

American Psychological Association's Apology to People of Color and Inaccurate IQ Tests

Jack A. Naglieri

Donna Y. Ford



drdonnayford@gmail.com www.drdonnayford.com
jnaglieri@gmail.com jacknaglieri.com NaglieriGiftedTests.com

The Topics for Today

A Fundamental Weakness of intelligence tests

Historical Context

The American Psychological Association Apology

How to Interpret Intelligence Tests

Closing remarks

The BIG picture

- Equitable Identification of gifted students is a critical issue
- Intelligence tests have played an important role in gifted identification and led to exclusion of students of color
- Understanding WHY we measure intelligence the way we do helps us understand what makes a test equitable
- It is important to differentiate test BIAS from test EQUITY
- Test EQUITY is about the CONTENT of the test questions
- Tests can be evaluated based on EQUITY
- The most equitable tests measure how well a student can THINK in a way that is not influenced by EXPOSURE; what they KNOW

3

Traditional IQ and Achievement Tests

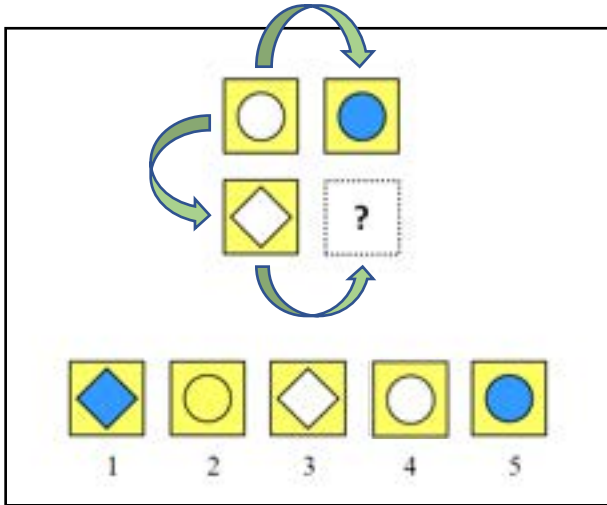
- Working as a school psychologist in 1975 noticed that some of the questions on the Wechsler intelligence tests were VERY similar to questions on the achievement tests (e.g., Vocabulary et al.,)
- It seemed wrong to measure 'intelligence' using questions that clearly demanded knowledge
- Shouldn't an intelligence test measure thinking rather than knowing?



1975 Charles Champagne Elementary, Bethpage, NY

4

Tests that Measure Thinking or Knowing?



Girl is woman as
boy is to _____?

3 is to 6 as
4 is to _____?

C⁷ is to F as
E⁷ is to _____?

Naglieri's Nonverbal Tests: 1985 to Present

- **Research on Six** Versions of the Naglieri Nonverbal Tests



MAT Short and Expanded Forms 1985 Naglieri Nonverbal Ability Test 1997 NNAT-Individual, 2003 NNAT-2 2008 NNAT3 2016

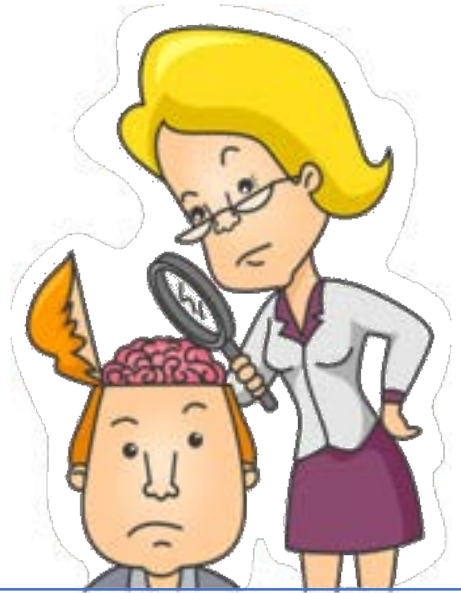
Each of these versions of the NNAT showed similar scores by RACE, ETHNICITY, & SEX and had strong correlation with achievement

This research convinced me that measuring intelligence using test questions that measured how well a student can think was a valid and equitable way to measure general intelligence 'g'.

