What are Standardized Tests?

Standardized Tests
- defining meaningful scores by comparison with the performance of a pre-tested “standardization sample”.

Issues related to “standardization sample”
- Population v. sample
- Size
- Representativeness
What are Standardized Tests?

- **Norm-Referenced Tests**
  - compare each test taker's scores with the performance of all the test takers.

- **Criterion-Referenced Tests**
  - measure a student’s performance relative to what the student should know, rather than to the performance of other students.

Types of Standardized Tests

- **Differential Aptitude Tests**
  - designed to measure such aptitudes as clerical speed and accuracy, mechanical reasoning, space relations, spelling and language usage.

- **Vocational Interest Test**
  - designed to help students decide where their vocational interests lie.

- **Achievement Tests**
  - designed to measure accomplishments in either single or multiple areas of endeavor, such as reading comprehension, mathematics, social studies, and science.
Example of Items on a Typical Achievement

<table>
<thead>
<tr>
<th>Content Area</th>
<th>Explanation of Task/Question</th>
<th>Example Task/Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence Completion</td>
<td>Show knowledge of meanings of words and their uses and relations.</td>
<td>Choose the word or set of words that best fits into a whole sentence, such as: “The most common use for a _____, of course, is to lock and unlock doors.” (Correct answer choice would be “key”)</td>
</tr>
<tr>
<td>Verbal Analogies</td>
<td></td>
<td>Choose the pair of words with the same relationship as a given pair (analogies). For example: plumber: sink = mechanic: engine</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>Demonstrate understanding of a text passage.</td>
<td>Correctly answer multiple-choice questions about a text passage, such as “What did the main character of this story hope to achieve?”</td>
</tr>
<tr>
<td>Quantitative Skills</td>
<td>Make calculations involving geometry, algebra, fractions, arithmetic, exponential numbers, etc.</td>
<td>Choose the correct answer from several possible answers in a multiple-choice format. For example, after viewing a picture of a polygon with some angles and sides labeled and others left blank, students might be asked to identify the correct number for one of the missing side measurements.</td>
</tr>
</tbody>
</table>

Intelligence Tests

- **Intelligence Test**
  - a method of assessing an individual’s mental aptitudes and comparing them to those of others, using numerical scores
  - First developed by Binet & Simon in France (≈1905)
Origins of Intelligence Testing

• Mental Age
  • a measure of intelligence test performance devised by Binet
  • chronological age that most typically corresponds to a given level of performance
  • child who does as well as the average 8-year-old is said to have a mental age of 8

Origins of IQ

• Intelligence Quotient (IQ)
  • defined originally the ratio of mental age (ma) to chronological age (ca) multiplied by 100
    • IQ = ma/ca x 100
  • on contemporary tests it is the average performance for a given age is assigned a score of 100
    • Based on normal distribution
The Normal Curve

Ninety-five percent of all people fall within 30 points of 100

Sixty-eight percent of people score within 15 points above or below 100

Tests of Intelligence

- **Standford-Binet**
  - Binet-Simon Scale
  - "IQ"
  - Ages 2-23
  - Four Subscales
    - Verbal
    - Abstract/Visual
    - Quantitative
    - STM

- **Weschler Scales**
  - WAIS-R (Adult)
  - WISC-III (5-16)
  - WPPSI (3-7)

- **Two Subscales**
  - Verbal
  - Performance
Assessing Intelligence

- **Wechsler Adult Intelligence Scale (WAIS)**
  - most widely used intelligence test
  - subtests
    - verbal
    - performance (nonverbal)

Assessing Intelligence - Sample Items from the WAIS

<table>
<thead>
<tr>
<th>VERBAL</th>
<th>PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Information</td>
<td>Picture Completion</td>
</tr>
<tr>
<td>Similarities</td>
<td>Picture Arrangement</td>
</tr>
<tr>
<td>Arithmetic Reasoning</td>
<td>Block Design</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Object Assembly</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Digit-Symbol Substitution</td>
</tr>
<tr>
<td>Digit Span</td>
<td></td>
</tr>
</tbody>
</table>

From Thorndike and Hagen, 1977
Evaluating Tests

**Reliability**
- the extent to which a test yields consistent results
- assessed by consistency of scores on:
  - “split-half”: comparing two halves of the test
  - “alternate forms”: comparing two forms of the test
  - “test-retest”: retesting the same individual

**Evaluating Tests**

**Validity**
- the extent to which a test measures what it is claims to measure

- **Content Validity**
  - the extent to which a test samples the behavior that is of interest
    - driving test that samples driving tasks

- **Construct Validity**
  - the extent to which a test completely and accurately captures the theoretical construct or attribute it is designed to measure.
Evaluating Tests

- **Criterion-Related Validity**
  - **Concurrent Validity**
    - The extent to which a test is correlated with another theoretically similar test
  - **Predictive Validity**
    - success with which a test predicts the behavior it is designed to predict
    - assessed by computing the correlation between test scores and the criterion behavior

Statistical Concepts

- **True Score**
  - the hypothetical score someone would get if he or she took a test an infinite number of times with no practice effects.
- **Observed Score**
  - the score someone actually receives on a test.
- **Confidence Interval**
  - the likely range of observed scores within which a person's true score lies.
Statistical Concepts

- **Raw Scores**
  - total number of items correct

- **Percentile Scores**
  - the proportion of other students’ scores that equal or fall below a given student’s score, multiplied by 100.

- **Standard Scores**
  - derive from converting a row score into units of standard deviation.

- **Grade Equivalent Scores**
  - a measure of grade-level achievement compared to a norm.

Statistical Concepts

- **Frequency Distribution**
- **Measures of Central Tendency**
  - Mean
  - Median
  - Mode

- **Measures of Dispersion**
  - Range
  - Standard Deviation
  - Variance

- **The Normal Distribution**
A Normal Distribution

Issues and Concerns in Standardized Testing

- Test Bias
- Culture - Relevant Testing
- Culture - Fair Testing
- Culture - Free Testing
Misuses of Standardized Testing

- High-Stakes Testing
- Overuse of Tests in Measuring Accountability
- New Directions in Standardized Testing