WAIS-IV and WISC-IV Structural Validity: Alternate Methods, Alternate Results. Commentary on Weiss et al. (2013a) and Weiss et al. (2013b)

Gary L. Canivez[1] and Joseph C. Kush[2]

Abstract

Weiss, Keith, Zhu, and Chen (2013a) and Weiss, Keith, Zhu, and Chen (2013b), this issue, report examinations of the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV) and Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV), respectively; comparing Wechsler Hierarchical Model (W-HM) and Cattell–Horn–Carroll Hierarchical Model (CHC-HM), subtest cross-loadings, and factorial invariance between clinical and normative groups from the respective standardizations. Both studies suffer from a number of theoretical, methodological, and practical problems that significantly limit conclusions and recommendations for practitioner interpretations of these instruments. Additional analyses and research are required to better inform practitioners for WAIS-IV and WISC-IV use.

We are pleased for the opportunity to provide comments regarding two articles (Weiss, Keith, Zhu, & Chen, 2013a; Weiss, Keith, Zhu, & Chen, 2013b) that examined the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV; Wechsler, 2008a) and Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003a), respectively, in the present issue. We are particularly appreciative of the editors of the *Journal of Psychoeducational Assessment* for inviting us and for devoting the page space necessary to critically review these detailed studies. While the role of the school psychologist continues to evolve to include a variety of tasks beyond assessment (e.g., designing interventions, counseling, consulting, program evaluation, etc.), the importance of psychoeducational assessment nevertheless remains an important aspect of the role and function. And while empirical studies examining assessment practices and the psychometric properties of psychological instruments is becoming less in vogue within

[1]Eastern Illinois University, Charleston, IL, USA
[2]Duquesne University, Pittsburgh, PA, USA

Corresponding Author:
Gary L. Canivez, Eastern Illinois University, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920, USA.
Email: glcanivez@eiu.edu

professional school psychology journals, it is refreshing to find the *Journal of Psychoeducational Assessment* encouraging the presentation and discussion of this important topic. Our review begins by acknowledging the strengths of these articles, continues with a section of what we consider limitations, and concludes with some final comments regarding implications of these findings for practitioners and for future research. Our commentary addresses what we consider to be a number of theoretical, methodological, and practical limitations of these studies, which we believe impact the interpretation of the results and ultimately the utility of the findings for practitioners.

In the present issue, Weiss et al. (2013a) and Weiss et al. (2013b) present examinations of the internal structure and factorial invariance between the clinical sample and the normative standardization sample gathered during respective standardizations for the WAIS-IV (Wechsler, 2008a) and the WISC-IV (Wechsler, 2003a). As theoretical consensus on the specific number of factors being measured by the WAIS-IV and WISC-IV does not exist, the authors compared four- and five-factor models of both instruments and subsequently completed additional analyses to determine potential improvements in model fit by including subtest cross-loadings. Finally, factorial invariance of the authors' purported best model was examined between the clinical sample and a random sample of the same size from the normative standardization sample. Specifically, multigroup confirmatory factor analyses compared WAIS-IV (Weiss et al., 2013a) and WISC-IV (Weiss et al., 2013b) models consisting of four first-order latent factors, hereafter referred to as the Wechsler Hierarchical Model (W-HM), with five first-order latent factors, hereafter referred to as the Cattell–Horn–Carroll Hierarchical Model (CHC-HM).

Results from both studies indicated that the initial WAIS-IV W-HM and the initial WISC-IV W-HM were well-fitting models based on current (but contentious [Marsh, Hau, & Wen, 2004]) standards (Hu & Bentler, 1998, 1999), which the authors claimed supported the respective W-HM structures. Further exploration led the authors to conclude that including various subtest cross-loadings on alternate factors (thereby abandoning simple structure and reducing parsimony) in the WAIS-IV W-HM and WISC-IV W-HM improved fit based on $\Delta\chi^2$, AIC, and aBIC criteria; and these models were forwarded for examination of invariance between the respective clinical and normative samples. It should be noted however, that had the authors used alternate criteria for meaningful or practical model differences (e.g., ΔCFI > +.01, Cheung & Rensvold, 2002; and ΔRMSEA > −.015, Chen, 2007) then both the WAIS-IV W-HM (ΔCFI = .01, ΔRMSEA = .008) and WISC-IV W-HM (ΔCFI = .000, ΔRMSEA = .008) initial and cross-loading models would have been found equivalent.

With respect to the initial WAIS-IV CHC-HM and WISC-IV CHC-HM models, the authors reported well-fitting models based on current standards (Hu & Bentler, 1998, 1999). Exploration of various multiple subtest cross-loadings on alternate factors (again abandoning simple structure and reducing parsimony) *and* introducing intermediary latent factors ("Quantitative Reasoning" for the WAIS-IV and "Inductive Reasoning" for the WISC-IV) between the G_f factor and indicators (subtests), led Weiss and colleagues to conclude improved fit based on $\Delta\chi^2$, AIC, and aBIC; and these models were then examined for invariance between the respective clinical and normative samples. However, as with the W-HMs, meaningful differences between these initial and cross-loading models were not observed based on Cheung and Rensvold (2002) and Chen (2007) criteria for both the WAIS-IV CHC-HM (ΔCFI = .000, ΔRMSEA = .006) and the WISC-IV CHC-HM (ΔCFI = .000, ΔRMSEA = .012).

