Glr Gs Gsm	Gc Gs Gsm	Gc Gsm
Reading Comprehension	Gc Ga Gsm	Gc Glr Gsm	Gc
Basic Math Skills	Gc Gf Gs	Gc Gf Gs	Gc Gf Gs
Math Reasoning	Gc Gf Gs	Gc Gf Gs	Gc Gf Gsm

Note. Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Short-Term Memory (Gsm), Long-Term Storage and Retrieval (Glr), Processing Speed (Gs), and Auditory Processing (Ga).

et al.'s literature review is highly subjective, and independent replication of their results is likely impossible.

Although McGrew and Wendling (2010) addressed a number of these limitations, neither they nor Flanagan et al. attempted to estimate the magnitude of the relations under investigation. In other words, the reviews by Flanagan et al. and McGrew and Wendling did not answer the question, “What is the size of the relationships between CHC cognitive abilities and academic achievement?” In contrast to Flanagan et al.'s review, which was a subjective research synthesis, McGrew and Wendling defined an “important” relationship in terms of the percentage of studies reporting statistically significant relations, rather than in terms of effect size (e.g., Cohen, 1988). Relations were classified as “high” when more than 80% of studies reported significant results, “medium” when 50%–79% reported significant results, and so on. While they acknowledged the arbitrariness of the categories in their classification system, they overlooked concerns associated with drawing conclusions based on *p*-values alone (e.g., Wasserstein & Lazar, 2016). With large samples, there is an increased likelihood of finding statistically significant relationships even when the effect size is trivial (Anderson, Burnham, & Thompson, 2000). The results of McGrew and Wendling's literature review, therefore, may be misleading, because virtually all of the research they reviewed was conducted with large norming samples, in particular the standardization sample from the Woodcock-Johnson III Normative Update (WJ III NU) Tests of Cognitive Abilities and Tests of Achievement (Woodcock, McGrew, Schrank, & Mather, 2007) or data from prior editions of these tests. Thus, the “important” relations between CHC abilities and academic achievement identified by McGrew and Wendling may in fact be unimportant and have limited or no practical application after examination of effect size.

1.2. Purpose of current study

Given the limitations of prior syntheses of research on the relations between the cognitive abilities in CHC theory and academic achievement, the purpose of this study was to conduct a meta-analysis to determine the magnitude of these relations across age. Due to the inadequacies of the literature review by Flanagan et al. (2006), this study was designed to match the original vote counting method employed by McGrew and Wendling (2010). To minimize potential specification error, they only included studies that used measures of five or more broad abilities at Stratum II of CHC theory. We therefore limited our meta-analysis to similar studies reporting relevant regression coefficients, path coefficients, and correlations in the prediction of BR, RC, BM, and MR. In addition, while McGrew and Wendling reported relations for both broad and narrow cognitive abilities of CHC theory, we only examined results for broad abilities so that our results would inform contemporary best practices in intelligence test interpretation. This is because of the current emphasis on the interpretation of index scores, which are intended to measure broad abilities, and the fact that narrow abilities are primarily measured by subtests. Results of this study will provide practitioners and researchers with a quantitative description of the cumulative research findings on CHC ability-academic achievement relations. Our results also may have important implications for practices that rely on analysis of the PSW of intelligence test scores and the interpretation of index scores as “marker” tests for the identification of learning difficulties.

2. Method

Full-text articles were collected for all studies included in McGrew and Wendling's (2010) review between 1988 and 2009. Studies were then added following a systematic review of the literature from 2009 to 2015. Articles were required to be published in English as peer-reviewed journal articles or dissertations, with the exception of McGrew (2007), which reported unpublished data in McGrew and Wendling. As in McGrew and Wendling, two databases, ProQuest Dissertations and Theses and PsycINFO, were used to identify studies. We also used the same search strings used by McGrew and Wendling: “Cattell-Horn-Carroll,” “CHC theory,” CHC, “Gf-Gc

Table 2
Characteristics of the studies included in the meta-analysis.

Study	Sample type[a]	Cognitive battery	Achievement battery	Sample age/grade range	Achievement domain				Type of analysis[b]
					BR	RC	BM	MR	
Armstrong (2011)	Clinical	WISC-IV	WIAT-II	6.6–16.9 years	X				Correlation
Bacal (2010)	School District	WJ III COG	WJ III ACH	Grades K–8			X	X	Regression
Beaujean, Parkin, and Parker (2014)	Linking Sample	WISC-IV	WIAT-II	6.0–16.0 years	X				Higher-order factor model
Beaujean et al. (2014)	Linking sample	WISC-IV	WIAT-II	6.0–16.0 years	X				Bifactor model
Brine (2013)	Clinical	WJ III COG	WRAT-4	11.0–17.0 years	X				Correlation
Evans, Floyd, McGrew, and Leforgee (2002)	WJ III	WJ III COG	WJ III ACH	6.0–19.0 years	X	X			Regression
Flanagan (2000)	WJ-R	WJ-R COG WISC-R	WJ-R ACH	Grades 3–8	X	X			SEM
Floyd, Bergeron, and Alfonso (2006)	WJ III	WJ III COG	WJ III ACH	7.0–18.0 years	X	X	X		ANOVA[c]
Floyd, Keith, Taub, and McGrew (2007)	WJ III	WJ III COG	WJ III ACH	5.0–19.0 years	X				SEM
Floyd, Meisinger, Gregg, and Keith (2012)	WJ III	WJ III COG	WJ III ACH	5.0–19.0 years		X			SEM
Hajovsky, Reynolds, Floyd, Turek, and Keith (2014)	Linking sample	KABC-II	KTEA-II	Grades 1–12		X			SEM
Hale et al. (2008)	Linking sample	DAS-II	WIAT-II	6.0–17.0 years			X	X	Regression
Ismailer (2014)	School district	WJ III COG	WJ III ACH	Grades K–8		X			Regression[d]
Juarez (2012)	School district	WJ III COG	WJ III ACH	5.0–15.0 years	X				SEM
Keith (1999)	WJ-R	WJ-R COG	WJ-R ACH	Grades 1–9	X	X	X		SEM
McGill and Busse (2015)	WJ III	WJ-R COG	WJ-R ACH	6.0–18.9 years	X	X	X	X	Regression
McGill (2015)	KABC-II	KABC-II	KTEA-II	7.0–18.9 years	X	X	X	X	Regression
McGrew (2007)	WJ III	WJ III COG	WJ III ACH	6.0–19.0 years			X	X	Regression
McGrew and Knopik (1993)	WJ-R	WJ-R Gf-Gc	WJ-R ACH	5.0–10.0 years	X	X			Regression
McGrew and Hessler (1995)	WJ-R	WJ-R Gf-Gc	WJ-R ACH	5.0–19.0 years			X	X	Regression
Miller (2001)	School district	WJ-R COG DAS SB-IV WISC-III	WJ-R ACH	Grades 6–8	X	X			Regression
Nielsen (2012)	School district	WJ III COG	WJ III ACH	Grades 1–3		X			Correlation
Parkin and Beaujean (2012)	Linking sample	WISC-IV	WIAT-II	6.0–16.0 years				X	SEM
Proctor, Floyd, and Shaver (2005)	WJ III	WJ III COG	WJ III ACH	6.0–18.0 years	X		X	X	ANOVA
Vanderwood, McGrew, Flanagan, and Keith (2002)	WJ-R	WJ-R COG	WJ-R ACH	Grades K–12	X	X			SEM

Note: Broad Reading (BR); Reading Comprehension (RC); Basic Math (BM); Math Reasoning (MR).

