This is an Accepted Manuscript of an article published by Taylor & Francis Group in Applied Measurement in Education on 17/06/2019, available online: https://doi.org/10.1080/08957347.2019.1619561

Please use the following citation when referencing this work:

McGill, R. J., & Dombrowski, S. C. (2019). Critically reflecting on the origins, evolution, and impact of the Cattell-Horn-Carroll (CHC) Model. *Applied Measurement in Education, 32,* 216-231. doi: 10.1080/08957347.2019.1619561

Critically Reflecting on the Origins, Evolution, and Impact of the Cattell-Horn-Carroll (CHC) Model

Ryan J. McGill

William & Mary

Stefan C. Dombrowski

Rider University

Author Note

A preliminary version of this paper was presented at the 2017 meeting of the American Psychological Association, Washington, D.C.

Correspondence concerning this paper should be addressed to Ryan J. McGill, William & Mary School of Education, P. O. Box 8795 Williamsburg, VA. 23187. E-Mail: rmcgill@wm.edu

Abstract

The Cattell-Horn-Carroll (CHC) model presently serves as a blueprint for both test development and a taxonomy for clinical interpretation of modern tests of cognitive ability. Accordingly, the trend among test publishers has been toward creating tests that provide users with an ever-increasing array of scores that comport with CHC. However, an accumulating body of independent research on modern intelligence tests has questioned many instruments' alignment with the CHC model. To shed potential insight on these discrepancies, we review the developmental history of CHC and its numerous modifications from 1997 to the present. Next, we identify and discuss several potential limitations in the CHC literature that may be responsible for this discrepancy. Finally, we encourage clinicians to consider the extant evidence currently available for engaging in CHC-inspired assessment applications (e.g., XBA, PSW).

Keywords: CHC, evidence-based assessment, intelligence testing

Critically Reflecting on the Origins, Evolution, and Impact of the Cattell-Horn-Carroll (CHC) Model

Essentially, all models are wrong, some are useful

Box and Draper (1987, p. 424)

Psychometric theories of intelligence have increasingly played a prominent role in the evolution of the measurement of intelligence (Beaujean, 2015).[1] Over the course of the last century, several influential psychometric models have been developed by researchers, based largely on the results furnished by systematic programs of structural research (particularly factor analytic models) on various datasets of mental tests (Lang, Kersting, & Beauducel, 2016). Due to its ability to shed insight about the covariation and structuring of variables, factor analysis is particularly well suited for studying the psychological variables (latent factors) that may account for relations among variables in cognitive data. The results of these analyses are commonly reported in test technical manuals because they provide the statistical rational for the scores that are developed for those measures (Bruner, Nagy, & Wilhelm, 2012; McGill & Dombrowski, 2017).[2]

A multitude of theories have been cited in the psychometric literature over the last 30 years and three in particular have had a tremendous influence on modern test development. These include Horn and Cattell's Fluid-Crystallized theory (*Gf-Gc*; 1966), Carroll's Three-Stratum theory (3S; 1993), and the Cattell-Horn-Carroll model of intelligence (CHC; Schneider & McGrew, 2018). CHC is presented as a synthesis of Gf-Gc and 3S, and is regarded by many in

---

[1] The term *intelligence* has no technical meaning; we use it out of convenience to denote a class of attributes related to cognitive functioning, not any particular attribute.
[2] Although much of our discussion is focused on the role that factor analysis has played in the development of CHC and other related psychometric models, we acknowledge that structural fidelity is only one element, albeit an important element, of psychological theory development (Borsboom, Kievit, Cervone, & Hood, 2009).

the assessment literature (e.g., McGrew, 2009a; Ortiz, 2015; Schneider & Newman, 2015) as *the* consensus model of human cognitive abilities.

Despite the pervasive influence of CHC in research and clinical practice, its application to test construction and clinical assessment has been the subject of criticism (e.g., Frazier & Youngstrom, 2007; Glutting, Watkins, & Youngstrom, 2003) and several scholars have begun to raise additional questions about the conceptual underpinnings of the model and the utility of its related assessment procedures (e.g., Beaujean, 2015; Cucina & Howardson, 2017; Dombrowski & Watkins, 2013; McGill, 2017). These concerns reflect longstanding global disagreements among prominent psychological scholars and have implications for test construction, the clinical assessment of intelligence, and broader research on the nature of cognitive abilities. To help contextualize these debates, we outline the origins, evolution, and impact of CHC on clinical assessment. From this, we identify and discuss several empirical and conceptual limitations within the CHC literature. Overall, the points highlighted in this article should raise questions and, we hope, promote fruitful discussion among researchers and practitioners alike regarding the clinical assessment of intelligence and the application of CHC in particular.

Factors Influencing the Development and Rise of CHC

According to McGrew (2005), “The CHC theory of intelligence is the tent that houses the two most prominent psychometric theoretical models of human cognitive abilities” (p. 137) and was the culmination of a process that began in the early 1990s to develop a consensus taxonomy of cognitive abilities that, like Mendeleev's periodic table of elements, could be used to classify all intelligence tests. Despite its popularity, the exact genesis of CHC has been difficult to discern given conflicting accounts as to when it was first referenced in the professional literature and the degree to which its namesakes were invested in its development and dissemination.

To help clarify these matters, we review several factors that appear to have influenced the development and rise of CHC.

***Gf-Gc* Theory**

The origins of *Gf-Gc* theory can be traced to a commentary by Cattell (1943) positing a theoretical model, derived from Hebb's theory of intelligence (Brown, 2016), in which intelligence was largely explained by the interactions between two broad abilities he termed Fluid and Crystallized Intelligence. Whereas Fluid Intelligence is described as the ability to reason and solve problems independent of prior knowledge, Crystallized Intelligence is dependent on the storage, retrieval, and use of knowledge accumulated from environmental experience. Cattell largely neglected the theory before conducting the first *Gf-Gc* experimental analysis nearly two decades later. Cattell (1963) factor analyzed data from a battery of cognitive tasks that were administered to a small sample of school-aged children and found that each of the tasks loaded primarily on one of the two intelligence factors. John Horn (1965) made a number of refinements to *Gf-Gc* theory in his doctoral dissertation that was chaired by Cattell at the University of Illinois. From a battery of 59 ability and personality tests, factor analyses revealed the presence of broad ability factors in addition to Fluid Intelligence (later termed "Fluid Reasoning" [Gf]) and Crystallized Intelligence (later termed "Crystallized Ability" [Gc]): General Visualization, General Speediness, Facility (F; a forerunner to Long-Term Retrieval).

For the next 30 years, Horn and colleagues (Horn, 1989; Horn & Cattell, 1966; Horn & Stankov, 1982) engaged in a systematic program of research aimed at identifying and clarifying relationships among broad abilities. One of the more consequential discoveries in this line of research is evidence for the differential growth and decline of various abilities over the course of the lifespan (McArdle & Hofer, 2014). In a substantive review of *Gf-Gc* theory, Horn and Blankson (2012) suggest that the accumulated evidence indicates that human abilities can be organized within a higher-order structure in which approximately 80 first-order primary mental abilities are explained by eight second-order abilities. The consensus second-order abilities include: Gc, Gf, Short-Term Apprehension and Retrieval (SAR), Fluency of Retrieval from Long-Term Storage (TSR), Processing Speed (Gs), Visual Processing (Gv), Auditory Processing (Ga), and Quantitative Knowledge (Gq).

