

Guide to Using a Research-Based Process to Review and Select Early Literacy Assessments

Jodi Jacobson Chernoff
Principal Researcher

Jennie Jiang
Researcher

Arielle Lentz
Research Associate

Matthew Farmer
Intern

Table of contents

01

A research-based assessment review process

The types of assessments and the purpose of each type

Reliability and validity

Early literacy literature and research recommendations

02

Building early literacy assessment databases

Assessment coding for the databases

The early literacy assessment databases

03

Putting it all together: Early literacy assessments

Characteristics of early literacy assessments

Identifying possible early literacy assessments

Discussion of sample characteristics

04

Conclusions and tools

Conclusions from our research-based review of early literacy assessments

Checklists and other tools

01 A research-based assessment review process

REL Midwest developed a research-based process for reviewing and selecting early literacy assessments for use by districts and schools seeking to improve the assessments they administer. This chapter outlines the properties essential to consider when determining whether an assessment is appropriate for your students and develops a common language for this discussion. These properties include the purpose and type of assessment, how the scores can be used, and whether the scores obtained are reliable and valid. The chapter presents research recommendations for which skills should be targeted, when, and how often.

In this section:

- The types of assessments and the purpose of each type
- Reliability and validity
- Early literacy literature and research recommendations

A research-based assessment review process

Building early literacy assessment databases

Putting it all together: Early literacy assessments

Conclusions and tools

This project was initiated to support Chicago Public Schools.

REL Midwest worked in partnership with Chicago Public Schools to set up a *research-based process for reviewing early literacy assessments*, to compile a database of early literacy assessment information, and to identify possible candidates for a common high-quality assessment.

Chicago Public Schools' vision for increasing reading proficiency among their youngest students includes leveraging early literacy assessments.



The project began by acknowledging that some **key factors must be considered to determine whether an assessment should be administered.**

The best assessment for your school depends on:

- What information you need. Many types of assessments yield different information.
- What you plan to do with the scores or how you plan to use them.
- Whether the scores the assessment yields are consistent and valid.
- Other considerations, such as the accommodations and languages needed, the data systems required, and other aspects of feasibility and cost.



Many kinds of assessments have different purposes and properties. Why you need an assessment determines the best type to use.

There is more than one kind of assessment. For example, some tests are designed to determine whether students met a goal. Other tests are designed to provide teachers with informal information. When conducting an assessment review, it is important to begin by considering **what type of assessment and corresponding information you need.**

Screening

Where are students in terms of skill level? Which students might be struggling?

Diagnostic

Where might additional support be needed?

Progress
Monitoring

How are students progressing?

Summative

How well do students meet yearly goals?

Consider who will use the data and how they will do so when determining which type of assessment is needed.

Universal Screener

Brief assessments, typically administered to all students, designed to identify those at risk of failing an outcome.

Performance on a screening assessment can be used to identify students who need further evaluation of skills as well as students who are expected to perform adequately or in an accelerated fashion on an outcome assessment (Foorman et al., 2013).

Diagnostic

An assessment that is typically given to those identified as at-risk on a screening assessment to provide specific information to practitioners about a student's strengths and weaknesses (Foorman et al., 2013).

Progress Monitoring^a

A process to determine the extent to which a student is responding to instruction either within the general curriculum or a specialized curriculum, placements, or instructional approaches.

Interim Assessment

An assessment that evaluates student knowledge and skills relative to a specific set of academic goals. These assessments are typically administered within a limited time frame.

Formative Assessment

A process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievements of intended instructional outcomes.

Benchmark

Benchmark assessments are administered periodically throughout the school year, at specified times during a curriculum sequence, to evaluate students' knowledge and skills relative to an explicit set of longer-term learning goals.

These categories are not mutually exclusive. An assessment can serve multiple purposes.

Summative

An assessment generally administered one time, usually at the end of a school year, to evaluate students' performance relative to a set of content standards. It may easiest to think about summative assessments as a test on learning rather than a test for learning.

a. We have included interim, formative, and benchmark assessments and progress monitoring into one category as each can be used for progress monitoring. However, these assessments types have distinct purposes and definitions.

Note: For more information on the evolution of formative assessments, as well as the definition and purpose of formative assessments, see CCSSO, 2018.