[a] Clinical: Outpatient or outpatient testing facility; Linking sample: Linking sample for the COG and ACH batteries reported; WJ III: Woodcock-Johnson III standardization sample; WJ-R: Woodcock-Johnson Revised technical manual; Kaufman-II: Kaufman-II standardization sample.

[b] Analysis from which the present study extracted data. SEM: Structural equation modeling; CFA: Confirmatory factor analysis.

[c] Only the average achievement sample was included.

[d] Analysis controlled for age/gender in the first steps of the regression.

theory," "Gf-Gc," "Woodcock-Johnson-Revised," "WJ-R," "Woodcock-Johnson III," "WJ III," "Kaufman Assessment Battery for Children-Second Edition," "KABC-II," "Differential Abilities Scales (DAS)," "Differential Abilities Scales-Second Edition (DAS-II)," "Stanford-Binet-Fifth Edition," and "SB-IV."

The first and third authors of this study reviewed articles for inclusion as well as extracted data from them. Additionally, we contacted authors of the included studies to retrieve data when necessary, including McGrew (2007) and Floyd et al. (2006), who provided them upon request. Two articles in McGrew and Wendling were excluded: Ganci (2004), because the data could not be transformed into correlation coefficients; and McGrew, Flanagan, Keith, and Vanderwood (1997), because it was a review of the literature and would have caused the present analysis to double-count data. In sum, 12 new articles were added to the research reviewed by McGrew and Wendling (2010), for a total of 25 studies. Table 2 presents a summary of the articles included in the meta-analysis and important study characteristics.

2.1. Data extraction

For all articles included in our analyses, the following study and sample characteristics were coded: (a) source of the sample (e.g., a test's standardization data), (b) total sample size, (c) sample size by age group, (d) independent and dependent variables, and (e) and type of variable used (i.e., latent or observed). The age groups we used corresponded to those in McGrew and Wendling's (2010) review: 6–8, 9–13, and 14–19 years. Next, we extracted statistics on the relations between general and broad cognitive abilities and achievement variables from these studies (e.g., mean differences, correlations, beta-weights, and path coefficients). For studies that did not include correlations or estimated structural models with path coefficients, procedures outlined by Peterson and Brown (2005) were followed. Specifically, all beta coefficients were extracted and converted to correlations using the following formula:

$$r_{y1} = \beta_1 + r_{12}(r_{y2} - \beta_1 r_{12})$$

where: y represents the dependent variable and 1 and 2 represent predictor variables.

Using this method, we extracted correlations for all reported relationships in McGrew and Wendling and for included studies published since then.

2.2. Data analysis

A series of meta-analysis models were estimated to aggregate the relationship (r) between CHC cognitive abilities and academic achievement. Correlations between abilities and achievement were extracted by age group and then aggregated following procedures outlined in Borenstein, Hedges, Higgins, and Rothstein (2009), which involves calculating the mean of the correlations and then adjusting the variance by the covariance between the effect sizes by age group. Fixed-effects meta-analysis models were estimated for all possible relationships for the aggregated data and by age group. Given the small number of studies available and the fact that most of them used large standardization samples, which tend to be representative of the United States population, fixed-effects models were chosen. All correlations were transformed to Fisher's z for modeling, and the results were transformed back to correlations following Borenstein et al.'s (2009) recommendations. Our meta-analysis was conducted using the metafor package (Viechtbauer, 2010) in R (R Core Team, 2017). To examine the relations between general and broad cognitive abilities across age, we compared the 95% confidence intervals (CIs) surrounding the aggregated correlations (rs) across age groups. Non-overlapping CIs were indicative of differences in ability-achievement relations across age groups. When comparing two overlapping CIs, the second CI's point estimate (i.e., aggregated r) was considered the null hypothesis value. Thus, statistical significance was inferred by determining whether the first CI contained the null value of the second (Cumming & Finch, 2005).

3. Results

Table 3 presents the aggregated rs for the general and broad cognitive abilities of CHC theory by academic achievement area for the entire sample, as well as the lower (LL) and upper (UL) limits of the 95% CI encapsulating those rs. The number of studies (K) included in each analysis ranged from 1 to 12, with $M = 4.9$; and the number of participants (N) ranged from 3519 to 18,489, with $M = 7079.0$. As can be seen in this table, psychometric g and one or more broad cognitive abilities were substantially related to each area of achievement. Psychometric g had by far the strongest relations across all domains. Aggregated rs for g ranged from 0.717 (0.704, 0.730) for BM to 0.756 (0.748, 0.765) for RC, with $M = 0.737$. The average effect size for g was $r^2 = 0.540$. Moreover, none of the CIs for g overlapped with those for the broad abilities. In every achievement domain, psychometric g accounted for more variance than all broad cognitive abilities *combined*. The strongest relation observed between the broad abilities and achievement was between Gc and BR ($r = 0.450$ [0.434, 0.466]) and RC ($r = 0.451$ [0.439, 0.462]). Although each broad ability had a least one CI that did not include zero, most of the aggregated rs explained less than 10% of the variance in achievement, and none explained more than 20%. The only broad abilities to explain 10% of the variance or more in achievement were: (a) Gc, which explained 20% for BR and RC, and 10% for MR; (b) Gf, which explained 15% for BM and 10% for MR; and (c) Ga, which accounted for 11% for BR. In addition to these results, Gv was, for all intents and purposes, unrelated to academic achievement; all CIs for Gv contained zero, with one exception. In this case, the r between Gv and MR was 0.090 (0.059, 0.121), which accounted for less than 1% of the variance.

Tables 4–6 present the results of the relations between the general and broad abilities of CHC theory and academic achievement by age group. As shown in Table 4, results for the 6- to 8-year-olds were similar to those for the entire sample. Psychometric g had the strongest relations across all domains of academic achievement. Aggregated rs for g ranged from 0.644, (0.617, 0.670) for BR to 0.772

Table 3
Meta-analysis results for broad cognitive abilities by academic achievement area for the entire sample.

