This is a pre-print version of following chapter: Miller, D. C., & McGill, R. J. (2016). Review of the WISC-V. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC-V* (pp. 645-662). Hoboken, NJ: Wiley, which has been published in final form at http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118589238.html. This chapter may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Chapter 17

Review of the WISC-V

Daniel C. Miller and Ryan J. McGill

Texas Woman's University

Denton, Texas

One of the major goals of the *Wechsler Intelligence Scale for Children - Fifth Edition* (Wechsler, 2014) was to incorporate contemporary intellectual assessment research into the revision. Advances on intellectual theory along with advances in theories of cognitive development, neurodevelopment, and cognitive neuroscience, all influence this current version of the Wechsler Scales. The purpose of this chapter is to provide an objective review of the strengths and weaknesses of the WISC-V. Table 1 provides an overview of these identified strengths and weaknesses of the test and the subsequent sections of this chapter will expound more of the details.

Table 1

Strengths and Weakness of the WISC-V

Strengths	Weaknesses
Theoretical Foundation	
• Integration of additional neuropsychological constructs (e.g., enhanced working memory, associative learning and recall, rapid automatized naming, etc.) is a welcome addition.	• Lack of a unified theory of intellectual ability for the entire test.
Family of Related Products	
• The WISC-V fits in the middle of a full range of cognitive assessment products designed for all ages including the WPPSI-IV and the WAIS-V). • The WIAT-III is a measure of academic achievement often used in conjunction with the WISC-V. • Digital version (Q-Interactive) of the full menu of the WISC-V subtests.	• Data are lacking on the relationship between the WISC-V and a comprehensive test of learning and memory. • Data are lacking on the relationship between the WISC-V and a comprehensive test of neuropsychological functioning (e.g., NEPSY-II).
Psychometric Properties	
• A representative standardization sample. • In general, relevant psychometrics for the instrument is strong. • The manual contains a wealth of information related to the development of the measure. • Adequate representation of relevant subpopulations (e.g., special education) within the normative sample.	• Confidence intervals based on true scores which may not be ecologically valid. • Lack of exploratory factor analysis (EFA) results. • The Arithmetic subtest still remains cognitively complex, which is hard to classify using factor analysis. • Use of coefficient alpha to estimate the

• Strong internal consistency reliability estimates. • Good convergent and divergent validity. • Improved floors and ceilings for individual tests. • Item biases based on race or ethnicity do not appear to be present.	reliability of multidimensional measures. • Further research needs to be conducted on the validity of using the WISC-V for determining cognitive strengths and weaknesses for diagnosing specific learning disabilities. • Decomposition procedures were not reported so that users can appropriately apportion higher-order and lower-order variances in the WISC-V subtests. • Failure to specify complementary measures in the structural model.
Quality of Testing Materials and Administration Issues	
• Significant number of test items were replaced or revised from the prior version for test security reasons. • Subtest arranged in stimulus booklets in a logical order. • Testing time was minimized by reducing the number of test items and modifying discontinuation rules. • Eight new subtests were added to the test. • Simplification of instructions for better ease in understanding. • Increase in practice items. • Succinct instructions. • Reduce number of items with time bonuses.	• Limitation of the plastic coil bindings. • The WISC-V no longer uses substitutes for invalid or contaminated subtests.
Interpretative Options	
• Multiple psychometric comparisons between indices provided. • Expanded significance level options for critical values. • Inclusion of base rates for several qualitative behaviors. • Attempting to adhere to a more "CHC" based structure. It is not perfect but it will help users with interpretation (e.g., splitting the PRI). • *Gf* Composite significantly improved with the inclusion of the Figure Weights test.	• Little information on supplementary measures and process scores. How they aid in diagnostic decision-making? • Little information on interpreting profiles of neurocognitive strengths and weaknesses.

Organization of the WISC-V

The organizational structure of the WISC-V is a significant departure from the previous version and now includes additional scales, batteries, and reference terminology; although many of these changes are consistent with those that have been made in recent revisions of instruments within the Wechsler family of products (e.g., WAIS-IV, WPPSI-IV). An outline of the subtest, scale, and composites scores contained within the WISC-V is provided in Tables 2 and 3. The WISC-V provides users with a multitude of scores including: subtest scores, index scores, composite scores, process scores, contrast scores, and base rate scores. In this chapter we focus primarily on the allocation and integrity of the traditional WISC-V standard scores (subtest, index, and composites), although some discussion regarding the process and base rate measures is provided.

Table 2

WISC-V Subtests and Subtest Categories

		Subtest Categories		
Subtest	Primary FSIQ	Primary	Secondary	Complimentary
Similarities	✓	✓		
Vocabulary	✓	✓		
Information			✓	
Comprehension			✓	
Block Design	✓	✓		
Visual Puzzles		✓		
Matrix Reasoning	✓	✓		
Figure Weights	✓	✓		
Picture Concepts			✓	
Arithmetic			✓	
Digit Span	✓	✓		
Picture Span		✓		
Letter-Number Sequencing			✓	
Coding	✓	✓		
Symbol Search		✓		
Cancellation			✓	
Naming Speed Literacy				✓
Naming Speed Quantity				✓
Immediate Symbol Translation				✓
Delayed Symbol Translation				✓

Recognition Symbol Translations				✓
Total	**7**	**10**	**6**	**5**

The WISC-IV is composed of a total of 21 subtest measures (*M* = 10, *SD* = 3, range = 1 to 19). Each subtest is grouped into three separate categories: primary, secondary, or complimentary. The primary subtests (*n* = 10) combine to form the Full Scale IQ Composite (FSIQ; *M* = 100, *SD* = 15) and the primary indexes. It should be noted that FSIQ is linearly derived from a combination of seven of the primary subtests, the remaining primary measures combining to form the primary index level scores. Users have the option of limiting administering to the seven primary FSIQ subtests if their only concern is obtaining an overall estimate of an examinee's general cognitive ability however, the WISC-V Technical and Interpretive Manual (Wechsler, Raiford, & Holdnack, 2014) encourages users to administer all 10 of the primary subtests to provide a broader sampling of cognitive functioning. Although users may substitute one secondary subtest to calculate the FSIQ, no substitutions are permitted at the index level. The five primary index scales include: Verbal Comprehension (VCI); Visual Spatial (VSI); Fluid Reasoning (FRI); Working Memory (WMI); and Processing Speed (PSI).