Finally, both studies reported general support for factorial invariance of the WAIS-IV W-HM and WISC-IV W-HM for the respective normative and clinical samples as well as support for factorial invariance of the WAIS-IV CHC-HM and WISC-IV CHC-HM for the respective normative and clinical samples. These invariant results are consistent with other studies that have reported the measurement invariance of the WISC-IV (Chen & Zhu, 2012; Chen, Keith, Weiss, Zhu, & Li, 2010). While the overall results appear to support previously reported WAIS-IV and

WISC-IV internal structures, the results presented are a disappointingly incomplete examination of important theoretical, methodological, and practical issues that must be considered in making an accurate characterization of the Wechsler Scales.

Theoretical Limitations

Problem #1: Selective Reporting and Review of WISC-IV and WAIS-IV Literature

Both articles suffer from an overly restricted, and somewhat selective, discussion of literature available on the internal structure of the WAIS-IV and WISC-IV specifically (and intelligence tests in general) that should have provided the basis for testing alternate rival models and conceptualizations rather than simply accepting the various hierarchical models. For example, Weiss et al. (2013b) simply listed studies by Watkins (2006), Watkins, Wilson, Kotz, Carbone, and Babula (2006), and Bodin, Pardini, Burns, and Stevens (2009) as providing support for, or replicating, the four WISC-IV first-order factors. However, Watkins (2006) and Watkins et al. (2006) were deeply concerned with the fact that the WISC-IV *Technical and Interpretive Manual* (Wechsler, 2003b) provided *no* higher order or hierarchical factor analyses despite the theoretical model of the WISC-IV and Wechsler's definition of intelligence (i.e., global capacity; Wechsler, 1939, p. 229). Further, Watkins and colleagues argued that the obliquely rotated and correlated factor indexes required explication of the higher-order factor (Gorsuch, 1983, 1988; Thompson, 2004). Further, as the WISC-IV was influenced by more "contemporary" CHC theory it was also important to apply a statistical method that Carroll (1993, 1995, 1997a, 2003) insisted upon and that others (Carretta & Ree, 2001; Gustafsson & Snow, 1997; McClain, 1996; Ree, Carretta, & Green, 2003; Thompson, 2004) recommended: the Schmid and Leiman (1957) orthogonal transformation to apportion subtest variance to *g* and to the first-order factors. Watkins (2006) and Watkins et al. (2006) applied the SL transformation and apportioned WISC-IV subtest variance to the higher-order *g* and first-order factors, illustrating that the overwhelming majority of subtest variance was associated with the higher-order *g* and *not* the first-order factors; and that the *g* factor accounted for considerably more total and common variance. *That* is *why* it was recommended that interpretation of the WISC-IV should "not discount the strong general factor" (Watkins, 2006, p. 123). This conclusion was further supported by Bodin et al. (2009) who also demonstrated through CFA that the overwhelming majority of WISC-IV subtest variance was associated with higher order *g* and *not* the first-order factors, and the *g* factor accounted for considerably more total and common variance.

Similarly, Weiss et al. (2013a) failed to cite results from research (Canivez & Watkins, 2010a, 2010b) that applied the SL transformation to the WAIS-IV standardization data; which, like the WISC-IV, demonstrated the overwhelming majority of WAIS-IV subtest variance was generally apportioned to the higher-order *g* and *not* the first-order factors; and the *g* factor accounted for substantially more total and common variance than that accounted for by the four first-order factors. Results such as these are not unusual or problematic because all intelligence tests seek to measure general intelligence (but also hope to adequately measure lower order dimensions) and *g* dominance has been consistently observed in numerous other tests of cognitive abilities (see Canivez, 2013).