Achievement area	*N*	*K*	*r*	LL	UL
Basic Reading					
g	5873	5	0.737	0.725	0.748
Gf	6871	5	0.070	0.047	0.094
Gc	9719	8	0.450	0.434	0.466
Glr	6871	6	0.227	0.206	0.247
Gv	6841	4	0.014	−0.009	0.038
Ga	6821	4	0.337	0.316	0.358
Gsm	9693	6	0.283	0.265	0.301
Gs	8153	6	0.207	0.187	0.228
Reading Comprehension					
g	10,041	4	0.756	0.748	0.765
Gf	8404	6	0.110	0.089	0.132
Gc	18,489	12	0.451	0.439	0.462
Glr	8354	6	0.237	0.217	0.257
Gv	8319	4	−0.013	−0.034	0.009
Ga	9499	6	0.180	0.160	0.199
Gsm	13,343	6	0.155	0.138	0.172
Gs	8570	7	0.121	0.100	0.142
Basic Mathematics					
g	5357	2	0.717	0.704	0.730
Gf	6465	5	0.387	0.366	0.408
Gc	9339	6	0.301	0.283	0.320
Glr	3982	4	0.091	0.060	0.122
Gv	3982	4	0.020	−0.011	0.051
Ga	3519	1	0.031	−0.002	0.064
Gsm	3982	4	0.095	0.064	0.126
Gs	9339	6	0.234	0.214	0.253
Mathematics Reasoning					
g	–	–	–	–	–
Gf	4005	4	0.317	0.289	0.345
Gc	4005	4	0.321	0.293	0.348
Glr	4005	4	0.074	0.043	0.105
Gv	4005	4	0.090	0.059	0.121
Ga	3592	2	0.029	−0.004	0.062
Gsm	4005	4	0.120	0.009	0.150
Gs	4005	4	0.211	0.181	0.240

Note. *r* = Aggregated correlation coefficient. 95% Confidence Interval: LL = Lower Limit, UL = Upper Limit. Fluid Reasoning (Gf), Comprehension-Knowledge (Gc), Long-Term Storage and Retrieval (Glr), Visual Processing (Gv), Auditory Processing (Ga), Short-Term Memory (Gsm), Processing Speed (Gs).

(0.752–0.790) for BM, with $M = 0.71$. The average effect size of for *g* was 0.52; and none of the CIs for the broad abilities overlapped with those for *g*. The broad ability with the highest aggregated *r* with BR was Ga ($r = 0.438$ [0.401, 0.474]), accounting for 19% of the variance. The only other broad ability to explain 10% of the variance or more in BR was Gc ($r = 0.319$ [0.282, 0.356]). Three broad abilities accounted for 10% or more variance in RC. The strongest of the relations was for Glr ($r = 0.489$ [0.454, 0.522]), which explained 24% of the variance, followed by Gc ($r = 0.374$ [0.346, 0.401]) and Ga ($r = 0.325$ [0.285, 0.364]). For BM, only Gf ($r = 0.370$ [0.337, 0.402]) accounted for at least 10% of the variance in achievement. None of the broad abilities accounted for 10% or more variance in MR.

As shown in Table 5, relations between *g* and each achievement domain for the 9- to 13-year old groups were generally stronger than they were for 6- to 8-year-olds. Aggregated *r*s for *g* ranged from 0.730 (0.709, 0.750) for BM to 0.764 (0.748–0.780) for RC, with $M = 0.740$. The average effect size for *g* was 0.55. For BR and RC, only Gc explained at least 10% of the variance in achievement (BR: $r = 0.487$ [0.462, 0.510]; RC: $r = 0.562$ [0.542, 0.581]), accounting for 24% of the variance in BR and 32% in RC. In addition, the only broad ability to account for 10% or more of the variance in BM was Gf ($r = 0.370$ [0.337, 0.402]); none did so for MR.

Table 6 presents results for the 14- to 19-year-olds. As with the other age groups, psychometric *g* had the strongest relationship across all domains of academic achievement. Aggregated *r*s for *g* ranged from 0.662 (0.632, 0.689) for BR to 0.844 (0.825, 0.862) for BM, with $M = 0.780$. The average effect size of for *g* was 0.61. Again, none of the CIs for the broad abilities overlapped with those for *g* in any achievement domain. For BR, two broad abilities explained more than 10% of the variance—Gc ($r = 0.543$ [0.515, 0.570]) and Gsm ($r = 0.405$ [0.350, 0.456])—accounting for 29% and 16% of the variance, respectively. Only one broad ability, Gc ($r = 0.572$ [0.555, 0.588]), accounted for at least 10% of the variance in RC. Gc attributed 33% of the variance to RC, the most of any broad ability across all achievement domains and age groups. Regarding mathematics, only Gc ($r = 0.337$ [0.300, 0.374], $r^2 = 0.11$) explained 10% or more of the variance in BM; and Gf ($r = 0.325$ [0.277, 0.371]) and Gc ($r = 0.312$ [0.264, 0.359]) accounted for 11% and 10% of the variance in MR, respectively.

In addition to these analyses, we also examined trends in the relations between the CHC abilities and achievement by academic

Table 4
Meta-analysis results of relations between CHC cognitive abilities and academic achievement for 6- to 8-year-olds.

Achievement area	*N*	*K*	*r*	LL	UL
Basic Reading					
g	1922	4	0.644	0.617	0.670
Gf	1922	2	0.107	0.063	0.015
Gc	2245	3	0.319	0.282	0.356
Glr	1177	2	0.260	0.206	0.313
Gv	1922	2	0.001	−0.043	0.046
Ga	1922	2	0.438	0.401	0.474
Gsm	2245	3	0.209	0.169	0.248
Gs	2245	3	0.306	0.268	0.343
Reading Comprehension					
g	1951	2	0.708	0.685	0.729
Gf	1913	2	0.098	0.053	0.142
Gc	3835	6	0.374	0.346	0.401
Glr	1913	2	0.489	0.454	0.522
Gv	1913	2	−0.026	−0.070	0.019
Ga	1968	3	0.325	0.285	0.364
Gsm	1913	2	0.108	0.063	0.152
Gs	1913	2	0.220	0.177	0.263
Basic Mathematics					
g	1824	2	0.772	0.752	0.790
Gf	2678	3	0.370	0.337	0.402
Gc	1999	2	0.254	0.213	0.295
Glr	854	1	0.037	−0.030	0.104
Gv	854	1	0.003	−0.064	0.070
Ga	854	1	0.107	0.040	0.173
Gsm	854	1	0.033	−0.034	0.100
Gs	2678	3	0.275	0.240	0.310
Mathematics Reasoning					
g	–	–	–	–	–
Gf	854	1	0.257	0.193	0.319
Gc	854	1	0.210	0.145	0.223
Glr	854	1	0.047	−0.020	0.114
Gv	854	1	0.073	−0.006	0.139
Ga	854	1	0.050	−0.017	0.117
Gsm	854	1	0.063	−0.004	0.129
Gs	2678	1	0.250	0.186	0.312

