100 Years of Intelligence Tests: We Can Do Better

Jack A. Naglieri[1]
University of Virginia
jnaglieri@gmail.com
www.jacknaglieri.com

Intelligence tests such as the Binet and Wechsler which are widely used in the US and other parts of the world have played a pivotal role in nearly all aspects of psychology and education. Despite the enormous contribution these tests have made to our ability to measure the construct of intelligence, they have important weaknesses. For example, (1) intelligence *tests* were not initially built on a *theory* of intelligence which led to an ill-defined blueprint for test development and interpretation; (2) the use of verbal and quantitative tests as measures of *ability* is hard to justify given that these test items are virtually indistinguishable from questions on tests of *achievement*; (3) the subtests that demand knowledge create considerable problems for equitable assessment of those with limited opportunity to learn which amplifies race and ethnic differences; (4) the inclusion of subscales on these tests have failed to account for more variance beyond general ability; (5) subtest and scale profiles for students with specific kinds of learning problems have not been validated; and (6) scores from subtests and scales continue to have little no relevance to instruction (see Naglieri (2015) and Naglieri and Otero (2017) for a summary of the evidence regarding these limitations). Some have argued that IQ tests should be abandoned, I think we need a more realistic view of the most widely used intelligence tests and to redefine the construct and its measurement.

A Realistic View of Intelligence Tests

Tradition intelligence tests measure general ability 'g', and little more (Benson, Beaujean, McGill & Dombrowski, 2018; Canivez, Watkins & Dombrowski, 2017). This is consistent with the intentions of the first test authors. Recall that when Binet published the 1905 edition of his new test, it yielded one score. Shortly thereafter, Yoakum and Yerkes published the *Army Mental Tests* (1920) upon which the Wechsler intellignce scales (originally published in 1939) were largely based. These tests all contained verbal, quantitative, and nonverbal test content and even though the first Wechsler intelligence test yielded Verbal IQ and Performance IQ scores it included a Full Scale score which represented general ability (*g*). "Dr. Wechsler remained a firm believer in Spearman's *g* theory throughout his liftime. He believed that his Verbal and Performance Scales represented different ways to access *g*, but he never believed in nonverbal [or verbal] intelligence as being separate from *g*" (Kaufman, 2006). It is also important to note that in the early 1900s "psychologists borrowed from every-day life a vague term implying all-round ability and knowledge, and in the process of trying to measure this trait and [we] are still attempting to define it more sharply and endow it with a stricter scientific connotation (Pintner, 1923, p. 53)." Recent efforts to improve 100 year-old tests like the *Stanford-Binet V* (Roid, 2003) and the *Wechsler*

[1] Jack A. Naglieri is the author of the first and second editions of the Cognitive Assessment System (CAS) and coauthor of the Wechsler Nonverbal Scale of Ability. I thank Gary Canivez, Alan S. Kaufman, and Cecil R. Reynolds for their thoughtful comments on an earlier draft of this article. 1.16.20

Intelligence Scale for Children Fifth Edition (Wechsler, 2014) have focused on increasing the number of scales the tests yeild.

Importantly, recent research on intelligence tests confirm that the most valid score on, for example, the *Wechsler Intelligence Scale for Children – Fifth Edition* (Canivez, Watkins, & Dombrowski, 2017), *Stanford-Binet Fifth Edition* (Canivez, 2008), *Differential Abilities Scales* (Canivez & McGill, 2016), and the *Woodcock-Johnson Fourth Edition* (Dombrowski, McGill & Canivez, 2017) is the total score that estimates *g*. That is, the scores which represent the factor based scales these tests provide do not have enough specific variance to be considered interpretable. In fact, a recent reanalysis of Carroll's survey of factor-analytic studies by Benson, Beaujean, McGill, and Dombrowski (2018) concluded that nearly all of the specified abilities presented by Carroll "have little-to-no interpretive relevance above and beyond that of general intelligence (p. 1028)." The only exception to these finding is research reported by Canivez (2011) regarding the *Cognitive Assessment System* (Naglieri & Das, 1997).

These important research findings pose a professional quandary for all those who rely on intelligence tests in applied and theoretical research. That is, if we only use the total score, then the factor based scales provided by authors or publishers get ignored. Can we make important decisions based on the profiles of subtest and scales based only on clinical judgement? Again, the answer from the most recent research (McGill, Dombrowski & Canivez, 2018) suggests, as McDermott, Fantuzzo & Glutting (1990) said 30 years ago, is no. McGill, et al., go on to state: "Given these complexities, it is imperative that practitioners develop a skill set that helps them to discern when claims made in the assessment literature are credible… (p. 118)."

How to Redefine the Construct of Intelligence and its Measurement

Some test authors have provided alternatives to traditional intelligence tests. An example is the *Kaufman Assessment Battery for Children* (K-ABC; Kaufman & Kaufman) first published in 1983 which emphasized the importance of having a theoretical prespective. They emphasized measuring cognitive processes and also bravely took the position that the verbal and quantitative subtests should be taken out of the measurement of ability. (Note that they modified this position in the second edition of that test when they provide a CHC interpretation of their test.) A second effort to advance the conceptualization and measurement of intelligence was provided in 1997 when Naglieri and Das published the *Cognitive Assessment System* (CAS). That approach was similar to the one taken by the Kaufmans in 1985 in so far as subtests requiring knowledge of vocabulary and arithmetic were excluded. The CAS was unique in that it contained four scales following A.R. Luria's (1980) view of four brain-based cognitive processing abilities. The goal was to provide a new way of defining ability based on a cognitive and neuropsychological theory, and to develop a test to measure these basic psychological processes. Both the K-ABC and the CAS departed from the traditional intelligence approach because of content differences and the strong theoretical basis. The concept behind the K-ABC was a convergence of two-dimensional understandings of ability (e.g., left and right brain, simultaneous and sequential, parallel and serial) and therefore designed to measure two dimensions. The CAS was the only test of intelligence to be explicitly developed based on one

conceptualization – that described by A. R. Luria.

