Intelligence 62 (2017) 31–47

Contents lists available at ScienceDirect

Intelligence

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

Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure?☆

Matthew R. Reynolds [a,*], Timothy Z. Keith [b]

[a] University of Kansas, United States
[b] The University of Texas at Austin, United States

ARTICLE INFO

Article history:
Received 29 June 2016
Received in revised form 3 December 2016
Accepted 17 February 2017
Available online 3 March 2017

Keywords:
Construct validity
Intelligence
Wechsler Intelligence Scale for Children, Fifth Edition

ABSTRACT

The purpose of this research was to test the consistency in measurement of Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014) constructs across the 6 through 16 age span and to understand the constructs measured by the WISC-V. First-order, higher-order, and bifactor confirmatory factor models were used. Results were compared with two recent studies using higher-order and bifactor exploratory factor analysis (Canivez, Watkins, & Dombrowski, 2015; Dombrowski, Canivez, Watkins, & Beaujean, 2015) and two using confirmatory factor analysis (Canivez, Watkins, & Dombrowski, 2016; Chen, Zhang, Raiford, Zhu, & Weiss, 2015). We found evidence of age-invariance for the constructs measured by the WISC-V. Further, both *g* and five distinct broad abilities (Verbal Comprehension, Visual Spatial Ability, Fluid Reasoning, Working Memory, and Processing Speed) were needed to explain the covariances among WISC-V subtests, although Fluid Reasoning was nearly equivalent to *g*. These findings were consistent whether a higher-order or a bifactor hierarchical model was used, but they were somewhat inconsistent with factor analyses from the prior studies. We found a correlation between Fluid Reasoning and Visual Spatial factors beyond a general factor (*g*) and that Arithmetic was primarily a direct indicator of *g*. Composite scores from the WISC-V correlated well with their corresponding underlying factors. For those concerned about the fewer numbers of subtests in the Full Scale IQ, the model implied relation between *g* and the FSIQ was very strong.

1. Introduction

The Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014) is the latest in this venerable and popular individually administered measure of intelligence for children and adolescents. The publishers probably felt like they were "walking a tightrope" (Keith & Witta, 1997, p. 89) during this revision as they balanced need for continuity with need for modernization based on contemporary theory and research. As in previous revisions since David Wechsler's death, the publishers have made numerous changes. For example, two subtests from the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003) have been eliminated (i.e., Word Reasoning and Picture Completion); thirteen subtests have been retained with modifications; and eight subtests have been added. A number of changes have been made to the composites derived from those subtests. The purpose of this research was to test the consistency in measurement of WISC-V constructs across the 6–16 age span and to understand the constructs measured by WISC-V subtests and composites. We used confirmatory factor analysis (CFA) to test for invariance and to test specific hypotheses about the constructs measured by the test. We also compared our findings with those from two recent exploratory factor analyses (EFA) (Canivez, Watkins, & Dombrowski, 2015; Dombrowski, Canivez, Watkins, & Beaujean, 2015) and two recent CFAs of the WISC-V (Canivez, Watkins, & Dombrowski, 2016; Chen, Zhang, Raiford, Zhu, & Weiss, 2015).

1.1. Wechsler Intelligence Scale for Children–Fourth Edition

The WISC-IV was described as "the best WISC ever" (Kaufman, Flanagan, Alfonso, & Mascolo, 2006, p. 293). It was lauded for its improved alignment with contemporary intelligence theories, but some researchers commented that the publishers did not take the last step and explicitly align the measure with a theory (e.g., Keith, Fine, Taub, Reynolds, & Kranzler, 2006). The WISC-V likewise has aligned itself more with contemporary hierarchical structural theories of intelligence (i.e., general construct at the highest level, broad cognitive abilities a level below, and narrower abilities at another level below) without explicitly endorsing any one theory

☆ Standardization data from the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-5). Copyright © 2014 NCS Pearson, Inc. Used with permission. All rights reserved.