Other studies not presented by Weiss et al. (2013a) or Weiss et al. (2013b) but that are directly related to the question regarding the latent structure of the WAIS-IV and WISC-IV were those conducted by Gignac (2005, 2006) where he described and compared a rival "direct hierarchical" CFA model (previously referred to as the bifactor model [Holzinger & Swineford, 1937] or the nested factors model [Gustafsson & Balke, 1993]) to the traditional hierarchical model. Keith (2005) also noted the similarity between the nested factors model and the SL transformation and that it provided better fit for the WISC-IV than did the indirect hierarchical model. The important
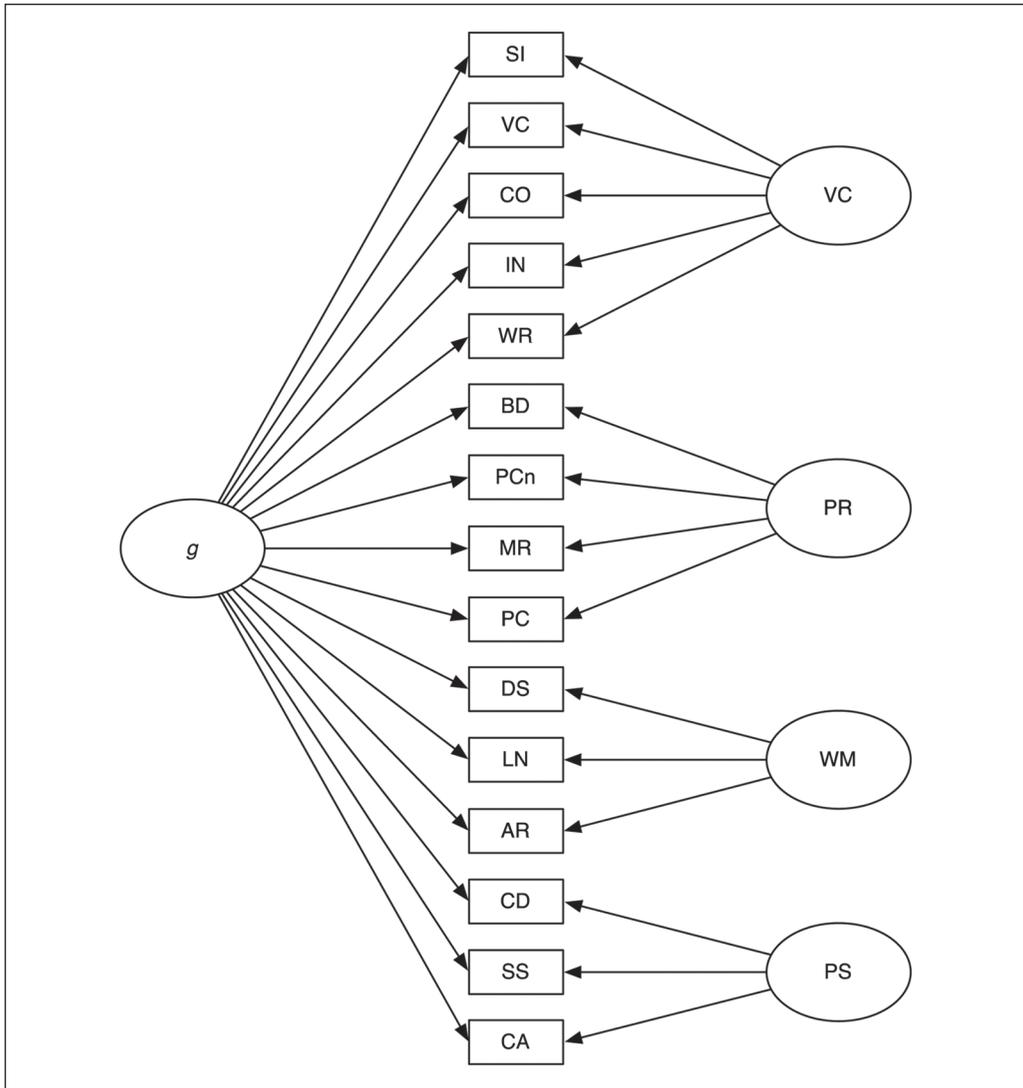
Figure 1. Direct hierarchical model for the Wechsler Intelligence Scale for Children–Fourth Edition (Wechsler, 2003a).
Note. SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; PC = Picture Completion; DS = Digit Span; LN = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search; CA = Cancellation; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; g = General Intelligence.