J. P. Das and I relied on the many works of A. R. Luria. For example, in *The Working Brain: An Introduction to Neuropsychology* (Luria, 1973) Luria described four neurocognitive processes associated with different parts of the brain. The first is **P**lanning, which is a mental activity that provides cognitive control; use of processes, knowledge, and skills; intentionality; organization; and self-monitoring and self-regulation. This processing ability is closely aligned with frontal lobe functioning (third functional unit). **A**ttention is the ability to demonstrate focused, selective, sustained, and effortful activity over time and resist distraction associated with the brain stem and other subcortical aspects (first functional unit). **S**imultaneous processing provides a person with the ability to integrate stimuli into interrelated groups or a whole usually found in tasks with strong visual-spatial demands. **S**uccessive processing ability involves working with stimuli in a specific serial order including the perception of stimuli in sequence and the linear execution of sounds and movements. This theory of brain function provided a blue print for test development and excluded tests that demand knowledge (e.g., Vocabulary, Arithmetic), thereby making it a more equitable measure as stated in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

Building a new test of intelligence that is based on a psychological processing conceptualization of brain function provides many advantages. Although the evidence of validity and reliability as well as clinical utility of the PASS conceptualization of intelligence is beyond the scope of this article, Naglieri and Otero (2017) have reported that PASS scores (a) are more predictive of achievement test scores than any other ability test; (b) show distinctive and stable profiles for students with different disabilities; (c) can be used for SLD eligibility determination consistent with Federal Law when the Discrepancy Consistency Method is applied to PASS and achievement test scores; (d) offer the most equitable way to measure diverse populations; and (e) can be readily used for instructional planning and interventions. These findings clearly suggest that we can do better than the latest editions of 100-year-old approaches to measuring intelligence.

Conclusions

Change in any field is not always easy. Perhaps the hardest part is looking at what we have with a fresh perspective. The two second-generation tests noted here illustrate that the K-ABC and CAS emphasis on measuring psychological processes provide new ways of thinking about and measuring intelligence. There is enough evidence to support consideration of a change in our field. I suggest that researchers and practitioners embrace this transition with the understanding that an evolutionary step in our field is most definitely needed considering all we have learned in the past 100 years. Only though substantial change can we improve the evaluation of human intelligence.

References

AERA, APA, NCME (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA

Benson, N. F., Beaujean, A. A., McGill, R. J, & Dombrowski, S. C. (2018). Revisiting Carroll's Survey of Factor-Analytic Studies: Implications for the Clinical Assessment of Intelligence. *Psychological Assessment,* 30, 8, 1028–1038.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29,* 458-472. http://dx.doi.org/10.1037/pas0000358

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales–Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment, 28,* 1475-1488. http://dx.doi.org/10.1037/pas0000279

Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales-Second Edition: Exploratory and hierarchical factor analyses with the core subtests. Psychological Assessment, 28, 1475–1488. https://doi.org/10.1037/pas0000279

Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales-Fifth Edition for children and adolescents. School Psychology Quarterly, 23, 533–541. https://doi.org/10.1037/a0012884.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ IV Cognitive at school age. *Psychological Assessment, 29,* 394-407. http://dx.doi.org/10.1037/pas0000350

Kaufman, A.S. (2006). Foreword to the *Wechsler Nonverbal Scale of Ability Administration and Scoring Manual* (p. iv). San Antonio: Pearson.

Kaufman, A.S., & Kaufman, N.L. (1983). *Kaufman Assessment Battery for Children.* Circle Pines: American Guidance.

Luria, A.R. (1973). *The Working Brain: An Introduction to Neuropsychology.* New York: Basic Books.

Luria, A. R. (1980). *Higher Cortical Functions in Man (2nd ed.).* New York: Basic Books.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. Journal of Psychoeducational Assessment, 8, 290–302. https://doi.org/10.1177/073428299000800307

Naglieri, J.A. (2015). One Hundred 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*. (295-316). New York: Springer.

Naglieri, J. A., & Das, J. P. (1997). Cognitive *Assessment System*. *Austin: ProEd*.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2012). *Cognitive Assessment System—Second Edition*. Austin, TX: PRO-ED.

Naglieri, J. A. & Otero, T. M. (2017). *Essentials of CAS2 Assessment*. New York: Wiley.

Pintner, R. (1923). Intelligence Testing; Methods and Results. New York: H. Holt and Company.

Roid, G. (2003). *Stanford Binet IntelligenceTest – Fifth Edition*. Austin, TX: PRO-ED.

Wechsler, D. (2014). *Wechsler IntelligenceScale for Children – Fifth Edition*. San Antonio: Pearson.

Yoakum, C. & Yerkes, R. M. (1920). *Amy Mental Tests*. New York: Holt and Company.

Yoakum, C. (1921). *Memoirs of the National Academy of Sciences, Volume XV Psychological Examining in the United States Army*. Washington: Goverment Printing Office.