Source: REL Southeast, 2017; Foorman et al., 2013

Another consideration is whether you wish to compare student scores and/or if you would just like to know whether a student mastered a set of learning objectives. These goals require different assessment properties.

To compare student scores, it is essential that the assessment is *standardized or given to all students the same way*.

To compare a student's scores with a set of learning objectives, you need a *criterion-referenced* assessment.

The student is compared with a set of criteria.

To observe how a student scores relative to other students in their grade, you need a *norm-referenced* assessment. *Students' percentile rank is provided.*

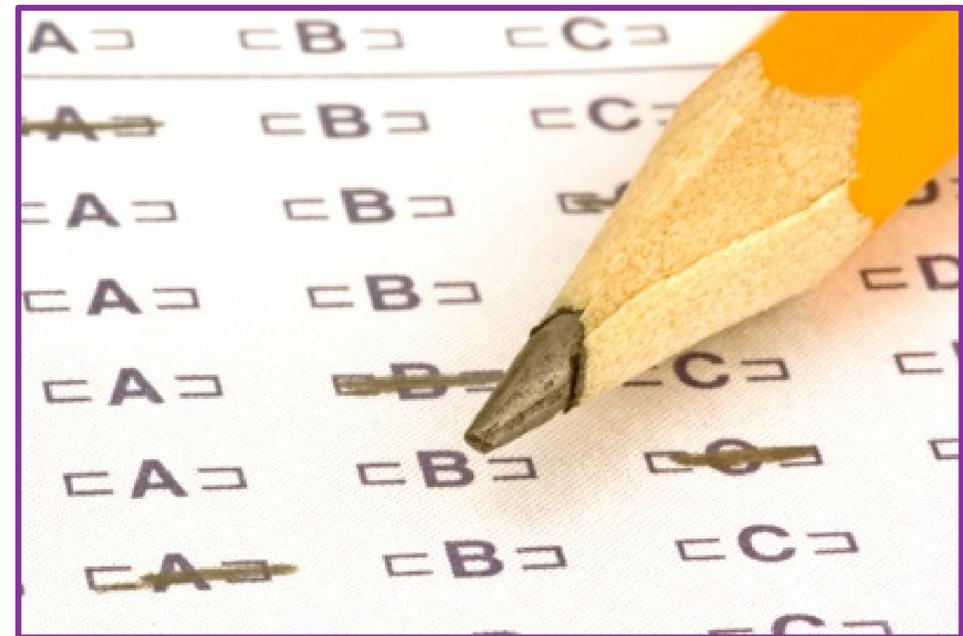


To compare student scores, it is essential that the assessment is *standardized or given to all students the same way.*

Standardized assessments are given to all test takers using the **same set of procedures**. The assessments have **established administration and scoring procedures**. Standardized tests typically include tests administered in a standard way with set guidelines. An example is a computerized test with items from a bank of questions that is administered in the same manner for all students.

Nonstandardized assessments allow for variation in administering, for example to accommodate disabilities, or are based on teacher observation, work samples, or simply have not been formally normed on a sample (for example, teacher-developed assessments).

Assessments that are administered in a “standard” or consistent manner make it possible to compare the relative performance of individual students or groups of students.



To compare a student's scores with a set of learning objectives, you need a **criterion-referenced** assessment.

The student is compared against a set of criteria.

To observe how a student scores relative to other students in their grade, you need a **norm-referenced** assessment.

Students' percentile rank is provided.

Criterion referenced	Norm referenced
Compares a person's knowledge or skills against a standard, a learning goal, or another criterion.	Compares a student's performance against the performance of a reference sample of peers.
Each student's performance is compared directly against the standard.	Does not indicate whether they met or exceeded a specific standard or criterion.
Performance of other students does not affect a student's score.	Distinguishes between high and low performers.
Often has cut scores to place students into categories.	Has a percentile rank for each student.

Some assessments can provide both criterion-referenced and norm-referenced results.

Reliability and validity indicate how confident you can be in the scores.

These psychometrics tell you how consistent and accurate the scores are.

An assessment may seem to be a perfect fit. It is the type of assessment you need, and it provides the information you seek. However, things are not always as they seem.

Every assessment needs to be independently evaluated to ensure that the scores and data it yields are consistent and true.

When looking at an assessment, be sure to have good evidence for reliability and validity.



Every assessment needs to be independently evaluated to ensure that the scores and data it yields are **reliable and valid**. Otherwise, the scores and data are meaningless.