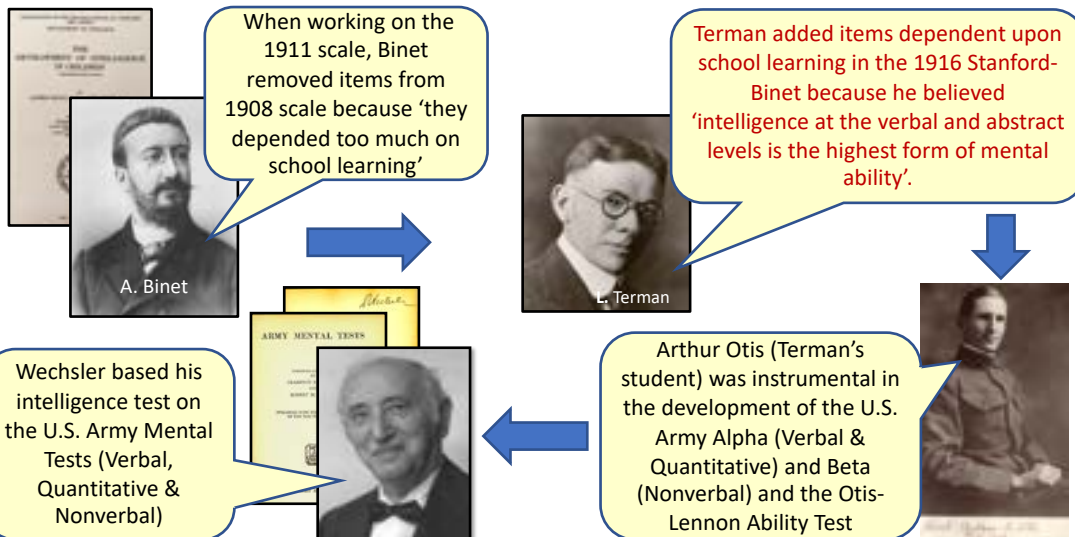
Two Questions:

1. Why do we measure ability the way we do?
2. Do the tests measure thinking or knowing?

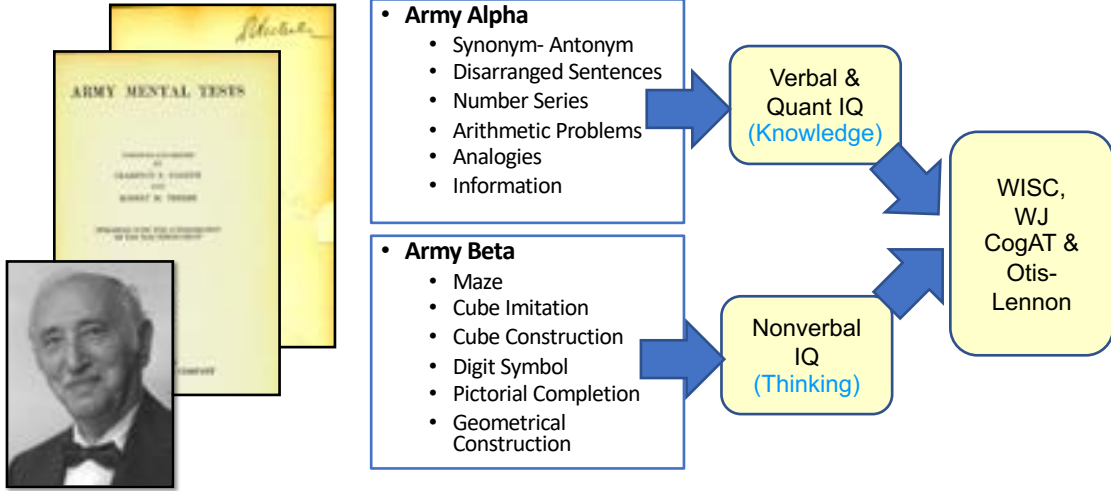
The early history of IQ tests provides the answers



Binet → Stanford-Binet → Army Mental Tests → WISC, CogAT, Olsat



Alpha & Beta → Wechsler Included Knowledge



Very Similar Items on "Different" Tests

Woodcock-Johnson Cognitive & Achievement Tests (CHC)

This block contains four screenshots of test items. Yellow arrows connect similar items across the different tests. The items are:

- Cognitive: Oral Vocabulary Subtest 1**: Item A asks for a synonym for 'big' (correct: large, gigantic, huge).
- Achievement: Reading Vocabulary-Synonyms Subtest 17**: Item A asks for a synonym for 'large' (correct: big, enormous, gigantic, huge).
- Test 17B Reading Vocabulary-Antonyms**: Item 1 asks for the opposite of 'big' (correct: tiny).
- Test 1C Verbal Comprehension-Antonyms**: Item 1 asks for the opposite of 'big' (correct: tiny, minuscule, minute).

 The arrows show that the 'big' item in the Oral Vocabulary test is similar to the 'large' item in the Reading Vocabulary test, and the 'big' item in the Reading Vocabulary test is similar to the 'big' item in the Verbal Comprehension test.

