Journal of School Psychology 51 (2013) 535–555

Contents lists available at SciVerse ScienceDirect

Journal of School Psychology

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

A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy☆

Matthew R. Reynolds [a,*], Timothy Z. Keith [b], Dawn P. Flanagan [c], Vincent C. Alfonso [d]

[a] University of Kansas, USA
[b] The University of Texas at Austin, USA
[c] St. John's University, USA
[d] Fordham University, USA

ARTICLE INFO

Article history:
Received 4 November 2011
Received in revised form 25 February 2013
Accepted 27 February 2013

Keywords:
Cattell–Horn–Carroll taxonomy
KABC-II
Planned missingness
Fluid intelligence
General intelligence
Flynn effect

ABSTRACT

The Cattell–Horn–Carroll (CHC) taxonomy has been used to classify and describe human cognitive abilities. The ability factors derived from the CHC taxonomy are often assumed to be invariant across multiple populations and intelligence batteries, which is an important assumption for research and assessment. In this study, data from five different test batteries that were collected during separate Kaufman Assessment Battery for Children—Second Edition (KABC-II; Kaufman & Kaufman, 2004) concurrent validity studies were factor-analyzed jointly. Because the KABC-II was administered to everyone in the validity studies, it was used as a reference battery to link the separate test batteries in a "cross-battery" confirmatory factor analysis. Some findings from this analysis were that CHC-based test classifications based on theory and prior research were straightforward and accurate, a first-order Fluid/Novel Reasoning (Gf) factor was equivalent to a second-order *g* factor, and sample heterogeneity related to SES and sex influenced factor loadings. It was also shown that a reference variable approach, used in studies that incorporate planned missingness into data collection, may be used successfully to analyze data from several test batteries and studies. One implication from these findings is that CHC theory should continue to serve as a useful guide that can be used for intelligence research, assessment, and test development.

1. Introduction

Cattell–Horn–Carroll (CHC) theory is an integration of Carroll's three-stratum theory (Carroll, 1993) and the Cattell–Horn extended *Gf-Gc* theory (Horn & Noll, 1997), representing a culmination of over 100 years of psychometric research in human intelligence. It is a multi-factor hierarchical model, which includes 8 to 10 broad cognitive ability factors (e.g., Verbal Comprehension-Knowledge [Gc], Fluid/Novel Reasoning [Gf], Visual–Spatial Ability [Gv], Short-Term Memory [Gsm], Long-Term Retrieval [Glr]), dozens of more specific narrow abilities, and a general factor at the apex (*g*).[1] CHC theory has been used as a template for intelligence test development (Keith & Reynolds, 2010), as a common language used to communicate intelligence test scores and research findings (Flanagan, Alfonso, & Ortiz, 2012; McGrew, 2009), as a model to investigate structural relations between intelligence factors of different order (Kvist & Gustafsson, 2008), and as a model to study relations between intelligence and external variables such as academic achievement (Keith,

☆ We thank Pearson for access to the data used in this research, and the Woodcock-Muñoz Foundation for access to the WJ-III data used to test for measurement invariance.
* Corresponding author at: University of Kansas, Psychology and Research in Education, Joseph R. Pearson Hall, 1122 West Campus Rd, Lawrence KS, 66045-3101, USA. Tel.: +1 785 864 9712.
E-mail address: mreynolds@ku.edu (M.R. Reynolds).
ACTION EDITOR: Craig Albers.
[1] There are differences in opinion on whether a *g* factor should be included, but this discussion is beyond the scope of the current article (cf. Carroll, 2003; Horn & McArdle, 2007).

0022-4405/$ – see front matter
http://dx.doi.org/10.1016/j.jsp.2013.02.003

1999) and personality (Baker & Bichsel, 2006). It is for these reasons, among others, that CHC theory has been recommended as a useful framework for intelligence research (McGrew, 2009). Nevertheless, it is a working model. There are numerous hypotheses that may be tested, including refinement or alterations to the framework itself (e.g., Kan, Kievit, Dolan, & van der Maas, 2011).

Both measurement and structural aspects of CHC theory are investigated in this study, with two primary aims. The first aim is to investigate the invariance of CHC broad ability factors across different intelligence batteries (Mulaik, 2010; Thurstone, 1947). Most new intelligence batteries specify a measurement structure so that two or three subtest scores are grouped to align with CHC broad abilities. If CHC theory is a useful framework for grouping subtest scores, subtests specified to load on CHC factors within a test battery should load on the same CHC factors when subtest scores from other test batteries are included in a factor analysis. The second research aim is to investigate the structural relations between first-order CHC broad ability factors and a second-order *g* factor. Of particular interest are the *g*–Gf relation and the influence of population heterogeneity on these first- and second-order relations (which are sometimes called *loadings*). Gf and *g* are often either perfectly or nearly perfectly related (Gustafsson, 1984), and population heterogeneity may be one reason that loadings vary across studies (Kvist & Gustafsson, 2008).