Reliability refers to the consistency of a measure. The three types of reliability are:

1. **test-retest reliability**, which measures consistency over time;
2. **internal reliability**, which measures consistency across items; and
3. **inter-rater reliability**, which measures consistency across different researchers.

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

- Assessments can be reliable but NOT valid if they consistently measure something other than what they were designed to measure.
- For example, a math test that has text-heavy word problems might consistently measure reading skill, not mathematics skills.

For more information about reliability and validity, including the types of reliability and validity, please see [REL Midwest's reliability and validity handout](#).



Three main types of reliability

Reliability type	Definition	Example
Test-retest reliability	The degree to which an individual's scores on an assessment are related to their scores on a subsequent administration of that assessment.	A single test is given to participants on two separate occasions. If the same or similar results are obtained, then test-retest reliability is established.
Internal reliability (Cronbach's alpha)	The relationship among items within an assessment or within assessment domains. The internal consistency measure estimates the degree to which items within an assessment or within assessment domains are related.	A question about the internal consistency might read, "How well do all of the items on the assessment, which are proposed to measure fluency, produce consistent results?" If all items on a test measure the same construct or idea, then the test has internal reliability. Cronbach's alpha is a statistical computation used on a single administration of a test.
Inter-rater reliability	The degree to which different raters produce the same rating on an assessment or assessment domain.	Two raters or observers observe the same behavior independently and compare their data. Reliability is established if their results are similar.

Three main types of validity

Validity type	Definition	Example
Construct validity	The degree to which the relationships among assessment items and domains (for example, components of the assessment) align with the intended construct(s) proposed for scoring and use. This category of analyses may examine the extent to which there is potential bias in items or domain scores across subgroups.	If a reading test is designed to assess oral reading fluency but includes popular nursery rhymes like Mary Had a Little Lamb, then the test may be assessing memory more than reading skills.
Concurrent validity	A type of criterion-related validity that estimates the degree to which the assessment correlates with other independent, reliable, and valid assessments of the same construct.	If a student's score on a brand-new vocabulary test is valid, then it should correlate highly with the same student's score on a well-established vocabulary test like the Peabody Picture Vocabulary Test.
Predictive validity	The degree to which assessment scores predict related measure scores at a later time. Predictive validity is a form of test-criterion validity.	Students' performance on phonological awareness assessments in preschool often predicts later reading achievement in elementary school.

REL Midwest engaged in a research-driven review process.

REL Midwest consulted the literature regarding early literacy skills to learn what skills to measure in young children, what type of assessment to use to do so, and how to evaluate reliability and validity evidence.

The literature explains what skills to measure when, what type of assessments to use, and how to ensure the assessments are reliable and valid.

This literature review helped REL Midwest determine what information to gather from assessments of interest.

The literature indicates that the following skills would best indicate how students in these early grades are doing in terms of becoming competent readers. Difficulty in obtaining these skills predicts difficulty with reading.

Assess these early literacy skills:

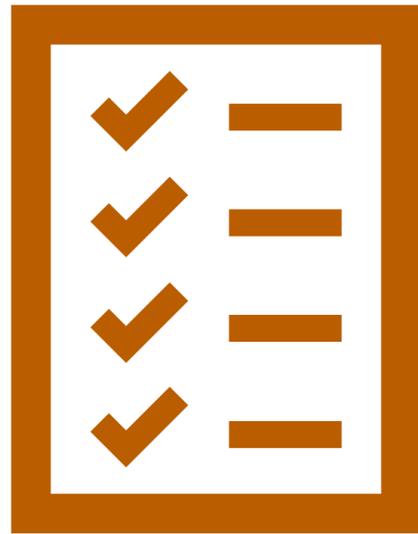
Foundational Reading Skills

- Listening comprehension
- Oral language skills
- Phonological awareness
- Letter identification
- Phonemic awareness
- Phonics
- Fluency

Reading Comprehension Skills

- Vocabulary development
- Broad conceptual knowledge
- Knowledge and abilities needed to comprehend text
- Thinking and reasoning skills
- Motivation to understand and work toward academic goals

Early literacy assessment research recommendations:



Screen all students for potential reading problems at the **beginning of the year**.



Conduct a second screening midyear in **kindergarten** and **grade 1**.

