



## **NCII TRC on Academic Progress Monitoring**

### **Glossary of Terms**

#### **Area Under the Curve (AUC)**

AUC is an overall indication of the diagnostic accuracy of a Receiver Operating Characteristic (ROC) curve (see definition below). AUC values closer to 1 indicate the screening measure reliably distinguishes among students with satisfactory and unsatisfactory reading performance, whereas values at 0.50 indicate the predictor is no better than chance.

#### **Benchmark**

A benchmark is a pre-determined level of performance on a screening test that is considered representative of proficiency or mastery of a certain set of skills.

#### **Classification Accuracy**

The classification accuracy indicates the extent to which a screening tool is able to accurately classify students into “at risk for reading disability” and “not at risk for reading disability” categories.

#### **Coefficient Alpha**

The coefficient alpha is a measure of the internal reliability of items in an index. Values of alpha coefficients can range from 0 to 1.0. Alpha coefficients that are closer to 1.0 indicate that the items are more likely to be measuring the same thing.

#### **Construct Validity**

Construct validity is a type of validity that assesses how well one measure correlates with another measure purported to represent a similar underlying construct.

#### **Content Validity**

Content Validity is a type of validity that uses expert judgment to assess how well items measure the universe they are intended to measure.

#### **Criterion Measure**

A criterion measure is a dependent variable, or outcome measure in a study.

#### **Cross-Validation**

Cross-validation is the process of validating the results of one study by performing the same analysis with another sample. In the cross-validation study, cut scores derived from the first study are applied to the administration of the same test and criterion measure with a different sample of students.

**Cut Score**

A cut score is a score on a screening test that divides students who are considered potentially at risk from those who considered not at risk.

**Disaggregated Data**

Data are disaggregated when they are calculated and reported separately for specific sub-populations (e.g., race, economic status, academic performance, etc.).

**Generalizability**

Generalizability is the extent to which results generated from one population can be applied to another population. A tool is considered more generalizable if studies have been conducted on larger, more representative samples.

**Inter-Rater Reliability**

Inter-rated reliability is the extent to which raters judge items in the same way.

**Kappa**

Kappa is an index which compares the agreement against that which might be expected by chance. Kappa can be thought of as the chance-corrected proportional agreement. Possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to -1 (complete disagreement).

**Norms**

Norms are a standard of performance on a test that is derived by administering the test to a large sample of students. Results from subsequent administrations of the test are then compared to the established norms.

**Predictive Validity**

Predictive Validity is a type of validity that assesses how well a measure predicts performance on some future, similar measure.

**Receiver Operating Characteristic (ROC) Curve**

A ROC curve is a generalization of the set of potential combinations of sensitivity and specificity possible for predictors. A ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cut-points of a diagnostic test. The Area under the Curve (AUC) represents an overall indication of the diagnostic accuracy of a ROC curve. AUC values closer to 1 indicate the screening measure reliably distinguishes among students with satisfactory and unsatisfactory reading performance, whereas values at 0.50 indicate the predictor is no better than chance.

**Reliability**

Reliability is the consistency with which a tool classifies students from one administration to the next. A tool is considered reliable if it produces the same results when administering the test under different conditions, at different times, or using different forms of the test.

**Response to Intervention (RTI)**

RTI integrates assessment and intervention within a multi-level prevention system to maximize student achievement and to reduce behavior problems. With RTI, schools identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student's responsiveness, and identify students with learning disabilities.

**Sensitivity**

Sensitivity is the extent to which a screening measure accurately identifies students at risk for the outcome of interest.

**Specificity**

Specificity is the extent to which a screening measure accurately identifies students not at risk for the outcome of interest.

**Split-Half Reliability**

Split-half reliability is a method of assessing internal reliability by correlating scores from one half of the items on an index or test with scores on the other half of the items.

**Test-Retest Reliability**

Test-retest reliability is a correlation of scores on a test given at one time to scores on the test given at another time to the same subjects.

**Validity**

Validity is the extent to which a tool accurately measures the underlying construct that it is intended to measure.