Including *Knowledge* in “Ability” Tests & Equity

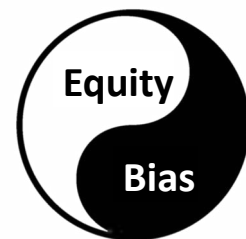
Stanford-Binet-5	WISC-V	WJ-IV	KABC-II	OLSAT	CogAT
<ul style="list-style-type: none"> • Verbal • Knowledge • Quantitative Reasoning • Vocabulary • Verbal Analogies 	<ul style="list-style-type: none"> • Verbal Comprehension • Vocabulary, Similarities, Information & Comprehension • Fluid Reasoning • Figure Weights, Arithmetic 	<ul style="list-style-type: none"> • Comprehension Knowledge: Vocabulary & General Information • Fluid Reasoning: Number Series & Concept Formation • Auditory Processing: Phonological Processing 	<ul style="list-style-type: none"> • Knowledge / GC • Riddles, Expressive Vocabulary, Verbal Knowledge 	<ul style="list-style-type: none"> • Verbal • Following directions • Verbal Reasoning • Quantitative • Verbal Arithmetic Reasoning 	<ul style="list-style-type: none"> • Verbal Scale • Analogies • Sentence Completion • Verbal Classification • Quantitative • 45 pages of oral instructions

Test Content, Test Bias and Test Equity

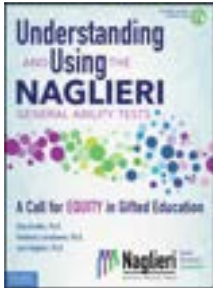
According to the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) Psychometric TEST BIAS and EQUITY are two different ways of measuring test fairness.



- ... if a person has had limited opportunities to learn the content in a test of intelligence, *that test may be considered unfair* ... even if there is no evidence of psychometric test bias.
- Evidence of EQUITY is examined by test content and mean score differences



Race and Ethnic Average Score Differences by Ability Test



Traditional tests that include knowledge and 2nd-Generation Ability Tests that minimize knowing

See Brulles, D., Lansdowne, K. & Naglieri, J. A. (2022). Understanding and Using the Naglieri General Ability Tests: A Call to Equity in Gifted Education. Minneapolis, MN: Free Spirit Publishing for more details.

Note: Even though a test may not show psychometric bias those tests with academic content that show large mean score differences are not equitable and are unfair.

	By Race	By Ethnicity
Tests that require knowledge		
Otto-Lennon School Ability Test (district wide)	Mn = 9.5	Mn = 5.2
Stanford-Binet IV (normative sample)	13.6	
WISC-V (normative sample)	12.6	
WI- HI (normative sample)	11.6	
WI- HI (normative sample)	10.9	10.7
CogAT7 (Nonverbal scale)	11.8	7.6
CogAT7 - Verbal	6.6	5.3
CogAT7-Quantitative	5.6	3.6
CogAT- Nonverbal	6.4	2.9
CogAT-Total (V, Q & NV)	7.0	4.5
WISC-V (statistical controls normative sample)	8.7	
Tests that require minimal knowledge		
K-ABC (normative sample)	Mn = 4.3	Mn = 2.9
K-ABC (matched samples)	7.0	
KABC-II (adjusted for gender & SES)	6.1	
KABC-II (adjusted for gender & SES)	6.7	5.4
CAS-2 (normative sample)	6.3	4.5
CAS (statistical controls normative sample)	4.8	4.8
CAS-2 (statistical controls normative sample)	4.3	3.8
CAS-2 Brief (normative samples)	2.0	2.8
NNAT (matched samples)	4.2	2.8
Naglieri General Ability Test-Verbal	2.2	1.6
Naglieri General Ability Test-Nonverbal	1.0	1.1
Naglieri General Ability Test-Quantitative	3.2	1.3

Note: The results summarized here were reported for the Otto-Lennon School Ability Test by Asant and O'Neal (2000), Stanford-Binet IV by Weissenberg (2000), Woodcock-Johnson III race differences by Edwards and Gelfand (2006) and ethnic differences by Smith-Sonuga-Barke, Smith, Thomas, and Gupta (2013), CogAT7 by Carlson, Mather and Barlock (2016) and Carlson (2016), WISC-V by Kaufman, Kaufman, and Coates (2016), Kaufman Assessment Battery for Children-II by Kaufman, Taylor, Kaufman & Kaufman (2016), CAS by Naglieri, Naglieri, Smith, and Naglieri (2016), CAS-2 and CAS-2 Brief by Naglieri, Day, and Robinson (2014) and (2016), Naglieri Nonverbal Ability Test by Naglieri and Manning (2000), and Naglieri General Ability Tests by Naglieri, Brulles, and Lansdowne (2022).

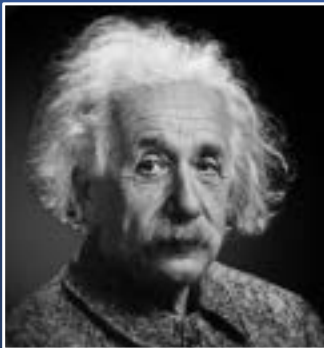


The test you choose determines the results you receive, the decisions you make, and the future of that student.