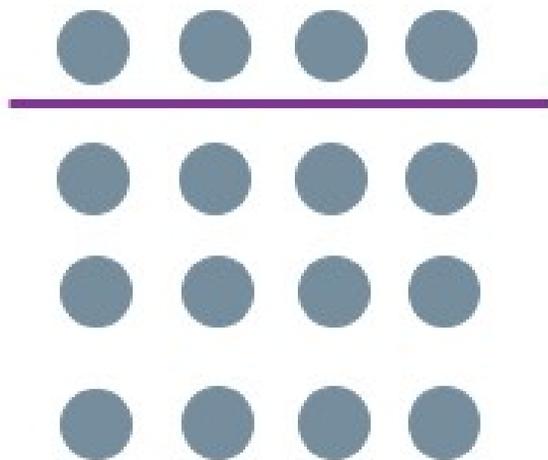


Use screeners with a predictive validity of **.60** or higher.

Seek a screening measure with **balanced degrees of sensitivity and specificity.**

Sensitivity

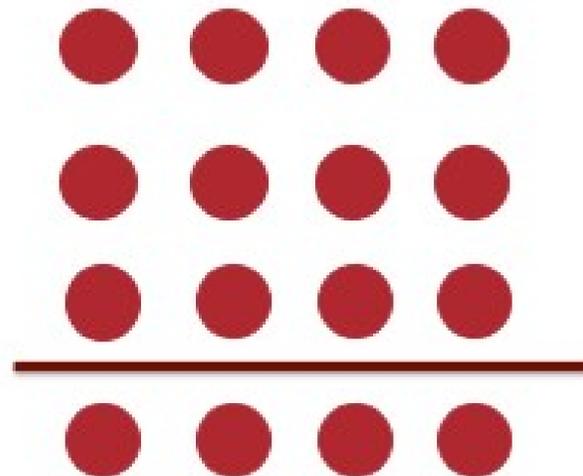
The number of students correctly identified as at risk.



If an assessment's cut point is **too high**, it will have high sensitivity but may have greater numbers of **false positives** and low specificity.

Specificity

The number of students correctly identified as **NOT** at risk.



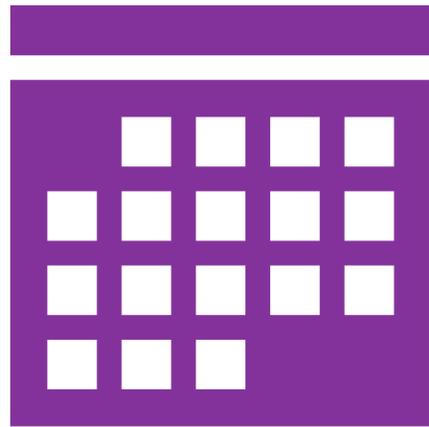
If an assessment's cut score is **too low**, it will have high specificity but may have greater numbers of **false negatives** and low sensitivity.

To balance **sensitivity** and **specificity**:

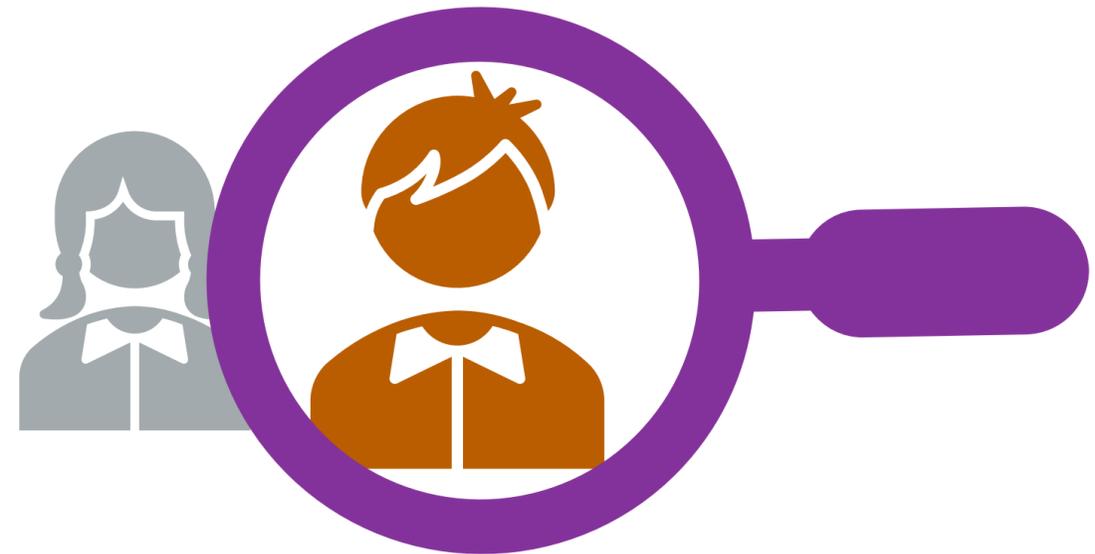
- Carefully **consider the cut score** used to determine risk.
- **Screen at the beginning of the year** for early intervention. Then, because efficient screeners tend to lack good sensitivity, screen **again midyear**.

