



2014

EPSY 8225
Operational Measurement

Standards for Validity

Standard 1.0

Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided.

Standards for Validity

Standard 1.1

The test developer should set forth clearly how test scores are intended to be interpreted and consequently used. The population(s) for which a test is intended should be delimited clearly, and the construct or constructs that the test is intended to assess should be described clearly.

Standards for Reliability/Precision

Standard 2.0

Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use.

Standards for Reliability/Precision

1. Specifications for replications of the testing procedure
2. Evaluating reliability/precision
3. Reliability/Generalizability coefficients
4. Factors affecting reliability/precision
5. Standard errors of measurement
6. Decision consistency
7. Reliability/precision of group means
8. Documenting reliability/precision

Standards for Fairness

Standard 3.0

All steps in the testing process, including test design, validation, development, administration, and scoring procedures, should be designed in such a manner as to minimize construct-irrelevant variance and to promote valid score interpretations for the intended uses for all examinees in the intended populations.

Standards for Test Design and Development

Standard 4.0

Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.

Standard 4.18

Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays.

Standard 4.20

The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring.

Standards for Scores, Scales...

1. Interpretation of scores
2. Norms
3. Score linking
4. Cut scores

Standards for Scores, Scales, Norms, Score Linking, and Cut Scores

Standard 5.0

Test scores should be derived in a way that supports the interpretations of test scores for the proposed uses of tests. Test developers and users should document evidence of fairness, reliability, and validity of test scores for their proposed use.

Standard 5.12

A clear rationale and supporting evidence should be provided for any claim that scale scores earned on alternate forms of a test may be used interchangeably.

Standard 5.21

When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly.

Standards for Test Administration, Scoring, Reporting, and Interpretation

Standard 6.0

To support useful interpretation of score results, assessment instruments should have established procedures for test administration, scoring, reporting, and interpretation. Those responsible for administering, scoring, reporting, and interpreting should have sufficient training and supports to help them follow the established procedures. Adherence to the established procedures should be monitored, and any material errors should be documented and, if possible, corrected.

Standards for Supporting Documentation

Contents of Test Documents:

1. Appropriate use
2. Test development
3. Test administration and scoring
4. Timeliness of delivery of test documents

Standards for Supporting Documentation

Standard 7.0

Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores.

Standards for Supporting Documentation

Standard 7.13

Supporting documents (e.g., test manuals, technical manuals, user's guides, and supplemental materials) should be made available to the appropriate people in a timely manner.

Standards for Test Users' Rights and Responsibilities

1. Validity of interpretations
2. Dissemination of information
3. Test security and protection of copyrights

Board of Testing and Assessment National Research Council

LESSONS LEARNED

- In many situations, standardized tests provide the most objective way to compare the performance of a large group of examinees across places and times.
- A test score is an estimate rather than an exact measure of what a person knows and can do.
- High-stakes decisions about individuals should not be made on the basis of a single test score.
- Tests should not be used for high-stakes decisions if test takers have not had an opportunity to learn the material on which they will be tested.

BOTA, Lessons Learned (cont.)

- States, districts, and schools should aim to maximize the participation of English-language learners and students with disabilities in large-scale tests.
- Teachers need professional development that helps them better understand core principles of assessment and how to apply these to their regular instruction and testing.
- In the design of tests, form must follow function.
- The design process must ensure that test score interpretations are valid.
- The design process must ensure that the test results are reliable and fair.

BOTA, Lessons Learned (cont.)

- Testing professionals should consider the relationships among cognition, observation, and interpretation—the “assessment triangle”—when evaluating the soundness of current educational tests or designing new ones.
- Advances in the cognitive sciences and measurement offer opportunities to develop educational assessments that better support learning.
- The people who design and mandate tests must be constantly vigilant about equity concerns, including opportunity to learn, cultural bias, or adverse impact.

BOTA, Lessons Learned (cont.)

- In the absence of effective services for low-performing students, better tests will not lead to better educational outcomes.
- Test results may be invalidated by teaching narrowly to a particular test.
- New testing programs should build in an evaluation component.
- Test developers and policy makers should clearly explain to the public the purpose for a test and the meaning of different levels of test performance.
- When test results are reported to students, teachers, and the public, the limitations of the test should be explained clearly to a lay audience.