

Chapter 18

Naglieri, J. A. (2016). Theoretical and Practical Considerations of the WISC-V. In A. S. Kaufman, S. Raiford & D. Coalson. *Intelligent Testing with the WISC-V (663-668)*. John Wiley & Sons.

Independent WISC-V Test Review: Theoretical and Practical Considerations

Jack A. Naglieri

I still have my copy of the original Wechsler Intelligence Scale for Children (WISC, Wechsler, 1949) which I was required to purchase (for about \$60) in 1973 for the first graduate course I took on assessment. I also still use my 1972 edition of *Wechsler's Measurement and Appraisal of Adult Intelligence* which Joe Matarrazo signed for me when I received an award for "outstanding achievements and contributions to psychology" from the Italian American Psychological Assembly during the 2011 APA convention. These tools formed the foundation of my learning about intelligence and its measurement and they guided my assessment of the many children I evaluated in schools and my clinical practice over the years that followed. My understanding of the Wechsler was greatly enhanced in 1977 when I took assessment for a second time as part of my Ph.D. program. By chance or fate, I was in Dr. Kaufman's class on how to interpret the WISC-R and he was assigned to be my advisor and later became my dear friend. I also still have my signed copy of *Intelligent Testing with the WISC-R* (Kaufman, 1979) and all the subsequent editions of that book that he has published. Those books have played a critical role in the field of intelligence testing with the Wechsler scales and it is an honor to have the opportunity to contribute to this edition.

I have always been a firm believer that the concept of general intelligence which the Wechsler scales represent remains, as noted by Anastasi and Urbina (1997), one of the most valuable contributions that psychology has made to society. It is with great confidence, as well as

research support (Bracken & McCallum; 2009; Naglieri & Otero, 2012; Brunnert, Naglieri & Hardy-Braz, 2009), that I have worked to publish nonverbal measures of general ability (Naglieri, 2008a; Wechsler & Naglieri, 2006). These tools meet the need to evaluate general ability using nonverbal content just as this method was originally intended by those who developed them - the U.S. Military. It is amply documented (Naglieri, 2008b, 2015) and clearly stated by Yoakum and Yerkes (1920) that the Army Beta test (the so called nonverbal portion) was intended for those who had limited English language skills as well as those who were illiterate (p. 51). The Army Alpha contained verbal and quantitative tests and was only appropriate for men who had the benefit of an education and could read and write English sufficiently. These testing procedures were intended to be fair for diverse populations, perhaps best illustrated when Yoakum and Yerkes stated: "men who fail in alpha are sent to beta in order that injustice by reason of relative unfamiliarity with English may be avoided (p. 19)". These two tests were developed for the practical need to evaluate a wide variety of people and not to measure different types of intelligence. Importantly, it was the Army Alpha and Beta which Wechsler used when he assembled the Wechsler-Bellevue in 1939.

Wechsler, like Yoakum and Yerkes, did not believe that his Verbal and Performance scales measured two different abilities. This point was emphasized by Kaufman in the Foreword of the Wechsler Nonverbal Scale of Ability (WNV) *Administration Manual* when he wrote:

The emphasis in the WNV Manual that the Full Scale measures general ability nonverbally—and not nonverbal ability—is an important distinction that further ties the WNV to Dr. Wechsler. Although his intelligence tests in the 1930s and 1940s departed from the one-score Stanford-Binet by offering separate Verbal and Performance IQs as well as a profile of scaled scores, Dr. Wechsler remained a firm believer in Spearman's *g* theory throughout his lifetime. He believed that his Verbal and Performance Scales represented different ways to access *g*, but he never believed in nonverbal intelligence as being separate from *g*. Rather, he saw the Performance Scale as the most sensible way to measure the general intelligence of people with hearing impairments, language disorders, or limited proficiency in English.

The use of nonverbal tests of general ability has increased in recent years as the percentage of people in the US with limited English language skills has increased (for more information about specific nonverbal tests see Naglieri & Goldstein, 2009) because it is clear that verbal tests are not appropriate for those with limited English language (and math) skills as Yoakum and Yerkes stated nearly 100 years ago. Tests like Vocabulary, Arithmetic, Similarities, and Comprehension do require knowledge. This point raises the critically important issue relating to all of Wechsler's tests (and traditional IQ tests); "Is it really a good idea to measure intelligence using subtests that also require knowledge?"