Table 3

Organizational Framework for the WISC-V

	Full Scale Level	Primary Index Level					Ancillary Index Level				
Subtest	FSIQ	VCI	VSI	FRI	WMI	PSI	QRI	AWMI	NVI	GAI	CPI
Similarities	✓	✓								✓	
Vocabulary	✓	✓								✓	
Information	*										
Comprehension	*										
Block Design	✓		✓						✓	✓	
Visual Puzzles	*		✓						✓		
Matrix Reasoning	✓			✓					✓	✓	
Figure Weights	✓			✓			✓		✓	✓	
Picture Concepts	*										
Arithmetic	*						✓				
Digit Span	✓				✓			✓			✓
Picture Span	*				✓				✓		✓
Letter-Number Sequencing	*							✓			
Coding	✓					✓			✓		✓
Symbol Search	*					✓					✓
Cancellation	*										

	Complementary Index Scale Level		
Complimentary Subtests	NSI	STI	SRI[1]
Naming Speed Literacy	✓		✓
Naming Speed Quantity	✓		✓
Immediate Symbol Translation		✓	✓
Delayed Symbol Translation		✓	✓
Recognition Symbol Translation		✓	✓

*Denotes allowable FSIQ subtest substitution.

[1]SRI is combination of NSI and STI standard scores and thus is a linear combination of the constituent subtest measures within these indexes.

Ancillary Index Scales are composed of various combinations of the primary and secondary subtests (*n* = 6). Ancillary Index Scores include: Quantitative Reasoning (QRI); Auditory Working Memory (AWMI); Nonverbal (NVI); General Ability (GAI); and Cognitive Proficiency (CPI). The remaining five complimentary subtests combine to form additional Complimentary Index Scales. These scales include: Naming Speed (NSI); Symbol Translation (STI); and Storage and Retrieval (SRI). All WISC-V index scores contain two or more subtest measures with the exception of the SRI, which is a combination of the NSI and STI standard scores. Taken as a whole, we believe the structural and design features of the WISC-V result in a more clinically useful instrument with broad applications for assessment psychologists as compared to its predecessor. Nevertheless, we will now proceed with a more in-depth discussion as it relates to the conceptual and technical properties of the measurement instrument.

Theoretical Foundation of the Test

Incorporating contemporary intellectual assessment research into the WISC-V was one of the goals of the most recent revision to the test. This goal was partially met by significantly enhancing the assessment of the following neuropsychological constructs: fluid reasoning, visual-spatial processing, working memory, naming fluency, and verbal-visual associative learning and recall. However, there are two dominate contemporary intellectual theories, one based on the work of Carroll-Horn-Cattell (CHC) (Schneider & McGrew, 2012) and the other based on Lurian theory (Luria, 1966, 1973, 1980); yet the WISC-V did not adopt either one of those theoretical approaches. Rather the WISC-V is simply a collection of tests; all designed to measure difference aspects of intellectual functioning. The test authors acknowledge that some have asserted that the Wechsler intelligence tests lack a unified theoretical foundation (Coalson, Raiford, Saklofske, & Weiss, 2010; Kaufman, 2010; Raiford & Colason, 2014). The authors contend that the WISC-V is consistent with Wechsler's view of intelligence, which is thought to encompass a variety of qualitatively different abilities (Wechsler, Raiford, & Holdnack, 2014).

It is important to recognize that even though the WISC-V may not be guided by an overall theory, the FSIQ does highly correlate with other the full scale intelligence test scores such as the *Kaufman Assessment Battery for Children – Second Edition* (KABC-2: Kaufman & Kaufman, 2004) or the *Woodcock-Johnson Tests of Cognitive Abilities – Fourth Edition* (WJ IV COG: Schrank, Mather, and McGrew, 2014). The WISC-V tests can easily be interpreted within a cross-battery assessment perspective (Flanagan, Ortiz, & Alfonso, 2013) or Miller's (2013) Integrated CHC / School Neuropsychology (SNP) Model.

Family of Related Products

One of the major advantages of the WISC-V is the integration of this particular test into an entire family of intellectual functioning measures that span early childhood through older adult age ranges. The WISC-V is designed to assess intellectual functioning in school-aged children, ages 6:0 through 16:11 years. The *Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition* (WPPSI-IV: Wechsler, 2012) is designed to measure intellectual functioning in young children ages 2:6 to 7:7 years, and the *Wechsler Adult Intelligence Scale – Fourth Edition* (WAIS-IV: Wechsler, 2008) is designed to measure intellectual functioning in individual ages 16:0 to 90:11 years. In the recent revisions of these three Wechsler products, the test developers have strived to measure comparable cognitive constructs across the developmental spectrum, and have been largely successful in doing so.

The WISC-V is a comprehensive intelligence test, but no one battery of tests is designed to measure all aspects of a person's cognitive, academic, and social emotional capabilities. The WISC-V will often be used in combination with a comprehensive test of achievement such as the

Wechsler Individual Achievement Test – Third Edition (WIAT-III: Wechsler, 2009) and a behavioral rating scale such as the *Behavior Assessment Scale for Children - Second Edition* (BASC: Reynolds & Kamphaus, 2009). The *WISC-V Technical and Interpretative Manual* (Wechsler, Raifird, & Holdnack, 2014) provides psychometric concurrent validity data for the WISC-V, WIAT-III, and the BASC-2 Parent Rating Scale comparisons.

In neuropsychological assessments, the WISC-V is often used in conjunction with other instruments such as the *NEPSY-II: A developmental neuropsychological assessment* (Korkman, Kirk, & Kemp, 2007) or a comprehensive test of learning and memory such as the *Children's Memory Scale* (CMS: Cohen, 1997). It is recognized that publishers cannot provide all possible test comparisons with the WISC-V as part of the initial validation, but with the inclusion of several new neuropsychologically-based tests on the WISC-V, the comparison of these tests to similar tests on the NEPSY-II would have been helpful. When the CMS is revised, it is hoped that a WISC-V concurrent validity study will be provided. Finally, the addition of the WISC-V *Integrated* test (Wechsler, in press) will strengthen the clinical utility of the WISC-V from a neuropsychological perspective.

One of the most innovative features of the WISC-V is the inclusion of the full battery of tests in Pearson's digital the Q-Interactive platform. The Q-Interactive software required the clinician to have two Apple iPads, one for the examiner and one of the examinee, linked electronically. The Q-Interactive allows the clinician to choose custom tests from a full array of Pearson assessment products, administer digital versions of the tests on the iPad, score the results electronically, and manage individual client records. In this day and age of tablets and smart phones and other advances in technology, digit versions of tests like the WISC-V are welcome additions to the profession. The Q-Interactive platform is relatively new to the field so practitioners and researchers are just starting to evaluate the digital versions of the products, compared to the paper-and-pencil versions (Dumont, Viezel, Kohlhagen, & Tabib, 2014).