Gf-Gc broad abilities are commonly portrayed in test technical manuals and interpretive guidebooks as if they are orthogonal components in the tradition of Thurstone's (1938) original conceptualization of Primary Mental Abilities, but this is incorrect. Horn and Cattell argued (1982) that the broad abilities should be modeled as overlapping correlated processes. Horn also maintained throughout the entirety of his career that *g* (Spearman, 1904) was likely a statistical artifact and lacked the requisite properties for measurement invariance. As a result, he never regarded it as a viable dimension in *Gf-Gc* theory.

Woodcock-Johnson Psycho-Educational Battery-Revised. The publication of the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R; Woodcock & Johnson, 1989) was a major development for Horn-Cattell theory as it was developed using *Gf-Gc* as a theoretical blueprint. Woodcock and Mather (1989) wrote "The battery provides a comprehensive representation of Horn-Cattell theory adapted to the applied work of psychoeducational diagnosticians" (p. 13). Horn (1991) contributed a chapter reviewing cognitive ability theory in the WJ-R Technical Manual and the test authors reported that he provided additional advice on content and technical matters. The organization and content of the WJ-R cognitive subtests is largely consistent with the descriptions provided by Horn and Blankson (2012) and the battery yields eight second-order broad ability cluster scores consistent with those *Gf-Gc* dimensions. At the time of its publication, the WJ-R was the only commercial ability instrument designed explicitly to align fully with the *Gf-Gc* conceptualization of intelligence and it served as a reference instrument for substantive *Gf-Gc* and early CHC research throughout the 1990s and early 2000s.

The results of several exploratory and confirmatory factor analyses (CFAs) of the WJ-R are reported in the Technical Manual (McGrew, Werder, & Woodcock, 1991) and in a supplemental interpretive handbook (McGrew, 1994). Fit statistics indicate that the *Gf-Gc* model fits the data relatively well across the age span and was superior to all of the rival models it was tested against, including a three-stratum variant with a higher-order general factor. The cognitive subtests aligned well with their respective factors and the patterns of loadings suggested that the subtests were best explained by seven first-order broad abilities (Long-Term Retrieval [Glr], Short-Term Memory [Gsm], Processing Speed, Auditory Processing, Visual Processing, Gc, and Gf). Additional analyses, incorporating several math subtests from the WJ-R achievement battery, indicated the presence of a Quantitative Knowledge factor. Later, Woodcock (1998) argued for adding a Reading-Writing factor (Grw) to the extended Gf-Gc model. This dimension, however, was not identified by Horn or Cattell in any of their seminal *Gf-Gc* publications or in a substantive review by Stankov (2000).

3S Model

In a watershed moment for the study of individual differences, Carroll (1993) published his magnum opus *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*, reporting the results of more than a decade's worth of work analyzing 480 cognitive datasets. As described by Carroll, "My intention has been simply to present what was in my opinion the most accurate, reasonable, and consistent picture of the total domain of cognitive abilities" (p. viii). For each dataset, Carroll employed exploratory factor analytic (EFA) procedures: oblique rotation of the principal factor matrices, followed by application of the Schmid-Leiman procedure (SL; 1957)[3]. The use of the SL transformation evinces an important theoretical and methodological distinction between Carroll and Horn-Cattell. According to Carroll (1993, 1995), all cognitive variables are multidimensional insofar as they are influenced simultaneously by an array of abilities at higher- and lower-strata, and failing to model the variables' variances appropriately can result in interpretive confusion of factor analytic solutions. According to Caretta and Ree (2001),

> When rotation occurs, the variance associated with [*g*] seems to disappear, but in reality, it has become the dominant source of variance in the now rotated factors. Thus, these other factors become *g* plus something else. However, it is usually the ‘something else’ that determines the label for the factor, while the general component is not acknowledged (p. 332).

As a variance apportionment procedure, SL helps to make these distinctions clearer (Wolff & Preising, 2005).

Carroll (1993) organized his results by domain and proposed the 3S model, which differentiates abilities as a function of their breadth. The most general ability resides at the apex of the model at Stratum III and is considered to be a general factor conceptually similar to *g*.[4] The next level (Stratum II) contains eight broad abilities similar to the second-order abilities posited by Horn-Cattell: Gf, Gc, Auditory Processing, Processing Speed, General Memory and Learning (Gy), Broad Visual Perception, Broad Retrieval Ability (Gr), and Decision Speed (Gt). Finally, at the base of the model (Stratum I) are approximately 70 narrow abilities that are organized according to their mapping onto the Stratum II dimensions.

---

[3] Carroll also used a variety of quantitative criteria to determine the number of factors to extract in EFA, a practice not commonly employed at the time.

[4] Carroll's (1993) original analysis precluded actually measuring *g*, but later clarified that he intended for the Stratum III attribute to be g (Carroll, 1996).

Cross-Battery Assessment: Reconciling Carroll and Horn-Cattell

An event that is often overlooked in the development of CHC is the birth of cross-battery assessment in the psychoeducational assessment literature and can be traced to Woodcock's (1990) call for clinicians to consider "crossing" batteries to obtain purer measures of relevant *Gf-Gc* attributes. At that time, the WJ-R was the only clinical battery available that provided measurements for all known *Gf-Gc* abilities. Examiners using other instruments (e.g., Wechsler Scales) who wished to assess all the *Gf-Gc* constructs needed to supplement their batteries with measures from other tests. Woodcock reported the results of several conjoint CFA analyses featuring the WJ-R and other commercial ability measures and argued that the results provided a roadmap for the development of a cross-battery assessment system based on Horn-Cattell theory that could be applied to all intelligence tests. At the same time, it was clear that Carroll and Horn both regarded their models as competing theories (e.g., Carroll, 1996; 2003). Thus, in order for cross-battery assessment (XBA) to flourish, proponents of these methods would have to endorse one model at the expense of the other or construct a unified framework reconciling the two approaches.

Integrated *Gf-Gc*. McGrew (1997) proposed what he described as a synthesized Carroll and Horn-Cattell *Gf-Gc* framework, the first in a series of integrated models that would eventually become CHC. Beyond ascertaining whether a general factor of intelligence was viable, a number of other issues needed to be resolved in reconciling 3S and Gf-Gc. These included the assignment of several narrow abilities to broad factors, clarifying the conceptualization of broad memory abilities, adopting a standard nomenclature to be used across tests, and determining whether it made sense to retain certain academic factors at Stratum II.

Carroll (1993) did not believe that Quantitative Knowledge was a viable broad ability noting "'Mathematical ability,' therefore, must be regarded as an inexact, unanalyzed, popular concept that has no scientific meaning unless it is referred to the structure of abilities that compose it. It cannot be expected to constitute a higher-level ability." (p. 627). Whereas Quantitative Reasoning was regarded in 3S as a narrow dimension subsumed within Gf, Horn (1989) identified Quantitative Knowledge as a Stratum II dimension. Similarly, although a conjoint Reading-Writing factor was posited as a potential broad ability by Woodcock (1998), Carroll located reading and writing variables as narrow abilities under Gc. To resolve these issues, McGrew (1997) reported the results of a CFA using the WJ-R normative data. Several models were examined including one which closely approximated Carroll's assignment of variables in 3S along with competing models, gradually implementing the assignment of indicators based on various configurations of Horn-Cattell theory. The final model, consistent with virtually all aspects of extended Gf-Gc theory, yielded the best fit to the data.