aspect is that in the direct hierarchical model the higher order *g* factor has direct paths to the subtests (and so technically not "hierarchical") as do the first–order factors (see Figure 1) and higher-order *g* is conceptualized as a *breadth* factor rather than higher-order *g* having paths to first-order factors that in turn have direct paths to the subtests (see Figure 2) and thus higher-order *g* conceptualized as a *superordinate* factor. Given this support, consideration of the direct hierarchical model as a competing model should have been examined.
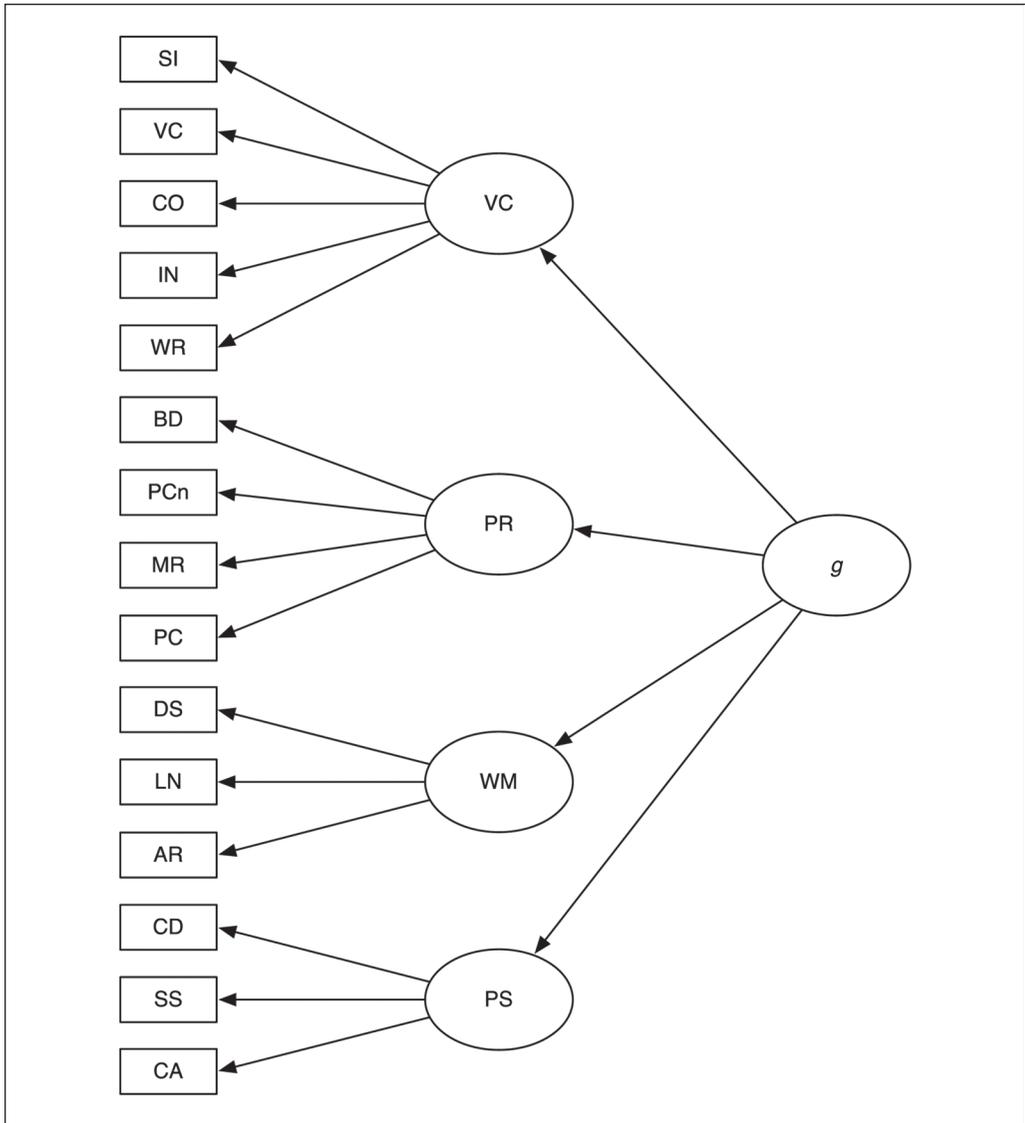
Figure 2. Indirect hierarchical model for the Wechsler Intelligence Scale for Children–Fourth Edition (Wechsler, 2003a).

Note. SI = Similarities; VC = Vocabulary; CO = Comprehension; IN = Information; WR = Word Reasoning; BD = Block Design; PCn = Picture Concepts; MR = Matrix Reasoning; PC = Picture Completion; DS = Digit Span; LN = Letter-Number Sequencing; AR = Arithmetic; CD = Coding; SS = Symbol Search, CA = Cancellation; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor; g = General Intelligence.

Problem #2: Problems With CHC Models

A number of problems are apparent in the CHC-HMs presented for both the WAIS-IV (Weiss et al., 2013a) and WISC-IV (Weiss et al. (2013b). First, while fit statistics indicated as good or somewhat better fit for CHC-HMs compared to W-HMs, the authors were forced to create an intermediary latent factor to support a CHC-HM structure to complete their analyses. In the case

of the WAIS-IV, G_f did not fit well enough and required the creation of the "Quantitative Reasoning" factor with hypothesized paths to the Figure Weights and Arithmetic subtests. In the case of the WISC-IV, G_f also did not fit well enough and required the creation of the "Inductive Reasoning" factor, with hypothesized paths to the Picture Concepts and Matrix Reasoning subtests. If a CHC-HM is to be a viable structure for the WAIS-IV and WISC-IV, why was it necessary to create intermediary factors? More troubling is the fact that standardized paths from g to G_f were 0.99 for the WAIS-IV and 1.0 for the WISC-IV, indicating that G_f was isomorphic with the higher-order g factor. If G_f is indistinguishable from higher-order g, what value is G_f in the model? It seems that this is an indication that a CHC-HM is an overfactored model and not truly viable. If the WISC-IV and WAIS-IV are to better measure CHC constructs it may well be that additional subtests will need to be created, standardized, and normed to provide for such measurement.