The idea that verbal tests of intelligence can perhaps be better thought of as tests of achievement, or at least tests of general ability that are confounded by knowledge, was first presented by Alan and Nadeen Kaufman when they began development of the *Kaufman Assessment Battery for Children*. (Kaufman & Kaufman, 1983) When Alan Kaufman described the WISC-R Verbal scale as achievement (he associated it with what would be the achievement portion of the K-ABC) I remember

thinking "that makes a lot of sense". His comment reminded me of my experiences giving the WISC-R and the *Peabody Individual Achievement Test* which both had an Information subtest, with questions that were essentially the same but the interpretation was very different; intelligence or achievement? The second point made by the Kaufmans was equally true and even more important- they wanted to build their test of intelligence on a theory of intelligence. Additionally, the test should be fair for diverse populations and inform instruction. These were revolutionary ideas in the late 1970s. It was clear to me that this was the way to advance the field of intelligence testing. I owe much of my understanding of how best to build a modern measure of ability to my experiences helping develop the K-ABC.

My view is that a measure of ability should be clearly based a neurocognitive theory, with the goal that the test should be as free of questions that require knowledge as possible so that the items would be accessible for diverse populations. Equally important, the test should be consistent with the theory. Evidence that the scales a test yields are predictive of achievement would be critical in order to understand the basic validity of the measure. Evidence must be found that the theoretically derived scales are sensitive to the cognitive problems experienced by those who have, for example, attention deficits and specific learning disabilities. Research must show that the scales are non-discriminatory as emphasized by IDEA. And finally, there should be a strong connection between test results and instruction. I have shown that all these criteria can be addressed with a brain-based neurocognitive theory that includes only four dimensions - Planning, Attention, Simultaneous and Successive neurocognitive processes (Naglieri & Otero, 2012; Naglieri & Conway, 2009; Naglieri, 2015). It is these standards and this perspective to which I have held my own tests (the *Cognitive Assessment System* (Naglieri & Das, 1997) and the CAS2 (Naglieri, Das & Goldstein, 2014) that will form the basis of my review of the WISC-V.

WISC-V and Theory

It was encouraging to see that the authors of the WISC-V manual included a section entitled “Update Theoretical Foundations” (pg. 22). I wondered if the development of the fifth edition was actually guided by a theory. That hope was quickly dashed by the very first sentence: “Various theories and models relevant to intellectual assessment influenced the WISC-V,” yet no specific theory was identified upon which the test was based. Next was a section about “Structural Intelligence Models” that begins with a discussion to clarify if Wechsler believed in general ability (as Kaufman has asserted) or that his tests measured specific abilities. There was no real resolution of that topic. Next the discussion continued with the position that because factor analytically derived models are widely accepted, the “verbal comprehension, visual spatial, fluid reasoning, working memory and processing speed abilities are important” (p. 23). So it seems clear that the factor analytic results obtained for this version of the Wechsler was used to define the ‘structural theory’. That is the method that has been used by Pearson and the previous publisher The Psychological Corporation for many years and why there were three scales in the WISC-III, four in the WISC-IV, and five in the WISC-V.

Next the manual states: “Theory is not the only consideration that drives development of the ...WISC-V” (p 23). This is followed by a long section on ‘Neurodevelopmental and Neurocognitive Research’. Reading this interesting summary of research leads one to wonder what relevance it has for the theory behind the WISC-V, but a hint to the answer appears on page 25. The authors appear to argue that because children use their brains to answer the questions on this test that somehow this supports the, as yet to be clearly defined, neuropsychological theory behind the WISC-V. One can only conclude that there really is no clearly defined theory that was used to develop the WISC-V and that users are, unfortunately, encouraged to find a model, hypothesis, or conceptualization that best fits the WISC-V scores a child or adolescent may earn. This would be a reasonable approach if there was

enough empirical support for the validity of all the possible interpretations.

The need for a theory is among the most practical of issues because it has tremendous implications for eligibility determination/diagnosis, and intervention. In the first chapter of the WISC-V manual is a section on subtest content where the authors provide considerable discussion about the “constructs [each] subtest is designed to measure (p. 7)”. If there was a specific theory, each subtest should represent the construct corresponding to the scale upon which the subtest was placed. For example, the Similarities, Vocabulary, Information and Comprehension subtests could have simply been deemed measures of Verbal Comprehension. (My preference would be to state, for example, that general ability is measured using these subtests that require verbal comprehension and expression.) Instead, the WISC-V authors give a list of as many as a dozen abilities or other factors that may be involved in answering the items on each of the WISC-V subtests. The list includes everything from verbal concept formation to cognitive flexibility, to auditory comprehension, and many more –which they justified because these interpretations appeared in a few books. This is clearest evidence that the theoretical construct underlying each subtest is undefined because there is no unifying theory upon which the WISC-V was built. Instead a few new subtests were added based on a rationale that is not well articulated and the factor structure was used to identify the concepts the scales represent. The end result is that without a clear theory to guide the interpretation of the WISC-V, the responsibility for understanding learning success and failure, connecting the test results to legal definition of a specific learning disability in IDEA, and determining what type of instructional intervention to use becomes much more difficult.

