Today's Plan

- Brief comments on standardized tests.

- Collect presentation folders. Groups finish presentations if needed. Others watch video on Smarter Balanced Consortium.

Standardized Tests
Fall 2012

- http://www.smarterbalanced.org/

- http://www.mistreamnet.com/vidflv.php?who=mac092812

A Definition

- A **standardized test** is a test that is administered, scored, and interpreted in the same way for all test-takers.

Types of Standardized Tests

- Achievement – assesses learning of what has been taught – used to measure progress, to judge instructional effectiveness, or to compare performance

- Aptitude – assesses a general capacity to learn – used mainly for prediction

Norm-referenced Scores

- A standardized test that yields norm-referenced test scores allows test-users to describe an individual's performance in terms of how it compares to the performance of others (typically those in a norm group). Examples are the SAT and the Terra Nova.

Criterion-referenced Scores

- A standardized test that yields criterion-referenced scores allows test-users to describe an individual's performance in terms of the tasks he or she can and cannot perform. A decision may be made about mastery or nonmastery of objectives. Examples are the PACT and the Praxis.

Raw Score

- The **raw score** is the score that is first obtained when scoring a test or performance task.

Derived Scores

- The raw score is frequently converted to percent correct, percentile ranks, grade equivalent scores, or standard scores for interpretation and/or comparison. These scores, which are obtained by applying a rule or mathematical formula to the raw score, are called **derived scores.**
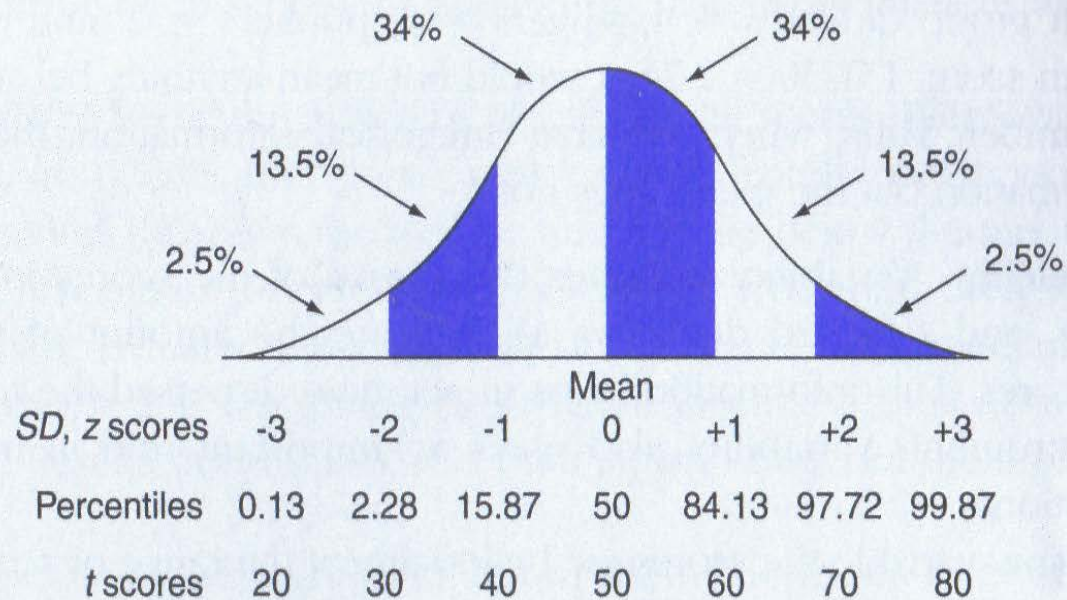
Percent Correct

- The percent of test items the examinee answered correctly.

- Divide the raw score by the total number of test items.

- Used to determine mastery of objectives. The cut-off typically varies from 70-80 percent correct.

The Normal Curve and Relative Standing

FIGURE 7.3

The Normal Curve

SD, z scores	-3	-2	-1	0 (Mean)	+1	+2	+3
Percentiles	0.13	2.28	15.87	50	84.13	97.72	99.87
t scores	20	30	40	50	60	70	80

34%   34%

13.5%   13.5%

2.5%   2.5%

Stanine

- Stanines are single-digit scores ranging from 1 to 9. The distribution of raw scores is divided into 9 parts.
- Range of percentiles in each stanine:

96 or above	9	23-39	4
89-95	8	11-22	3
77-88	7	4-10	2
60-76	6	Below 4	1
40-59	5		

Grade Equivalent Score

- The basis of comparison is students in the norm group at specified grade levels.

