Understanding Standardized Tests

Susan C. Anthony, P.O. Box 111704, Anchorage, AK 99511. 907-345-6689. www.SusanCAnthony.com

Steps in Teaching

1. Diagnose and assess the needs of the learner.
2. Choose appropriate instructional objectives.
3. Select materials and activities, then offer relevant instruction.
4. Evaluate the outcomes.
5. Use evaluation results to decide on the next step: go on to new objectives or review old ones.

Measurement / Evaluation

Measurement is the process of taking a test.

Evaluation is judging what a test score *means* and deciding what action to take next.

Evaluation decisions should *never* be based on a single test score.

Informal / Formal Evaluation

Informal evaluation takes place constantly in the context of instruction. Does the child understand the material?

Formal evaluation is *testing* to see if an objective has been mastered.

A combination of formal and informal evaluation yields the best results.

Reliability / Validity

Reliability means the test gives *consistent* results.

Validity means the test measures what it sets out to measure. A multiple-choice test cannot measure speaking or writing ability.

No test is perfectly valid or reliable.

Objective Tests / Subjective Tests

Objective test questions have only *one* right answer (multiple-choice, true-false, matching, fill in the blank). In the world of sports, racers are judged objectively.

Subjective tests may be scored differently by different people (essay tests). In the world of sports, gymnasts are judged subjectively.

Both objective and subjective tests have value, depending on what is being measured. Objective tests are more *reliable.*

Aptitude Tests / Achievement Tests

Aptitude tests are predictive. They are designed to predict future performance and to measure the ability to learn. Intelligence (IQ) tests are designed to measure aptitude.

Achievement tests are reflective. They are designed to measure the skills and knowledge a child has attained.

In reality, both kinds of tests measure both things, but they differ in emphasis.

Teacher-Made Tests / Standardized Tests

Teacher-made tests are specifically geared to the students and the objectives taught in a specific situation. They yield information on individual student progress as well as on the general effectiveness of instruction.

Standardized tests are administered under standard conditions wherever and whenever given. All test-takers are given the same instructions, the same amount of time to work, the same scoring. Standardized tests are more useful for making judgments about large groups of people than about individuals.

Criterion-Referenced Tests / Norm-Referenced Tests

Criterion-referenced tests, also called mastery tests, compare a person's performance to a set of objectives. Anyone who meets the criterion can get a high score.

Norm-referenced tests compare an individual's performance with the performance of others. They are designed to yield a normal curve.

Prepare for Testing

Factors which correlate with high test scores.
— Ample recreational reading.
— Small class size (tutorial is best).
— Stability at home, two parents.
— Family discussions around the dinner table.
— Good health and nutrition.
— No more than four hours of TV a day.
— Little time spent playing video games.

Schedule the test.
— See the C.H.E.C. Handbook or this handout for information about available resources.
— Schedule the test for the middle of a week, not just before or after a holiday or special event. Choose a time of day when your child is alert.
— Don't test after heavy physical activity.
— Allow no interruptions such as ringing phones.
— Allow for breaks between subtests.

Take a relaxed, positive approach to testing.
— Help kids anticipate tests with interest.
— Consider giving frequent timed tests throughout the year (on math facts, for example) to help kids overcome test anxiety.
— Tell children the test measures only one small part of them. It doesn't measure character, talent, ability to learn or happiness!

Explain to students why they are being tested.
— It will help them understand their special strengths and weaknesses.
— It will help them see how they are progressing from year to year.
— It will help us improve instruction.
— It is required by law.

Tell students what to expect during the test.
— Share the planned schedule.
— The test is timed. There is plenty of time to finish, but don't waste time or daydream.
— This is not like a classroom test. You are not expected to know all the answers. It is virtually impossible to get a perfect score! The important thing is to *do your best.*

Tell students what will be done with the results.
— The results will not affect their grades.
— We will go over the results together.
— The results will be sent to ____ school to fulfill the requirements of the law.

Teach "testwiseness" tips.
— Before starting a test, relax, breathe, focus your mind and pray that you will do your personal best. Some test anxiety is normal.
— Listen carefully to *all* of the directions and make sure you understand them exactly.
— *Always* check to see that you are marking the correct line on the answer sheet.
— Read all of the choices. Usually one answer will be *almost* right. Choose what you think is the *best* answer. If you are not sure, eliminate the ones you know are wrong and choose from those which remain. Do not skip questions, but don't waste time on very hard questions.
— If you make a mistake, erase *completely.* Your first response is usually best. Change an answer only if you are *sure* you were wrong.
— Use all of the time. If you finish before time is called, go back to review and recheck your answers carefully.

Answer Sheet

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ	Ⓐ
B	○	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ	Ⓑ
C	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ	Ⓒ
D	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ	Ⓓ
E	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ	Ⓔ

Interpreting Test Scores

Raw Score. This is the number of items the student answered correctly. It is used to calculate the other, more useful scores.

Stanine. One of nine equal sections of the normal curve. Stanines can be easily averaged and compared from test to test, but are less precise than other scores.

Normal Curve Equivalent (NCE). For these scores, the normal curve is divided into equal units ranging from 1 to 99, with an average of 50. These can be averaged and compared from test to test or year to year.

National Percentile (NP). Percentile scores range from 1 to 99 with an average of 50, but the units are *not* of equal size. A score of 74 means the student scored higher than 74% of children in the norm group. Percentile scores can *not* be averaged or compared from test to test. The difference between 45 and 49 is *much* smaller than the difference between 95 and 99! (See illustration.)

Grade Equivalent (GE). Grade equivalent scores can be misleading. A third grader with a GE of 5.3 has performed as well as the average fifth grader in the third month of school would probably perform *on the third grade test.* This does *not* mean the child is ready for fifth grade material, just that she is above average. A 5.3 in math is *not* equivalent to a 5.3 in reading.

Stanine		*NCE*	*NP*	*% of students*
9	Highest	86–99	96–99	4%
8	High	76–85	90–95	7%
7	Well above average	66–75	78–89	12%
6	Slightly above average	56–65	60–77	17%
5	Average	45–55	41–59	20%
4	Slightly below average	35–44	23–40	17%
3	Well below average	25–34	11–22	12%
2	Low	15–24	5–10	7%
1	Lowest	1–14	1–4	4%

When Are Percentile Score Differences Significant?

— If a student's composite percentile rank is between the 25th and 75th percentile, differences of at least **18** points between the composite score and a subtest score are significant.

— If a student's composite percentile rank is between the 76th and 87th or the 13th and 24th percentile, differences of at least **12** points between the composite and a subtest score are significant.

— If a student's composite rank is between 88th and 99th or the 1st and 12th percentiles, differences of at least **6** points between the composite score and a subtest score are significant.