

ASSESSING INTELLIGENCE

Chapter 11 (Myers 7th Ed.)



What is the difference between an aptitude test and an achievement test? What does it mean for such tests to be standardized, reliable, and valid? How do we assess intelligence? Movie hero Forrest Gump's answer- "Stupid is as stupid does" catches the spirit of psychology's simplest answer: Intelligence is as intelligence does on an IQ test. In other words, intelligence is whatever intelligence tests measure. So what are these tests and what makes a test credible?

A) Modern Tests of Mental Abilities (p 432)

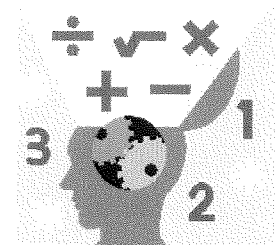
Psychologists classify tests as either :

- 1) **APTITUDE TESTS** (intended to *predict* your ability to learn a new skill)
- 2) **ACHEIVEMENT TESTS** (intended to *reflect* what you have learned)

The most widely used intelligence test, the **Wechsler Adult Intelligence Scale (WAIS)**, was created by psychologists David Weschler, who, as a 6 year old Romanian boy, was among the Eastern European immigrants who came over in the early 1900's and was considered 'feeble minded'. He later developed a test for children **Weschler Intelligence Scale for Children (WISC)**.

The WAIS consists of 11 sub-tests. It yields not only an overall intelligence score, as does the Stanford-Binet test, but also separate verbal and performance (nonverbal) scores. (see page 433 of text for test example-Fig. 11.2)

Striking differences between the two scores alert the Examiner to possible learning problems. The tests also provide clues to cognition strengths that a teacher can build upon.



B) Principles of Test Construction (p.434)

To be widely accepted, psychological tests must meet three criteria:

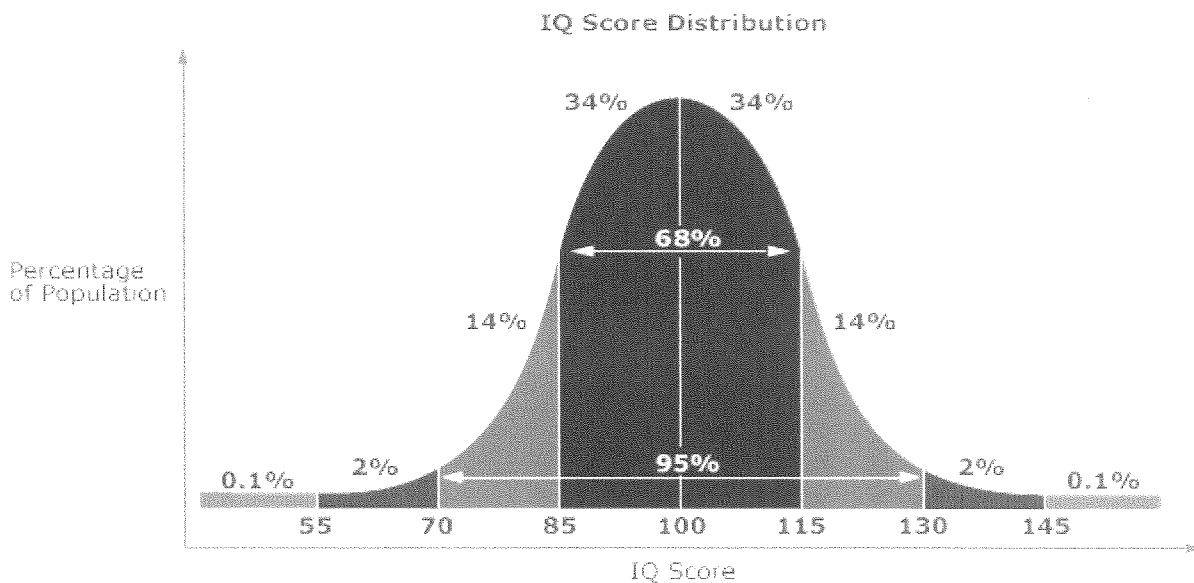
- 1) They must be **standardized**
- 2) They must be **reliable**
- 3) They must be **valid**

STANDARDIZATION:

Defining meaningful scores by comparison with the performance of a pretested 'standardized group'

The number of question you answered correctly on an intelligence test would tell us almost nothing. To evaluate your performance, we need a basis for comparing it with others' performance. To enable meaningful comparisons, test-makers first give the test to a representative sample of people. When others take the test following the same procedures, their scores can be compared with the standards defined by the sample.

Standardized test results typically form a normal distribution, a bell-shaped pattern of scores that forms the normal curve.



No matter what we measure- height, weight or mental aptitudes- scores often form a roughly symmetrical, bell shaped distribution clustered around the average. On an intelligence test, we call this average score 100. As we move out from the average, we find fewer and fewer people.

(The Flynn Effect) p.435- is there a problem with the tests or is intelligence changeable

RELIABILITY

The extent to which a test yields consistent results, as assessed by the consistency of scores on two halves of the test, on alternate forms of the test, or on retesting.

A good test must yield dependably consistent scores. To check a test's reliability, researchers retest people using either the same test or another form of it. If the two scores generally agree, or CORRELATE, the test is reliable. As an alternative, the researcher may split a test in half and see whether scores derived from odd and even questions agree.

The higher the correlation between the test-retest or the split-half scores, the higher the test's reliability. The tests we have seen (*The Stanford-Binet*, *WAIS*, *WISC*) all have reliabilities of about **+0.9** which is very high. (which means, when they are re-tested, people's scores are generally the same as their first test)



VALIDITY

Validity is the extent to which the test actually measures what it is supposed to measure or predicts what it is supposed to predict. If you use an inaccurate tape measure to measure your height, your height report would have a high reliability (consistency) but low validity. It is enough for some test that they have **content validity** (the extent to which a test samples the behaviour that is of interest ie: like a driving test asking driving questions)

- **Criterion**
- **Predictive Validity**

Summing Up:

- Aptitude tests tend to be highly reliable, but are weak predictors of success in life.
- However, their predictive validity for academic success in the early grades is very strong
- Test validity weakens for predicting grades in college as the range of student abilities becomes more restricted.

``**Almost all of the joyful things of life are outside the measure of IQ tests.**``

- Madeleine L'Engle 1972