* Corresponding author at: Educational Psychology, Joseph R. Pearson Hall, 1122 West Campus Rd, University of Kansas, Lawrence, KS 66045, United States.
E-mail address: mreynolds@ku.edu (M.R. Reynolds).

http://dx.doi.org/10.1016/j.intell.2017.02.005
0160-2896/

(Wechsler, 2014, p. 23). Nevertheless, broad cognitive abilities referenced in the WISC-V appear closely aligned with Cattell-Horn-Carroll theory (Carroll, 1993; Horn & Blankson, 2005). This alignment seems predictable given that the revisions of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV; Wechsler, 2008) and the Wechsler Preschool and Primary Intelligence Scale–Fourth Edition (WPPSI-IV; Wechsler, 2012) have also moved closer to alignment with CHC theory.

Factor analysis of WISC-IV data generally supported a structure with a hierarchical *g* factor along with four to five first-order factors (e.g., Canivez, 2014; Keith et al., 2006; Watkins, 2006; Weiss, Keith, Zhu, & Chen, 2013b). The five first-order factors were consistent with CHC broad abilities: Verbal Comprehension (i.e., Comprehension-Knowledge or Gc), acquired verbal knowledge and the ability to express that knowledge; Visual Spatial (i.e., Visual Processing or Gv), the ability to use simulated mental imagery to solve problems, Fluid Reasoning (or Gf), novel reasoning; Working Memory (i.e., Short-Term Memory or Gsm), the ability to maintain and use information in one's immediate awareness; and Processing Speed (Gs), the ability to perform simple repetitive tasks quickly (Schneider & McGrew, 2012). Research was mixed, however, with regard to Gv and Gf being two separate factors versus one Perceptual Reasoning factor, and, relatedly, the abilities measured by the Arithmetic subtest. A special issue of the *Journal of Psychoeducational Assessment* explored these issues in more depth (Tobin, 2013).

1.2. WISC-V: questions about what it measures

1.2.1. Primary indexes

In contrast to the WISC-IV, the WISC-V is designed to measure five CHC-type factors, with its scoring structure including five primary indexes (i.e., Verbal Comprehension, Fluid Reasoning, Visual Spatial, Working Memory, and Processing Speed). Toward this end, one new Gf measure and one new Gv measure were included (Figure Weights and Visual Puzzles respectively–similar to WAIS-IV subtests). The Perceptual Reasoning Index was eliminated. The Working Memory Index was changed; it pairs Digit Span with Picture Span, a new subtest adapted from the WPPSI-IV that was designed to measure visual working memory. Further, a third task, Digit Span Sequencing, was added to Digit Span. Last, two extended primary indexes, each composed of four subtests, have been included since the test was published; an Extended Verbal Comprehension Index and an Extended Fluid Reasoning Index (Raiford, Drozdick, Zhang, & Zhou, 2015). Arithmetic, traditionally associated with Working Memory (or in earlier versions Verbal intelligence), is now associated with Fluid Reasoning and included in the Extended Fluid Reasoning Index. Thus two pertinent questions concerning the WISC-V are whether the five-factor measurement model is viable and if the structure is invariant across the age range. Further, what are the correlations between the primary indexes and their respective factors?

1.2.2. Full Scale IQ (FSIQ) and General Ability Index (GAI)

Two options are provided as measures of general intelligence on the WISC-V: FSIQ and GAI. Perhaps the most unforeseen and important change to the WISC-V is that the FSIQ includes only seven subtests compared to the traditional 10. Further, some subtests included in the FSIQ are new. The seven subtests are weighted toward constructs more closely aligned with psychometric *g*. For example, 40% of the subtests included in the WISC-IV FSIQ were from the Processing Speed and Working Memory Indexes. That percentage has been reduced to 29%, with only one subtest from the Working Memory Index and one subtest from the Processing Speed Index included on the WISC-V FSIQ. The remaining five FSIQ subtests include two from the Verbal Comprehension Index, two from the Fluid Reasoning Index, and one from the Visual-Spatial Index. Psychometric *g* tends to be most closely related to fluid intelligence and verbal comprehension (Carroll, 1993),[1] so it seems appropriate to give subtests from these two domains more weight (Reynolds & Hadorn, 2016).