Note. *r* = Aggregated correlation coefficient. 95% Confidence Interval: LL = Lower Limit, UL = Upper Limit. Fluid Reasoning (Gf), Comprehension-Knowledge (Gc), Long-Term Storage and Retrieval (Glr), Visual Processing (Gv), Auditory Processing (Ga), Short-Term Memory (Gsm), Processing Speed (Gs).

domain and age. Here, we discuss only those age-related differences for cognitive abilities that explained at least 10% of the variance in one or more achievement domains for at least one age group. Results of these analyses revealed monotonically increasing trends in the aggregated *r*s across age groups for psychometric *g* and BR and RC. For BR, the *r*s increased from 0.644 (0.617, 0.670) for the 6–8 years age group, to 0.734 [0.715, 0.752] for the 9–13 years age group, and to 0.844 (0.825, 0.862) for the 14–19 years age group. For RC, *r*s increased from 0.708 (0.685, 0.729) for the 6–8 years age group, to 0.764 (0.748, 0.780) for the 9–13 years age group, and to 0.820 (0.807, 0.833) for the 14–19 years age group. The only other monotonically increasing trend in *r* across age groups was for Gc and BR, where *r*s increased from 0.319 (0.282, 0.356) for the 6–8 years age group, to 0.487 (0.462, 0.510) for the 9–13 years age group, and to 0.543 (0.515, 0.570) for the 14–19 years age group. One monotonically decreasing trend was observed for *g* and BM. In this case, the *r*s decreased from 0.772 (0.752, 0.790) for the 6–8 years age group, to 0.730 (0.709, 0.750) for the 9–13 years age group, and to 0.622 (0.632, 0.689) for the 14–19 years age group.

In addition to these results, several differences in ability-achievement relations were noted between adjacent age groups. For Gsm and BR, the CI for the aggregated *r* for the 14–19 years age group (0.405 [0.350, 0.456]) did not overlap with those for the 6–8 years (0.209 [0.169, 0.248]) and 9–13 years (0.268 [0.238, 0.297]) age groups. Only one other increase in *r* was observed between one age group and the next. This was for Gc and BM. Here, the CI for *r* for the 14–19 years age group (0.337 [0.300, 0.374]) did not overlap with those for the 6–8 years (0.254 [0.213, 0.295]) and 9–13 years (0.262 [0.230, 0.294]) age groups. Decreases in *r*s were also observed for several broad abilities across age groups. For Glr and RC, the CI for the 6–8 years age group (0.489, [0.454, 0.522]) did not overlap with that of 9–13 years age group (0.162, [0.128, 0.195]), and the CI for the 9–13 age group did not contain the point estimate for the 14–19 years age group (0.098, [0.062, 0.133]). A similar pattern was observed for the *r*s between Ga and RC and between Gf and MR. For Ga and RC, the *r*s decreased from 0.325 (0.285, 0.364) for the 6–8 years age group to 0.095 (0.060, 0.129) for the 9–13 years age group and 0.059 (0.023, 0.095) for the 14–19 years age group. For Gf and MR, the *r* for the 6–8 years group (0.370, [0.337, 0.402]) was not contained within the 9–13 years (0.337, [0.306, 0.367]) or the 14–19 years (0.110, [0.062, 0.158]) age groups.

Table 5
Meta-analysis results of relations between CHC cognitive abilities and academic achievement for 9- to 13-year-olds.

Achievement area	*N*	*K*	*r*	LL	UL
Basic Reading					
g	2455	5	0.734	0.715	0.752
Gf	3246	3	0.045	0.011	0.080
Gc	3794	4	0.487	0.462	0.510
Glr	3246	3	0.210	0.177	0.242
Gv	3216	2	−0.040	−0.075	−0.006
Ga	3246	3	0.216	0.183	0.248
Gsm	3764	3	0.268	0.238	0.297
Gs	3246	3	0.154	0.120	0.187
Reading Comprehension					
g	2552	2	0.764	0.748	0.780
Gf	3246	3	0.129	0.095	0.163
Gc	4679	7	0.562	0.542	0.581
Glr	3246	3	0.162	0.128	0.195
Gv	3216	2	−0.052	−0.086	−0.018
Ga	3246	3	0.095	0.060	0.129
Gsm	3216	2	0.085	0.051	0.120
Gs	3412	4	0.110	0.076	0.143
Basic Mathematics					
g	2019	2	0.730	0.709	0.750
Gf	3306	3	0.337	0.306	0.367
Gc	3306	3	0.262	0.230	0.294
Glr	1995	1	0.120	0.077	0.163
Gv	1287	1	−0.082	−0.136	−0.028
Ga	1287	1	0.020	−0.035	0.075
Gsm	1287	1	0.076	0.021	0.130
Gs	3306	3	0.155	0.121	0.188
Mathematics Reasoning					
g	–	–	–	–	–
Gf	856	1	0.257	0.193	0.319
Gc	856	1	0.210	0.145	0.273
Glr	856	1	0.047	−0.020	0.114
Gv	856	1	0.073	0.006	0.139
Ga	856	1	0.050	−0.017	0.117
Gsm	856	1	0.063	−0.004	0.130
Gs	856	1	0.250	0.186	0.312

Note. *r* = Aggregated correlation coefficient. 95% Confidence Interval: LL = Lower Limit, UL = Upper Limit. Fluid Reasoning (Gf), Comprehension-Knowledge (Gc), Long-Term Storage and Retrieval (Glr), Visual Processing (Gv), Auditory Processing (Ga), Short-Term Memory (Gsm), Processing Speed (Gs).

4. Discussion

Today, the primary focus of intelligence test interpretation is widely recommended to be on normative strengths and weaknesses in an individual's index scores (e.g., Flanagan & Alfonso, 2017), which are seen to measure broad cognitive abilities at Stratum II in CHC theory (e.g., Schneider & McGrew, 2012). Decisions about which abilities should be assessed for the diagnosis of academic learning difficulties are based on literature reviews by Flanagan et al. (2006) and by McGrew and Wendling (2010). Both of these research syntheses, however, were traditional narrative reviews, which have been soundly criticized over the past two decades (for a review, see Gurevitch, Koricheva, Nakagawa, & Stewart, 2018). In addition, neither of these reviews attempted to estimate of the size of the relations between CHC cognitive abilities and academic achievement. The purpose of this study, therefore, was to conduct a meta-analysis to determine the magnitude of these relations across age.