Quality of Testing Materials

The overall production quality of the materials is very good. The WISC-V test kit includes: Administration and Scoring Manual, Administration and Scoring Manual Supplement, Technical and Interpretative Manual, 3 stimulus booklets, Symbol Search scoring key, Cancellation scoring template, Coding A scoring template, set of 9 red and white blocks, a red and yellow pencil, a set of record forms, a set of Response Booklet 1 record forms, and a set of Response Booklet 2 record forms. The only picky criticism of the production quality of the test is the use of plastic coils to bind the stimulus booklets and manuals. The publisher does acknowledge that after repeated uses of the bound booklets, the plastic coils will twist off and require the user to adjust them accordingly. This is a minor annoyance but one that could be fixed through better engineering of the bindings. Of course this would be a moot point if the digital version of the tests were administered.

The subtests are arranged in the stimulus booklets in a logical order to make administration easier. The test authors did a good job in reducing the total test time required by reducing the number of test items and modifying discontinuation rules. These changes were made in recognition of the increased time constraints on practitioners and minimizing the sustained attention requirements for children who are being assessed.

Due to copyright laws, and prior test items becoming more widely known to the public, many of the test items on the WISC-V are new or were revised in some fashion. These changes were made to increase the security of the test. Another major goal of the test revision was to increase the developmental appropriateness of the instrument. The test developers seem to have accomplished this by simplifying the test instructions for easier understanding and making the instructions more succinct. To ensure that children understand the task requirements more practice items were added to the tests, as appropriate. Finally, the idea that quick task completion is always essential was de-emphasized somewhat in the WISC-V by reducing the number of tests with time bonus points.

New WISC-V Tests. The WISC-V includes eight new tests: Figure Weights, Visual Puzzles, Picture Span, Naming Speed Literacy, Naming Speed Quantity, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation. Figure Weights was originally introduced on the WAIS-IV (Wechsler, 2008) and is designed to measure aspects of fluid and quantitative reasoning. Figure Weights and the Matrix Reasoning tests now form the FRI, which significantly improves the quality of that index.

Visual Puzzles is another test adapted from the WAIS-IV version (Wechsler, 2008). The test is designed to measure visual-spatial reasoning during a non-motor construction task. The test also requires some mental rotations, visual working memory, understanding of part-to-whole relationships, and visual analysis and synthesis. Visual Puzzles and Block design now form the VSI. Splitting the WISC-IV PRI into the Visual-Spatial and Fluid Reasoning Indices strengthens the WISC-V considerably. In an effort to improve the quality of the PSI, the Picture Span test was added. Picture Span is designed to measure visual working memory and visual working memory capacity.

The Naming Speed Literacy, Naming Speed Quantity, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation tests are referred to by the test authors as complementary tests. These tests were specifically included in the WISC-V for use with special clinical populations such as the assessment of specific learning disabilities.

Speed naming tasks require a child to name colors, words, or letters as quickly as possible. These tasks are often referred to in the neuropsychology literature as rapid automatized naming (Miller, 2013). These types of speeded naming tasks have been shown to predict, or be associated with disorders of reading and spelling (Crews & D'Amato, 2009) and to disorders of mathematics (McGrew & Wendling, 2010). The Naming Speed Literacy and the Naming Speed Quantity are not intended to be measures of intelligence, and as a result as not included in any of the indices; however, they should prove to be useful additions to the test for assessing children with suspended processing disorders.

The Immediate Symbol Translation, Delayed Symbol Translation, and the Recognition Symbol Translation tests measure different aspects of visual-visual associative learning and recall. These tests are also not intended to be measures of intelligence, but rather used as supplemental measures for evaluating potential learning disorders in children. These types of tasks often predict performance on reading decoding, reading accuracy, reading fluency, and reading comprehension tests (Litt, de Jong, van Bergen, & Nation, 2013).

Subtest Modifications. Word Reasoning and Picture Completion from the WISC-IV were dropped in this revision. The following tests had modifications made to their recording and scoring of items: Similarities, Vocabulary, Information, Comprehension, Block Design, Digit Span, Letter-Number Sequencing, Coding, and Symbol Search. In another revision, test items were added to Similarities, Vocabulary, Information, Comprehension, Block Design, Matrix Reasoning, Picture Concepts, Arithmetic, Digit Span, Letter-Number Sequencing, Coding, Symbol Search, and Cancellation. In total, these subtest modifications in combination with the addition of the new tests, reflect a major revision to the test.

Interpretative Options

The Technical and Interpretive Manual encourages examiner's to interpret the WISC-V in a top down fashion, beginning with the FSIQ, using a series of iterative steps designed to provide users with multiple levels of information about an individual's performance. The FSIQ is the most reliable score on the WISC-V and is considered to be score that is most representative of *g*. The FSIQ is best interpreted after considering the degree of variability in the profile of primary index scores. Comparisons can be made between the FSIQ and each primary index score using *a priori* critical values to determine if the observed differences are statistically significant. WISC-V critical value options have been expanded relative to the WISC-IV with the number of options increased from two to four (now includes .01, .05, .10, and .15). Additionally, examiners can then determine the relative clinical significance of the difference value using base rates provided in the Administration and Scoring Manual.

The Technical and Interpretive Manual suggests that primary interpretation of the WISC-V should focus on the profile of obtained primary index scores in order to determine the presence of individual cognitive strengths and weaknesses. Profile variability can be examined both within index (e.g., subtest differences) and across indexes using similar procedures as previously described with the FSIQ. It is suggested that examiner's begin by describing the overall index score profile and then proceed to evaluating level of performance and degree of variability for each measure individually. Although the implication is that profile variability and scatter is potentially clinically relevant, limited evidence is provided within the Technical and Interpretive Manual to support these claims.

Similar evaluation procedures can also be used to examine cognitive strengths and weaknesses at the subtest level. However, due to the fact that subtest variability is common within the population (see Watkins, Glutting, & Youngstrom, 2005 for a review), inferences at this level of interpretation should be made cautiously. Accordingly, the Technical and Interpretive Manual warns that subtest level profile analysis should only be conducted when the examiner has a clear rationale for doing so.

Although administration of the primary battery yields a comprehensive evaluation of intellectual ability, supplementing the 10 primary subtests with the five complimentary subtests may be warranted depending on the clinical needs of the client. The Technical and Interpretive Manual denotes that profile analysis with the ancillary and complementary scales is optional. That is, examiners should administer these measures only when there is a specific clinical purpose to do so (e.g., suspected memory or other related neurocognitive impairment). If these measures are administered, examiners may employ the procedures described above for examining individual cognitive strengths and weaknesses. As would be expected, the empirical literature regarding the technical properties and potential clinical applications is in its infancy. We encourage users of the WISC-V to keep abreast of subsequent developments in that regard and to modify or supplement their interpretations of the measurement instrument accordingly.