Next, a standardized taxonomy of broad and narrow abilities was outlined based primarily on the results of three WJ-R cross-battery CFAs (Flanagan & McGrew, 1998; McGhee, 1993; Woodcock, 1990). This taxonomy was then applied to classify tests from several commercial ability measures to illustrate potential XBA applications that were described in more detail in a complementary chapter by Flanagan and McGrew (1997). Following these developments, McGrew and colleagues (McGrew & Flanagan, 1998) published *The Intelligence Test Desk Reference* (ITDR) which outlined a formal system of XBA using an integrated *Gf-Gc* framework as well as a related handbook for the Wechsler Intelligence Scales (Flanagan, McGrew, & Ortiz, 2000).

Curiously, despite noting the influence of Carroll, a general factor was not modeled or accounted for in any of the early fusion model studies or chapters. Accordingly, McGrew (1997) stated, "No attempt was made to resolve the existence of a *g* factor, a long standing theoretical debate that does not bear directly on the purpose of this chapter" (p. 153). Going a step further, McGrew and Flanagan (1998) suggested that *g* had no practical value for clinical assessment, yet, stipulated that "the exclusion of *g* does not mean that the integrated model does not subscribe to a separate general human ability or that *g* does not exist" (p. 14).

Cattell-Horn-Carroll (CHC) Theory

It is difficult to discern when the integrated *Gf-Gc* framework officially became CHC. There have been numerous references to a meeting that took place sometime in 1999 that resulted in an informal CHC branding. The first references to CHC that can be located in the empirical literature are a pair of articles that were published in 2000 (Flanagan, 2000; Kranzler, Keith, & Flanagan, 2000). For example, in an independent CFA of the Cognitive Assessment System by Kranzler, Keith, and Flanagan (2000) the authors invoke the term "the Cattell-Horn-Carroll (CHC) Theory of Cognitive Abilities" (p. 144) with no attributing citation. It has also been sourced by McGrew (2005) to the Wechsler *Gf-Gc* text by Flanagan and colleagues (2000) and in other publications (e.g. Ortiz, 2015) to the WJ-III Technical Manual (McGrew & Woodcock, 2001). Nevertheless, the publication of the WJ-III (Woodcock, McGrew, & Mather, 2001) was a pivotal moment in the rise of CHC as the instrument was revised explicitly to comport with that model. CFA evidence in the technical manual supported a three-stratum model, with a general factor at the apex, followed by nine broad abilities, and several narrow intermediary dimensions. A standalone XBA handbook predicated on CHC was also published that same year (Flanagan & Ortiz, 2001), ushering the CHC era into existence.

Conceptually, CHC is portrayed as a three-stratum model (i.e., 3S) borrowing much of the terminology and organization of broad and narrow abilities at Stratums I and II proposed by McGrew (1997). Elaborations and revisions to the model are described in a series of chapters authored by McGrew (2005) and Schneider and McGrew (2012) that are regarded as citation classics in the field (556 and 394 citations respectively [Google Scholar on May 18, 2018]). The majority of these refinements have been focused on modifying the operational definitions of various abilities, expanding the number of attributes in Stratum I and II, and speculating on potential reconciliation of CHC with other theories emanating from the cognitive neuroscience literature.

Impact of CHC on Clinical Assessment

At the time of its publication, the WJ III was the only commercial ability measure designed to align with the CHC model; although subsequently a number of tests have been revised using CHC as a blueprint, the WJ-III/WJ-IV remains the only individually administered cognitive test to yield scores for the Stratum II broad abilities (Gc, Gf, Visual Processing, Auditory Processing, Short-Term Memory, Long-Term Retrieval, Processing Speed, Quantitative Knowledge, and Reading-Writing) that are most commonly referenced in the CHC literature. As illustrated in Figure 1, there are now six comprehensive intelligence tests that have been revised and now yield scores consistent with CHC. The publication of the Wechsler Intelligence Scale for Children-Fifth Edition (Wechsler, 2014) is particularly noteworthy as it is the first variant of the vaunted Wechsler Scales to explicitly incorporate a CHC inspired factor structure.

Although prominent CHC chapters and commentaries (e.g., McGrew, 2005, 2009a; Schneider & McGrew, 2012) continue to cite large quantities of extended *Gf-Gc*, 3S, and early integrated *Gf-Gc* studies as empirical support for CHC, a theory-specific research program has emerged over the course of the last 15 years. Like its predecessor, the WJ III and WJ IV have served as the preeminent reference instrument for the preponderance of this research and the subsequent refinements that have been made to the taxonomy. For example, Taub & McGrew (2014) conducted a CFA with participants from the WJ III normative sample and found evidence to support the presence of a series of intermediary factors consistent with the Cognitive

Performance Model (CPM) described by Woodcock (1998). McGrew and colleagues have also posited a number of possible extensions and theoretical refinements to the CHC model in a series of white papers reporting the results of literature reviews and internal validation studies (e.g., McGrew, 2009b; McGrew and Evans, 2004). Additionally, a series of studies have been conducted examining CHC cognitive-achievement relations over the lifespan (McGrew & Wendling, 2010) and several conjoint CFAs with other instruments have been reported (e.g., Keith & Reynolds, 2010).

Since 2001, references to CHC in the empirical literature have risen dramatically over the course of 15 years. As illustrated in Figure 2, results from a Web of Science search indicate that as of 2016 there were approximately 400 references *per year* to CHC in the professional literature, and the slope of the yield line portends that the number will continue to grow steeply in the coming years. For example, Pesta's (2018) bibliometric analysis of the journal *Intelligence* found that McGrew's (2009a) CHC editorial was the most cited article over an eight-year span (2008-2015).

A panoply of CHC inspired interpretive systems and products have been developed over the last twenty years that now are widely disseminated in books, chapters, and workshops devoted to clinical assessment. These include the most recent iterations of XBA (Flanagan, Ortiz, & Alfonso, 2013), the Culture-Language Interpretive Matrix (C-LIM; Ortiz, Melo, & Terzulli, 2018), and multiple diagnostic schemes for the identification of specific learning disability based on patterns of performance on CHC indicators (e.g., Flanagan et al., 2018; Schultz & Stephens, 2015). By any measure CHC has had a substantial impact on assessment training and practice over a relatively short time span.

Potential Limitations Identified in the CHC Assessment Literature

Although CHC has been characterized as “the most comprehensive and empirically supported psychometric theory of the structure of cognitive and achievement abilities” (McGrew, 2005, p. 185), Horn and Blankson (2012) presciently noted that, “All scientific theory is wrong. It is the job of science to improve theory” (p. 73). In spite of this caveat, CHC has been uncritically accepted in many assessment circles. For example, some of the first independent exploratory investigations of CHC (i.e., Dombrowksi, 2013; Dombrowski & Watkins, 2013) did not come about until a full decade after the WJ-III was published, and were one of the first times that the veracity of CHC was directly called into question.

In the following section we identify and describe a series of provisional limitations (see Table 1 for a summary) identified in the CHC literature. Whereas some of these concerns are elements that are drawn from the evidence-based practice literature (e.g., Lilienfeld, Ammirati, & David, 2012), others relate to more focal concerns with the quality of the psychometric evidence furnished to support the use of CHC in clinical practice.