Results of our analyses show that psychometric *g* and one or more broad cognitive abilities are substantially related to each area of academic achievement. For the overall sample, *g* had by far the strongest relations across all domains of academic achievement. On average, *g* explained more than 50% of the variance in achievement, which was more than all of the broad cognitive abilities taken as a whole. In addition, the pattern of its relations with achievement is consistent with what is known about psychometric *g*. Because *g* is related to the complexity of information processing (see Jensen, 1998), educational outcomes that involve perceiving abstract or conceptual relationships are predicted best by it, such as those required by RC and MR. In contrast, outcomes that primarily involve rote learning or associative memory (e.g., BR and BM) are predicted less well by *g*.

Cohen (1988) asserted that $|rs|$ = 0.10, 0.30, and 0.50 can be classified as "small," "medium," and "large," respectively. According to these benchmarks, the relations between psychometric *g* and academic achievement can be described as large. Although the vast majority of relations between the broad abilities and achievement had CIs that did not contain zero, their effect sizes were generally small. For the total sample, only one or two broad abilities had what can be classified as medium effects per achievement domain. Moreover, Gc was the only broad ability that had a medium effect across all four domains. The largest effect sizes for the

Table 6
Meta-analysis results of relations between CHC cognitive abilities and academic achievement for 14- to 19-year-olds.

Achievement area	*N*	*K*	*r*	LL	UL
Basic Reading					
g	946	2	0.844	0.825	0.862
Gf	1585	1	0.047	−0.002	0.096
Gc	2546	4	0.543	0.515	0.570
Glr	1585	1	0.095	0.046	0.144
Gv	1585	1	0.121	0.072	0.169
Ga	2961	2	0.059	0.023	0.095
Gsm	961	3	0.405	0.350	0.456
Gs	1585	1	0.063	0.014	0.112
Reading Comprehension					
g	2582	2	0.820	0.807	0.833
Gf	2961	2	0.101	0.066	0.137
Gc	3750	5	0.572	0.555	0.588
Glr	2961	2	0.098	0.062	0.133
Gv	2961	2	0.032	−0.005	0.068
Ga	2961	2	0.059	0.023	0.095
Gsm	2961	2	0.116	0.090	0.142
Gs	2961	2	0.027	−0.009	0.063
Basic Mathematics					
g	1514	2	0.662	0.632	0.689
Gf	2185	2	0.110	0.062	0.158
Gc	2892	3	0.337	0.300	0.374
Glr	1378	1	0.108	0.056	0.160
Gv	1378	1	−0.060	−0.113	−0.007
Ga	1378	1	−0.035	−0.088	0.018
Gsm	1378	1	0.040	−0.013	0.093
Gs	2085	2	0.236	0.195	0.277
Mathematics Reasoning					
g	–	–	–	–	–
Gf	1378	1	0.325	0.277	0.371
Gc	1378	1	0.312	0.264	0.359
Glr	1378	1	0.040	−0.013	0.093
Gv	1378	1	−0.002	−0.055	0.051
Ga	1378	1	−0.002	−0.055	0.051
Gsm	1378	1	0.020	−0.033	0.073
Gs	1378	1	0.180	0.128	0.231

Note. *r* = Aggregated correlation coefficient. 95% Confidence Interval: LL = Lower Limit, UL = Upper Limit. Fluid Reasoning (Gf), Comprehension-Knowledge (Gc), Long-Term Storage and Retrieval (Glr), Visual Processing (Gv), Auditory Processing (Ga), Short-Term Memory (Gsm), Processing Speed (Gs).

broad abilities were between Gc and BR and RC. In both cases, Gc accounted for 20% of the variance. In addition, Gf had a medium effect on BM and MR; and Ga had a medium effect on BR. All of the other CHC cognitive abilities had either a small effect across achievement domains or none at all. Results for the total sample, therefore, are consistent with prior research showing that the predictive validity of intelligence tests in educational settings is primarily a function of *g* (e.g., Jensen, 1998).

Within each of the three age groups, psychometric *g* had a similarly strong effect across all domains of academic achievement. All of the effects sizes for *g* were large and none of the CIs for *g* and the broad abilities overlapped. With regard to the effects of the broad abilities, only Gc had effect sizes that can be described as large and these were only observed with BR and RC for the older age groups. For the 6–8 years old group, Ga, Gc, and Gs had medium effects for BR. For RC, Glr had a medium effect in addition to Gc. The only broad ability to have a medium effect for BM within this age group was Gf; and none had at least a medium effect on MR. For the 9–13 years old group, Gc had a medium effect on BR and a large effect on RC; and Gf had a medium effect for BM. Similar results were observed for the 14–19 years old group. Gc had a large effect for BR and RC and a medium effect for BM and MR; and Gf also had a medium effect for MR. Although the CIs for the other broad abilities tended not contain zero across achievement domains, with few exceptions, their effects were generally small.

We also examined trends in the relations between the general and broad abilities and academic achievement across age. Results of these analyses revealed monotonically increasing trends across age groups for *g* and BR and RC. This trend is likely related to the so-called "Matthew effect" in reading (e.g., Stanovich, 1986), which stems from individual differences in the acquisition of literacy skills. Over time, individuals with higher general ability tend to have a cumulative advantage over those with lower ability, because they benefit more from new educational experiences. As Walberg, Strykowski, Roval, and Hung (1984) stated:

> Those who did well at the start may have been more often, or more intensively, rewarded for their early accomplishments; early intellectual and motivational capital may grow for longer periods and at greater rates; and large funds and continuing high growth rates of information and motivation may be more intensely rewarded. (p. 92)

The Matthew effect may also explain the monotonically increasing trend observed for Gc and BR. In addition to these results, one monotonically decreasing trend was observed between *g* and BM. While this may imply the decreasing importance of *g* in basic mathematics, perhaps due to increased restriction of range in the automatization of calculation skills over time, it is important to note that the *r* between *g* and BM for the 9–14 years group was still twice the size of the largest *r* for all the broad abilities.

In addition to these monotonically increasing and decreasing trends, differences between some, but not all, age groups were also observed. These included both decreases and increases in broad ability-achievement relations. For example, the *r*s between Glr and RC and between Ga and RC deceased from the 6–8 years group to the older two age groups; and the *r* between MR and Gf was substantially smaller for the 9–14 years group than those for the two younger age groups. Such patterns of decreasing relations across age suggest that these cognitive abilities are more essential during the early stages of learning and become less so after mastery of basic skills. The opposite may be the case for those relations that increased at older ages.