Early literacy assessments research recommendations

Monitor the progress of struggling readers.



Conduct progress monitoring at least **eight times a year**, ideally monthly (may include informal and formal assessments).



Use progress monitoring data to **identify students** who need additional instruction.

When should these skills be mastered for later reading proficiency? Here is a guide to when to screen and when to employ progress monitoring.

After these skills are mastered, they need not be monitored any longer. Thus, the skills are recommended to be screened and monitored in a developmental progression.

Students should be proficient at (sample skills)	Recommended grade levels to screen			Recommended grade levels for progress monitoring		
	Kindergarten	Grade 1	Grade 2	Kindergarten	Grade 1	Grade 2
Letter naming fluency (Letter name identification and the ability to rapidly retrieve abstract information)	X	X				
Phoneme segmentation (Phonemic awareness)	X	X		X		
Nonsense word fluency (Proficiency and automaticity with basic phonics rule)		X			X	
Word identification (Word reading)		X	X		X	
Oral reading fluency (Reading connected text accurately and fluently)		X	X		X	X

Reading comprehension is an important skill that develops as children become more proficient readers. The Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 framework focuses more on basic skills and vocabulary in kindergarten and grade 1, and then begins to increase the number of reading comprehension items starting in grade 2. However, to best identify struggling readers, it helps to focus on those aspects of decoding text that permit later comprehension, as shown in the table above.

Sources: Gersten et al., 2009; Najarian et al., 2018; Shanahan et al., 2010.

The literature recommends thresholds for reliability and validity. Thresholds enabled REL Midwest to determine which assessments can be considered reliable and valid.

Research recommends examining reliability and validity evidence of early literacy assessments.



A reliability of **.70** or higher



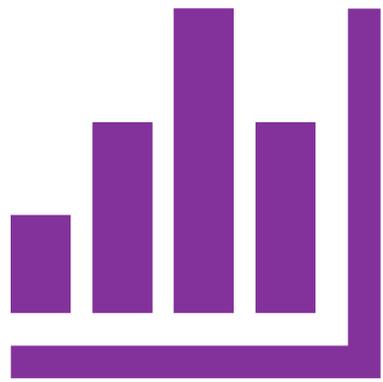
A predictive validity of **.60** or higher

Summary of research on early literacy assessments

- Key foundational reading and reading comprehension skills to measure include:
 - Letter naming fluency
 - Phoneme segmentation
 - Nonsense word fluency
 - Word identification
 - Oral reading fluency
- In the early grades, identifying struggling readers is most important to improving early literacy. **Screening and progress monitoring assessments are key to early identification and support.**
- **When to screen for these skills and when to monitor progress** with respect to these skills in terms of kindergarten, grade 1, and grade 2 depends on the skill.
- Research-based thresholds for **reliability are .70** and **predictive validity are .60.**



In summary, when reviewing assessments there are several things to consider:



The **purpose** of the assessment and **how the data will be used**. **Different types of assessments** yield different information.

- Universal screener
- Progress monitoring
- Diagnostic
- Summative



Are the scores meaningful?

- **Reliability**: measures the consistency of the assessment.
- **Validity**: measures what it intends to measure.



Last, the research recommends that five key skills be screened for and monitored in the early grades using assessments with reliability at .70 or higher and predictive validity a .60 and higher:

- Letter naming fluency
- Phoneme segmentation
- Nonsense word fluency
- Word identification
- Oral reading fluency

02

Building early literacy assessment databases

REL Midwest compiled a list of early literacy assessments.