(1) Given the pivotal role that various iterations of the WJ battery have played in the development and subsequent refinements to CHC as well as its progenitors (e.g., extended/integrated *Gf-Gc*), there may be an element of circularity in this line of research wherein the model is used initially to develop the test and then the test is used as the primary vehicle for further elaborating on the model. For example, the WJ-R was the focus of a CFA conducted by McGrew (1997) that was instrumental in resolving the substantive theoretical differences that beget the integrated *Gf-Gc* taxonomy. In doing so, virtually all of Carroll’s descriptions in 3S that diverged from *Gf-Gc* theory were dismissed. Given the alignment of the WJ-R with extended Horn-Cattell theory, however, use of that instrument for these purposes can be regarded as a kind of psychometric tautology—akin to using the same data to build a model and then test its validity. The same argument can be made for the use of the WJ-III/WJ-IV for

CHC research. Although this limitation has been acknowledged by some proponents of CHC (e.g., Keith & Reynolds, 2010), and efforts have been made to include additional batteries in recent cross-battery CFAs (e.g., Reynolds et al., 2013), the vast majority of internal and external CHC validation projects continue to utilize the WJ.

Proposed Remedy: Utilize large scale, publicly available datasets of intelligence tests (e.g., Johnson & Bouchard, 2011) as the focal point for validating and extending CHC in the future.

(2) Factor analysis can be a useful technique for theory building and it is used widely in the CHC literature. Nonetheless, many of the CFA investigations that are cited as being instrumental in the development of CHC suffer from substantive methodological limitations, giving rise to concerns about the empirical foundations of CHC. These include (a) consistent use of small unrepresentative samples resulting in insufficient power (e.g., McGhee & Lieberman, 1994), (b) retention of models despite encountering issues with global and local fit (e.g., Flanagan & McGrew, 1998; McGhee, 1993; Tusing & Ford, 2004; Woodcock, 1990), (c) failing to examine plausible rival models, (d) focusing only on inter-individual differences (e.g., Gomes, de Arújo, Ferreira, & Golino, 2014) and (e) imposing post-hoc model adjustments in order to get preferred solutions to work. For example, McGhee (1993) examined the potential *Gf-Gc* alignment of 31 tests from different batteries in a sample of 100 school-aged children. The fit indices (Goodness of fit index [GFI] and the adjusted GFI) for the retained *Gf-Gc* model were .68 and .61 respectively[5]. These values fall well below established standards for acceptable

[5] Identification of this model was only possible after using modification indices to impose a series of post-hoc model tweaks, none of which were specified *a priori*. Although this is a common practice in the CHC literature, it has been regarded as equivalent to hypothesizing after results are known (i.e., HARKing; Cucina & Byle, 2017). According to Meehl (1978), all *post-hoc* specification searches automatically become exploratory studies, so are best used for theory generation rather than theory confirmation.

model fit (Shevlin & Miles, 1998). Furthermore, the inadequate sample size to variable ratio (*N*:*p* = 3.12) suggest that these results are unlikely to replicate (DiStefano & Hess, 2005). In a subsequent investigation (McGhee, Buckhalt, & Phillips, 1994), application of a similar model using many of the same measures produced a Heywood case indicating an impermissible solution for the data.

To be fair, there have been a number of CFA studies in which a CHC structure has been found to fit various datasets well[6] (see Schneider & McGrew, 2012). Adequate fit indices, alone, cannot *confirm* a model as the tenability of a preferred model can only be ascertained when other plausible rival models have been explored and found wanting (Lewis, 2017). In one of the most substantive CHC investigations to date, Reynolds et al. (2013) conducted a cross-battery CFA featuring five different cognitive tests. Multiple CHC measurement models were explored with each configuration showing good fit to the data, but no rival non-CHC models were explored. Thus, the situation was designed for researchers to “confirm” *a priori* preferred models (Ropovik, 2015). Interestingly, it was noted that future research examining alternative models such as the VPR model would be instructive. The VPR (verbal-perceptual-image rotation) is a more parsimonious model of abilities based on Vernon’s (1950) original hierarchical model that has been found to fit several large samples of cognitive data better than the CHC model (Major, Johnson, & Deary, 2012). Although these findings have been acknowledged in some CHC writings (e.g., McGrew, 2009a), relevant explorations of the VPR model remain noticeably absent in this literature.

[6] We were able to replicate Carroll’s (2003) factor analytic results of the WJ-R supporting an eight-factor structure consistent with Horn-Cattell theory using EFA procedures for approximating a bifactor structure in **R** (https://osf.io/vfbyx/). Analyses indicate that Quantitative Knowledge, Gc, and Gf were found to have interpretive relevance beyond *g*. Thus, we agree with Carroll (1993, 2003) that existing evidence supports the presence of a general intelligence factor and an as yet undetermined number of broad abilities.

Although the results furnished by cross-battery CFAs are frequently invoked to support the validity of CHC, these results must be reconciled with the accumulating body of independent factor analytic research calling into question the structural validity and alignment of many modern intelligence tests with CHC (Canivez & Youngstrom, this issue). Several of these investigations involve the WJ-III and WJ-IV (e.g., Dombrowski, 2013; Dombrowski, McGill, & Canivez, 2017, 2018; Dombrowski & Watkins, 2013). For example, in a recent CFA of the WJ-IV, Dombrowski, McGill, and Canivez (2018) used EFA to guide their CFA and found that a four-factor structure—one that was not consistent with CHC—provided the best fit to the WJ-IV normative data. Of concern, attempts to specify the seven-factor CHC solution posited by the test publisher resulted in a series of model specification errors indicating that solution is not tenable for what is regarded as the preeminent reference instrument for the model.

Whether the inability to locate posited CHC dimensions is due to underlying issues with the model or merely reflects local measurement problems with specific instruments has yet to be determined. For this to be determined, however, will require a moving beyond using omnibus taxonomies to classify tests.

> If we simply list the tests or traits which have been shown to be saturated with the "factor" or which belong to the cluster, no construct is employed. As soon as we even summarize the properties of this group of indicators—we are already making some guesses. Intensional characterization of a domain is hazardous since it selects (abstracts) properties and implies that new tests sharing those properties will behave as do the known tests in the cluster (Cronbach & Meehl, 1955, p. 292).

Instead, it will require using a variety of methods looking purposefully to disconfirm the CHC model (Cattell, 1988).

From a different angle, the replication crisis in scientific psychology has illustrated the need for reevaluating the evidence-base for widely accepted theories and recommended applications of those theories. Since its publication, Carroll's (1993) work has been regarded in the psychometric canon as almost beyond reproach. Yet, this was not how Carroll (1998) viewed his study; he conceded that his methods for classifying several thousand token factors was akin to a qualitative meta-analysis and "inevitably entailed a considerable degree of subjectivity" (p. 8). Further, he acknowledged that future research would likely reveal "errors of commission and omission" (Carroll, 1997, p. 128) and that the structure he identified inevitably would become more parsimonious once the importance and relevance of the various abilities it specified became clear. His prognostication seems to have been correct. Benson and colleagues (2018) re-analyzed several datasets from Carroll's (1993) survey using CFA procedures and found that Carroll likely extracted too many factors representing Stratum II abilities; virtually all of the identified broad abilities had no interpretive relevance beyond *g*.

Proposed Remedy 2.1: Include plausible rival models (e.g., bifactor, VPR) in CHC factor studies (e.g., Reynolds & Keith, 2017).

