

ITBS/ITED INTERPRETATION INFORMATION

What is a Standardized Test?

A standardized test is an assessment tool that has the sameness to it in terms of the tasks students are asked to complete, the procedures used to administer it, and the methods used to score it. That is, all the conditions are the same for all students so that scores can be compared from one time to the next, so that the scores of individuals can be combined meaningfully to describe group performance, and so that fair comparisons can be made among students and groups.

What is an Achievement Battery?

An achievement battery is a collection of tests from several areas that have been standardized with the same group of students. That is, the norms for all tests have been obtained from a single group of students at each grade level. This unique aspect of the battery makes it possible to use the scores from the separate tests to determine areas of relative strength and weakness in skill development and to estimate year-to-year growth.

What is the Purpose of Testing?

The primary purpose of using a standardized achievement test is to provide information that can be used to improve instruction. The information from this test is unique in that it is not available from other sources. It is valuable to the extent that it contributes to better instructional decisions than educators could make without having that information available. Though standardized achievement scores cannot and should not replace teacher observations and classroom assessment information, they can provide unique supplementary information that bears on decisions about selecting learning objectives and procedures, designing or choosing instructional materials, and creating an effective learning environment.

At all test levels, the ITBS/ITED has been designed to fulfill three main purposes: 1) to obtain information that can support instructional decisions made by teachers in the classroom, 2) to provide information to students and their parents for monitoring the student's growth from grade to grade, 3) to examine the yearly progress of grade groups as they pass through the school's curriculum.

How should we interpret the scores?

Scores such as percentile ranks, stanines, and grade equivalents differ from one another in the purposes they can serve, the precision with which they describe achievement, and the kind of information they provide.

Percentile Rank (PR):

A student's percentile rank is a score that tells the percent of students in a particular group that got lower raw scores on a test than the student did. It shows the relative position or rank in a group of 100 students who are in the same grade and who were tested at the same time of year. For example, if Tom earned a percentile rank of 72 on the Math Problem Solving Test, it means that he scored higher than 72 percent of the students in the group with which he is being compared. It also means that 28 percent of the group tested higher than Tom. Percentile ranks range from 1-99. Percentile ranks are especially useful for determining the areas of relative strength and weakness for an individual student, a class, or a grade group. "On Grade Level" or average is defined as performance between the 25th and 75th percentiles. Percentile ranks are different from *percent-correct* scores, but the two are often confused. The percentile rank shows a student's relative standing or rank in a group of 100. The *percent-correct* score indicates what percent of the items a student got right on a test. So this score simply tells how close the student got to a perfect score—all items correct. It does not involve comparing with the scores of other students. Thus, these two types of scores

provide quite different information about a student's test performance. Percentile ranks should not be averaged.

Stanines (S):

Stanines are course groupings of percentile ranks. Stanines range from 1-9 (one being low, 9 being high) with an average value of 5. Because stanines are coarse groupings of percentile ranks, they are less precise indicators of student achievement than percentile ranks. For example, percentile ranks of 24 and 39 are both in the 4th Stanine. Further more, the fact that 23 and 24 are consecutive percentile ranks that are in different stanines (3 and 4 respectively) points out the potential misconceptions that could arise when using stanines. Nonetheless, stanines are convenient scores to use to help students and parents identify areas of strength and weakness that might be represented by a set of test scores. Stanines should not be used to describe a student's developmental level or to measure growth.

PR: 1-4	5-11	12-23	24-40	41-59	60-76	77-88	89-95	96-99
S: 1	2	3	4	5	6	7	8	9

Grade Equivalent (GE):

The GE is a decimal number that describes performance in terms of grade level and months. For example, if a sixth grade student obtains a GE of 7.8 on the Vocabulary test, her score is like the one a typical student at the end of the eighth month of seventh grade would likely get on that same sixth grade Vocabulary test. The average yearly growth is 10 months. Grade equivalents are particularly useful for measuring individual growth from one year to the next and for estimating a student's developmental status in terms of grade level. It should not be misused or misinterpreted though. It is important to know that the GE only estimates a student's developmental level; it does not provide a prescription for grade placement. These scores are ill-suited to identifying a student's standing within a group or to diagnosing areas of relative strength or weakness.

Points to Ponder:

Test results provide one dimension of a variety of information about a student which may be obtained and studied in an effort to better understand his/her instructional needs.

Essentially, a test score indicates how well the student performed in the testing situation, but does not reveal why the student performed as he/she did.

Test scores should not be presumed to reflect constant levels of performance.

Since these tests attempt to measure achievement in several areas, only a relatively small number of concepts can be measured, and usually there are only a few questions about each concept.

No derived scores connote value judgments such as "good" or "poor". Rather, it is an indication of a level of performance in relation to the reference population. (e.g. the national norm group)

Low scores should be interpreted with caution. Though low scores indicate that the student did not score at least in the average range, they do not indicate the student's ability to do so.

High scores are not likely to be obtained by chance and may be regarded as more significant.