

GLOSSARY OF FREQUENTLY USED TESTING TERMS

Achievement Levels - represent levels of mastery based on the professional judgments of panels of educators and citizens. For example, on the *FCAT Sunshine State Standards (SSS)* there are five levels of achievement with 1 representing the lowest level and 5 representing the highest level.

Benchmark - a specific statement that describes what a student should know and be able to do, as part of the *Sunshine State Standards*.

Content Cluster Scores - the number of points earned in each grouping, cluster, or strand of related statements that describe what students should know and be able to do. For example, on the *FCAT SSS Reading* there are content cluster scores for Words and Phrases in Context; Main Idea, Plot, and Purpose; Comparisons and Cause/Effect; and Reference and Research.

Criterion Referenced Test - a test that is intended to measure how well a student has learned specific content area skills. Examples of criterion referenced tests, also referred to as CRTs, include the *FCAT SSS* in reading, mathematics, science, and writing.

Developmental Scale Score - a score used to determine a student's annual progress from one grade to the next grade. For example, the *FCAT SSS* developmental scale ranges from 86 to 3008 across grades 3 through 10.

Expository Writing - writing that provides information, explains how or why, clarifies a process, or defines a concept.

Gridded Response Item - an item that requires a student to bubble in a numeric answer on a number grid.

Item - another term for a test question.

Mean - the arithmetic average of a group of scores, computed by adding all scores together and dividing by the number of scores in a group.

Median- a score that identifies the middle value of a group of data. One half of the scores fall above the median; the other half fall below the median.

Median Percentile Rank - the midpoint of a group of percentile rank scores.

Mode- the most frequently occurring score or a set of scores.

Multiple Choice Item - an item that presents a student with several options from which to choose the correct answer.

Narrative Writing - writing that recounts a personal experience or tells a story based on a real or an imagined event.

Nationally Mandated Test - a test that is required by the federal government. An example of a nationally mandated test is the *National Assessment of Educational Progress (NAEP)*.

Non-secure Testing Materials - materials that can be previewed by test staff or other personnel before, during, and/or after testing. Examples of non-secure testing materials are teacher-developed tests and tests that are included in textbooks.

Norm Referenced Test - a test that compares a student's performance against how other students in a norm group did on the test. An example of a norm referenced test, referred to as an NRT, is the *Stanford Achievement Test (SAT)*.

Number Correct Score- a score that reports the raw number of points a student earned in each test question.

Percentile Rank - indicates the percentage of a reference group obtaining scores equal to or less than the score achieved by an individual. This rank indicates the relative standing of one student in comparison to students in the same grade who took the test. Percentile ranks can be used to report student performance across subject areas.

Performance Task - an item that requires students to write an answer instead of choosing one from several choices or gridding a response. Two types of performance tasks are used in the *FCAT SSS* which differ in terms of the length or complexity of the expected answer (short-response items and extended-response items).

Persuasive Writing - writing that attempts to convince the reader that a point of view is valid or that the reader should take a specific action.

Prompt- the topic on which a student is to write an essay response. On the *FCAT Writing+*, the prompt has 2 parts: the writing situation and the directions for writing.

Quartile Score Analysis - provides the percent of students whose percentile ranks fall within each of the four quartile ranges (1-25, 26-50, 51-75, and 76-99). In a normal distribution, one would expect to have 25% of the students scoring in each quartile range.

Reliability - the degree to which a test consistently measures what it is intended to measure. For example, reliability can be measured by administering a test to the same group of individuals on two different occasions. If the test is reliable, the results from both administrations will be consistent.

Rubric - scoring tool or a set of criteria used to evaluate a student's response to a constructed response item.

Scale Score - raw score that has been converted to a standardized scale. Scale scores are suitable for comparing different test levels or test forms of the same subject area.

Secure Testing Materials - materials that must be accounted for before, during, and after testing; kept in locked storage when not in use; and cannot be previewed and/or copied by school staff and other personnel. Examples of secure testing materials are *FCAT SSS* test booklets and completed answer sheets.

Standardized Test - a test in which the directions, time limits, materials, and scoring procedures are designed to remain constant each time the test is administered in order to ensure comparability of scores. Standardized tests can be either criterion referenced or norm referenced.

Stanines - represent a student's relative standing with respect to a reference group. Stanine scores range from a low of 1 to a high of 9, with 5 representing average performance.

State Mandated Test - a test that is required by the State of Florida. In some cases, the particular test is required by the state such as the *FCAT SSS*. In other cases, a type of test is required by the state and the district selects or develops the particular test, for example the *Grade 3 Reading Student Portfolio*, that is administered during the second semester of each school year to third graders as well as selected a norm referenced test, *Alternative Assessment for Grade 3 Promotion*, that is administered in the summer of each school year to retained third graders.

Validity - the degree to which a test actually measures what it is intended to measure. For example, the *FCAT SSS* must directly and reliably measure proficiency on the state's Sunshine State Standards to be a valid measure of the curriculum. In order for a test to be valid, it must first be reliable.