It is important to note that the research included in our meta-analysis on the relations between CHC cognitive abilities and academic achievement did not examine incremental predictive validity and consequently did not report semi-partial *r*s. Nevertheless, several studies have been conducted on the incremental validity of different intelligence tests in the prediction of academic achievement (e.g., Canivez, 2013; Freberg, Vandiver, Watkins, & Canivez, 2008; Glutting et al., 2006; Kranzler, Benson, & Floyd, 2015). For example, Canivez (2013) found that composite scores on the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler, 2008) rarely accounted for more than 2% of the variance in academic achievement beyond that explained by the overall score and often less than 1%. Kranzler et al. (2015), in contrast, used orthogonal factor scores derived from a bifactor model to examine the predictiveness of latent constructs measured by the WAIS-IV. They found that, in addition to the ubiquitous effect of *g*, Gc was the only ability to explain more than 15% of the variance and only for achievement in reading, spelling, and oral communication skills. Results of their study, therefore, which are consistent with our meta-analysis results, indicate that only *g* and Gc are importantly related to academic achievement, and for Gc only in some areas of achievement.

4.1. Comparison with McGrew and Wendling (2010)

Before comparing the results of our meta-analysis to McGrew and Wendling's (2010) narrative research synthesis, it is important to note that we included 12 studies that were not available to them, because they were published after their review. Thus, some differences are to be expected based on our analyses of a somewhat different set of studies. That said, McGrew and Wendling identified "important" relations between CHC abilities and academic achievement based on the percentage of studies reporting statistically significant results and not in terms of effect size. This may have misleading results because most of the studies they reviewed used large standardization samples, whereas we estimated the actual magnitude of those effects.

Results of our meta-analysis of CHC ability-achievement relations both refute some of their conclusions and support others. First and foremost, our findings do not support de-emphasis of the interpretation of the overall score on intelligence tests, which measures psychometric *g* (e.g., Farmer, Floyd, Reynolds, & Kranzler, 2014). Across all academic domains and for all age groups, *g* had by far the strongest relations, with large effect sizes. The variance explained by *g* was typically more than that accounted for by all the broad abilities combined. Only Gc had a medium to large effect for each achievement domain, with its largest effects for reading with older students. For the other broad abilities, with few exceptions, effect sizes tended to be small and explained less than 10% of the variance in achievement.

In addition, results of our meta-analysis provide mixed support for the Ability by Achievement by Age three-way interaction effects identified by McGrew and Wendling (2010). Table 7 displays the broad cognitive abilities with at least medium effects on academic achievement by age group according to our results. As can be seen here, our findings support the "importance" of Gc across all achievement domains, but not all age levels. Gf is also importantly related to mathematics, but its relations vary by domain and age group. In addition, our results support the importance of the relation between Gsm and BR for the 9- to 14-year-old age group, but not for younger children. The relations between Glr and reading also differ from McGrew and Wending. We did not find Glr to be importantly related to BR for 6- to 8-year-olds, but it appears to be so for RC. Last, we found that Ga is importantly related to BR. In

Table 7
Broad cognitive abilities in the Cattell-Horn-Carroll theory with medium to large effects on academic achievement by age group.

Achievement area	Age group		
	6–8 years	9–13 years	14–19 years
Basic Reading Skills	Gc Ga Gs	Gc	Gc Gsm
Reading Comprehension	Gc Ga Glr	Gc	Gc
Basic Math Skills	Gf	Gf	Gc
Math Reasoning			Gc Gf

Note. Comprehension-Knowledge (Gc), Fluid Reasoning (Gf), Short-Term Memory (Gsm), Long-Term Storage and Retrieval (Glr), Processing Speed (Gs), and Auditory Processing (Ga).

sum, results of our meta-analysis provide only partial support for the three-way interactions identified by McGrew and Wendling.

4.2. Limitations

Results of our meta-analyses of the relations between the broad cognitive abilities of CHC theory and academic achievement are susceptible to specification error. As McGrew and Wendling (2010) stated,

> *specification error*, sometimes called omitted-variable bias (Keith, 2006), occurs when potentially important variables in predictive or explanatory research are not included in a study's research design. Specification error can lead to biased estimates of the effects (i.e., relative importance) of individual predictor variables on the DV of interest. (p. 653, emphasis in the original)

Most of the studies included in our meta-analysis used correlation or regression to examine relations between academic achievement and composite scores on intelligence tests as measures of broad cognitive abilities. Many of these studies omitted measurement of psychometric *g*. Composite scores on intelligence tests are not "pure" measures of the broad ability constructs they are intended to measure. Rather, they reflect variance related to these constructs, but also variance related to psychometric *g*—to a rather large extent in most cases—specific abilities, and measurement error. Because of their factorial complexity, and the failure to control for *g* in these studies, our estimates of the relations between broad abilities and academic achievement are likely somewhat inflated. Further research on the relations between CHC broad cognitive abilities and academic achievement after *g* is accounted for is needed.

Another limitation of this study concerns the independence of the research included in our meta-analysis. Most of the investigations on CHC cognitive ability-achievement relations were conducted with the norming sample data from the WJ III NU (Woodcock et al., 2007). In meta-analysis, it is best to combine the results of independent samples to estimate a pooled effect size. The generalizability of our meta-analysis results may be influenced by samples contributing more than one effect size to the model. Thus, due to widespread use of the WJ standardization data, results of our meta-analysis – as well as virtually all prior research on CHC ability-achievement relations – may lack generalizability to the CHC cognitive abilities assessed by other intelligence tests.

Research on the relations between CHC abilities and academic achievement assumes that the composite scores on intelligences are valid measures of the cognitive constructs that they are intended to assess. As Messick (1995) stated, "the internal structure of the assessment (i.e., interrelations among the scored aspects of tasks and subtask performance) should be consistent with what is known about the internal structure of the domain" (p. 746). When the results of factor analysis indicate that the observed variables do not covary as predicted by the theoretical structure of the test, then those variables lack structural fidelity. Independent research by Dombrowski (2013) on the internal structure of the WJ-III, which was used in the majority of studies on CHC ability-achievement relations, suggests that it is over-factored. In this study, he used higher-order exploratory factor analysis to examine the structural fidelity of the WJ-III with groups of children aged 9–13 and 14–19 years. Results these analyses indicated the presence of a robust second-order psychometric *g* factor, but only three first-order factors for the 9- to 13-year-olds (Gc, Gs, and Glr) and two for the 14- to 19-year-olds (Gs and Glr). Results of Dombrowski's study suggest that the composite scores of the WJ-III are not completely aligned with the test's underlying theoretical structure. Dombrowski, McGill, and Canivez (2018b) obtained similar results for the Woodcock-Johnson Tests of Cognitive Ability-Fourth Edition (WJ IV; Schrank, McGrew, & Mather, 2014). Thus, given concerns surrounding the fidelity of its internal structure, results of research using the WJ to examine relations between broad cognitive abilities and academic achievement are questionable, at least for some abilities at Stratum II of CHC theory.

