Journal of School Psychology 58 (2016) 1–19

Contents lists available at ScienceDirect

Journal of School Psychology

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

Examining the integrity of measurement of cognitive abilities in the prediction of achievement: Comparisons and contrasts across variables from higher-order and bifactor models☆

Nicholas F. Benson [a,*], John H. Kranzler [b], Randy G. Floyd [c]

[a] Baylor University, United States
[b] University of Florida, United States
[c] University of Memphis, United States

ARTICLE INFO

Article history:
Received 12 August 2015
Received in revised form 10 June 2016
Accepted 27 June 2016
Available online xxxx

Keywords:
General intelligence
Broad cognitive abilities
Achievement
Factor scores

ABSTRACT

Prior research examining cognitive ability and academic achievement relations have been based on different theoretical models, have employed both latent variables as well as observed variables, and have used a variety of analytic methods. Not surprisingly, results have been inconsistent across studies. The aims of this study were to (a) examine how relations between psychometric *g*, Cattell–Horn–Carroll (CHC) broad abilities, and academic achievement differ across higher-order and bifactor models; (b) examine how well various types of observed scores corresponded with latent variables; and (c) compare two types of observed scores (i.e., refined and non-refined factor scores) as predictors of academic achievement. Results suggest that cognitive–achievement relations vary across theoretical models and that both types of factor scores tend to correspond well with the models on which they are based. However, orthogonal refined factor scores (derived from a bifactor model) have the advantage of controlling for multicollinearity arising from the measurement of psychometric *g* across all measures of cognitive abilities. Results indicate that the refined factor scores provide more precise representations of their targeted constructs than non-refined factor scores and maintain close correspondence with the cognitive–achievement relations observed for latent variables. Thus, we argue that orthogonal refined factor scores provide more accurate representations of the relations between CHC broad abilities and achievement outcomes than non-refined scores do. Further, the use of refined factor scores addresses calls for the application of scores based on latent variable models.

1. Introduction

A psychological construct is defined as an "attribute of people, assumed to be reflected in test performance" (Cronbach and Meehl, 1955, p. 283). Two domains are held to underlie the constructs measured on psychological tests: a theoretical domain

☆ The authors thank NCS Pearson for providing the standardization data for the Kaufman Assessment Battery for Children—Second Edition (KABC-II; A. Kaufman & N. Kaufman, 2004a) and the Kaufman Tests of Educational Achievement—Second Edition (KTEA-II; A. Kaufman & N. Kaufman, 2004b). Copyrights by NCS Pearson, Inc. used with permission. The authors also thank the guest editor, Dr. Joel Schneider, for his helpful insights regarding this research, including creating Fig. 1, and calculating reliability coefficients for construct scores.
* Corresponding author at: Department of Educational Psychology, Baylor University, One Bear Place #97301, Waco, TX 76798, United States.
E-mail address: Nicholas_Benson@baylor.edu (N.F. Benson).
ACTION EDITOR: Joel Schneider.

http://dx.doi.org/10.1016/j.jsp.2016.06.001
0022-4405/

and an empirical domain (see J. Benson, 1998, for discussion). The theoretical domain represents all that is known about the construct, whereas the empirical domain concerns its operationalization in behavioral terms (i.e., observable variables). Ideally, the empirical domain is an accurate reflection of the theoretical domain. Validation of the use of a psychological test involves determining the extent to which the empirical evidence supports interpretation of its scores in terms of the constructs the test was designed to measure. As Messick (1989) stated, construct validity involves substantiation of "the adequacy and appropriateness of inferences and actions based on test scores" (p. 6).