Proposed Remedy 2.2: Reexamine seminal CHC datasets to determine whether previously published results can be replicated.

Proposed Remedy 2.3: Move beyond studying inter-individual variability to see if the same attributes are found in other types of experiments (e.g., Molenaar & Campbell, 2009).

(3) One of the core features of developed sciences is "connectivity" in which claims build on or connect with prior research. To a large extent, CHC exhibits many of the characteristics of a developed science. For example, McGrew (2005) systematically described the theoretical lineage of CHC within the context of previous research. At the same time, however, an amalgam

of loosely connected constructs and refinements (e.g., correlated residual terms between various broad abilities) have also been postulated in CHC models over the course of the last decade and the degree to which these refinements are incorporated in subsequent research has been inconsistent. Although Taub and McGrew (2014) found that an alternative CHC structure with intermediary factors best fit the WJ-III normative data, this model does not appear to have been explicated in any meaningful way since. Numerous other elements have been added to the taxonomy based on the results of isolated factor analytic studies using different configurations of variables and some elements seem to emerge out of thin air. For example, in the most recent iteration of XBA, Flanagan and colleagues (2013) discuss the practical value of partitioning Broad Reading-Writing into distinct reading and writing factors even though this separation has yet to be validated empirically. Horn (1967) argued long ago against retaining factors solely for their practical value.

Additionally, the way in which CHC indicators are described and aligned in the literature is not consistent. For example, in the XBA system, several measures of mathematics achievement are speculated to align with a narrow Quantitative Reasoning (RQ) ability within Gf. Yet, measures with similar content have been used predominately to identify Quantitative Knowledge in the factor analytic literature (e.g., Woodcock, 1998). As a result, Carroll's (1993) reservations about the veracity of a Quantitative Knowledge factor appear to be apt.

Proposed Remedy: Establish consensus guidelines for the identification of constructs and elaborations to the CHC taxonomy (e.g., Schneider, Mayer, & Newman, 2016).

(4) The trend among test publishers has been toward creating tests that provide users with an ever-increasing array of scores that ostensibly comport with CHC, which increases the time and cost of assessing intelligence in clinical practice (Frazier & Youngstrom, 2007; Williams & Miciak, 2018). Empirical evidence for the utility of various CHC clinical applications is less than

compelling, however, and the results of several investigations indicate that use of CHC-derived diagnostic and interpretive procedures is potentially contraindicated (e.g., Kranzler et al., 2016; Schneider & McGrew, 2011; Stuebing et al., 2012; Styck & Watkins, 2013). Many researchers contend that these results may be due to the paltry interpretive relevance and incremental validity of many CHC derived indices (e.g., Beaujean & Benson, 2018; Benson, Kranzler, & Floyd, 2016, Watkins, 2017). Unfortunately, this body of evidence is rarely disclosed in the CHC literature (Dombrowski et al., 2017).

Proposed Remedy: Disclose *all* relevant literature in empirical studies, manuscripts, and presentations detailing CHC clinical applications.

(5) Proponents of the CHC framework make regular contributions to the peer reviewed literature; still, non-empirical books, chapters, commentaries, internet posts, and workshop proceedings are increasingly becoming the primary means for disseminating refinements to CHC. In many of these forums, unpublished data and personal communications are cited to support the empirical and conceptual underpinnings of the taxonomy. Of particular concern is the role that a series of non-refereed white papers reporting results of internal validation studies and *ad hoc* literature reviews, run by a small cadre of researchers, appears to have played in more recent iterations of the model (e.g., McGrew, 2009b; McGrew & Evans, 2004). This insularity and evasion of peer review raise concern about potential allegiance effects (Lilienfeld & Jones, 2008).

Relatedly, various depictions of the circumstances surrounding the seminal 1999 branding meeting raise questions about the potential role that commercial interests may have played in the rise of CHC. Based on available reports (Kaufman, 2009; McGrew, 2005), it appears that the meeting was arranged and attended by several commercial test authors and staff members of a prominent test publisher.

> Talk about the tail wagging the dog! What had begun back in the late 1970s and early 1980s as a search for the best theories on which to build an IQ test had come full circle: Two decades later, the *needs* of test publishers and test authors forged the theory that underlies almost all current-day IQ tests (Kaufman, 2009, p. 99, emphasis added).

The important role that the Woodcock-Johnson and XBA played in the decade long development of the CHC taxonomy further buttress this concern.

> **Proposed Remedy 5.1:** Encourage pre-registration of empirical CHC investigations.
>
> **Proposed Remedy 5.2:** Adopt a "many labs" framework in which investigators with different viewpoints—hopefully some of which have no financial conflicts of interest—conduct similar investigations simultaneously and agree to share results in open source venues.
>
> **Proposed remedy 5.3:** Disclose all information needed for the independent replication of CHC studies irrespective of the medium published (e.g., peer-reviewed outlets, white papers, internet posts, technical manuals).

(6) Although Carroll's name is invoked frequently in CHC writings, it is difficult to locate much, if any, of his influence in modern CHC incarnations. Virtually all of Carroll's conceptualizations of broad and narrow abilities were excised in favor of Horn-Cattell theory when creating the *Gf-Gc* integrated model and the relevance of *g* has been a source of vigorous debate since the birth of CHC. Numerous depictions of the CHC structure omit a Stratum III general factor and its relevance has been questioned in seminal CHC writings (e.g., Schneider & McGrew, 2012)[7]. Many CHC factor analytic investigations continue to employ correlated factors models (e.g., Jewsbury, Bowden, & Duff, 2017), which omit a general factor altogether and its

---

[7] More recently, Schneider and McGrew (2018) argue that a general factor exists but that it is likely not an ability.

effects have been accounted for inconsistently in other CHC investigations[8]. Cucina and Howardson (2017) speculated that these vacillations may be an attempt to promote broad ability interpretation at the expense of *g*, illustrating well the pervasive influence of Horn-Cattell (Beaujean & Benson, this issue).

> Indeed, an important element that is missing in CHC but present in Carroll's work is the incorporation of the magnitude of the unique factor loadings. Under the Three Stratum Theory, the magnitude of the non-*g* loadings is low and this is made quite clear. We are unaware of any CHC publications that recognize these low magnitudes (p. 1013).

Carroll (1993, 1996) regarded the status of *g* to be the most salient distinction between 3S and Horn-Cattell theory and he argued vigorously for the existence of *g* in much of his published work. In his last publication, Carroll (2003) noted that his position on this issue was not reconcilable with Horn's views on the matter. As a result, the degree to which he was invested in the development of CHC can be questioned. Whereas Carroll (2003) acknowledged what he termed "a so-called CHC (Cattell-Horn-Carroll) theory of cognitive abilities" he further elaborated that he was "not sure what caused or motivated it" (p. 18). Additionally, there is a published footnote (McGrew, 2005, p.174 [11]) detailing a private conversation in which Carroll noted that the CHC umbrella did not reflect his agreement with Horn on various aspects of the structure of intelligence, a position that he planned to make clear in a future publication that was not completed.