WISC-V and Achievement

One of the most important types of evidence of an ability test’s validity is its correlation with achievement. This kind of validity is important because we use tests like the WISC-V to help explain why a student

referred for an evaluation is having trouble in school. The answer to this question helps us understand how well the intelligence test we use is related to current academic performance, and we hope it also provides good prediction of future performance. The *WISC-V Manual* includes an important section on the relationships with the *Wechsler Individual Achievement Test – Third Edition* (WIAT-III; Wechsler, 2009). The results are most informative, especially when the correlation between the WISC-V and WIAT-III is understood in relation to the results previously reported for the WISC-IV and WIAT-II (Wechsler, 2003).

Before the correlations between the WISC-V and WIAT-III can be evaluated we first have to consider the similarity in content between these two seemingly different tests. We assume that because the WISC-V is a measure of intelligence and the WIAT-III is a measure of ability that the content of these tests is different. That assumption has been questioned and the similarities between the tests documented (Naglieri & Bornstein, 2003). It must be recognized that both tests have items requiring math facts (Arithmetic on the WISC-V and Mathematics and Math Fluency on the WIAT-III) and questions requiring knowledge of words (Vocabulary and Similarities on the WISC-V and Oral Language, Basic Reading and Reading Comprehension on the WIAT-III). The similarity in content across these two tests which are intended to measure two distinct constructs (intelligence vs. achievement) ensures that they will be correlated because of the similarity of the knowledge both demand. This is a significant conceptual and methodological issue that should be addressed in the WISC-V (or any of the traditional IQ test) manual.

The similarity in content between Wechsler's intelligence and achievement tests has an effect on any study of the validity of the WISC-V; the obtained correlation should be considered an overestimate of the relationship between general ability and achievement. Some indication of how much of an overestimate the shared content creates can be gleaned from the correlations of the several WISC-V scales with the WIAT-III. Examination of Table 5.13 of the

WISC-V Manual provides these correlations. The WISC-V Full Scale IQ correlation is .81 with the Total Achievement score from the WIAT-III. The Verbal Comprehension Index correlated the highest (.74) with the WIAT-III Total Achievement scale and the remaining scales' correlated substantially lower (Visual Spatial = .46; Fluid Reasoning = .40; Working Memory = .63; and Processing Speed = .34). The best explanation for why the Verbal Comprehension scale and the WIAT-III were so highly correlated is the similarity in content across the two tests. The correlation between the four remaining scales yields a good estimate of the relationship between the WISC-V and achievement without the overlap in test content. The average of those values is .47, which gives a very different view of validity.

A similar pattern involving correlations with the WIAT-II (Wechsler, 2001) is found in the WISC-IV Manual (Table 5.15). The highest correlation between the WIAT-II Total Achievement and WISC-IV was found for the Verbal Comprehension Scale and all other correlations were lower. Most interesting is the difference between the Total Achievement and Full Scale score correlations for the WISC-IV (.87) and WISC-V (.81). These findings beg the questions: Why isn't the WISC-V correlation with the WIAT-III higher than the WISC-IV / WIAT-II correlation given that the new version now has five rather than four scales? Is this a sampling issue or has the structure and new items on the WISC-V rendered the new test less effective than the previous edition?

WISC-V Profiles and Eligibility Determination

The WISC-V authors rightfully remind the reader "Intelligence tests were not originally designed to serve as neuropsychological measures" but they also suggest that "the WISC-V primary index scores represent cognitive processes of interest in neuropsychological assessment...and [they can be used] to generate hypotheses about neuropsychological processing deficits..." (p. 34). This statement suggests that the scales measure cognitive processes that could be used for eligibility determination. There are at least two important pre-requisites

for recommending that scores on a test of ability be used to measure cognitive processes. First, the test should be built on a theory of brain-based cognition, which was addressed earlier in this review; and second, that there should be evidence of distinct profiles for students with specific learning disabilities. Examination of the profiles for students with specific learning disabilities in reading (Table 5.28) and written expression (5.30) does not suggest that practitioners should anticipate specific Primary or Ancillary scale profiles. Equally important is the implication that the WISC-V can be used to measure processing strengths and weaknesses to meet criteria in IDEA 2004.