Psychometric Adequacy of the WISC-V

Standardization Sample. The Technical and Interpretive Manual presents extensive and detailed information on the standardization procedures for the instrument and the development of the normative sample. The normative sample included 2,200 children and adolescents divided into 11 age groups. The standardization sample was obtained through proportional sampling and stratified across key demographic variables such as age, sex, ethnicity, geographic region, and parent educational level.

Inspection of the normative tables provided in the Technical and Interpretive Manual revealed a close match between obtained proportions and parameter estimates from the 2012

U.S. Census. Additionally, an effort was made to include participants with relevant special education classifications in the normative sample. As a result, the normative sample closely matches U.S. population estimates for several relevant special education classifications (e.g., specific learning disability, intellectual disability, and attention deficit/hyperactivity disorder). A list of exclusionary criteria is also provided. Some of the factors that were exclusionary include, language and primary method of communication limitations, disruptive behavior or inability to test, motor difficulties that would impact test performance, taking medications that would impact cognitive performance, and diagnoses of a neurological or psychological condition that would impact test performance (e.g., epilepsy, mood disorder).

Subtest scaled scores were developed using the inferential norming method (Zhu & Chen, 2011). This procedure examines obtained means, standard deviations, and skewness estimates using linear to 4^{th} degree polynomials to determine the best fitting curve for each age group based upon theoretical conjecture and the pattern of growth curves observed in the WISC-V. The selected curves were then used to estimate population parameters and generate theoretical distributions for each age group. The percentages for each raw score were then converted to scaled or standard scores using the mid-interval percentile method.

Composite scores (e.g., FSIQ) are based on the respective sums of age-based scaled or standard scores. As previously mentioned, the lone exception is the SRI, which is derived from summing the NSI plus the STI. Tables provided within the Technical and Interpretative Manual indicated the means, standard deviations, and sum of scaled scores for each composite were relatively consistent across age groups. More importantly, evidence was provided that suggests that the distributions of the scaled score sums approximate the normal distribution. For each scale, the distribution of scaled scores was used to convert obtained percentiles to standard scores. The Technical and Interpretative Manual indicated that standard score distributions were smoothed visually to ensure consistency with the normal distribution. As a result of obtaining non-normal distributions for several scores on the WISC-V (e.g., span and sequence, error, and process scores), standard scores could not be developed and these measures are reported as base rates or cumulative percentages. The cumulative percentages reflect the base rate of an occurrence of a behavior that was observed in the normative sample.

Item Gradients, Floors, and Ceilings. All WISC-V index and composite score ranges are adequate, generally reflecting a range of values that is sufficient for estimating the broad spectrum of cognitive performance. Index level scores (e.g., VCI, VSI, FRI, WMI, PSI) ranged from 45-155 whereas composite level scores (i.e., FSIQ) ranged from 40-160. Additional items were added to several subtests (e.g., Digit Span, Vocabulary, Information) to expand the range of ability sampled by these measures. Inspection of the conversion tables for subtests, index, and composite scores provided in the Administration and Scoring Manual revealed that each of the WISC-V measures generally met the guidelines suggested by Bracken (2007) for floors, ceilings, and item gradients. These results suggest that WISC-V measures contain a sufficient number of items for ensuring adequate construct variation.

Reliability Evidence. The WISC-V Technical and Interpretative Manual reports three methods of estimating reliability: internal consistency, test-retest stability, and interscorer agreement. Internal consistency estimates were obtained using the split-half method, using the Spearman-Brown correction formula for all subtests except Coding, Symbol Search, Cancellation, Naming Speed Literacy, Naming Speed Quantity, Immediate Symbol Translation, and Delayed Symbol Translation. Due to the speeded nature of the aforementioned measures, test-retest coefficients were used as reliability estimates for these measures. A table in the

Technical and Interpretive Manual presents subtest, process, and composite score reliability coefficients for each of the 11 age groups as well as the average coefficients across the age groups. Internal consistency estimates across the age groups ranged were .96 to .97 for the FSIQ, and ranged from .88 to .95 for index scores, and .81 to .94 for subtest scores. Coefficients for all of the indexes, with the exception of the PSI, exceeded .90 at all age levels. As would be expected, the range of subtest level coefficients (.76 to .95) was slightly more expansive across age groups. It should be noted that the coefficients for the VCI are lower than those that were reported for that same index in the WISC-IV (Wechsler, 2008). It is suggested that this is the result of the fact that the WISC-V VCI contains only two subtest measures whereas the WISC-IV VCI contained three.

Standard errors of measurement (*SEM*), based on the reliability coefficients are also reported in the Technical and Interpretative Manual. Overall average *SEM* for the composite and index level scores ranged from 2.90 (FQIQ) to 5.24 (PSI) and subtest level values ranged from .73 (Figure Weights) to 1.34 (Symbol Search). Though Hanna, Bradley, and Holen (1981) note that these estimates should be considered optimistic given that they do not account for potential sources of error such as administration or scoring errors.

The WISC-V Administration and Scoring Manual provides estimated true score confidence intervals (90% and 95%) that correspond to the observed standard score obtained for indexes and composites. In contrast to estimation methods that utilize the observed score and *SEM*, the true score estimation method utilizes an estimated true score (transformation of observed standard score) and the standard error of the estimate (*SEE*), resulting in an asymmetrical confidence interval (McDonald, 1999). This asymmetry occurs because the estimated true score is closer to the mean than the observed score. The estimation method using the *SEE* serves as a correction for regression to the mean. However, the bands reported in the Administration and Scoring Manual utilized the average reliability coefficient across ages rather than age-based coefficients in the estimation equations. Thus, if users wish to report more precise confidence bands that correspond more closely to the examinee's age, they will have to use observed level estimation methods to hand calculate them on a case by case basis. According to Glutting, McDermott, and Stanley (1987), these procedures are appropriate for individual decision-making.

Test-retest stability was estimated by administering the WISC-V twice to a stratified subsample of 218 participants comprising five age bands from the normative sample. Retest intervals ranged from 9 to 82 days with a mean interval of 26 days. Uncorrected stability coefficients for all ages were .91 for the FSIQ, .68 to .91 for index scores, and .63 to .89 for subtest scores. Corrected coefficients were slightly higher.

In order to examine interscorer agreement, all of WISC-V standardization sample record forms were double scored by two independent examiners and bivariate correlations were used as an index of agreement between the two forms. While the Technical and Interpretative Manual indicates that not all subtests were examined, it does not specify the subtests that were selected for inclusion. Overall, coefficients ranged from .98 to .99. Given the fact that the Verbal Comprehension subtests require more judgment in scoring, these measures were selected for additional examination. A sample of 60 record forms were randomly selected from the standardization sample and independently scored by nine raters who were in the process of completing clinical assessment training. None of the raters had any previous experience with the WISC-V measurement instrument. Coefficients were .98 for Similarities, .97 for Vocabulary, .99 for Information, and .97 for Comprehension.