Our conclusion is supported by Benson, Hulac, and Kranzler (2010) who compared the Wechsler hierarchical model and a CHC model in higher-order CFA with the WAIS-IV standardization sample and argued for superiority of the CHC model with a variety of subtest cross-loadings. They, however, did not create an intermediary factor to support their CHC model but also observed and reported a standardized path of 1.0 from g to G_f in the initial model and 0.99 in the final model. This was pointed out as a major problem by Ward, Bergman, and Hebert (2012); however, their proposed CHC solution was incomplete due to a lack of inclusion of higher-order g. Similarly, Chen, Keith, Chen, and Chang (2009) compared the Wechsler hierarchical model versus a CHC model in higher-order CFA using the Taiwan WISC-IV standardization sample and argued for the superiority of the CHC model; but again, the standardized coefficient of 0.98 between higher-order g and G_f is problematic. Such high coefficients indicated isomorphism and thus G_f was indistinguishable from higher order g, thereby abandoning parsimony. These results of isomorphism of G_f and g may, however, be an artifact of CFA methods as illustrated by Golay, Reverte, Rossier, Favez, & Lecerf (2012) who used Bayesian Structural Equation Modeling with the French WISC-IV.

Methodological Limitations

Problem #3: Fishing Expeditions, Abandoning Simple Structure, and the Search for Cross-Loadings and Less Parsimony for Miniscule Improvements in Fit. What Do They Provide?

We have a number of methodological concerns regarding the present studies, some technical and others related to what we would consider to be inappropriate utilization of the CFA technique. While the sample sizes of these studies are adequate, there is no discussion of how missing cases were handled, nor any discussion of normality, outliers, linearity, or multicollinearity. Similarly, no justification for their choice of statistical software or estimation methods were provided. A discussion of each of these issues is considered basic, required information in manuscripts that report CFA as the primary statistical technique (Schreiber et al., 2006).

More importantly, CFA is a *confirmatory* statistical technique and is designed to be theoretically driven. For the most part, the authors attempted to put their discussion of the evolution of the Wechsler scales within the framework of the Cattell-Horn-Carroll theory (McGrew, 1997, 2005) that combines the work of Carroll (1993, 1997b) and Cattell and Horn (Cattell & Horn, 1966; Horn, 1988). Specifically, "Wechsler's tests have gone through progressive revisions to update the theoretical foundations based on cumulative research findings in neuropsychology, executive functions and working memory" (Weiss et al., 2013a, p. 95). However, it is important to remember that intelligence is not the same thing as IQ tests, a critical distinction that gets muddled at times in these articles. We believe that the construction of cognitive tests should be based on a sound, established theoretical foundation whereas methodology used by these

researchers almost reads as if it were taken from Kline's (2010) book, "Chapter 13: How to Fool Yourself With SEM."

The basic foundation of both of these articles is a comparison of four- and five-factor Wechsler models—neither of which are truly theoretically grounded. After completing the analyses, the authors respecified their models in an attempt to improve the fit of the model to the data (and not the theory). However, as MacCallum and colleagues (1992) warned, "When an initial model fits well, it is probably unwise to modify it to achieve even better fit because modifications may simply be fitting small idiosyncratic characteristics of the sample" (p. 501).

As CFA is a confirmatory technique, any modifications of the fit indices should be theoretically driven, otherwise the modifications become a fishing expedition, which increases the probability of a Type 1 error. In this regard, as Schreiber and colleagues (2006) pointed out, researchers all too often become fascinated with escalating fit indexes that ultimately leads to an abuse of these modifications. As Ullman (2001) cautioned, adding post hoc paths is like eating salted peanuts: "One is never enough" (p. 750). And adding paths (cross-loadings) almost always results in improvement in model fit.

Problem #4: Failure to Consider and Examine Rival Direct Hierarchical/Bifactor/Nested Factor Models

Determining the best fitting model among several *rival* theoretically based models should assist in better understanding the structural validity as well as determining the most appropriate way to interpret a test (Kline, 2005). For example, one rival model for the WAIS-IV and WISC-IV is the *direct* hierarchical (Gignac, 2005, 2006, 2008; Watkins, 2010) model (see Figure 1), also termed the bifactor model (Holzinger & Swineford, 1937) or the nested factors model (Gustafsson & Balke, 1993; Keith, 2005). The models proposed by Weiss et al. (2013a) and Weiss et al. (2013b) are *indirect* hierarchical models (see Figure 2) with a general intelligence (*g*) factor at the structure's apex and first-order factors below. Gignac (2005, 2006, 2008) criticized the indirect hierarchical model because he argued it is unrealistic to assume that *g*'s influence on subtest performance is fully mediated by the first-order factors and having no direct influence. The direct hierarchical model makes no such assumptions, allowing the *g* factor and the first-order factors to be orthogonal and to simultaneously directly influence subtest performance (Gignac, 2005). The direct hierarchical model has provided better fit than did the indirect hierarchical model in factor analytic studies of Wechsler scales (Gignac, 2005, 2006; Golay et al., 2012; Watkins, 2010).

