

**Twelve Aspirations:
Objectives for Development of Tests
Intended for High Ability/Gifted Individuals**

**Exploring Ideal Elements of Tests of Ability
for Gifted Students**

John D. Wasserman

**National Association for Gifted Children
Saturday, November 13, 2010, Atlanta, Georgia**

About 15 Years Ago ...

- ❑ ... the head of psychological test development at The Psychological Corporation said, “We develop tests backwards. First we standardize and norm tests, and then we research how they work in special populations. We should do it the other way.”
- ❑ **These test development psychometric aspirations propose that test development should proceed with giftedness carefully studied based on data researched as a forethought rather than an afterthought.**

Norms and Sampling: Aspiration 1

Develop High Ability Norms

- ❑ Intelligence tests (with $M=100$, $SD=15$) typically base their upper range norms ($IQ \geq 130$) on extrapolation with the assumption of a normal distribution, with only 2% of the sample earning these high scores.
- ❑ **Test developers should explore methodologies to define high ability performance using actual representative samples from individuals known to be intellectually gifted through some independent criteria.**
 - ❑ Norm (and weight) the high end independently, and then link to the main part of the distribution?
 - ❑ Rasch Scaling with High Ability Samples?
 - ❑ Special Gifted Reference Group Norms?
 - ❑ A variety of methods need to be researched ...

Norms and Sampling: Aspiration 2

Raise Test and Subtest Ceilings

- ❑ The WISC-IV has broken its ceiling subtest scaled score of 19 (+3 SD) and composite score of 160 (+4 SD) (Zhu, Cayton, Weiss, & Gabel, 2008)**
- ❑ The SB5 EXIQ is intended for FSIQ > 150**
- ❑ *The Woodcock-Johnson III NU Cog and Ach scores extend to 200 or more***
- ❑ There is no longer any psychometric reason why test score ceilings cannot be extended using techniques from item response theory.**

Norms and Sampling: Aspiration 3

Calibrate Items on High Ability Samples

- ❑ *Calibrate the most difficult subtest and test items on high ability samples.***
- ❑ **The psychometric advances needed to raise test ceilings can readily be accomplished through calibration of test items on highly gifted samples or chronologically older samples (who therefore have higher raw ability as a function of their age).****

Norms and Sampling: Aspiration 4

Ensure Adequate Difficulty Gradients

- ❑ Test (and subtest) difficulty gradients should be sufficiently fine and gradual so as to facilitate identification at varying degrees (or gradations) of intellectual giftedness.
- ❑ **A minimum of three items per Rasch logit or standard deviation equivalent has been suggested as a minimally adequate gradient.**
- ❑ Gaps in item difficulties may lead to larger-than-necessary jumps in estimated abilities.

Test Score Validity: Aspiration 5

Discriminate Levels of Giftedness

- ❑ Tests should provide evidence supporting the discriminate validity of the test in differentiating between gifted and nongifted samples, as well as between different levels of giftedness.
- ❑ ... assuming that students with different levels of intellectual giftedness have some different educational needs

Test Score Validity: Aspiration 6

Discriminate Types of Giftedness

- Provide evidence of discriminative validity if the test purports to differentiate between different types of giftedness.***
- Test (and subtest) capacities to identify different types of giftedness (i.e., mathematically precocious, verbally advanced, spatially gifted) should be documented, in conjunction with evidence of validity for any such claim.**

Test Score Validity: Aspiration 7

Validity of Discontinue Rules

- Test developers should document evidence supporting the formulation of subtest and test discontinue rules with gifted samples.***
- Test discontinue rules should be independently validated for gifted samples, who are thought by some to show greater scatter in abilities, answering very difficult items even after having missed many easier items.**

Test Score Validity: Aspiration 8

Academic Consequential Validity

- Provide evidence of consequential validity by meaningfully linking test performance with success in a gifted curriculum.*
- Are students identified by specific criteria likely to fare successfully in gifted education curriculums?
- Example: *Do nonverbal tests identified students who succeed in highly verbal gifted curriculums?*

Test Score Reliability: Aspiration 9

Reliability with Gifted Samples

- Provide evidence of classical forms of reliability with high ability samples.*
- Classical measures of test score reliability, including internal scale consistency, temporal test-retest stability, and interscorer reliability, should be independently computed and verified for samples of gifted individuals.**

Test Score Reliability: Aspiration 10

Gifted Decision-Making Reliability

- Provide evidence of decision-making consistency for high ability samples.***
- Decision-making consistency, or the degree to which tests consistently identify individuals falling in a given category (such as gifted), may be helpful in studying the stability of gifted classification systems.**
- Classification of other exceptionalities, such as specific learning disabilities, have not proven to be adequately stable after testing with a two to three year intervening interval.***

Test Score Fairness: Aspiration 11

Fairness in High Ability Samples

- *Provide evidence of test fairness using traditional techniques with high ability samples.***
- **Traditional test fairness and bias techniques (e.g., bias review panels, differential item function, differential prediction of achievement) need to be extended to high ability samples.****

Test Score Fairness: Aspiration 12

Prop. Identification of Minorities

- States are mandated by the Office of Civil Rights to show proportional identification of minorities for gifted and talented placements.***
- Evidence needs to be presented as to the use of intelligence tests in identifying similar proportions of minority or culturally and linguistically diverse individuals as census findings would predict.**

Twelve Aspirations

Concluding Comments

- ❑ Tests intended for use with gifted student identification should provide rigorous psychometric evidence pertaining to that use – beyond the samples of convenience found in the WISC-IV ($n=63$) and Stanford-Binet 5 ($n=96$).
- ❑ *Without further evidence to support their validity, reliability, and fairness, intelligence tests are at risk for being replaced by RTI in the assessment of academic giftedness.*