That is the *Practical Impact* of test selection

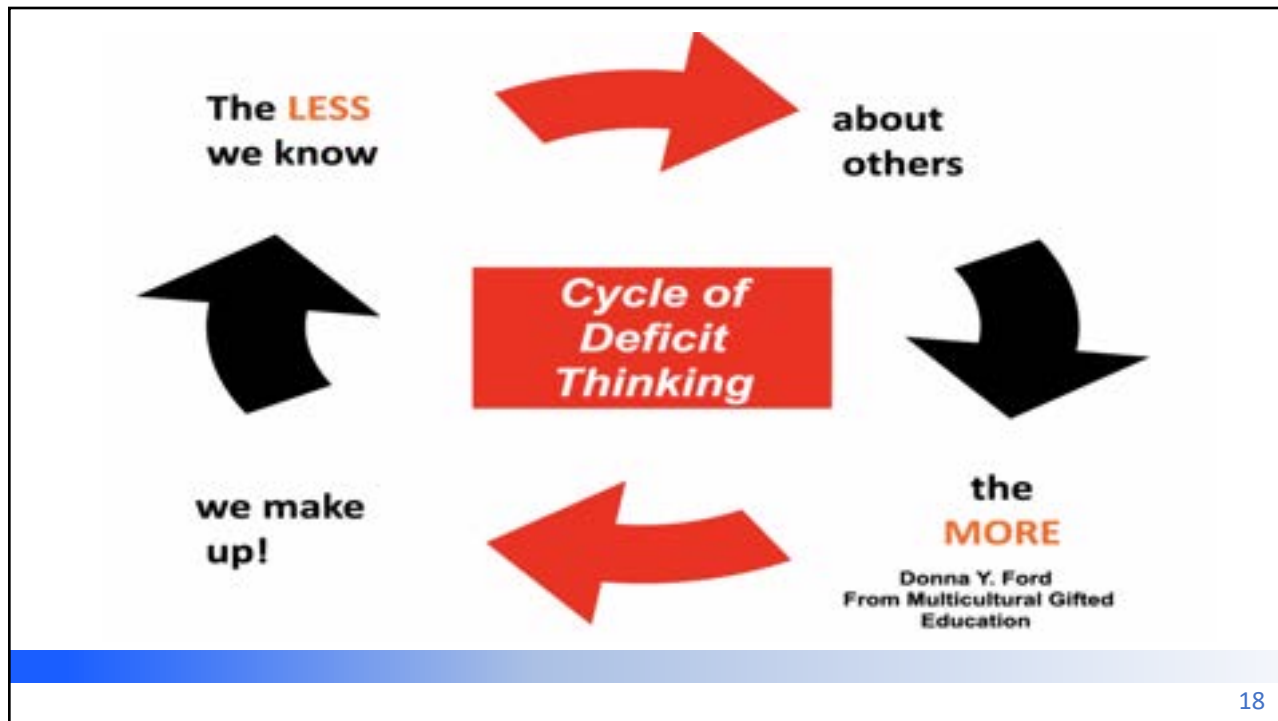
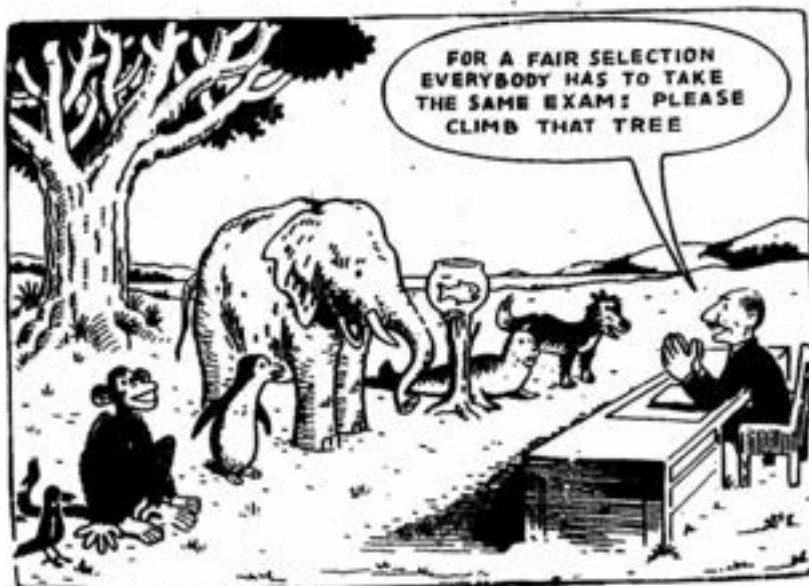
Are There Any Questions or Thoughts?





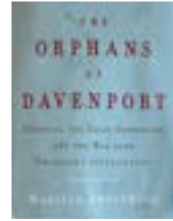
If you ask a fish to climb a tree, it will spend its entire life thinking it is stupid.