Evidence of Validity. Consistent with the most current version of the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), validity evidence was provided in the areas of test content, response processes, internal structure, relations with other variables, special group studies, and the potential consequences of testing.

Content Validity. Content validity was estimated by surveying the relevant technical literature to substantiate the use of the WISC-V subtests for each latent trait estimated by each measure. An expert advisory panel was also formed to evaluate new items, as well as, to ensure improved subtest content coverage and theoretical relevance. Individual members of the advisory panel are listed in the Technical and Interpretive Manual.

Construct Validity. As expected, subtest intercorrelations were all positive across age groups, reflecting Spearman's (1904) positive manifold and measurement of the general ability factor (*g*). Consistent with current and previous iterations of the Wechsler Scales (e.g., Canivez, 2014b; Watkins, 2006), moderate to high correlations between the WISC-V index scores was also observed. Despite the significant content and structural modifications specified in the WISC-V revision plan, results from exploratory factor analysis (EFA) was not reported in the Technical and Interpretive Manual, a departure from previous versions of this instrument. The structural validity of the WISC-V was largely estimated using confirmatory factor analytic (CFA) procedures. CFA is generally preferred to EFA when the theory underlying the structure of a measurement instrument such as the WISC-V is known or has been well established in the technical literature (Schmitt, 2011). Though it should be noted that many researchers (e.g., Gorsuch, 1983; Haig, 2005) highlight the complimentary nature of EFA as it relates to CFA and advocate the use of multiple factor analytic procedures to obtain a clear picture of the most optimal measurement model explaining cognitive test data.

Due to recent investigations suggesting that a five-factor measurement model provided a better fit to other versions of the Wechsler Scales (e.g., Weiss, Keith, Zhu, & Chen, 2013a; 2013b), the WISC-V was developed under the theoretical assumption that the scale provides an estimate of general ability (*g*) along with five additional second-order cognitive factors (e.g., Verbal Comprehension, Visual-Spatial Processing, Fluid Reasoning, Working Memory, and Processing Speed). CFA procedures were utilized to examine the tenability of the five-factor model for all 16 of the WISC-V subtests when compared to competing one, two, three, and four-factor hierarchical models. The results of the CFA examinations indicated that a five-factor model adequately fit the WISC-V dataset and provided for statistically significant improvements to model fit when compared to several competing four-factor measurement models. However, additional clarification with respect to determining how to appropriately constrain the Arithmetic subtest was needed.

As a result of the multidimensional nature of the Arithmetic measure, conflicting results have been obtained in previous CFA examinations of the WISC-IV. Specifically, Keith, Goldenring Fine, Taub, and Kranzler (2006) found that Arithmetic best loaded on a hypothetical Fluid Reasoning factor; whereas, Weiss et al. (2013b) found that Arithmetic cross-loaded on both the Perceptual Reasoning Index (PRI) and the WMI within a four-factor model and loaded solely on a Fluid Reasoning factor in a five-factor model. Interestingly, in a CFA analysis of the WAIS-IV, Weiss and colleagues (2013a) found that Arithmetic cross-loaded on the VCI and WMI in a four-factor model and cross-loaded on the WMI and FRI (indirectly through an intermediate Quantitative Reasoning factor) in a five-factor measurement model.

Accordingly, contrasting five-factor models were examined in which a) Arithmetic was constrained to load only on the WMI; b) Arithmetic was constrained to load only on the FRI; c) Arithmetic was freed to cross-load on the FRI and WMI; d) Arithmetic was freed to cross-load on the VCI and WMI; and e) Arithmetic was freed to cross-load on the VCI, WMI, and FRI. Results indicated that a constrained loading on the FRI alone was not tenable due to a *g* loading for Fluid Reasoning (1.03) that was greater than 1.0, suggesting an improper solution (Brown, 2015). Ultimately, it was determined that the model in which Arithmetic was specified to cross-load on the VCI, WMI, and FRI best fit the WISC-V across five age groups and thus severed as the final validation model (see Figure 1). Subsequent analysis indicated that the validation model also provided excellent fit for the primary battery composed of the 10 core subtests (see Figure 2). Additional commentary in the Technical and Interpretive Manual revealed that incremental improvement in fit was obtained with a slight modification to the final validation model in which Figure Weights was unconstrained to cross-load on both the FRI and VSI. However, it was argued that this cross-loading made little sense theoretically and ultimately was not retained. Interestingly, inspection of the standardized coefficients in the final validation model again reveals isomorphism between *g* and Fluid Reasoning (1.00). Golay and colleagues (2013) argue that this common observation in CFA research with the Wechsler Scales is potentially an artifact of constraining non-trivial cross-loadings to zero, which has been shown to distort the underlying structure of measurement models (see Asparouhov & Muthen, 2009). Unfortunately, ancillary and complementary measures on the WISC-V were not specified in any of the validation models thus the relationship of these measures within the WISC-V structural/interpretive model is not known.