1.1. Intelligence theory and the structure of cognitive abilities

For the first 100 years of standardized intelligence ("IQ") testing, IQ tests could be criticized on the grounds that the constructs they measured were not derived from sound theory (e.g., Brody, 1992; Esters, Ittenbach, and Han, 1997). In recent years, however, test developers have increasingly used contemporary theory—and the Cattell–Horn–Carroll theory (CHC; see McGrew, 2005; Schneider and McGrew, 2012) in particular—to guide test development. CHC theory is an integration of Carroll's (1993) three-stratum theory and the Horn–Cattell fluid and crystallized theory of intelligence (*Gf-Gc*; Horn and Noll, 1997). Essentially a taxonomy of the structure of human cognitive abilities, CHC theory, specifies the number of factors (i.e., latent variables) underlying individual differences in test performance and their relation to each other. In CHC theory, "intelligence" is conceptualized as a multidimensional construct that consists of a number of different cognitive abilities. CHC theory posits a hierarchical structure with three levels of cognitive abilities that reflect different degrees of referent generality, defined as "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). The factor with the most referent generality is psychometric *g*, which is located at the apex of the hierarchy at stratum III. In CHC theory, psychometric *g* is typically conceptualized and modeled as a higher-order factor, and its effects on all variables are indirect. That is, lower-order factors completely mediate the effects of the higher-order psychometric *g* factor on all test scores (e.g., Gignac, 2008). Stratum II consists of 15 broad cognitive abilities (e.g., Fluid Reasoning [*Gf*], Comprehension-Knowledge [*Gc*], Short-Term Memory [*Gsm*], Long-Term Storage and Retrieval [*Glr*], Visual Processing [*Gv*]), and the third and last stratum consists of over 80 narrow cognitive abilities (McGrew, 2005; Schneider and McGrew, 2012).

The importance of CHC theory lies in the fact that it provides researchers and test developers with a common theoretical framework and nomenclature for understanding individual differences in cognitive abilities and their measurement. As a result, CHC theory "has formed the foundation for most contemporary IQ tests" (A. Kaufman, 2009, p. 91). Some IQ tests, such as the Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV; Schrank, McGrew, and Mather, 2014) and the Kaufman Assessment Battery for Children, Second Edition (KABC-II; A. Kaufman and N. Kaufman, 2004a), are explicitly based on CHC theory; whereas other IQ tests, such as the Wechsler Intelligence Scales for Children, Fifth Edition (WISC-V; Wechsler, 2014), can be seen to have been influenced by CHC theory.

Despite the fact that CHC theory has been increasingly used over the past decade as the foundation for the development of new IQ tests or the revision of extant tests, it is important to note that scientific truth is not determined by the prevailing scientific zeitgeist. During the past century, a consensus has never been reached among researchers on the structure of cognitive abilities. A number of contemporary intelligence theories are not hierarchical and do not include a general ability, such as Sternberg's triarchic theory (1985), Gardner's theory of multiple intelligences, Ceci's bioecological framework (e.g., Ceci, 1990), and the planning, attention, and simultaneous-successive processes theory (e.g., Naglieri and Otero, 2012). Moreover, during the past decade, the bifactor model has been advocated as a viable alternative to the higher-order model underlying CHC theory (e.g., Beaujean, Parkin, and Parker, 2014; Canivez, 2016; Gignac, 2008; Gignac and Watkins, 2013; Kranzler, Benson, and Floyd, 2015; Reise, 2012; Watkins and Beaujean, 2014).

This bifactor model (also referred to as a direct hierarchical or nested-factors model) differs from the higher-order model in at least two important ways that suggest the constructs targeted in each model vary somewhat. First, bifactor models specify psychometric *g* as a first-order "breadth" factor that directly affects scores on all tests, rather than as a higher-order factor that directly affects lower-order factors and indirectly affects test scores. As a result, interpretation of the general factor in bifactor models is more straightforward than it is in the higher-order models. In higher-order models, lower-order factors are specified as explaining all the covariance across test scores, and the higher-order general factor is defined by these lower-order factors and not by test scores (Beaujean et al., 2014). These higher-order factors are thus "abstractions of abstraction even more removed from the measured variables" (Thompson, 2004, pp. 73–74). In contrast, the general factor in a bifactor model reflects the direct effects of the construct of psychometric *g* on all test scores.

Second, more specific cognitive ability factors (often called group factors in bifactor models) are specified as reflecting the covariation of reliable residual variance in test scores after accounting for the effects of psychometric *g*. Like the general factor in the bifactor model, they are also first-order factors, but they affect only subsets of test scores. They represent specific cognitive abilities that are (a) orthogonal to (i.e., negligibly correlated with) psychometric *g*—essentially controlling for its influence—and (b) orthogonal to one another. As a result, each factor represents only one cognitive ability—regardless of its stratum level. This difference in how specific abilities are represented across higher-order models and bifactor models has received little attention, but is important to understand.

The benefits of the bifactor model are appealing to many of its advocates for three reasons associated with independence of measurement and prediction. One reason is that variation in a test score can be explained by as few as three or four influences—psychometric *g*, specific abilities consistent with group factors, and an admixture of test-specific influences of the test itself and random error. As a result, it is easier to understand the unique (perhaps confounding) effects across higher- and lower-order stratum cognitive abilities (see Gustafsson, 2002) using a bifactor model. For example, in a higher-order model,