- "On the reading vocabulary test, Mary scored as well as the typical student in a national group did in the spring of the 5th grade."

Standard Scores

- Scores that reflect a student's position in the "normal distribution." Many psychologists believe that educational and psychological characteristics fit this pattern.

Confidence Bands or Intervals

- A confidence band or interval is the range around a score that reflects the probable amount of error in the score. On most standardized score reports, confidence intervals show a 68% likelihood that the student's "true" score lies within the indicated interval.
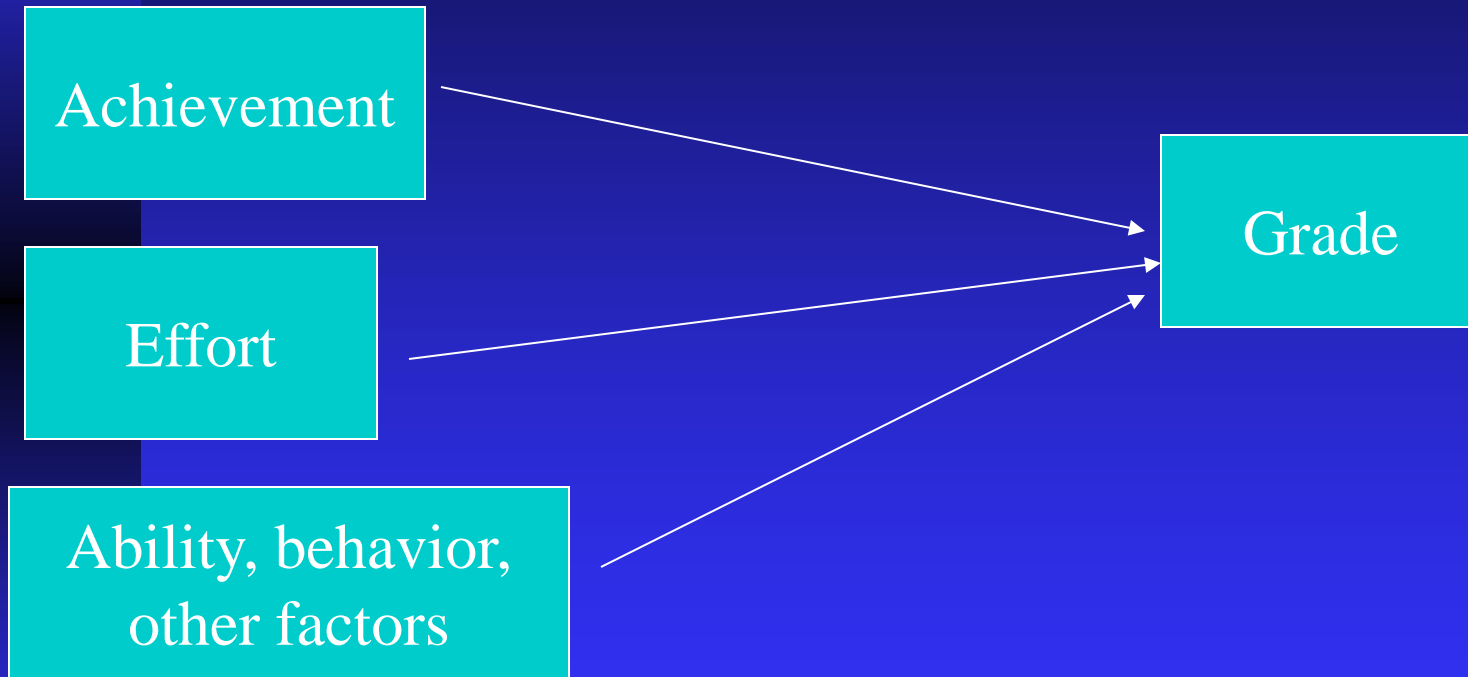
Assessment Bias

- "An assessment instrument has cultural bias if any of its items either offend or unfairly penalize some students on the basis of their ethnicity, gender, or SES." http://www.youtube.com/watch?v=HjID39cqmzw