In 1999, I suggested that evidence for a specific learning disability could be found if a pattern of strengths and weaknesses in basic psychological processes which corresponded to similar variability in achievement test scores was obtained during a comprehensive evaluation. This approach unites the definitional criteria found in IDEA 2004 for specific learning disability with the method for making the eligibility determination (Hale, Kaufman, Naglieri & Kavale, 2006). The authors of the *WISC-V Manual* describe this approach as a “legally acceptable and clinically sound approach for helping practitioners identify SLDs and develop intervention plans based on a child’s strengths and weaknesses (p. 183)” but the evidence that this method applies to the scores from the WISC-V is lacking. What would be needed is (a) a theory based definition of the basic psychological processes, preferably defined according to a neuropsychological framework (e.g., Das, 2015; Otero, 2015); (b) evidence of distinct profiles for students with specific learning disorders; and (c) research evidence that profiles have relevance for instructional decision making. Much research is needed to demonstrate the utility of WISC-V profiles for eligibility determination and instructional relevance.

Determining eligibility for educational services as well as diagnosis of childhood disorders certainly involves thoughtful integration of information from a variety of sources and a wide variety of tools. I agree completely with the statement in the WISC-V

Manual that “The practitioner, using tests as part of the assessment activity, is responsible for interpreting results and making diagnostic, treatment, or intervention decisions (p. 186).” The critical issue is, however, how the profile of scores and interpretations offered in the WISC-V Manual help or hinder accurate decisions which can have a profound influence on a young person’s life.

WISC-V and Race/Ethnic

The WISC-V Manual has a section on the “Consequences of Testing” (p. 147) that discusses, for example, the importance of the clinical diagnostic utility at the individual level and the item bias procedures used during test development. It is also reported (see Chapter 2) that “problematic items were deleted on the basis of formal expert review of the items and empirical data from statistical analyses of differential item function (p. 32). This is good test development. Some (Braden & Niebling, 2012) also stress the importance of looking at test score differences between groups (as noted in the WISC-V Manual) but WISC-V mean score differences for diverse populations was given little attention. The one study that is reported involves a group of Asian and Hispanic English Language Learners (ELL); and that study only has 16 participants. The results of this study are consistent with expectations discussed earlier in this review – these children earned low scores on the verbal (i.e., achievement based) subtests Similarities, Vocabulary, and Information and the scale which they comprise (Verbal Comprehension Index). The VCI mean (85.6) is one standard deviation below the normative mean and substantially lower than the matched control group. It is certainly expected that children learning English would earn low scores on tests that require comprehension and expression of their second language. What is concerning about the description of these findings is the manner in which the findings are interpreted. My concern is with the statement: “subtests requiring minimal expressive language and reduced receptive language *abilities* [emphasis added]...” (p. 141); implies that the verbal tests on which these children did poorly are measures of verbal ability. This can lead consumers to conclude that a child learning

English has low verbal ability when they may not. It would be in the best interest of students who are learning English that all subtests that demand knowledge and use of words not be interpreted as a measure of *ability* (Fagan, 2000; Suzuki & Valencia, 1997). The authors of the WISC-V should have made it clear that when this test is given to those with limited knowledge of English the Verbal Comprehension Index should be considered spoiled (or more a measure of verbal expression and comprehension of English) and not used to create a Full Scale. Failing to recognize the confounding influence of knowledge of English has led many professionals to incorrectly evaluate the intellectual levels of countless students.

Conclusions

Any review of the fifth edition of the Wechsler Scales has to begin by recognizing the enormous impact this tool has had on the field of psychology and education. The test's use is unsurpassed by any measure of ability and it has come to represent the very definition of intelligence. As the description and number of abilities Wechsler's test measures has changed, there has been a growing awareness that the essential ingredients initially developed in 1917 by the U.S. Army do not sufficiently meet the needs of the field today. Efforts on behalf of the publisher for sustaining the Wechsler brand reflect the recognition that more information is needed from the test. Their solution to the dilemma, make it better but do not change it too much, apparently has led to reliance on factor analysis and diverse interpretive solutions. The authors of the WISC-V have made a valiant attempt to strengthen the integrity of this time honored test. Even given all the efforts to inform users of the various ways to interpret the scores the WISC-V yields, these ideas are constrained by the historically determined content. All this raises the question, "How can the field move forward and achieve a more effective way to measure intelligence?" I believe we can add to what the WISC-V measures and meet the demands of our current educational and psychological professions by including additional measures in our comprehensive assessments.