1.1. Cattell–Horn–Carroll taxonomy

A second-order *g* factor and first-order broad cognitive ability factors are often used to represent an underlying system of common factors that produces variation in intelligence test scores. Almost all tests measure *g*, one or more broad cognitive abilities, and something specific to that test. CHC theory is one such descriptive system (Schneider & McGrew, 2012). It is also often referred to as a taxonomy and has been used extensively to classify abilities and to select, organize, and interpret intelligence test scores (Flanagan et al., 2012). Several intelligence batteries have subtests organized into composites based on CHC theory (e.g., Kaufman Assessment Battery for Children, Second Edition [KABC-II], Kaufman & Kaufman, 2004, and the Woodcock–Johnson III [WJ III] Tests of Cognitive Abilities, Woodcock, McGrew, & Mather, 2001). Each battery assesses various CHC broad abilities, with several abilities (Gf, Gc, and Gv) measured in each. Although CHC factors common to these intelligence batteries are similar in name, they are measured by subtests varying in their task demands, response format, and test stimuli. Are the CHC factors measured by these subtests the same? This question is one of *factorial invariance*.

1.1.1. Factorial invariance

The replicability of common factors across various conditions speaks to the basic nature and usefulness of those factors. Such replicability is encapsulated in factorial invariance. Two types of factorial invariance have been described: invariance under selection of populations and invariance under selection of variables (Mulaik, 2010; Thurstone, 1947). If factors are invariant under selection of populations, then the same common factors should emerge when the same battery is administered to other populations. If factors are invariant under selection of variables, then the same common factors should emerge when other indicator variables (e.g., subtests) are selected from a broader domain of indicators (e.g., different intelligence batteries).

Invariance under selection of populations is often investigated via multi-group confirmatory factor analysis (MG-CFA) and has been formulated within the broader scope of measurement invariance (Meredith, 1993). This approach is often used to investigate internal test bias. Measurement invariance of CHC broad ability factors has been investigated and for the most part supported across age (e.g., Taub & McGrew, 2004) and sex (e.g., Reynolds, Keith, Ridley, & Patel, 2008) in several popular intelligence batteries. The lesser known type of invariance—although implicitly assumed in almost all research and assessment—is invariance under selection of variables (Thurstone, 1947). This type of invariance implies that the factorial composition of a subtest, described via factor loadings, should not differ when a subtest is moved to a new battery, which also includes that common factor. Tucker (1958) succinctly summarized this issue by posing the question "Do factors transcend batteries" (p. 112)? Factors should transcend batteries.

There have been attempts to establish the invariance of *g* under different selections of tests. Thorndike (1987) inserted subtests into different battery groupings, which were then submitted to factor analysis, and found correlations ranging from .52 to .94 between factor loadings of those subtests. He concluded that *g* was invariant because the rank ordering of the loadings was fairly stable across analyses. Other researchers have studied *g* loadings across different test batteries, and have estimated the influences of battery size and composition, factor-extraction technique, and interactions among these influences on the dependability of *g* loadings (cf., Floyd, Shands, Rafael, Bergeron, & McGrew, 2009; Jensen & Weng, 1994; Major, Johnson, & Bouchard, 2011). Despite some similar findings across these studies, the interpretations of those findings have varied.

One argument often levied against *g* invariance is that *g* depends on the characteristics and composition of subtests included in a battery (Horn, 1991). For example, *g* may be biased toward Gc because Gc tests are often overrepresented in intelligence batteries (Ashton & Lee, 2006). Or, the way in which *g* is extracted or modeled, especially related to Gc, may also bias *g*. For example, compared to Gf, "a larger proportion of the systematic variance in the Gc-tests is turned into common factor variance," thus in many studies in which a *g* factor has been represented by a first principal factor, the *g* factor is actually more Gc-like (Kvist & Gustafsson, 2008, p. 434). Because of this concern, among others, higher-order factor models have often been recommended to model *g* (Jensen, 1998; Keith, 2005). When *g* has been modeled as a second-order factor in CFA models using data from examinees who were administered more than one intelligence test battery, the two *g* factors correlate either perfectly or near perfectly (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004; Keith, Kranzler, & Flanagan, 2001). Some researchers consider this correlation as evidence of *g* invariance. The substantive interpretation of that correlation differs, however. Some researchers suggest that it is strong because the tests measure the same "thing," whereas other researchers suggest it is strong because they measure the same "things" (cf., Horn & McArdle, 2007, p. 221; Johnson et al., 2004; van der Maas, Dolan, Grasman, Wicherts,