-Albert Einstein



IQ Tests Role in Promoting Racism

- Lewis Terman – promoter of eugenics (Greek for good birth) and Stanford-Binet (1916) author wrote that his test would reveal “significant racial differences in general intelligence...which cannot be wiped out by any culture”
- He advocated that identification of low-intelligence children and adults who would be involuntarily institutionalized and *sterilized would improve society*. (p. 68, Brookwood, 2021)
- His emphasis on VERBAL as the highest form of intelligence distorted the evaluation of intelligence for countless numbers of people



19

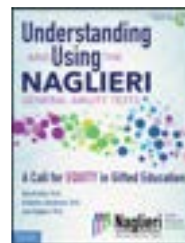
- ‘APA recognizes the roles of psychology in promoting...racism, and the harms that have been inflicted on communities of color ... and the ways measurement of intelligence has been systemically used to create the ideology of White supremacy’
 - Throughout the 1900s prominent psychologists involved in IQ test development supported eugenics
 - In 1916 Lewis Terman *Stanford-Binet* author advocated an educational system which separated white children from Blacks, Mexicans and Native Americans
 - 1933 Raymond Cattell (CHC & WJ) spoke out against race mixing and he lobbied to overturn the 1954 Brown v. Board Education
- What impact has this had on identification of GIFTED STUDENTS?



20

Numbers of Gifted Students Missed = 1,235,434

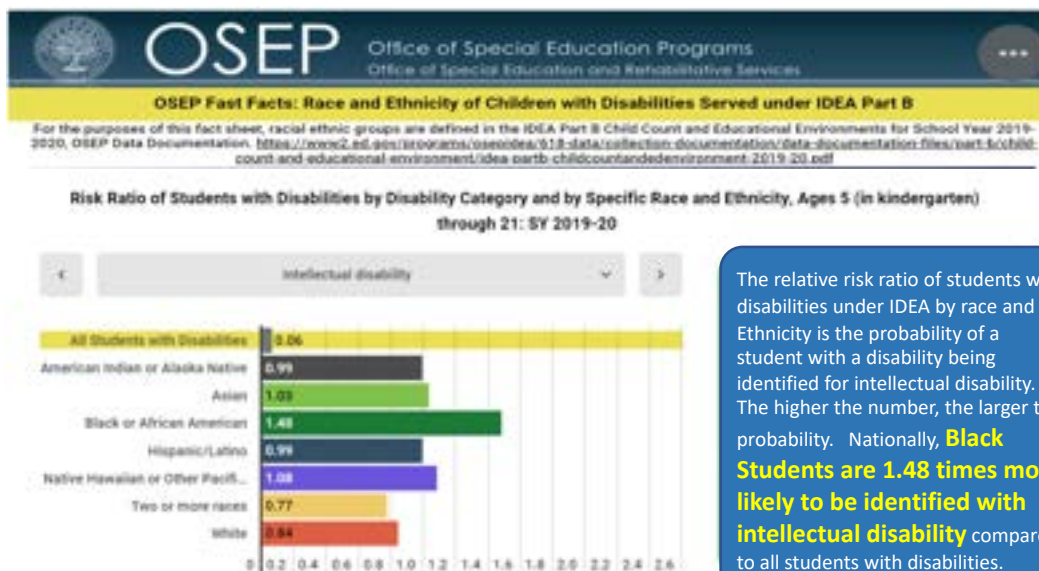
	N in Public Education K-12 in 2020	N Potentially Gifted (8%; 92 %tile)	N Students in gifted programs	Difference Between Potential and Identified
White	23,834,458	1,906,757	1,937,350	30,593
Black	7,754,506	620,360	330,774	-289,586
Hispanic	14,337,467	1,146,997	600,498	-546,499
Native American/ Alaska Native	484,766	38,781	27,712	-11,069
Two or More Races	1,641,817	131,345	105,371	-25,974
Total Non-Whites	24,218,556	1,937,484	1,064,355	-873,129



873,129 +



Percent of Schools that do not Identify	41.5%
Additional non-white gifted students = 41.5% of 873,129	N = 362,305
Total non-white gifted students missed	N = 1,235,434



The relative risk ratio of students with disabilities under IDEA by race and Ethnicity is the probability of a student with a disability being identified for intellectual disability. The higher the number, the larger the probability. Nationally, **Black Students are 1.48 times more likely to be identified with intellectual disability** compared to all students with disabilities.