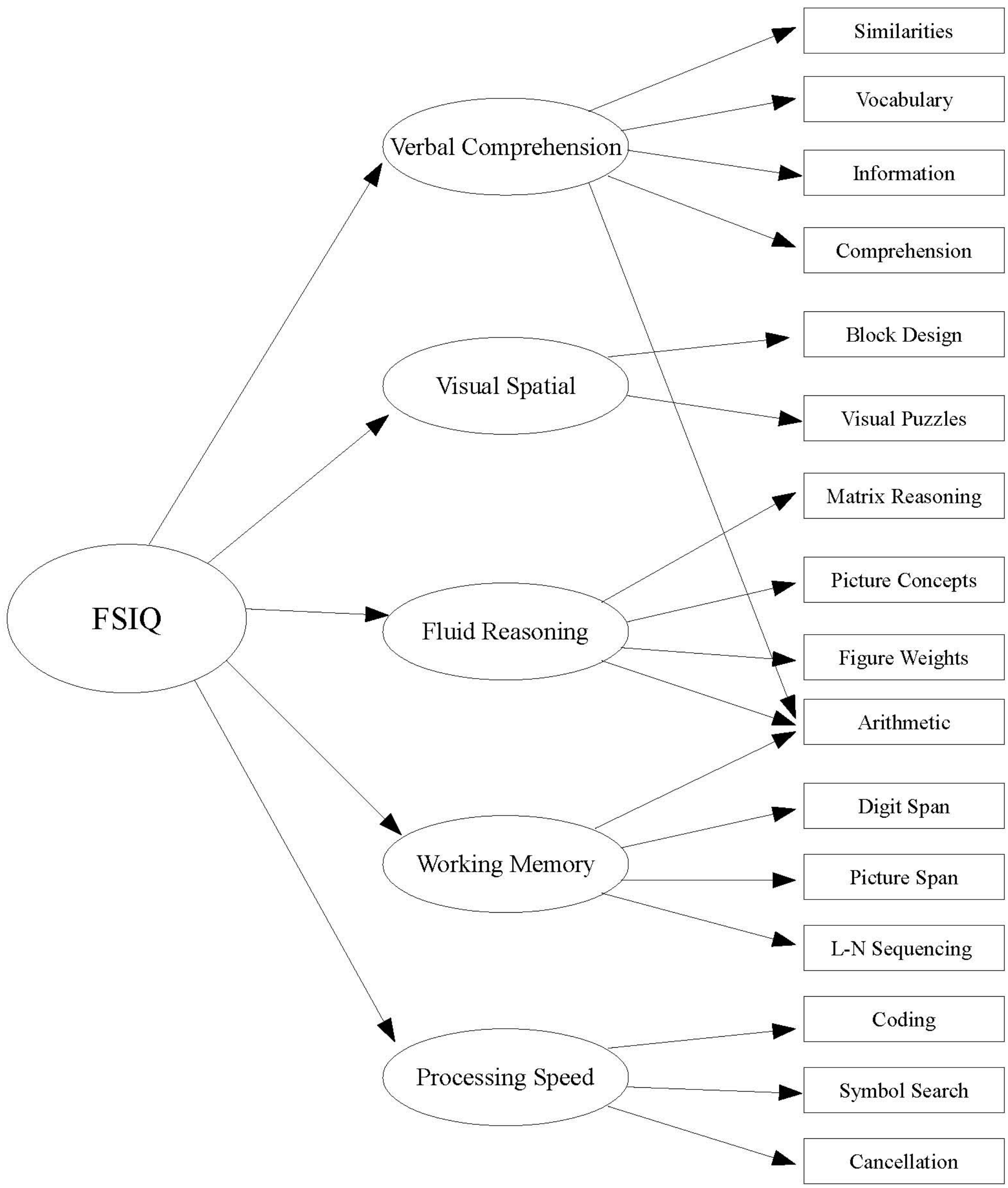

Figure 1. Final Five-Factor WISC-IV Validation Model for Primary and Secondary Subtests.

Additionally, the aforementioned cross-loadings (both specified and implied) also create a potential confound as it relates to estimating model-based reliability of some of the WISC-V subtest measures. As discussed previously, coefficient alpha was the primary metric utilized to estimate the internal consistency of the non-speeded WISC-V measures. According to Nunnally and Bernstein (1994), coefficient alpha can broadly be defined as a measure of the interclass correlation between all the items contained within a measure and is commonly (albeit incorrectly see Yang & Green, 2011) interpreted as an index for estimating the degree to which a set of items measures a single unidimensional latent construct. The assumption that all true score variance is attributable to a single latent dimension is critically important when determining whether the use of coefficient alpha is appropriate, as the coefficient cannot account for multiple sources of influence on the observed interclass correlation in psychological measures that are inherently multidimensional (Reise, Bonifay, Haviland, 2013). Although most of the research examining the effects of multidimensionality on the usefulness of coefficient alpha has been concerned with extricating higher-order variance (*g*) from lower-order variance (group factors), a Monte Carlo simulation conducted by Zinbarg, Revelle, and Yovel (2007) revealed that coefficient alpha may overestimate the reliability of a measure even more when items within a measure are influenced by multiple common or group factors (e.g., WISC-V index level abilities). In such circumstances, the use of alternative omega coefficients has been advised (Dunn, Baguley, & Brunsden, 2013; Yang & Green, 2011). Until such coefficients are calculated for WISC-V measures (e.g., Arithmetic, Figure Weights) that are suspected of being influenced by multiple group factors, users have no way of appropriately determining the mechanism(s) underlying the reliable variance that is observed within these measures.

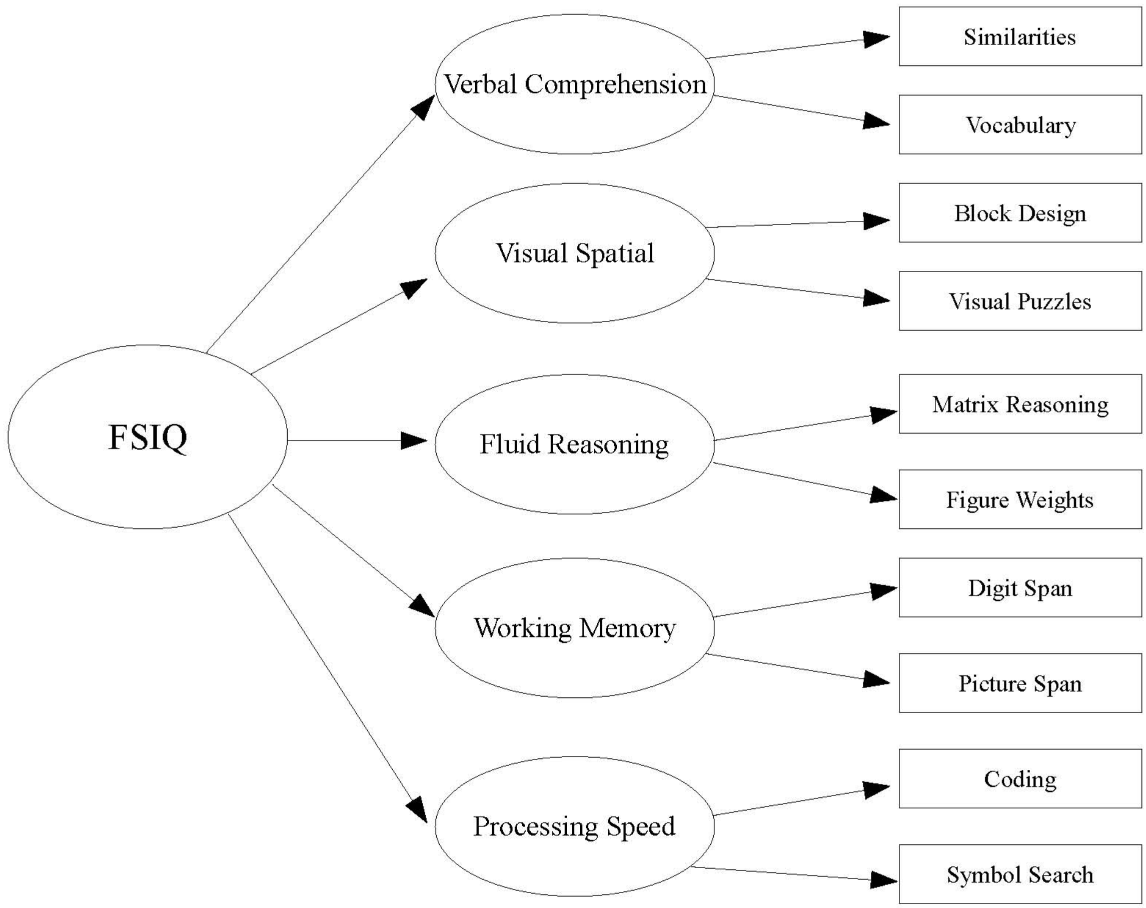

Figure 2. Five-Factor Validation Model for the Primary Subtests.

