

# RELIABILITY

Form of reliability	Defining features
Internal consistency	How well a set of item scores correlate with other item scores on the same test.
Alternate form reliability	How well scores on two different versions, or forms, of the same test correlate with each other .
Test-retest reliability	The stability of a test score over a fixed period of time.
Inter-rater reliability	The consistency with which different observers rate the same behavior.

Each of the types of reliability are distinct and useful for their own purposes, but should not be used interchangeably. For example, alternate form reliability is an important consideration for screeners with multiple forms, such as many curriculum-based measures, but may not be relevant to other types of tests. When evaluating the reliability of a screener, ensure that the type of reliability that is reported is appropriate for the type of screener and that two or more forms of reliability are reported.

## What is it?

Reliability is the consistency of a set of scores that are designed to measure the same thing.

## Example

Suppose that a family is shopping at a supermarket. As the family makes its way through the produce section, the children decide to weigh a watermelon on five of the scales to figure out how much it weighs. Reliability in measurement refers to how consistently the five scales provide the same weight for the watermelon. If each scale gave a different weight for the fruit, the family would not know the cost of the watermelon, and thus may not be able to make an informed decision about whether to buy it.

## Why does reliability matter?

Decisions about what to do with data require that the information is trustworthy. When teachers, school psychologists, or other school personnel administer screeners of reading, there is typically an implicit trust or assumption that the obtained scores from the screener accurately reflect a student's ability, and there is little to no error in the score. In reality, the reliability of screening measures may vary considerably and should not be taken for granted. To make judgements about whether a student is truly at risk for reading problems, and therefore needs intervention, we should ensure that the screening tool we are using provides consistent information.

## Where do I go from here?

For more information about the reliability of screening and progress monitoring measures, visit the National Center on Intensive Intervention's (NCII's) [academic](#) and [behavior](#) screening and [academic](#) and [behavior](#) progress monitoring tools charts. NCII publishes these charts to assist educators and families in becoming informed consumers who can select screening and progress monitoring tools that best meet their needs.

For more information on literacy screening processes, see resources from the National Center on Improving Literacy: <https://improvingliteracy.org/>.

# Academic Screening Tools Chart

Universal screening can be used to identify which children will need the most intensive intervention. In some cases, children with the weakest initial skills may bypass Tier 2 intervention and move directly into intensive intervention. The tools on the academic screening tools chart can be used to identify students at risk for poor academic outcomes, including students who require intensive intervention.

This tools chart has three tabs that include ratings on the technical rigor of the tools: (1) Classification Accuracy, (2) Technical Standards, and (3) Usability Features.

Last updated: July 2019. [Learn more about the content and structural changes to the academic screening tools chart during the most recent update.](#)

## Legend

- Convincing evidence
- ◐ Partially convincing evidence
- Unconvincing evidence
- Data unavailable

<sup>d</sup> Disaggregated data available

[View Chart Resources](#)

# Academic Progress Monitoring Tools Chart

This tools chart presents information about academic progress monitoring tools. The following three tabs include ratings on the technical rigor of the tools:

- Performance Level Standards
- Growth Standards
- Usability

Last updated: October 2018

## Legend

- Convincing evidence
- ◐ Partially convincing evidence
- Unconvincing evidence
- Data unavailable

<sup>d</sup> Disaggregated data available

[View Chart Resources](#)

**FILTER RESULTS**

**Subject**  Reading  Mathematics

**Grade**  Pre-K  Elementary (K-4)  Middle School (5-8)  High School (9-12)

[Apply](#) [Print Chart](#)

[Hide/Show Advanced Filters](#) [Clear Filters](#)

Reset Chart [Compare Tools](#) [Prev Tab](#) [Next Tab](#)

Classification Accuracy **Technical Standards** Usability Features

All	Title	Area	Grade	Reliability	Validity	Sample Representativeness	Bias Analysis Conducted
<input type="checkbox"/>	Acadience Reading (aka DIBELS Next)	Composite Score	K	●	○	Regional without Cross-Validation	Yes

**FILTER RESULTS**

**Subject**  Mathematics  Reading  Spelling & Written Expression

**Grade**  Elementary (K-5)  Middle School (6-8)  High School (9-12)  Pre-K

[Apply](#) [Print Chart](#)

[Hide/Show Advanced Filters](#) [Clear Filters](#)

Reset Chart [Compare Tools](#) [Prev Tab](#) [Next Tab](#)

**Performance Level Standards** Growth Standards Usability

All	Title	Area	Grade	Measure	Reliability	Validity	Bias Analysis Conducted
<input type="checkbox"/>	aimswebPlus Math	Math Facts Fluency-1 Digit	1	Short Term Skill	○	◐	No

This handout was produced under the U.S. Department of Education, Office of Special Education Programs, Award No. H326Q160001. Celia Rosenquist serves as the project officer. The views expressed herein do not necessarily represent the positions or policies of the U.S. Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service, or enterprise mentioned in this website is intended or should be inferred.