REL Midwest consulted the early literacy development research to guide their review while simultaneously compiling all relevant information in two databases. One database consists of only psychometric evidence. The other database consists of all other relevant information one might need to consider in choosing an assessment: how long it takes to do so, the type of assessment, the skills assessed, and so on.

In this section:

- Early literacy assessment databases
- Assessment coding for the databases

**A research-based assessment
review process**

**Building early literacy
assessment databases**

**Putting it all together:
Early literacy assessments**

Conclusions and tools

Early literacy assessment databases

To identify early literacy assessments that satisfy research criteria, it was necessary to compile all relevant information in one place.

Our databases provide information on 25 early literacy assessments.

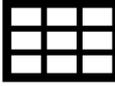
The databases are useful because assessments target different skills in different ways. Each assessment provides different accommodations and requires different data systems and resources. With this information in one place makes, it is easier to review the assessments and determine which ones best align with the research criteria.

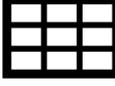


REL Midwest compiled all publisher-provided information for select early literacy assessments in a database so that one could easily compare and contrast assessments.

Assessment name
ACCESS for English Learners
Fountas & Pinnell Benchmark Assessment System (F&P BAS)
Kindergarten Individual Development Survey (KIDS)
mCLASS DIBELS 8th edition
mCLASS DIBELS/IDEL
mCLASS Reading 3D
NWEA Measures of Academic Progress Growth K-2 (NWEA MAP Growth)
NWEA Measures of Academic Progress Reading Fluency (NWEA MAP Reading Fluency)
Strategic Teaching and Evaluation of Progress Assessment (STEP)
GOLD by Teaching Strategies (TS GOLD)
Text Reading and Comprehension (TRC) English/Spanish
Brigance
Education Software for Guiding Instruction (ESGI)
Fast Bridge Early Reading
Fast Bridge CBM Reading
Student Annual Needs Determination Inventory (SANDI)
DYnamic Learning Maps (DLM)
Standardized Test or the Assessment of Reading (STAR) Early Literacy
Acadience (DIBELS Next)
Brigance III IED Norm Ref
Brigance III IED Standardized
Brigance Screens K-1st grade
Developmental Reading Assessment 2nd Edition PLUS (DRA2+)

Two databases were created as part of this project:

 **Early literacy assessments database**

 **Early literacy assessments psychometric database**

These two databases are available upon request. Please email relmidwest@air.org

How were assessments coded and corresponding information entered into the databases?

Developer and publisher administration and technical manuals were the source of determination. For example, to determine the type of assessments, coders relied on what was stated in the administration and technical manuals.

ACCESS Fact Sheet

The purpose of the assessment is to monitor student progress in English language proficiency on a yearly basis and to serve as criterion to aid in determining when English learner students have attained full language proficiency. The assessment provides information to enhance instruction and learning for English learner students.

F&P BAS

The F&P BAS is a formative reading assessment composed of 58 high-quality, original titles, or “little books” divided evenly between fiction and nonfiction. The assessment measures decoding, fluency, vocabulary, and comprehension skills for students in kindergarten through grade 8.

mCLASS: DIBELS Next

Research-based universal screening and progress monitoring assessment that measures the acquisition of early literacy skills from kindergarten through grade 6. The current version was released in 2010 by DMG.

The REL Midwest early literacy assessment database includes information on assessment type.

Assessment	Type of assessment					
	Standardized	Universal screener	Diagnostic ^a	Progress monitoring, formative, interim ^b	Summative	Other (specify)
F&P BAS	Yes	Yes	Yes	Yes	No ^c	Benchmark assessment
mCLASS DIBELS, 8th ed. ^d	Yes	Yes	Yes ^d	Yes	No	Benchmark assessment
NWEA MAP Growth K–2	Yes	Yes	Yes ^e	Yes	No	Benchmark assessment, adaptive
NWEA MAP Reading Fluency	Yes	Yes	No	Yes	No	Benchmark assessment, adaptive
ACCESS for English Learners	Yes	Yes ^f	No	Yes	Yes ^g	
KIDS	No ^h	Yes	No	No ^h	No	Observational, developmental
mCLASS DIBELS/IDEL ^d	Yes	Yes	Yes ^d	Yes	No	
mCLASS Reading 3D (TRC)	Yes	Yes	Yes	Yes	No	Benchmark assessment
Acadience (DIBELS Next) ^d	Yes	Yes	Yes ^d	Yes	No	
STEP	No ⁱ	No	No	Yes	No	
GOLD by Teaching Strategies	No ^h	No	No	Yes	No	Observational, developmental

a. The broader definition of diagnostic assessments was used for this project rather than the more clinical/evaluative definition of diagnostic assessments as those used to diagnose a disability or developmental delay. Diagnostic assessments are defined as assessments typically given to students identified as at risk on a screening assessment to provide specific information to practitioners about a student's strengths and weaknesses. See slide 12 for definitions of assessment types and sources.