Subtest *g* loadings ranged from .21 (Cancellation) to .72 (Vocabulary). With the exception of Arithmetic (.70), measures from the VCI loaded highest on the general factor. The results are consistent with previous research (e.g., Keith et al., 2006). However, decomposition procedures (e.g., Schmid-Leiman, 1957) whereby subtest variance is appropriately apportioned to higher-order and lower-order dimensions was not reported. Given the hierarchical model nature of the structural model, such analyses are crucial for guiding the interpretative focus of users of this measurement instrument within clinical settings (Canivez, 2013).

Despite the ambitious structural validation procedures that were employed, the absence of several plausible measurement models (e.g., correlated factors, bifactor) from the CFA analyses are noteworthy. In the Technical and Interpretive Manual it was noted that validation studies was constrained to facilitate the examination of various hierarchical iterations. As it relates to measures of cognitive ability, the hierarchical or indirect hierarchical model implies that a higher-order construct (e.g., *g*) has indirect effects on subtest measures whereas lower-order broad abilities have direct effects. Thus, in the WISC-V, *g*-factor effects on the subtests are hypothesized to channel through the latent abilities estimated by the index scores. Alternatively, the bifactor or direct hierarchical model (Holzinger & Swineford, 1937) suggests that both the higher order *g*-factor and the broad second-order abilities have direct effects simultaneously on the subtests. Recently rediscovered (see Reise, 2012), the bifactor model has been found to provide better fit to data from multiple versions of the Wechsler Scales (Canivez, 2014a; Gignac, 2006; Gignac & Watkins, 2013; Golay et al., 2013; Nelson, Canivez, Watkins, 2013; Watkins & Beaujean, 2013) when compared to rival measurement models such as the correlated factors model and the indirect hierarchical model.

Ideally in CFA, a hypothesized measurement model is examined to determine how well it fits the data in relationship to all relevant competing models. Failing to specify a model that has been found to fit the data in previous researches is akin to using a convenience sample to make inferences regarding population parameters. This is not to suggest that the final validation presented in the Technical and Interpretive Manual is wrong however, the absence of relevant measurement models from the WISC-V structural analyses points to the need for additional research to be conducted so that users can be confidant in the factor structure implied by the configuration of the measurement instrument.

Relationships with Other Measures and Variables. Convergent and divergent validity was estimated by examining correlations between the WISC-V and a number of other measures, including commonly used measures of intellectual functioning and achievement. Overall conclusions indicate that the WISC-V correlated highly with instruments purported to measure similar cognitive and intellectual constructs. Of particular importance, scores on the WISC-V demonstrated high consistency with those from the previous edition, with correlations (corrected) ranging from .63 to .86 for composites and indexes and .57 to .82 for subtests. Of particular note, given the bifurcation of the WISC-IV's PRI into separate Visual Spatial and Fluid Reasoning indices in the current edition, correlations between the PRI and VSI (.66) and the PRI and FRI (.63) were similar. Correlations between the WISC-V indexes and theoretically consistent scores on the KABC-II were generally moderate to strong. With a strong correlation observed between the VCI and Crystallized Ability Composite (.74) and moderate correlations observed between the WMI and Short-Term Memory Composite (.63), VSI and Visual Processing Composite (.53), and FRI and Fluid Reasoning Composite (.50). Predictive relationships between the WISC-V and the WIAT-III and KTEA-3 achievement batteries were commensurate with estimates obtained from other measures of intellectual functioning. Consistent with previous research (e.g., Keith, Fehrmann, Harrison, & Pottebaum, 1987), preliminary evidence for divergent validity was established as a result of trivial or negative correlations between WISC-V scores and measures from behavior rating scales such as the BASC-2 and Vineland-2.

Small Group Studies. Small special subsamples (20 to 95 participants) and matched controls were compared to test for clinically significant group differences. Groups included individuals identified with giftedness, various levels of intellectual disability, specific learning disability, attention-deficit/hyperactivity disorder, traumatic brain injury, and autism spectrum disorder. Observed mean differences were consistent with theoretical expectations. Although the Technical and Interpretative Manual suggests that the WISC-V is useful for determining individual cognitive strengths and weaknesses that may be relevant for diagnosing specific learning disabilities, the evidence provided in the specific learning disability tables suggests that this conclusion may be optimistic. Generally, the most discernable discrepancy between learning disability subgroups was consistently lower scores across indexes when compared to matched controls. Limited evidence of breakout scores was observed. For instance, in the reading disability group, WISC-V index score means only fluctuated by four standard score points with all scores falling within the low average to average range. The lone exception was the QRI (M = 79.9) in the math disability group which is theoretically consistent given the traits purported to be sampled by that measure. Overall the WISC-V appears to be an adequate instrument for discriminating between individuals suspected of giftedness and intellectual disability although additional evidence is needed for establishing the potential diagnostic utility of the instrument (Canivez & Gaboury, 2011; Styck & Watkins, 2013).

Consequences of Testing. According to Braden and Niebling (2012), evidence based on the consequences that result from testing should include evaluations of diagnostic utility at the individual level. Accordingly, differential item functioning was used to examine potential item bias and content fairness. Inspections of item characteristic curves provided in the Technical and Interpretative Manual indicate that WISC-V items do not appear to discriminate between individual examinees on the basis of race or ethnicity. However, examiners must remain vigilant with respect to the intended and unintended consequences that may result from clinical use of the WISC-V (Hubley & Zumbo, 2011).

What Contributions will the WISC-V Make to the Field

The WISC-V is a significant and positive revision from its predecessor. The integration of additional neuropsychological constructs, which have been shown to predict various aspects of academic achievement, is a welcome addition to the test. The move from a four-factor model of interpretation better reflects current conceptualizations of intelligence. The test offers multiple psychometric comparisons between indices and subtests, which should enhance the test's clinical utility. The digital version of the test is a significant advancement for the assessment field. Like any new major test that is published, assessment specialists are encouraged to read future research studies that continue to validate the psychometric properties and clinical applications of the WISC-V.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bracken, B. A. (2007). Creating the optimal preschool testing situation. In B. A. Bracken & R. J. Nagle (Eds.), *The psychoeducational assessment of preschool children* (4th ed., pp. 137-154). Mahwah, NJ: Erlbaum.

