

VALIDITY

Form of Validity	Key Questions	Example Evidence
Content	Does the content of the test match what the test is trying to measure?	Experts in reading instruction agree that the content of a reading screener matches the stated purpose of the test.
Substantive	What explains the consistency in one's responses to test items?	A high proportion of experts agree that the items on a test match the objectives they were designed to measure.
Structural	Do respondents' scores on a test group together in ways that we would expect based on theory?	Factor analysis shows that students' scores on items measuring the same skill (e.g., phonological awareness) are strongly correlated.
Generalizability	How well does the interpretation of scores generalize across different samples and different time points?	When analyzing the validity of the screener, researchers conducted a study using a large, nationally representative sample.
External	How strongly do scores correlate that should correlate? Do scores correlate that should not correlate? Do scores at one time point predict scores at another time point?	Scores on the reading screener correlate with scores on a high-stakes statewide assessment to ensure that scores that should be related are correlated.
Consequential	What are the implications for test takers when test scores are used for decision making?	A school district using a reading screener provides a clear description and rationale for what will happen as a result of using screening data.

What is it?

Validity is broadly defined as how well something measures what it is supposed to measure. While reliability is evaluated through the consistency of scores, validity is concerned with how well a set of scores reflects the intended construct or domain being assessed.

Example

A speedometer that shows you are traveling at 25 miles per hour, when your actual speed is 65, is not a valid measure of your car's speed. If the speedometer repeatedly shows 25 when you are traveling 65, the speedometer may be reliable, but not a valid test of speed. The speedometer does not accurately reflect what it is designed to measure (the true speed of the car). It is imperative that assessments used to screen and progress monitor students are valid, so we can make accurate assessments of students' skills across constructs.

Where do I go from here?

For more information about the validity 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

^d 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

^d 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.