- "Would you rather swim in the ocean, a lake, or a swimming pool? Write a two-page essay defending your choice" (Ormrod, 2006, p. 594).
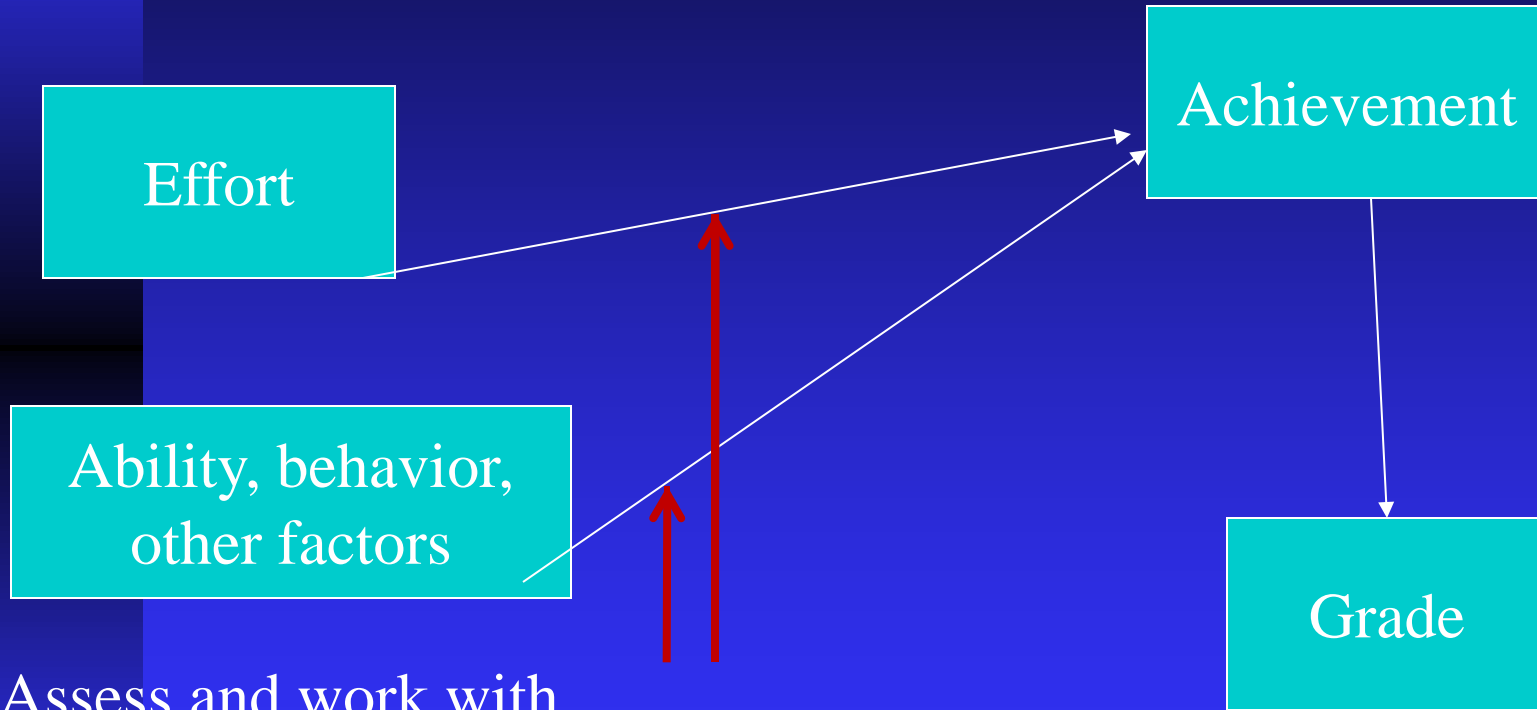
Issues in Grading—Brookhart

Achievement → Grade

Effort → Grade

Ability, behavior, other factors → Grade

NO!

Issues in Grading

Achievement → Grade

TOO SIMPLE

Issues in Grading

Effort

Ability, behavior, other factors

Achievement

Grade

Assess and work with
these separately from the grade,
so they indirectly contribute
 to quality of final products

Ethical Dilemmas in Grading

- DO NO HARM

- Avoid score pollution

- Use best assessment practices