Last, we examined trends in the relations between cognitive abilities and academic achievement across age groups. To do so, we compared the aggregated *r*s and CIs for three age groups. Because we conducted a meta-analysis, our analyses were based on the data from prior research, most of which was based on norming samples from standardized tests of intelligence, particularly the WJ-III. Thus, the examination of age-related change in this study was cross-sectional, not longitudinal. Because cross-sectional research examines different groups of participants at one point in time, rather than the same participants over time, the trends in cognitive ability-academic achievement relations across age observed in this study may reflect to some degree differences between groups of participants sampled from the population. To the best of our knowledge, no longitudinal research has been conducted on the relations between CHC abilities and academic across age. Research using a longitudinal design is therefore needed to substantiate our meta-analysis results.

4.3. Implications for practice

McGrew and Wendling (2010) asked, "What CHC broad or narrow cognitive abilities hold promise either as early screening markers or collectively as pattern indicators of a potential SLD process disorder?" (p. 652). A marker test is used to identify the presence of something. For example, in medicine, blood tests are often conducted to look for certain proteins and circulating tumor cells that are indicative of cancer. If the blood test is positive, then the presence of a cancer tumor is highly likely. If CHC broad abilities are to be used as "markers," then those abilities should be closely related to academic achievement; and if those abilities are to be used as "pattern indicators," specific patterns of those abilities should be capable of reliably grouping children and adolescents with and without SLD.

Use of Cohen's (1988) benchmarks for the interpretation of the effect size of correlations is controversial and their generalizability to applied practice is arguable. With regard to the use of CHC abilities as markers, it is important to note that Cohen's (1988) benchmarks were intended for interpretation of effect sizes *in the social and behavioral sciences*. Further, as Gignac and Szodorai (2016) asserted, Cohen did not substantiate his benchmarks with a quantitative analysis of the data. Nevertheless, as our meta-analysis showed, almost all relations between the broad abilities and academic achievement were lower than 0.50, and often much lower. Given these small to moderate correlations for the broad abilities, interpretation of composite scores on intelligence tests as "markers" will result in many false negatives and positives. Regarding their interpretation as "pattern indicators," empirical research on the PSW methods of SLD identification has found that these methods do not reliably classify children and youth with and without SLD (Kranzler, Floyd, Benson, Zaboski, & Thibodaux, 2016a, 2016b; Miciak, Fletcher, Stuebing, Vaughn, & Tolar, 2014; Miciak, Taylor, Stuebing, & Fletcher, 2018; Stuebing, Fletcher, Branum-Martin, & Francis, 2012; Taylor, Miciak, Fletcher, & Francis, 2017). Results of the extant empirical research suggest that school psychologists using the PSW methods will spend a great deal of time conducting assessments that have a very low probability of accurately identifying true SLD (e.g., Kranzler et al., 2016b).

5. Conclusion

We advocate for constrained and evidence-based interpretation of intelligence tests. Although an improvement over the ipsative analysis of cognitive profiles, results of our meta-analysis do not support the widely recommended focus on index scores as the primary level of clinical interpretation. Instead, results of our analyses support the interpretation of the overall score on intelligence tests as a measure of psychometric *g* for diagnosing difficulties in academic achievement. Given the likelihood of specification error resulting from the failure to control for psychometric *g* in most prior research on CHC ability-academic achievement relations, at the current time the empirical research only supports the interpretation of index scores representing Gc when assessing learning difficulties in reading. Further research substantiating the adequacy and appropriateness of inferences and actions based on other intelligence test index scores is needed before other specific recommendations for practice can be made.

Declaration of interest

None.

References[1]

Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *The Journal of Wildlife Management,* 912–923.

*Armstrong, M. M. (2011). *The relationship between intelligence, achievement, and psychological diagnosis: Applying the CHC theory to clinical diagnosis.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).

*Bacal, E. (2010). *The Woodcock Johnson Three and math learning disabilities.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).

*Beaujean, A. A., Parkin, J., & Parker, S. (2014). Comparing Cattell-Horn-Carroll factor models: Differences between bifactor and higher order factor models in predicting language achievement. *Psychological Assessment, 26*, 789–805.

Benson, N., Floyd, R. G., Kranzler, J. H., Eckert, T. L., & Fefer, S. (2018, February). *Contemporary assessment practices in school psychology: National survey results. Paper presented at the annual convention of the National Association of School Psychologists. Chicago, IL*.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, England: Wiley.

Braden, J. P., & Shaw, S. R. (2009). Intervention validity of cognitive assessment: Knowns, unknowables, and unknowns. *Assessment for Effective Intervention, 34*, 106–115.

*Brine, A. (2013). *Linking cognitive and academic profiles to children with Asperger's disorder.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Burns, M. K., Peterson-Brown, S., Haegele, K., Rodriguez, M., Schmitt, B., Cooper, M., ... VanDerHeyden, A. M. (2016). Meta-analysis of academic interventions derived from neuropsychological data. *School Psychology Quarterly, 31*, 28–42.

Canivez, G. L. (2013). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment, 25*, 484–495.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 28*, 975–986.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological assessment, 29*, 458–472.

Canivez, G. L., Dombrowski, S. C., & Watkins, M. W. (2018). Factor structure of the WISC-V in four standardization age groups: Exploratory and hierarchical factor analyses with the 16 primary and secondary subtests. *Psychology in the Schools, 55*, 741–769.

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

Cattell, R. B. (1983). The clinical use of difference scores: Some psychometric problems. *Multivariate Experimental Clinical Research, 6*, 87–98.

Coan, R. W. (1964). Facts, factors, and artifacts: The quest for psychological meaning. *Psychological Review, 71*, 123–140.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach.* Thousand Oaks, CA: Sage.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *The Psychological Bulletin, 50*, 456–473.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* Oxford, England: Irvington.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170–180.

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III Cognitive at school age. *School Psychology Quarterly, 28*, 154–169.

Dombrowski, S. C., Canivez, G. L., & Watkins, M. W. (2018a). Factor structure of the 10 WISC-V primary subtests across four standardization age groups. *Contemporary School Psychology, 22*, 90–104.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018b). An alternative conceptualization of the theoretical structure of the Woodcock-Johnson IV Tests of Cognitive Abilities at school age: A confirmatory factor analytic investigation. *Archives of Scientific Psychology, 6*, 1–13.

Elliott, J. G., & Resing, W. C. (2015). Can intelligence testing inform educational intervention for children with reading disability? *Journal of Intelligence, 3*, 137–157.

*Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31*, 246–262.