It is true that Carroll (1993) noted that Horn-Cattell theory "appears to offer the most well-founded and reasonable approach to an acceptable theory of the structure of cognitive abilities," (p. 62), and this quote is featured in many CHC writings. However, the way in which it

---

[8] Although authors identify these models as "CHC models", they are virtually identical to those identified in extended *Gf-Gc* research.

is invoked is somewhat misleading. This comment is extracted from a chapter surveying historical foundations of intelligence theory prior to Carroll's depiction of *his* model. A full reading of the passage makes this and his reservations about Horn-Cattell theory clear since the sentence following the previous quote is "The major reservation I would make about it is that it appears not to provide for a third-order *g* factor to account for correlations among the broad second-order factors." Further, it is worth mentioning that Carroll analyzed several *Gf-Gc* datasets, compared 3S to competing models that included *Gf-Gc*, and indicated that he viewed his model as an extension and elaboration of Spearman's (1927a, b) in his survey.

> I find myself in agreement with much of Spearman's thinking, as expressed in his various works . . .Spearman's work can be viewed as contributing a glorious and honorable first approximation to the true structure of cognitive abilities-an approximation that in many ways has stood the test of time, but that contrarily, has over the years since then been shown to be grossly inadequate. (Carroll, 1996, pp. 1,5)

Thus, it is clear that 3S was intended to extend Spearman's work and supplant *Gf-Gc* theory; the fact that Carroll continued to promote and reference 3S well into the CHC era is compelling.

Proposed Remedy: Recognize that important aspects of 3S and *Gf-Gc* theory are potentially irreconcilable and that the parameters of CHC are more consistent with the latter rather than the former.

Conclusion

There are many positives associated with CHC taxonomy. Namely, it has provided practitioners and researchers with a relatively standard nomenclature for discussing cognitive abilities. Although advancing CHC is a desideratum of many assessment professionals, the potential limitations identified in the present review will likely have to be overcome in order for CHC to realize its potential as a viable theory.

We suggest that the process by which it was conceptualized and developed was more akin to a clash of divergent theoretical viewpoints (Canivez & Youngstrom, this issue). Carroll and Horn-Cattell cater to different epistemological traditions challenging their reconciliation (Beaujean & Benson, this issue). To that end, we have proposed several potential remedies that we believe merit consideration, and, hopefully, will aid in the development of future studies.

References

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3*, 121-136. doi: 10.3390/jintelligence3040121

Beaujean, A. A., & Benson, N. F. (2018). Theoretically-consistent cognitive ability test development and score interpretation. *Contemporary School Psychology*. Advance online publication. doi: 10.1007/s40688-018-0182-1

Benson, N. F., Beaujean, A. A., McGill, R. J., & Dombrowski, S. C. (2018). Revisiting Carroll's survey of factor-analytic studies: Implications for the clinical assessment of intelligence. *Psychological Assessment, 30,* 1028-1038. doi: 10.1037/pas0000556

Benson, N. F., Kranzler, J. H., & Floyd, R. G. (2016). Examining the integrity of measurement of cognitive abilities in the prediction of achievement: Comparisons and contrasts across variables from higher-order and bifactor models. *Journal of School Psychology, 58,* 1-19. doi: 10.1016/j.jsp.2016.06.001

Borsboom, D., Kievit, R. A., Cervone, D., & Hood, S. B. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra & N. Chaudhary (Eds.), *Dynamic Process Methodology in the Social and Developmental Sciences* (pp. 67-97). New York, NY: Springer US.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: Wiley.

Brown, R. E. (2016). Hebb and Cattell: The genesis of the theory of fluid and crystallized intelligence. *Frontiers in Human Neuroscience, 10,* 606. doi:10.3389/fnhum.2016.00606

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80,* 796-846. doi: 10.1111/j.1467-6494.2011.00749.x

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research: *International Journal of Selection and Assessment, 9,* 325-335. doi: 10.1111/1468-2389.00184

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

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429-452. doi: 10.1207/s15327906mbr3003_6

Carroll, J. B. (1996). A three-stratum theory of intelligence: Spearman's contribution. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 1-17). Mahwah, NJ: Erlbaum.

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

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5-23). Mahwah, NJ: Erlbaum.

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

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153-193. doi: 10.1037/h0059973

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,*1-22. doi: 10.1037/h0046743

Cattell, R. B. (1988). Psychological theory and scientific method. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2^{nd} ed., pp. 3-20). Boston, MA: Springer.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302. doi: 10.1037/h0040957

Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence, 5 (3),* 27. doi: 10.3390/jintelligence5030027

Cucina, J. M., & Howardson, G. N. (2017). Woodcock-Johnson–III, Kaufman Adolescent and Adult Intelligence Test (KAIT), Kaufman Assessment Battery for Children (KABC), and Differential Ability Scales (DAS) Support Carroll but Not Cattell-Horn. *Psychological Assessment, 29,* 1001-1015. doi:10.1037/pas0000389.

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23,* 225-241. doi: 10.1177/073428290502300303

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III Cognitive at school age. *School Psychology Quarterly, 28,* 154-169. doi: 10.1037/spq0000010

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age. *Psychological Assessment, 29,* 394-407. doi: 10.1037/pas0000350

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). 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. doi: 10.1037/arc0000039

Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ-III full test battery: A school-ages analysis. *Psychological Assessment, 25,* 442-455. doi: 10.1037/a0031335

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. doi: 10.1037/h0088789

Flanagan, D. P., Alfonso, V. C., Sy, M. C., Mascolo, J. T., McDonough, E. M., & Ortiz, S. O. (2018). Dual discrepancy/consistency operational definition of SLD: Integrating multiple data sources and multiple data-gathering methods. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed.; pp. 329-430). 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 (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314-325). New York: Guilford.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc theory: Joint confirmatory analysis of the WJ-R and KAIT in a non-white sample. *Journal of School Psychology, 36,* 151-182. doi: 10.1016/S0022-4405(98)00003-X

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gf theory: A contemporary approach to interpretation*. Needham Heights, MA: Allyn and Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169-182. doi: 10.1016/j.intell.2006.07.002

Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multifactored and cross-battery assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence aptitude, and achievement* (2nd ed., pp. 343-374). New York: Guilford.

Gomes, C. M. A., de Arújo, J., Ferreira, M. G., & Golino, H. F. (2014). The validity of the Cattel-Horn-Carroll model on the intraindividual approach. *Behavioral Development Bulletin, 19*, 22-30. doi:10.1037/h0101078

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign. (AAT 6507113)

Horn, J. L. (1967). On subjectivity in factor analysis. *Educational and Psychological Measurement, 27,* 811-820. doi: 10.1177/001316446702700404

Horn, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29-73). Urbana, IL: University of Illinois Press.

