

Northwest Evaluation Association Measures of Academic Progress (MAP)

Glossary of Terms

The terms defined are the more common or basic ones such as occur in test manuals and educational journals. In the definitions, certain technicalities and niceties of usage have been sacrificed for the sake of brevity, and, it is hoped, clarity.

achievement test - A test that measures the extent to which a person has "achieved" something, acquired certain information, or mastered certain skills - usually as a result of planned instruction or training.

item analysis - The process of evaluating single test items in respect to certain characteristics. It usually involves determining the difficulty value and the discriminating power of the item, and often its correlation with some external criterion.

multiple-choice item - A test item in which the examinee's task is to choose the correct or best answer from several given answers or options.

norms - Statistics that supply a frame of reference by which meaning may be given to obtained test scores. Norms are based upon the actual performance of pupils of various grades or ages in the standardization group for the test. Since they represent average or typical performance, they should not be regarded as standards or as universally desirable levels of attainment. The most common types of norms are deviation IQ, percentile rank, grade equivalent, and stanine. Reference groups are usually those of specified age or grade.

percentile - A point (score) in a distribution at or below the percent of cases indicated by the percentile. Thus a score coinciding with the 35th percentile is regarded as equaling or surpassing 35 percent of the persons in the group. It also means that 65 percent of the performances exceed this score. "Percentile" has nothing to do with the percent of correct answers an examinee makes on a test.

reliability - The extent to which a test is consistent in measuring whatever it does measure; dependability, stability, trustworthiness, relative freedom from errors of measurement. Reliability is usually expressed by some form of reliability coefficient or by the standard error of measurement derived from it.

RIT Scale - Short for Rasch Unit, honoring George Rasch, the Danish mathematician who developed the underlying theory for this type of measurement. The RIT Scale is a curriculum scale developed by NWEA that uses the individual item difficulty values to estimate student achievement. Advantages to the RIT Scale are that it can relate the numbers on the scale directly to the difficulty of items on the tests and it is equal interval. Equal interval means that the difference between scores is the same regardless of whether a student is at the top, bottom or middle of the RIT Scale, and it has the same meaning regardless of grade level.

standardized test (standard test) - A test designed to provide a systematic sample of individual performance, administered according to prescribed directions, scored in conformance with definite rules, and interpreted in reference to certain normative information. Some would further restrict the usage of the term "standardized" to those tests for which the tests have been chosen on the basis of experimental evaluation, and for which data on reliability and validity are provided. Others would add "commercially published" and/or for "general use".