Braden, J. P., & Niebling, B. C. (2012). Using the joint testing standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 739-757; 3rd ed.). New York: Guilford Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford press.

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

Canivez, G. L. (2014a). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly, 29,* 38-51. doi: 10.1037/spq0000032

Canivez, G. L. (2014b). Review of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The nineteenth mental measurements yearbook* (pp. 732-737). Lincoln, NE: Buros Institute of Mental

Measurements.

Canivez, G. L., & Gaboury, A. R. (2013). Construct validity and diagnostic utility of the Cognitive Assessment System for ADHD. *Journal of Attention Disorders*. Advance online publication. doi: 10.1177/1087054713489021

Coalson, D. L., Raiford, S. E., Saklofske, D. H., & Weiss, L. G. (2010). WAIS-IV: Advances in the assessment of intelligence. In L. G. Weiss, D. H. Saklofske, D. L. Coalson, & S. E. Raiford (eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 3-23). Amsterdam: Elsevier Academic Press.

Cohen, M. J. (1997). *Children's Memory Scale*. San Antonio, TX: Harcourt Assessment, Inc.

Crews, K. J., & D'Amato, R. C. (2009). Subtyping children's reading disabilities using a comprehensive neuropsychological measure. *International Journal of Neuroscience, 119*, 1615-1639. doi: 10.1080/00207450802319960

Dumont, R., Viezel, K. D., Kohlhagen, J., & Tabib, S. (2014). A review of Q-interactive assessment technology. *Communıqué, 43(1),* 8-12. Retrieved from http//:www.nasponline.org

Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105,* 399-412. doi: 10.1111/bjop.12046

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

Fletcher-Janzen, E. (2014). *Foreward*. In D. Wechsler, S. E. Raiford, & J. A. Holdnack. *Wechsler Intelligence Scale for Children – Fifth Edition: Technical and interpretative manual* (pp.xiii-xv). Bloomington, MN: Pearson.

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. doi: 10.1027/1614-0001.27.2.73

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability on the WAIS-IV. *Multivariate Behavioral Research, 48,* 639-632. doi: 10.1080/00273171.2013.804398

Glutting, J. J., McDermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement, 47,* 607-614. doi: 10.1177/001316448704700307

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structure equation modeling. *Psychological Assessment, 25,* 496-508. doi: 10.1037/a0030676

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

Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research, 40,* 303-329. doi: 10.1207/s15327906mbr4003_2

Hanna, G. S., Bradley, F. O., & Holen, M. C. (1981). Estimating major sources of measurement error in individual intelligence tests. *Journal of School Psychology, 19,* 370-376. doi: 10.1016/0022-4405(81)90031-5.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor model. *Psychometrika, 2,* 41-54. doi: 10.1007/BF02287965

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research, 103,* 219-230. doi: 10.1007/s11205-011-9843-4

Kaufman, A. S. (2010). Foreward. In L. G. Weiss, D. H. Saklofske, D. L. Coalson, & S. E. Raiford (eds.), *WAIS-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. xiii-xxi). Amsterdam: Elsevier Academic Press.

Kaufman, A. S., Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2006). Test review: Wechsler Intelligence Scale for Children: Fourth Edition (WISC-IV*). Journal of Psychoeducational Assessment, 24*, 278-295.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children – Second Edition*. Circle Pines, MN: American Guidance Service Publishing.

Keith, T. Z., Fehrmann, P. G., Harrison, P. L., & Pottebaum, S. M. (1987). The relation between adaptive behavior and intelligence: Testing alternative explanations. *Journal of School Psychology, 25,* 31-43. doi: 10.1016/0022-4405(87)90058-6

Keith, T. Z., Goldenring Fine, J., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review, 35,* 108-127. Retrieved from http://www.nasponline.org

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

Litt, R. A., de Jong, P. F., van Bergen, E., & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associative learning (PAL): What drives the PAL – reading relationship? *Journal of Experimental Child Psychology, 115*, 137-149. doi: 10.1016/j.jecp.2012.11.012

Luria, A. R. (1966). *The working brain: An introduction to neuropsychology*. NY: Basic Books.

Luria, A. R. (1973). *Higher cortical function in man*. NY: Basic Books.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). NY: Basic Books.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

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*, 51-675.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment - second edition*. Hoboken, NJ: John Wiley & Sons.

Nelson, J. M., Canivez, G. L., Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale-Fourth Edition with a clinical sample. *Psychological Assessment, 25,* 618-630. doi: 10.1037/a0032086

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill

Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV Assessment*. Hoboken, NJ: John Wiley & Sons.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667-696. doi: 10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95,* 129-140. doi: 10.1080/00223891.2012.725437

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* (pp. 99-144). New York: The Guilford Press.

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

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29,* 304-321. doi: 10.1177/0734282911406653

Schrank, F. A.; Mather, N., McGrew, K. S. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities Examiner's Manual, Standard and Extended Batteries*. Itasca, IL: Riverside.

Spearman, C. (1904). "General intelligence": Objectively determined and measured. *American Journal of Psychology, 15,* 201-293. Retrieved from http://www.jstor.org/stable/1412107

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

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children-Fourth Edition. *Psychological Assessment, 18,* 123-125. doi: 10.1037/1040-3590.18.1.123

Watkins, M. W., & Beujean, A. A. (2013). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. *School Psychology Quarterly, 29,* 52-63. doi: 10.1037/spq0000038

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds*.), Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 251-268). New York: The Guilford Press.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition*. Bloomington, MN: Pearson.

Wechsler, D., (2012). *Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition* Bloomington, MN: Pearson.

Wechsler, D., & Kaplan, E. (2015). *Wechsler Intelligence Scale for Children - Fifth Edition Integrated*. Bloomington, MN: Pearson.

Wechsler, D., Raifird, S. E., & Holdnack, J A. (2014). *Wechsler Intelligence Scale for Children – Fifth Edition: Technical and interpretative manual*. Bloomington, MN: Pearson.

Weiss, L. G., Keith, T., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31,* 94-113. doi: 10.1177/0734282913478030

Weiss, L. G., Keith, T., Zhu, J., & Chen, H. (2013b). WISC-IV and clinical validation of the four and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31,* 114-131. doi: 10.1177/0734282913478032

Yang Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29,* 377-392. doi: 10.1177/0734282911406668

Zhu, J., & Chen, H. (2011). Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment, 29,* 57-580. doi: 10.1177/0734282910396323

Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating ω_h for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement, 31,*135-157. doi: 10.1177/0146621606291558