The direct hierarchical model has been described as a more defensible model than the indirect hierarchical model because the full mediation of the narrow group factors of the indirect model "should probably be considered unreasonable in the area of psychology, where test developers are not likely adept enough to develop subtests in such a way that they will share variance with the general factor to the extent that they will load onto a group level factor" (Gignac, 2006, p. 85). In addition, general intelligence is modeled as a *breadth* factor in a direct hierarchical model rather than as a *superordinate* factor in an indirect hierarchical model. Gignac (2008) argued that breadth rather than superordination has historically been a more fundamental aspect of general intelligence from a theoretical perspective. This was also the conclusion of Golay et al. (2012). It is clear that the direct hierarchical model should have been examined in both studies as rival structures in both the W-HM and CHC-HM comparisons.

Problem #5: Failure to Present Decomposed Orthogonal Subtest Variance Estimates to g and First-Order Factors

The present WAIS-IV (Weiss et al., 2013a) and WISC-IV (Weiss et al., 2013b) studies, like an earlier similar study of the WISC-IV (Keith, Fine, Taub, Reynolds, & Kranzler, 2006); *did not*

present decomposed subtest variance estimates to both the higher-order *and* first-order latent factors. Table 3 in the present WAIS-IV and WISC-IV studies present subtest *g* loadings (higher-order factor associations) that were mostly fair to good (only WAIS-IV Cancellation and WISC-IV Coding and Cancellation were poor) using Kaufman's (1994) criteria. As Keith (2005) noted, such obtained *g* loadings are similar to those produced by a nested factor (bifactor or direct hierarchical) CFA model. However, failure to report the portions of subtest variance *remaining* in the first-order factors does not allow for direct comparison of relative importance of subtest measurement of the higher-order *versus* first-order factors. Later, both the WAIS-IV and WISC-IV studies included tables of "primary and secondary abilities" of the subtests but these relate *only* to primary *first-order* and secondary *first-order* abilities measured by the WAIS-IV subtests (see Table 6 in Weiss et al., 2013a) and the WISC-IV subtests (see Table 6 in Weiss et al., 2013b). Without presentation of decomposed subtest variance estimates to the higher-order *g* factor in comparison to variance apportioned to the first-order factors (as insisted by Carroll), one does not "see" that the primary ability measured by most subtests is higher-order *g*.

Previous studies presenting decomposed apportions of subtest variance have been published (Bodin et al., 2009; Canivez & Watkins, 2010a, 2010b; Watkins, 2006; Watkins et al., 2006) and have consistently demonstrated that for most subtests the overwhelming majority of Wechsler subtest variance is apportioned to higher-order *g* and substantially less subtest variance is associated with the first-order factors. This is what led Watkins (2006), Watkins et al. (2006), Canivez and Watkins (2010a, 2010b), and Golay et al. (2012) to recommend *primary* interpretation of WAIS-IV and WISC-IV at the FSIQ (*g*) level *rather* than the factor index score level. While it is indeed true that the present WAIS-IV and WISC-IV studies found the correct assignment of subtests to their proposed latent first-order factors and supported the factor structures, the WAIS-IV and WISC-IV *primarily* measured the higher-order *g* dimension.

The Chen et al. (2009) study described earlier also presented decomposed orthogonal higher-order CFA variance estimates, and demonstrated that G_f accounted for only 0.2%, 1%, and 2% of Similarities, Picture Concepts, and Matrix Reasoning variance, respectively; while higher-order *g* accounted for 49%, 54%, and 71% variance, respectively. Why are these variance proportions for the WAIS-IV and WISC-IV absent in the present studies? If similarly poor portions of G_f factor variance are apportioned to these subtests, how can this be a viable dimension apart from the higher-order *g* factor and how is CHC theory based G_f supported in the Wechsler scales?

Practical Limitations

Problem #6: Ascribing Clinical Importance to Results Based Solely on Internal Structure

Factor-based scores can be considered at either of two levels: latent constructs or observed scores. Theory-development is supported by latent constructs, the underlying foundation of a theory, comprised of unobservable, theoretical relationships. Clinicians, on the other hand, are typically more concerned with observed scores as they are directly observed and offer more applied utility. The four, or five, factors identified by the authors represent latent constructs that should not be confused with observed factor scores.

Weiss et al. (2013a) and Weiss et al. (2013b) concluded that results provide for clinical interpretation, however the only clinical "interpretation" that the present results could offer is explanatory and without presentation of decomposed subtest variance, results would likely lead to misguided inferences. Further, the present studies simply examine the internal structure of selected models and there are many plausible models *not tested*. Examinations of the internal structure of tests using EFA *or* CFA are insufficient as they cannot fully answer questions of validity or utility (Carroll, 1997a; Kline, 1994; Lubinski & Dawis, 1992). The internal structure

of any measure must be further examined in relation to external criteria to see how well they perform. The present studies provide no such examination of model predictive validity, incremental validity, diagnostic efficiency, or differential treatment recommendations and effects. Without these sorts of studies clinicians cannot know anything more than the explanatory hypotheses from the various proposed models.