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

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.; pp. 73-98). New York: Guilford.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57,* 253-270. doi: 10.1037/h0023816

Horn, J. L., & Cattell, R. B. (1982). Whimsy and misunderstanding of gf-gc theory: A comment on Guilford. *Psychological Bulletin, 91,* 623-622. doi: 10.1037/0033-2909.91.3.623

Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence, 6,* 165-185. doi: 10.1016/0160-2896(82)90012-5

Jewsbury, P. A., Bowden, S. C., & Duff, K. (2017). The Cattell-Horn-Carroll model of cognition for clinical assessment. *Journal of Psychoeducational Assessment, 35,* 547-567. doi: 10.1177/0734282916651360

Johnson, W., & Bouchard, T. J. (2011). The MISTRA data: Forty-two mental ability tests in three batteries. *Intelligence, 39,* 82-88. doi: 10.1016/j.intell.2011.02.010

Kaufman, A. S. (2009). *IQ testing 101*. New York: Springer.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What have we learned from 20 years of research? *Psychology in the Schools, 47,* 635-650. doi: 10.1002/pits.20496

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School & Educational Psychology, 4,* 124-136. doi: 10.1080/21683603.2016.1155515

Kranzler, J. K., Keith, T. Z., Flanagan, D. P. (2000). Independent examination of the factor structure of the Cognitive Assessment System (CAS): Further evidence challenging the construct validity of the CAS. *Journal of Psychoeducational Assessment, 18,* 143-159. doi: 10.1177/073428290001800204

Lang, J. W. B., Kersting, M., & Beauducel, A. (2016). Hierarchies of factor solutions in the intelligence domain: Applying methodology from personality psychology to gain insights into the nature of intelligence. *Learning and Individual Differences, 47,* 37-50. doi: 10.1016/j.lindif.2015.12.003

Lewis, T. F. (2017). Evidence regarding the internal structure: Confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development, 50,* 239-247. doi: 10.1080/07481756.2017.1336929

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing be- tween science pseudoscience in school psychology: science and scientific thinking as safeguards against human error. *Journal of School Psychology, 50,* 7-36. doi: 10.1016/j.jsp.2011.09.006.

Lilienfeld, S. O., & Jones, M. K. (2008). Allegiance effects in assessment: Unresolved questions, potential explanations, and constructive remedies. *Clinical Psychology: Science and Practice, 15,* 361-365. doi: 10.1111/j.1468-2850.2008.00148.x

Major, J. T., Johnson, W., & Deary, I. J. (2012). Comparing models of intelligence in Project TALENT: The VPR model fits better than the CHC and extended Gf-Gc models. *Intelligence, 40,* 543-559. doi: 10.1016/j.intell.2012.07.006

McArdle, J. J., & Hofer, S. M. (2014). Fighting for intelligence: A brief overview of the academic work of John L. Horn. *Multivariate Behavioral Research, 49,* 1-16. doi: 10.1080/00273171.2013.841089

McGhee, R., & Lieberman, L. (1994). Gf-Gc theory of human cognition: Differentiation of short-term auditory and visual memory factors. *Psychology in the Schools, 31,* 297-304. doi: 10.1002/1520-6807(199410)31:4<297::AID-PITS2310310408>3.0.CO;2-L

McGhee, R., Buckhalt, J. & Phillips, C. (1994, November). *Joint factor analysis of the Woodcock-Johnson Psychoeducational Battery-Revised and the Detroit Tests of Learning Aptitude*. Paper presented at the meeting of the Mid-South Educational Research

Association, Nashville, TN.

McGhee, R. L. (1993). Fluid and crystallized intelligence: A factor analytic study of second order abilities from three cognitive batteries. *Journal of Psychoeducational Assessment, Monograph Series: WJ-R,* 20-38.

McGill, R. J. (2017). Re(examining) relations between CHC broad and narrow cognitive abilities and reading achievement. *Journal of Educational and Developmental Psychology, 7,* 265-282. doi: 10.5539/jedp.v7n1p265

McGill, R. J., & Dombrowski, S. C. (2017). School psychologists as consumers of research: What school psychologists need to know about factor analysis. *Communiqué, 46 (1),* 16-18. Retrieved from http://www.nasponline.org

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

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

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

McGrew K. S. (2009a). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1-10. doi: 10.1016/j.intell.2008.08.004

McGrew, K. S. (2009b). *Cluster analysis of the WJ III Battery: Implications for CHC test interpretation and possible CHC model extensions (Applied Psychometrics 101 #4)*.

Institute of Applied Psychometrics. Retrieved from http://www.iapsych.com/iapap101/iapap1014.pdf

McGrew, K. S., & Evans, J. (2004). *Internal and external extensions to the Cattell-Horn-Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal treatise* (Carroll Human Cognitive Abilities Project: Research Report No. 2). St. Cloud, MN: Institute for Applied Psychometrics.

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

McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*, 651-675. doi: 10.1002/pits.20497

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

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, Il: Riverside Publishing.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806-834. doi: 10.1037/0022-006X.46.4.806

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences, 25,* 85-90. doi: 10.1016/S0191-8869(98)00055-5

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*, 112-117. doi:10.1111/j.1467-8721.2009.01619.x

Ortiz, S. O. (2015). CHC theory of intelligence. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 209-227). New York: Springer.

Ortiz, S. O., Melo, K. E., & Terzulli, M. A. (2018). Difference or disorder: Assessment of SLD with an English learner. In V. C. Alfonso & D. P. Flanagan, *Essentials of specific learning disability identification* (2nd ed.; pp. 505-548). Hoboken, NJ: Wiley.

Pesta, B. J. (2018). Bibliometric analysis across eight years 2008-2015 of *Intelligence* articles: An updating of Wicherts (2009). *Intelligence, 67,* 26-32. doi: 10.1016/j.intell.2018.01.001

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. doi: 10.1016/j.intell.2017.02.005

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51,* 535-555. doi: 10.106/j.jsp.2013.02.003

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology, 6,* 1715. doi: 10.3389/fpsyg.2015.01715

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53-61. doi: 10.1007/BF02289209

Schneider, W. J., Mayer, J. D., & Newman, D. A. (2016). Integrating hot and cool intelligences: Thinking broadly about broad abilities. *Journal of Intelligence, 4 (1),* 1. doi: 10.3390/jintelligence4010001

Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resources Management Review,* *25,* 12-27. doi: 10.1016/j.hrmr.2014.09.004

Schneider, W. J., & McGrew, K. S. (2011). *"Just say no" to averaging IQ subtest scores (Applied Psychometrics 101 #10)*. Institute of Applied Psychometrics. Retrieved from http://www.iqscorner.com/2011/03/iap-applied-psychometrics-101-report-10.html

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

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73-163). New York: Guilford.

Schultz, E. K., & Stephens, T. L. (2015). Core-selective evaluation process: an efficient & comprehensive approach to identify students with SLD using the WJ IV. *The DiaLog, 44 (2),* 5-12.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201-292. doi: 10.2307/1412107

Spearman, C. E. (1927a). *The abilities of man: Their nature and measurement*. New York, NY: Blackburn Press.

Spearman, C. E. (1927b). *The nature of ``intelligence'' and principles of cognition* (2nd ed.). London, England: Macmillan.

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12,* 35-51. doi: 10.1016/S1041-6080(00)00037-6

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. Retrieved from http://www.nasponline.org

Styck, K. M., & Watkins, M. W. (2013). Diagnostic utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children-Fourth Edition among referred students. *School Psychology Review, 42,* 367-382. Retrieved from http://www.nasponline.org

Taub, G. E., & McGrew, K. S. (2014). The Woodcock-Johnson Tests of Cognitive Abilities III's cognitive performance model: Empirical support for intermediate factors within CHC theory. *Journal of Psychoeducational Assessment, 32,* 187-201. doi: 10.1177/0734282913504808

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Tusing, M. E., & Ford, L. (2004). Examining preschool cognitive abilities using a CHC framework. *International Journal of Testing, 4,* 91-114. doi: 10.1207/s15327574ijt0402_1

Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist, 31,* 1113-1126. doi: 10.1080/13854046.2017.1317364

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children-Fifth Edition*. San Antonion, TX: NCS Pearson.