Farmer, R. L., Floyd, R. G., Reynolds, M. R., & Kranzler, J. H. (2014). IQs are very strong but imperfect indicators of psychometric g: Results from joint confirmatory

[1] * denotes studies included in the Meta-Analysis.

factor analysis. *Psychology in the Schools, 51*, 801–813.
*Flanagan, D. P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly, 15*, 295–329.
Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment.* Hoboken, NJ: Wiley.
Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, P. L. Harrison, & L. Patti (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314–325). New York, NY: Guilford.
Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.
Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. (2006). *The achievement test desk reference (ATDR): A guide to learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.
Flanagan, D., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). New York, NY: Wiley.
*Floyd, R. G., Bergeron, R., & Alfonso, V. C. (2006). Cattell-Horn-Carroll cognitive ability profiles of poor comprehenders. *Reading and Writing, 19*, 427–456.
*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., Meisinger, E., Gregg, N., & Keith, T. (2012). An explanation of reading comprehension across development using models from Cattell-Horn-Carroll theory: Support for integrative models of reading. *Psychology in the Schools, 49*, 725–743.
Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied Neuropsychology, 15*, 131–139.
Ganci, M. (2004). *The diagnostic validity of a developmental neuropsychological assessment (NEPSY)-Wechsler Intelligence Scale for Children-Third Edition (WISC-III)-based cross-battery assessment.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).
Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78.
Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40*, 103–114.
Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*, 175–182.
*Hajovsky, D., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review, 43*, 385–406.
*Hale, J. B., Fiorello, C. A., Dumont, R., Willis, J. O., Rackley, C., & Elliott, C. (2008). Differential Ability Scales–Second Edition (neuro) psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools, 45*, 838–858.
Horn, J. L., & Noll, J. (1997). Human cognitive abilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). (2nd ed.). New York, NY: Guilford Press.
*Ismailer, S. S. (2014). *Relationships between Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement within a sample of K–8 students with learning disabilities.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).
Jensen, A. R. (1998). *The g factor: The science of mental ability. Westport, CT.* Praeger.
*Juarez, B. M. (2012). *Effects of general and specific cognitive abilities on reading achievement in a referred sample.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).
Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A history of intelligence test interpretation. In D. P. Flanagan, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 56–70). (3rd ed.). New York, NY: Guilford Press.
Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York, NY: Wiley.
Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York, NY: Wiley.
Kearns, D. M., & Fuchs, D. (2013). Does cognitively-focused instruction improve the academic performance of low-achieving students? *Exceptional Children, 79*, 263–290.
*Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly, 14*, 239–262.
Keith, T. Z. (2006). *Multiple regression and beyond.* Boston: Allyn & Bacon.
Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide.* New York, NY: Guilford Press.
Kranzler, J. H., Benson, N., & Floyd, R. G. (2015). Using estimated factor scores from a bifactor analysis to examine the unique effects of the latent variables measured by the WAIS-IV on academic achievement. *Psychological Assessment, 27*, 1402–1416.
Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016a). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School and Educational Psychology, 3*, 124–136.
Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016b). Cross-battery assessment pattern of strengths and weaknesses approach to the identification of specific learning disorders: Evidence-based practice or pseudoscience? *International Journal of School and Educational Psychology, 3*, 146–157.
Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "intelligent testing" approach to the WISC-III. *School Psychology Quarterly, 12*, 197–234.
Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly, 30*, 457–469.
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.
*McGill, R. J. (2015). Interpretation of KABC-II scores: An evaluation of the incremental validity of Cattell-Horn-Carroll (CHC) factor scores in predicting achievement. *Psychological Assessment, 27*, 1417–1426.
*McGill, R. J., & Busse, R. T. (2015). Incremental validity of the WJ III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly, 30*, 353–365.
*McGrew, K. (2007, January). *Prediction of WJ III reading and math achievement by WJ III cognitive and language tests.* (Unpublished Raw Data).
*McGrew, K. S., & Hessler, G. L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment, 13*, 21–38.
*McGrew, K. S., & Knopik, S. N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life-span. *School Psychology Review, 22*, 687–695.
McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-ability achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*, 651–675.
McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26*, 189–210.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.
Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly, 29*, 21–37.
Miciak, J., Taylor, W. P., Stuebing, K. K., & Fletcher, J. M. (2018). Simulation of LD identification accuracy using a pattern of processing strengths and weaknesses method with multiple measures. *Journal of Psychoeducational Assessment, 36*, 21–33.
*Miller, B. D. (2001). *Using Cattell–Horn–Carroll cross-battery assessment to predict reading achievement in learning disabled middle school students.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).
*Nielsen, L. M. (2012). *An investigation of cognitive measures of elementary students identified as having a specific learning disability.* ((Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global).

*Parkin, J. R., & Beaujean, A. A. (2012). The effects of Wechsler Intelligence Scale for Children-Fourth Edition cognitive abilities on math achievement. *Journal of School Psychology, 50*, 113–128.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*, 175–181.

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

R Core Team (2017). R: A language and environment for statistical computing [Programming language]. Retrieved from https://www.r-project.org.

Rapaport, D., Gil, R., & Schafer, R. (1945–1946). *Manual of psychological testing-I: Diagnostic testing of intelligence and concept formation.* New York: The Josiah Macy Jr. Foundation.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fifth edition: What does it measure? *Intelligence, 62*, 31–47.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan, & P. Harrison (Eds.). *Contemporary intellectual assessment* (pp. 99–144). (3rd ed.). New York, NY: Guilford Press.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Smith, C. B., & Watkins, M. W. (2004). Diagnostic utility of the Bannatyne WISC-III Pattern. *Learning Disabilities Research and Practice, 19*, 49–56.

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51*, 1031–1045.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41*, 3–22.

Stuebing, K. K., Barth, A. E., Trahan, L. H., Reddy, R. R., Miciak, J., & Fletcher, J. M. (2015). Are child cognitive characteristics strong predictors of responses to intervention? A meta-analysis. *Review of Educational Research, 85*, 395–429.

Taylor, W. P., Miciak, J., Fletcher, J. M., & Francis, D. J. (2017). Cognitive discrepancy models for specific learning disabilities identification: Simulations of psychometric limitations. *Psychological Assessment, 29*, 446–457.

Terman, T. L. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale.* Boston: Houghton Mifflin.

*Vanderwood, M., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences, 13*, 159–188.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Walberg, H. J., Strykowski, B. E., Roval, E., & Hung, S. S. (1984). Exceptional performance. *Review of Educational Research, 54*, 87–112.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70*, 129–133.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15*, 465–479.

Watkins, M. W., Kush, J. C., & Glutting, J. J. (1997). Discriminant and predictive validity of the WISC-III ACID profile among children with learning disabilities. *Psychology in the Schools, 34*, 309–319.

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. Flanagan, & P. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 251–268). New York, NY: Guilford.

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

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson Assessment.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children-Fifth Edition technical and interpretive manual.* San Antonio: NCS Pearson.

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2007). *Woodcock-Johnson III Normative update.* Rolling Meadows, IL: Riverside.