Weiss et al. (2013a) noted, "For adults with consistent subtest scores within each of the four WAIS-IV composites, the current four WAIS-IV Index scores constitute an appropriate level of interpretation. For adults with discrepant subtest scores within some of the four composites the five-factor model suggests a likely interpretive reorganization" (p. 108). Decide which model to interpret based on subtest scatter?! Where are the data in these articles, or any other for that matter, to support this recommendation? The practice of interpreting subtest scatter has an abundant literature (see Canivez, 2013; Watkins, 2003) illustrating poor reliability, validity, and diagnostic utility that would argue against this.

Glutting, Watkins, Konold, and McDermott (2006) demonstrated that the WISC-IV FSIQ predicted substantial portions of variance in reading and mathematics scores on the WIAT-II although the four factor index scores did not contribute additional meaningful prediction beyond the FSIQ. Structural equation modeling (SEM) analyses by Glutting et al. (2006) found only the higher-order *g* and the latent VC construct offered significant *explanations* of WIAT-II reading and mathematics constructs (PR, WM, and PS constructs provided no increases in explanation). Weiss et al. (2013a) and Weiss et al. (2013b) also suggested that different interpretations might be made when there was significant scatter or deviation among WISC-IV or WAIS-IV scores but did not examine this, thereby recommending practices without evidence. With respect to external validity, however, even in situations where significant factor score variability exists, the FSIQ is still a valid and superior predictor of academic achievement (Daniel, 2007; Freberg, Vandiver, Watkins, & Canivez, 2008; Watkins, Glutting, & Lei, 2007) and numerous studies have illustrated the psychometric inferiority of scatter and related deviation scores (Canivez, 2013).

Oh, Glutting, Watkins, Youngstrom, and McDermott (2004) cogently pointed out latent constructs are not directly observable, have different distributions, *and* latent construct scores are difficult to calculate and not readily available. Also, while decomposed subtest variance can be done for group data, it cannot be done for individuals. Thus, there are no direct *practical* clinical applications from the present CFAs.

Problem #7: Arithmetic! What to Do About Arithmetic?

Problems with the Arithmetic subtest and its inclusion in Wechsler scales have been noted in numerous studies and while it is a supplemental subtest for the WISC-IV as an indicator of Working Memory it is a core Working Memory subtest for the WAIS-IV. Arithmetic is the 3-card Monte subtest of the Wechsler scales. In the transition from the 2-factor WISC to the WISC-R, factor analyses indicated that Arithmetic, Digit Span, and Coding splintered into a non–theoretically based third factor, then called Freedom from Distractibility, which was questioned for its relationship to attention (Kaufman, 1994; Keith, 2005; Keith & Witta, 1997) as well as working memory (Kranzler, 1997). During the revision of the WISC-R, it was thought that the creation of a new subtest, Symbol Search, would strengthen the Freedom from Distractibility factor but it did not. The inclusion of Symbol Search actually pulled the Coding subtest away from the Freedom from Distractibility factor and a fourth (again, non-theoretically derived) factor was created and labeled Processing Speed (a name questioned by Keith [1997] and Kranzler [1997]). This left Arithmetic and Digit Span to represent the third factor, which was then relabeled Working Memory for the WAIS-III (Wechsler, 1997), WISC-IV (Wechsler, 2003b), and WAIS-IV (Wechsler, 2008b). Certainly Working Memory is involved in mental arithmetic tasks (and all cognitive activities) but it is also, and perhaps more appropriately considered a task of

quantitative skills and a fair to good measure of *g*. A factor analytic problem is that there are no similar quantitative tasks in the WISC-IV or WAIS-IV to better measure a latent Quantitative Reasoning dimension that is part of CHC conceptualization. As evidenced by the numerous cross loadings reported in the two present studies, it may be better to remove Arithmetic altogether, or create additional quantitative reasoning tasks in subsequent versions of the Wechsler scales. Golay et al. (2012) found Arithmetic not to be a measure of G_f in the French WISC-IV. Ironically, Weiss et al. (2013a, p. 109) noted that "Keith reported that 'The WAIS–IV four factor structure fits better than the CHC model when Arithmetic is excluded' (Lichtenberger & Kaufman, 2009, p. 32)" which begs the question, why not test a model that did not include Arithmetic?

Conclusion

Weiss et al. (2013a) and Weiss et al. (2013b) present interesting analyses and results for the WAIS-IV and WISC-IV internal structures and indicated structural invariance between the clinical and normative groups from the respective standardizations. These results, however, must be challenged due to a number of theoretical, methodological, and practical limitations. We strongly believe that the substantial theoretical, methodological, and practical limitations greatly limit any interpretations of the results, particularly those suggesting *utility* of the findings for practitioners. It is hoped that the authors will provide the necessary analyses to address the highlighted concerns to better inform practitioners about just what the WAIS-IV and WISC-IV actually measure, and measure well enough for clinical practice.

Declaration of Conflicting Interests

Funding

References

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*, 121-130.