Williams, J., & Miciak, J. (2018). Adoption costs associated with processing strengths and weaknesses methods for learning disabilities identification. *School Psychology Forum, 12,* 17-19. Retrieved from http://www.nasponline.org

Wolff, H., & Preising, K. (2005). Exploring item and higher order factor structure with the Schmid-Leiman solution: Syntax codes for SPSS and SAS. *Behavior Research Methods,* *37,* 48-58. doi: 10.3758/BF03206397

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258. doi: 10.1177/073428299000800303

Woodcock, R. W. (1998). Extending Gf-Gc theory into practice. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 137-156). Mahwah, NJ: Erlbaum.

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

Woodcock, R. W., & Mather, N. (1989). *WJ-R Tests of Cognitive Ability-Standard and supplemental batteries: Examiner's manual*. Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, Il: Riverside Publishing.

Table 1

List of Provisional Empirical and Conceptual Limitations Identified in the Cattell-Horn-Carroll (CHC) literature

(1)	Use of various iterations of the Woodcock-Johnson to advance CHC and the inherent circularity between the instrument and the theory.
(2)	Strong inference and the use of factor analysis to build elaborate theories.
(3)	Absence of connectivity in the CHC literature.
(4)	Clinical utility and cost-benefit associated with CHC clinical applications.
(5)	Evidential quality of the CHC literature and the forums in which it is disseminated.
(6)	Wherefore art thou Carroll?

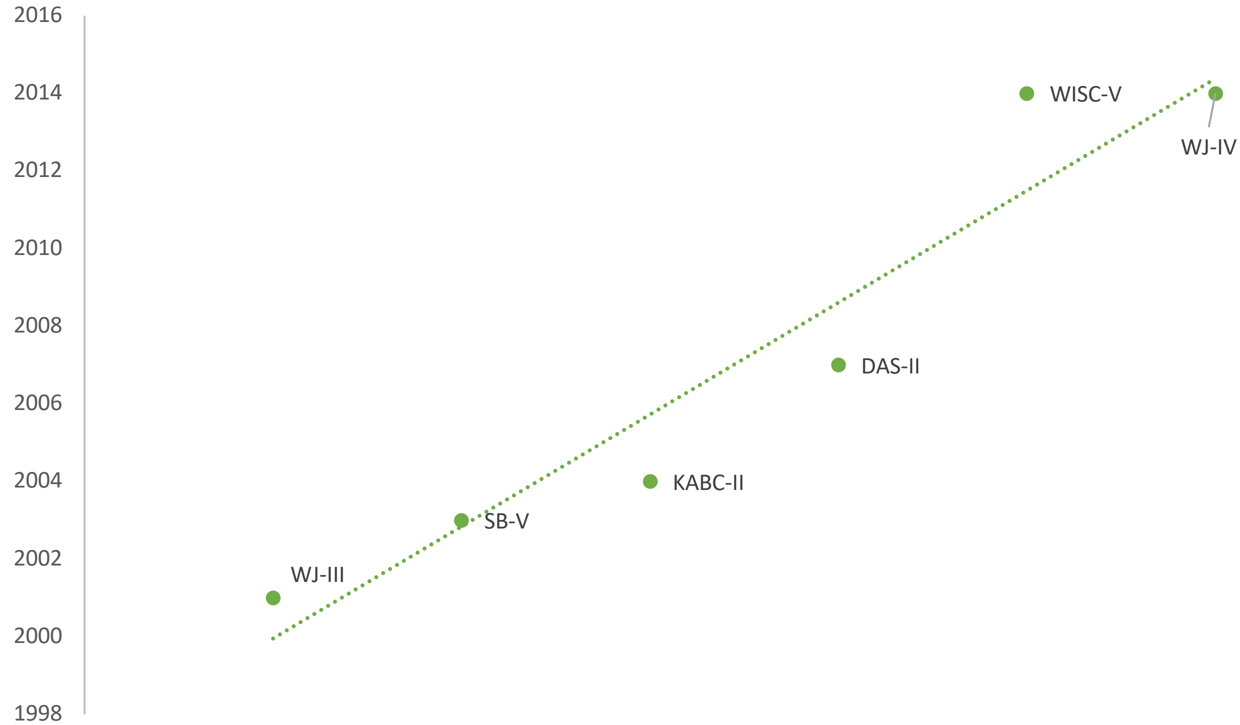

Figure 1. Evolution of major commercial ability measures incorporating Cattell-Horn-Carroll (CHC) inspired factor structures.

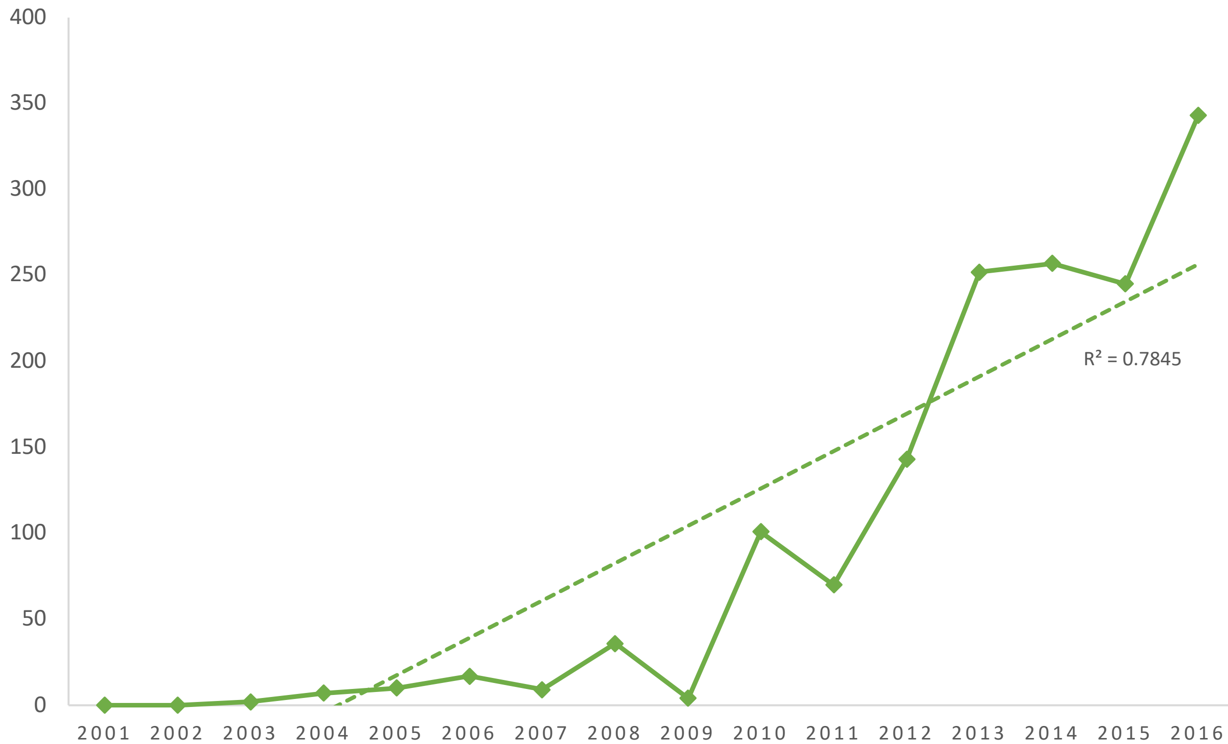

Figure 2. Cattell-Horn-Carroll (CHC) citations from 2000-2016 (Web of Science: July 10, 2017).