

---

## **NCRTI TRC on Progress Monitoring Glossary of Terms**

### **Alternate Forms**

Alternate forms are parallel versions of the measure within a grade level, of comparable difficulty (or with Item Response Theory (IRT) based, item or ability invariance).

### **Benchmark**

A benchmark is an established level of performance on a test. A benchmark can be used for screening if it predicts important outcomes in the future. Alternatively, a benchmark can be used as a cut-score that designates proficiency or mastery of skills.

### **Coefficient Alpha**

Coefficient alpha is a measure of the internal consistency of items within a measure. Values of alpha coefficients can range from 0 to 1.0. Alpha coefficients that are closer to 1.0 indicate items are more likely to be measuring the same thing.

### **Criterion Validity**

Criterion validity indexes how well one measure correlates with another measure purported to represent a similar underlying construct. It can be concurrent or predictive.

### **Content Validity**

Content Validity relies on expert judgment to assess how well items measure the universe they are intended to measure.

### **Criterion Measure**

A criterion measure is the measure against which criterion validity is judged.

### **Cross-Validation**

Cross-validation is the process of validating the results of one study by performing the same analysis with another sample under similar conditions.

### **Direct Evidence**

Direct evidence is a term used on the Center's Tools Charts to refer to data from a study based on the tool submitted for evaluation.

### **Disaggregated Data**

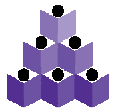
Disaggregated data is a term used on the Center's Tools Charts to indicate that a tool reports information separately for specific sub-populations (e.g., race, economic status, special education status, etc.).

### **End-of-Year Benchmarks**

End-of-year benchmarks specify the level of performance expected at the end of the grade, by grade level.

### **General Outcome Measure (GOM)**

A GOM is a measure that reflects overall competence in the annual curriculum.



**Generalizability**

Generalizability is the extent to which results generated on a sample are pertinent to a larger population. A tool is considered more generalizable if studies have been conducted on large representative samples.

**Growth**

Growth refers to the slope of improvement or the average weekly increase in scores by grade level.

**Indirect Evidence**

Indirect evidence is a term used on the Center's Tools Charts to refer to data from studies conducted using other tools that have similar test construction principles.

**Inter-Scorer Agreement**

Inter-scorer agreement is the extent to which raters judge items in the same way.

**Kappa**

Kappa is an index that compares the agreement against what 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).

**Mastery Measurement (MM)**

MM indexes a student's successive mastery of a hierarchy of objectives.

**Norms**

Norms are standards of test performance derived by administering the test to a large representative sample of students. Individual student results are compared to the established norms.

**Pass/Fail Decisions**

Pass/fail decisions are the metric in which mastery measurement scores are reported.

**Performance Level Score**

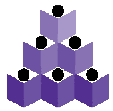
Performance level score is the score (often the average or median for 2-3 scores); it indicates the student's level of performance.

**Predictive Criterion Validity**

Predictive validity indexes how well a measure predicts future performance on a highly valued outcome.

**Progress Monitoring**

Progress monitoring is repeated measurement of academic performance used to inform instruction of individual students in general and special education in grades K-8. It is conducted at least monthly to (a) estimate rates of improvement, (b) identify students who are not demonstrating adequate progress and/or (c) compare the efficacy of different forms of instruction to design more effective, individualized instruction.



**Rate of Improvement**

Rates of improvement specify the slopes of improvement or average weekly increases, based on a line of best fit through the student's scores.

**Reliability**

Reliability is the extent to which scores are accurate and consistent.

**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 measure reveals improvement over time, when improvement actually occurs.

**Skill Sequence**

The skills sequence is the series of objectives that correspond to the instructional hierarchy through which mastery is assessed.

**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 indexes a test's internal reliability by correlating scores from one half of items with scores on the other half of items.

**Standard Error of the Mean (SEM)**

The standard error of the mean (SEM) is the standard deviation of the sample mean estimate of a population mean.

**Technical Adequacy**

Technical adequacy implies that the psychometric properties such validity and reliability meet strong standards.

**Test-Retest Reliability**

Test-retest reliability is the consistency with which a tool indexes student performance from one administration to the next.

**Validity**

Validity is the extent to which scores represent the underlying construct.