In my recent book chapter entitled, "100 Years of Intelligence Testing: Moving from Traditional IQ to Second-Generation Intelligence Tests (Naglieri, 2015)," I have provided both the rationale and research evidence which shows the advantages of new approaches to understanding human functioning (intelligence). The two tools I describe as second-generation are the *Kaufman Assessment Battery for Children* (first and second editions) and the *Cognitive Assessment System* (first and second editions). Considering these modern approaches to measuring ability along with an objective analysis of the Wechsler (and other traditional IQ tests) gives a greater understanding of learning and learning difficulties. We can improve our view of intelligence using tools that are defined within a neurocognitive theory and for the sake of the children and adolescents we work so diligently to help, we should.

References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper saddle River, NJ: Prentice Hall.
- Bracken, B. A. & McCallum, R. S. (2009). Universal Nonverbal Intelligence Test. In J. A. Naglieri & S. Goldstein. *A Practitioner's Guide to Assessment of Intelligence and Achievement* (pp. 291-313). New York: Wiley.
- Braden, J. P. & Niebling, B. C. (2012). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 739-757). New York, NY: Guilford Press.
- Brunnert, K. A., Naglieri, J. A., & Hardy-Braz, S. T. (2009). *Essentials of WNV assessment*. Hoboken, NJ: John Wiley & Sons, Inc.
- Das, J.P. (2015). *Cognitive Planning and Executive Functions*. New Delhi: SAGE Intl Publishers
- Fagan, J. R. (2000). A theory of intelligence as processing: Implications for society.

- Psychology, Public Policy, and Law*, 6, 168–179.
- Hale, J. B., Kaufman, A. S., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Using RTI and Cognitive Assessment Methods. *Psychology in the Schools*, 753-770.
- Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.
- Naglieri, J. A. (2008a). *Naglieri Nonverbal Ability Test – 2nd Edition*. San Antonio, TX: Pearson.
- Naglieri, J. A. (2008b). Traditional IQ: 100 Years of Misconception and its Relationship to Minority Representation in Gifted Programs in J. VanTassel-Baska (Ed.) *Critical issues in equity and excellence in gifted education* (pp. 67-88). Waco, TX: Prufrock Press.
- Naglieri, J. A. (2012). Psychological Assessment by School Psychologists: Opportunities and Challenges of A Changing Landscape. In K. Geisinger & B. A. Bracken (Eds.) *APA handbook of testing and assessment in psychology*. Washington, D.C.: APA.
- Naglieri, J. A. (2015). 100 Years of Intelligence Testing: Moving from Traditional IQ to Second-Generation Intelligence Tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.) *Handbook of intelligence: Evolutionary theory, historical perspective and current concepts*. New York: Springer.
- Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Austin: ProEd.
- Naglieri, J. A., & Goldstein, S. (2009). *A practitioner's guide to assessment of intelligence and achievement*. New York: Wiley.
- Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and Achievement: Just How Correlated Are They? *Journal of Psychoeducational Assessment*, 21, 244-260.
- Naglieri, J. A., & Conway, C. (2009). The Cognitive Assessment System. In J. A. Naglieri & S. Goldstein. *A Practitioner's Guide to Assessment of Intelligence and Achievement* (pp. 3-10). New York: Wiley.
- Naglieri, J. A. & Otero, T. M. (2012). The Cognitive Assessment system: From Theory to Practice. In D. P. Flanagan & P.L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 376-399). New York: Guilford.
- Naglieri, J. A. & Otero, T. M. (2012). The Wechsler Nonverbal Scale of Ability: Assessment of diverse populations. In D. P. Flanagan & P.L. Harrison (Eds.). *Contemporary intellectual assessment, third edition: Theories, tests, and issues* (436-455). New York: Guilford.
- Naglieri, J. A., Das, J. P. & Goldstein, S. (2014). *Cognitive Assessment System – Second Edition*. Austin, TX: ProEd.
- Otero, T. M. (2015). Intelligence: Defined as neurocognitive processing. In S. Goldstein, D. Princiotta, & J. A. Naglieri (Eds.) *Handbook of intelligence: Evolutionary theory, historical perspective and current concepts*. New York: Springer.
- Suzuki, L. A., & Valencia, R. R. (1997). Race–ethnicity and measured intelligence. *American Psychologist*, 52, 1103–1114.
- Wechsler, D. (2001). *Wechsler Individual Achievement Test – Second Edition*. San Antonio, Psychological Corporation.
- Wechsler, D. (2009). *Wechsler Individual Achievement Test- Third Edition*. San Antonio, Pearson.
- Wechsler, D. (2003). *Wechsler Intelligence Test for Children – Fourth Edition*. Bloomington, MN: Pearson.
- Wechsler, D. (2015). *Wechsler Intelligence Test for Children – Fifth Edition*. Bloomington, MN: Pearson.
- Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Pearson.