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology*, *15*, 417-424.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, V. L. Schwean, & C. R. Reynolds (Eds.), *The Oxford Handbook of Child Psychological Assessments* (pp. 84–112). New York: Oxford University Press.

Canivez, G. L., & Watkins, M. W. (2010a). Investigation of the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment*, *22*, 827-836.

Canivez, G. L., & Watkins, M. W. (2010b). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly*, *25*, 223-235.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504.

Chen, H. -Y., Keith, T. Z., Chen, Y. -H., & Chang, B.–S. (2009). What does the WISC-IV measure? Validation of the scoring and CHC-based interpretive approaches. *Journal of Research in Education Sciences, 54*, 85-108.

Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010, November). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, *49*, 677-682. doi:10.1016/j.paid.2010.06.004

Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences*, *52*, 161-166.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

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

Carroll, J. B. (1993). *Human cognitive abilities abilities*. Cambridge, UK: Cambridge University Press.

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

Carroll, J. B. (1997a). Theoretical and technical issues in identifying a factor of general intelligence. In B. Devlin, S. E. Fienberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes, and success: Scientists respond to the bell curve* (pp. 125-156). New York, NY: Springer-Verlag.

Carroll, J. B. (1997b). 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. 183-208). New York, NY: Guilford.

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

Daniel, M. H. (2007). "Scatter" and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology*, *14*, 291-295.

Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC–III Full Scale IQ in predicting later academic achievement. *Applied Neuropsychology*, *15*, 131-139.

Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment*, *12*, 320-329.

Gignac, G. E. (2006). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences*, *27*, 73-86.

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science Quarterly*, *50*, 21-43.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education*, *40*, 103-114.

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2012, November 12). Further insights on the French WISC–IV factor structure through Bayesian structural equation modeling (BSEM). *Psychological Assessment*. Advance online publication. doi:10.1037/a0030676

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade, & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 231-258). New York, NY: Plenum.

Gustafsson, J. -E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 107-135). Westport, CT: Greenwood Press.

Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407-434.

Holzinger, K. J., & Swineford, F. (1937). The bifactor method. *Psychometrika*, *46*(1), 41-45.

Horn, J. L. (1988). Thinking about human abilities. In R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York: Plenum Press.

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.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York, NY: Wiley.

Keith, T. Z. (1997). What does the WISC-III measure? A reply to Carroll and Kranzler. *School Psychology Quarterly*, *12*, 117-118.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581-614). New York, NY: Guilford.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher-order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children–Fourth Edition: What does it measure. *School Psychology Review*, *35*(1), 108-127.

Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly*, *12*, 89-107.

Kline, P. (1994). *An easy guide to factor analysis*. London, UK: Routledge.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

Kranzler, J. H. (1997). What does the WISC-III measure? Comments on the relationship between intelligence, working memory capacity, and information processing speed and efficiency. *School Psychology Quarterly*, *12*, 110-116.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment*. New York, NY: Wiley.

Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills, and proficiencies. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 1-59). Palo Alto, CA: Consulting Psychology Press.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.

Marsh, H. W., Hau, K. -T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341.

McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: Aids to interpretation. *Advances in Social Science Methodology*, *4*, 229-240.

McGrew, K. S. (1997). Analysis of the major intelligence 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, NY: Guilford.

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

Oh, H. J., Glutting, J. J., Watkins, M. W., Youngstrom, E. A., & McDermott, P. A. (2004). Correct interpretation of latent versus observed abilities: Implications from structural equation modeling applied to the WISC-III and WIAT linking sample. *Journal of Special Education*, *38*, 159-173.

Ree, M. J., Carretta, T. R., & Green, M. T. (2003). The ubiquitous role of *g* in training. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 262-274). New York, NY: Pergamon Press.

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

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*, 323-337.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick, & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Ward, L. C., Bergman, M. A., & Hebert, K. R. (2012). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment*, *24*, 328-340.

Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion? *The Scientific Review of Mental Health Practice*, *2*, 118-141.

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children–Fourth Edition. *Psychological Assessment*, *18*, 123-125.

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children–Fourth Edition among a national sample of referred students. *Psychological Assessment*, *22*, 782-787.

Watkins, M. W., Glutting, J. J., & Lei, P. -W. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*, 13-20.

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition among referred students. *Educational and Psychological Measurement*, *66*, 975-983.

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

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–Third Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children–Fourth Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children–Fourth Edition Technical and Interpretive Manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2005). Wechsler Intelligence Scale for Children (WISC IV)–translated. *ECPA, Paris*.

Wechsler, D. (2008a). *Wechsler Adult Intelligence Scale–Fourth Edition*. San Antonio, TX: Pearson.

Wechsler, D. (2008b). *Wechsler Adult Intelligence Scale–Fourth Edition Technical and Interpretive Manual*. San Antonio, TX: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psycheducational Assessment*, *31*, 94-113.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013b). WISC-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psycheducational Assessment*, *31*, 114-131.