<https://sites.ed.gov/idea/osep-fast-facts-race-and-ethnicity-of-children-with-disabilities-served-under-idea-part-b/>
https://ldaamerica.org/lda_today/disproportionate-identification-of-students-of-color-in-special-education/

Bridging Two Fields

The diagram features two overlapping circles. The left circle is orange and labeled "Urban/ Multicultural Education". The right circle is black and labeled "Gifted/AP Education". An arrow points from the intersection of the two circles down to a white box with a black border. Inside this box, the text reads: "Needs and Development", "Academic and Cognitive", "Affective and Psychological", and "Social and Cultural".

D.Y. Ford

23

Systemic... Achievement Gap 5

The diagram is set against a dark red background. At the top, a dark grey bar contains the text "Systemic..." on the left and "Achievement Gap" in a white box in the center. To the right of this bar is an orange square with the number "5". Three orange arrows point downwards from the "Achievement Gap" box to three white boxes below. The first box on the left is labeled "SPECIAL EDUCATION Over-Representation". The middle box is labeled "DISCIPLINE Over-Representation". The box on the right is labeled "GIFTED EDUCATION & AP Under-Representation".

24

Academic Learning Loss & COVID

- COVID-19 has increased the impact of disparities in access and opportunity for students of color and they are even further behind than they were before.
- Their **scores on traditional intelligence tests** which demand knowledge **are even more inaccurate.**
- **Solutions:**
 - For traditional tests, use post-COVID norms only.
 - Use intelligence tests that are not dependent upon knowledge



Education in a Pandemic: The Disparate Impacts of COVID-19 on America's Students. US Dept. of Ed- Office of Civil Rights. June, 21, 2021. <https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.p>

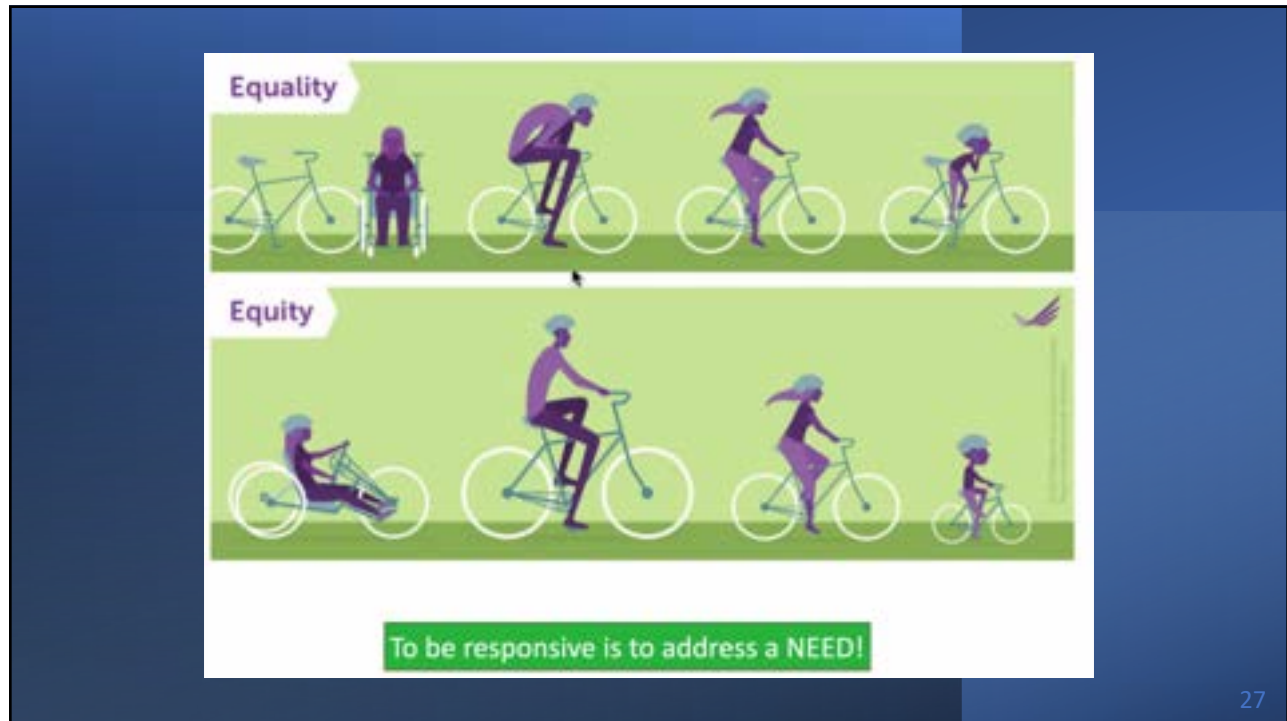
25



Psychologists who studied race and ethnic differences attributed IQ test results to the **people** instead of the **tests**