The averaged internal consistency estimate for the WISC-V FSIQ was reported in the technical and interpretative manual as 0.96 (Wechsler, 2014), compared with 0.97 on the WISC-IV (Wechsler, 2003), suggesting that there is not a difference in the average correlations among the items when using seven subtests versus ten subtests. From a validity standpoint, however, it is important to understand the extent to which the FSIQ indexes psychometric *g* (Daniel, 1997), or how much FSIQ variance is explained by the general factor–omega hierarchical (Gustafsson & Åberg-Bengtsson, 2010; McDonald, 1999; Zinbarg, Yovel, Revelle, & McDonald, 2006).[2] Model-based omega hierarchical estimates, when averaged across ages, indicated that about 82% of the FSIQ variance on the WISC-IV was explained by *g* (Reynolds, Floyd, & Niileksela, 2013). Or alternatively, a model implied correlation between *g* and the FSIQ was 0.91 (i.e., square root of 0.82). This correlation has been referred to as the "*g* loading" of the FSIQ (Farmer, Floyd, Reynolds, & Kranzler, 2014; Jensen, 1998). Those estimates were based on the 10 subtest WISC-IV FSIQ. Some have described moving away from a 10 subtest composite to the briefer seven subtest composite as a limitation, suggesting the FSIQ lacks breadth and is basically a screening score (Sattler, Dumont, & Coalson, 2016). Does a seven subtest composite measure *g* as well a 10 subtest composite? In addition, the subtests that compose the FSIQ have changed considerably. With fewer and different subtests included in the WISC-V FSIQ, is the model-implied correlation between the FSIQ and *g* (or *g* loading of the FSIQ) as strong as with previous versions of the instrument? Further, what are the model implied *g* correlations, and how much do they change with alternative global composites made up of different subtests?

Last, the GAI is an alternative measure of general intelligence, minus working memory and processing speed influences. It includes five subtests: the FSIQ subtests minus Digit Span and Coding. What is the model implied correlation between the GAI and *g*? These questions with regard to *g*, the FSIQ, and the GAI have yet to be answered.

1.2.3. WISC-V subtests

CFAs in the WISC-V technical and interpretative manual indicated that a higher-order model with five first-order factors fit best out of several competing models (Wechsler, 2014).[3] Subtests were for the most part associated with their predicted group factor. Arithmetic, as in the past, was less clear, loading on Fluid Reasoning (0.32), Working Memory (0.31), and Verbal Comprehension factors (0.16) (Wechsler, 2014). Further, Figure Weights loaded on the Visual-Spatial factor in addition to the Fluid Reasoning factor, but there was no clear theoretical rationale so it was dismissed (Wechsler, 2014, p. 83).

We had additional questions about the constructs measured by the subtests. For example, Similarities was designed to measure verbal concept formation and abstract reasoning (Flanagan & Kaufman, 2009; Wechsler, 2014). Does Similarities load on Fluid Reasoning in addition to Verbal Comprehension? Further, subtests that avoid verbal stimuli may also invoke visual spatial abilities even though they were designed to measure a different broad ability such as reasoning or processing speed. For example, Matrix Reasoning and Symbol Search also loaded on a weakly-defined Gv factor in the WISC-IV (Keith et al., 2006; Weiss et al., 2013b). Do these subtests also measure visual spatial ability? A potentially better defined Gv factor on the WISC-V compared with the WISC-IV may help to answer that question. Last, a new subtest, Picture Span, requires visual memory skills. Visual memory, defined as

[1] Three-stratum theory aligns broad abilities from left to right in the order that they are associated with the second-order *g* factor. Fluid Reasoning and Crystallized Intelligence are the first two in order from left to right (Carroll, 1993).

[2] These are often referred to as composite reliability estimates, but omega hierarchical seems more consistent with a validity estimate—or somewhere in between because of the focus on the specific construct-related variance.

[3] See Beaujean (2016) for a critique of the factor scaling method used in these CFAs.