Functions of Standardized Tests

• Student Assessment
  • Diagnosis
• Placement and Selection
  • Accountability
  • Predictive Validity

What is a standardized test?

• A test that has standard procedures for administration, scoring, and interpretation.
Advantages of Standardized Tests

• Evaluating students’ general educational development in the basic skills & in learning outcomes common to many courses of study
• Evaluating student progress during the school year or over a period of years
• Determining strengths & weaknesses

Weaknesses of Standardized Tests

• Evaluating the learning outcomes and content unique to a particular class or school
• Evaluating students’ day-to-day progress
• Evaluating knowledge of current developments in rapidly changing content areas such as science and social studies
Types of Achievement Tests

• **State Developed Tests**
  N.C. EOG’s, FL FCAT’s, VA’s SOL’s

• **Publisher-Developed Norm-referenced batteries**
  ITBS, CAT, GRE, Stanf. Ach. Test

• **Publisher-Developed Norm-referenced content area tests**
  Nelson-Denny, Gates-McGinitie

• **Publisher-Developed Criterion-referenced tests**

Some popular tests to be aware of:

• California Achievement Test (CAT)
• Iowa Tests of Basic Skills (ITBS)
• Metropolitan Achievement Tests (MAT)
• Stanford Achievement Tests
• TerraNova
Why do we use these?

- High technical quality
- Standard directions for administration and scoring
- Norms based upon national large national samples
- Equivalent forms
- Comprehensive manuals

Using Achievement Tests:

- Be wary of using subtests for diagnostic purposes unless enough items are included
- What is a norm group and what is the benefit of having a norm group?
- Norm groups provide a standard frame of reference
- Equivalent forms
Judging the Adequacy of Norms

• Should be relevant (Who do you want to compare your scores to?)
• Should be representative
• Should be up to date
• Should be comparable
• Should be adequately described

Achievement Batteries

• Consists of a series of individual tests all standardized on the same national sample
Where to find info on tests:

- **Mental Measurement Yearbook (MMY)**
- **MMY published by Buros Center for Testing**
  [http://www.unl.edu/buros/](http://www.unl.edu/buros/)

Some Aptitude/Intelligence tests to be aware of:

- **Wechsler Intelligence Tests (WISC, WAIS)**
- **Stanford-Binet**
- **Raven’s Advanced Progressive Matrices**
- **Cognitive Abilities Test (CogAT)**
- **Graduate Record Examination (GRE)**
- **Otis-Lennon School Ability Test (OLSAT)**
- **Cattell Culture-Fair Intelligence Tests**
- **Armed Services Vocational Aptitude Battery (ASVAB)**
- **Differential Aptitude Test (DAT)**
Aptitude Tests

• Do not measure fixed capacity but rather a different type of ability used to predict future performance
• Common distinction: achievement tests measure what a student has learned and that aptitude tests measure the ability to learn new tasks

Why use aptitude tests when you have achievement tests?

• Can be administered in a relatively short time
• Can be used with students of more widely varying educational backgrounds
• Can be used before any training or instruction
Specific theories of importance:

- Spearman vs. Thurstone
- Guilford’s 120 abilities
- Crystallized & Fluid intelligence
- Gardner’s multiple intelligences
- Others include David Perkins & Robert Sternberg

What is general intelligence?

- General ability typically measured via standardized tests--symbolized as $g$
- Predictive power strongest when facing novel tasks or beginning competence
- Considered to be reasoning ability (typically inductive) that is highly dependent upon working memory--making transformations in your head
Normal distributions (bell shaped) are a family of distributions that have the same general shape. They are symmetric (the left side is an exact mirror of the right side) with scores more concentrated in the middle than in the tails. Examples of normal distributions are shown to the right. Notice that they differ in how spread out they are. The area under each curve is the same.

If your data fits a normal distribution, approximately 68% of your subjects will fall within one standard deviation of the mean.

Approximately 95% of your subjects will fall within two standard deviations of the mean.

Over 99% of your subjects will fall within three standard deviations of the mean.
The mean and standard deviation are useful ways to describe a set of scores. If the scores are grouped closely together, they will have a smaller standard deviation than if they are spread farther apart.

When you have a subject's raw score, you can use the mean and standard deviation to calculate his or her standardized score if the distribution of scores is normal. Standardized scores are useful when comparing a student's performance across different tests, or when comparing students with each other.
The number of points that one standard deviation equals varies from distribution to distribution. On one math test, a standard deviation may be 7 points. If the mean were 45, then we would know that 68% of the students scored from 38 to 52.

On another test, a standard deviation may equal 5 points. If the mean were 45, then 68% of the students would score from 40 to 50 points.

Data do not always form a normal distribution. When most of the scores are high, the distributions is not normal, but negatively (left) skewed.

Skew refers to the tail of the distribution.

Because the tail is on the negative (left) side of the graph, the distribution has a negative (left) skew.
When most of the scores are low, the distributions is not normal, but positively (right) skewed.

Because the tail is on the positive (right) side of the graph, the distribution has a positive (right) skew.

When data are skewed, they do not possess the characteristics of the normal curve (distribution). For example, 68% of the subjects do not fall within one standard deviation above or below the mean. The mean, mode, and median do not fall on the same score. The mode will still be represented by the highest point of the distribution, but the mean will be toward the side with the tail and the median will fall between the mode and mean.
Standard Scores

- A calculated score that enables a researcher to compare scores from different scales
- Z-score most popular (mean of zero and SD of one)
- $z = \frac{(X-M)}{SD}$
Some Popular Standard Scores

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>Scale Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>100</td>
<td>GRE; SAT; GMAT</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>Wechsler IQ</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>ACT</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>T-scale MMPI</td>
</tr>
</tbody>
</table>

T Scores

- Another type of standard score -- sometimes preferred because all numbers are positive
- T score = 50 + 10(z)
- Example: If you scored 2 SD’s above the mean on a reading test your T score would be 50 + 10(2) = 70
Stanines

• Another type of standard score -- preferred for interpretability
• Score between 1 and 9 with each stanine covering 1/2 standard deviation unit (e.g., stanine of 5 = 40%ile-59%ile)

Avoid Misconceptions with Grade Equivalent Scores

• Don’t confuse norms with standards of what should be
• Don’t interpret a grade equivalent as an estimate of the grade where a student should be placed
• Don’t expect that all students should gain 1.0 grade equivalent each year
• Don’t assume that the units are equal at different parts of the scale
Avoid Misconceptions with Grade Equivalent Scores

• Don’t assume that scores on different tests are comparable
• Don’t interpret extreme scores as dependable estimates of student’s performance level

National Testing Program: NAEP

• NAEP – National Assessment of Educational Progress
• Formulated in the 1960s to provide benchmarks of educational attainment
• Now tests at grades 4, 8, and 12
• Subject areas change depending on year
• Includes multiple-choice and open-ended test items