That is the Practical Impact of flawed intelligence tests

26



27

Tests with Equity as a Goal 1985-Present

Traditional Tests

1. Naglieri, J. A. (1985). *Matrix Analogies Test - Expanded Form*. San Antonio: The Psychological Corporation.
2. Naglieri, J. A. (1985). *Matrix Analogies Test - Short Form*. San Antonio: The Psychological Corporation.
3. Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test*. San Antonio, TX: The Psychological Corporation.
4. Naglieri, J. A., & Bardos, A. N. (1997). *General Ability Scale for Adults*. San Antonio, TX: Pearson.
5. Naglieri, J. A. (2003). *Naglieri Nonverbal Ability Test - Individual Form*. San Antonio, TX: Pearson.
6. Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX: Pearson.
7. Naglieri, J. A. (2008). *Naglieri Nonverbal Ability Test - 2nd Edition*. San Antonio, TX: Pearson.
8. Naglieri, J. A. (2016). *Naglieri Nonverbal Ability Test - Third Edition*. San Antonio, TX: Pearson.

Second Generation

9. Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Austin: ProEd
10. Naglieri, J. A., Das, J. P., Goldstein, S. (2014). *Cognitive Assessment System Second Edition*. Austin, ProEd.
11. Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System Second Edition - Brief*. Austin, ProEd.
12. Naglieri, J. A., Moreno, M. A., & Otero, T. M. (2017). *Cognitive Assessment System - Español*. Austin, ProEd.
13. Naglieri, J. A. (2022). *Naglieri General Ability Test: Nonverbal*. Markham, Canada: MHS.
14. Naglieri, J. A. & Brulles, D. (2022). *Naglieri Ability Test: Verbal*. Markham, Canada: MHS.
15. Naglieri, J. A. & Lansdowne, K. (2022). *Naglieri Ability Test: Quantitative*. Markham, Canada: MHS.

28

Can a Traditional Intelligence Test of General Ability be Equitable?

Measure 'Thinking' with minimal influence of 'Knowing'

The Naglieri General Ability Tests: Verbal, Nonverbal and Quantitative

VERBAL - Dina Brulles, Ph.D. dbrulles@gmail.com

NONVEBAL - Jack A. Naglieri, Ph.D. jnaglieri@gmail.com

QUANTITATIVE - Kim Lansdowne, Ph.D. Kimberly.Lansdowne@asu.edu



29

Naglieri General Ability Tests

Jack A. Naglieri, Dina Brulles & Kimerly Lansdowne (2022)



- We **explicitly made tests for equitable identification** of students from diverse cultural, linguistic, or socioeconomic backgrounds using the traditional Verbal, Nonverbal and Quantitative formats to **measure general ability:**

- Animated instructions remove the need for verbal comprehension of directions,
- Test questions that do not require academic knowledge,
- Verbal and Quantitative test questions that can be solved using any language,
- A multiple-choice response removes the need for verbal expression.



30

Naglieri General Ability Tests

Three tests of general ability that measure how well a student can **think** to arrive at the answer rather than what they **know**.



Three Research Studies (2022)

Selvamenan, M., Paolozza, A., Solomon, J., Naglieri, J. A., & Schmidt, M. T. (submitted for publication, 2022). Race, Ethnic, Gender, and Parental Education Level Differences on Verbal, Nonverbal, and Quantitative Naglieri General Ability Tests: Achieving Equity.

- **VERBAL SAMPLE**
 - 2,482 That closely matches the US population on key demographics
- **GENDER**
 - **No differences** between **males** and **females** for raw score across all forms
- **RACE/ETHNICITY**
 - **No differences** among **White, Black, & Hispanic** for raw score across all forms
- **PARENTAL EDUCATION LEVEL**
 - **No differences** among five education levels (**No high school diploma; High School graduate; Some college/Associate's degree; Bachelor's degree; Graduate/professional degree**) for raw score across all forms

- **NONVERBAL SAMPLE**
 - 3,630 That closely matches the US population on key demographics
- **GENDER**
 - **No differences** between **males** and **females** for raw score across all forms
- **RACE/ETHNICITY**
 - **No differences** among **White, Black, & Hispanic** for raw score across all forms
- **PARENTAL EDUCATION LEVEL**
 - **No differences** among five education levels (**No high school diploma; High School graduate; Some college/Associate's degree; Bachelor's degree; Graduate/professional degree**) for raw score across all forms

- **QUANTITATIVE SAMPLE**
 - 2,841 That closely matches the US population on key demographics
- **GENDER**
 - **No differences** between **males** and **females** for raw score across all forms
- **RACE/ETHNICITY**
 - **No differences** among **White, Black, & Hispanic** for raw score across all forms
- **PARENTAL EDUCATION LEVEL**
 - **No differences** among five education levels (**No high school diploma; High School graduate; Some college/Associate's degree; Bachelor's degree; Graduate/professional degree**) for raw score across all forms

General Ability Tests



"we did not start with a clear definition of general intelligence... [but] borrowed from every-day life a vague term implying all-round ability and... we [are] still attempting to define it more sharply and endow it with a stricter scientific connotation (Pintner, 1923 p. 53)".



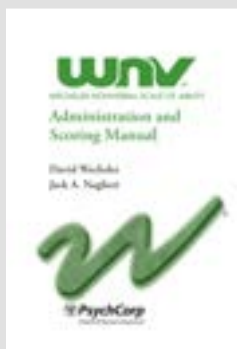
"The aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment (1939)"

General Ability
not verbal or nonverbal intelligences !



33

The emphasis in the *WNV Manual* that the Full Scale measures *general ability nonverbally*—and *not nonverbal ability*—ties the WNV to Dr. Wechsler



Dr. Wechsler remained a firm believer in Spearman's g theory ... 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.

He saw the Performance Scale as the most sensible way to measure the general intelligence of people with ... limited proficiency in English.

Quotes from Alan S. Kaufman in the Wechsler Nonverbal Manual; Wechsler & Naglieri (2006)

34



Support for 'g'



- ...The small portions of variance uniquely captured by [subtests]... render the group factors [scales] of questionable interpretive value independent of g (FSIQ general intelligence)
- Present CFA results confirm the EFA results (Canivez, Watkins, & Dombrowski, 2015); Dombrowski, Canivez, Watkins, & Beaujean (2015); and Canivez, Dombrowski, & Watkins (2015).

- The results of this study indicate that most **cognitive abilities specified in John Carroll's three-stratum theory have little-to-no interpretive relevance above and beyond that of general intelligence.**

35

Research Supports 'g' but little More

Watkins, M. W., & Canivez, G. L. (2021). Assessing the psychometric utility of IQ scores: A tutorial using the Wechsler intelligence scale for children—fifth edition. *School Psychology Review*, 1-15.

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.

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. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales-Fifth Edition for children and adolescents. *School Psychology Quarterly*, 23, 533–541.

Dombrowski, S. C., Canivez, G. L., & Watkins, M. W. (2017, May). Factor structure of the 10 WISC–V primary subtests across four standardization age groups. *Contemporary School Psychology*. Advance online publication.

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.

McGill, R. J., & Canivez, G. L. (2017, October). Confirmatory factor analyses of the WISC–IV Spanish core and supplemental Subtests: Validation evidence of the Wechsler and CHC models. *International Journal of School and Educational Psychology*. Advance online publication.

Watkins, M. W., Dombrowski, S. C., & Canivez, G. L. (2017, October). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children—Fifth Edition. *International Journal of School and Educational Psychology*.

36




What is the Practical Impact?

Verbal, Nonverbal, and Quantitative scales are NOT different types of intelligence and do NOT reflect different ways of learning

37







Don't just tell a different version of the same story.
Change The Story!




EQUALITY

EQUITY

LIBERATION

38

NAGC Professional Standards



2.2. Identification. Students with gifts and talents are identified for services that match their interests, strengths, and needs.

2.2.5. Educators select assessments that minimize bias by including information in the technical manual that describes content in terms of potential bias, includes norms that match national census information or local populations, shows how items discriminate equally well for each group, and provides separate reliability and validity information for each group.

2.3. Identification. Students with identified gifts and talents represent diverse backgrounds.


2.3.1. Educators select and use equitable approaches and assessments that minimize bias for referring and identifying students with gifts and talents, attending to segments of the population that are frequently hidden or underidentified. Approaches and tools may include front-loading talent development activities, universal screening, using locally developed norms, assuring assessment tools are in the child's preferred language for communication, or nonverbal formats.

39

Summary: Equitable Assessment of Intelligence

- **Equitable evaluation of intelligence** demands test questions that can be solved regardless of the amount of academic knowledge and facility with language a student has
- We have shown that
 - General ability (*g*) **can be measured equitably** across Verbal, Quantitative and Nonverbal content if the tests do not require academic knowledge
- Verbal, Quantitative and Nonverbal are **a description of the content of the tests'** questions **NOT** different types of intelligence
- Equitable tests measure THINKING in a manner that is minimally influenced by KNOWING

40



We do the best we can with what we know, and when we know better, we do better.

— Maya Angelou —

Change Demands Courage to Think Differently

Socially just identification of gifted students requires self-reflection and self-correction in response to current research

41

WE CAN DO
BETTER

We Must do Better



The graphic is divided into two main sections. The left section has a pink background with the text "FINAL THOUGHTS!" in large, light blue, stylized letters. Below the text is the word "dreamstime" in a smaller, lighter font. At the bottom of this section is a row of ten colorful, stylized hands of various colors (black, green, blue, purple, grey, green, grey, purple, grey, blue) raised in a gesture of participation or agreement. The right section has a black background with a large, glowing red question mark in the center. A blue horizontal bar is located at the bottom of the graphic area.