b. Progress monitoring includes formative and interim assessments, although there are distinctions in definitions of formative, interim, and progress monitoring assessments.

c. F&P BAS notes on website assessment is summative, however assessment does not meet summative definition

d. All three assessments (DIBELS 8th ed., Acadience Reading/DIBELS Next and IDEL) state that their key purposes are for universal screening, benchmark assessment, and progress monitoring. However, they can provide teachers with information on students' skills in foundational early literacy skills. DIBELS 8th edition can screen for dyslexia risk, but the publisher explicitly states that it does not claim the use of DIBELS 8th ed. in dyslexia diagnosis.

e. Only the NWEA MAP Growth skills checklist is considered diagnostic. A skills checklist is available only for grades K–2.

f. ACCESS is used as a universal screener only for students who are identified as English learner students.

g. ACCESS considered summative assessment as it determines student proficiency in English language skills.

h. KIDS and TS GOLD are not standardized in terms of administration because they rely on observational and portfolio evidence but do have standardized coding. KIDS can be formative/progress monitoring if administered multiple times per year.

i. STEP incorporates both formal and informal observational assessments. The formal assessment is considered standardized; however, the observational assessment is not.

The REL Midwest Early Literacy database also contains evidence for each assessment in terms of reliability and validity. The research standards were used to indicate how strong the evidence is for each assessment.

Assessment	Psychometrics	
	Reliability	Validity
DIBELS 8th ed.	Promising evidence	Promising evidence
Fastbridge Early Reading	Convincing evidence	Promising evidence
Acadience (DIBELS NEXT)	Convincing evidence	Convincing evidence for grades 1–6 No evidence for kindergarten
NWEA Reading Fluency	Promising evidence	Weak evidence
F&P BAS	Promising evidence	Promising evidence
mCLASS DIBELS/IDEL	Weak evidence	Convincing evidence
mCLASS Reading 3D	Convincing evidence	Convincing evidence
NWEA MAP Growth K–1	Convincing evidence	Convincing evidence

Convincing: Two different types of psychometric evidence provided at or above the .70/.60 threshold

Promising: One type of psychometric evidence provided at or above the .70/.60 threshold

Weak: Little psychometric evidence is provided

Promising evidence means that the evidence provided looks good, but more evidence is needed to be convincing.

Other information is provided in the databases including accommodations provided, languages, and what data systems might be needed. Other aspects of feasibility are also included.

Assessment	Factors to consider			
	Accommodations	Languages	Data systems used/needed	Types of professional development (PD) available
F&P BAS	Yes	English Spanish (Systema)	Computer/tablet with Internet connection for teacher to input scores and run reports	Onsite, offsite, and online PD available
mCLASS DIBELS, 8th ed.	Yes	English	Computer/tablet with Internet connection	Onsite, offsite, and online PD available
NWEA MAP Growth K–2	Yes	English/Spanish	Computer/tablet with Internet connection	1–4 hours of training required; online or in-person training available
NWEA MAP Reading Fluency	Yes	English/Spanish Progress monitoring English only	Computer/tablet with Internet connection	1–4 hours of training required; online or in-person training available
ACCESS (for English Learners)	Yes	English	Computer/tablet with Internet connection	Onsite, offsite, and online PD available
KIDS	Yes	English	Access to KIDSTech an online data entry system to enter data	ISBE KIDS training required for all kindergarten teachers
mCLASS DIBELS/IDEL	Yes	Spanish	Computer/tablet with Internet connection	Onsite, offsite, and online PD available
mCLASS Reading 3D (TRC)	Yes	English & Spanish	Computer/tablet with Internet connection	4–8 hours of training required
Acadience (DIBELS Next)	Yes	Spanish Reading	Computer/tablet with Internet connection	4–8 hours of training required
STEP	No info	English K–5/Spanish K–3	Computer/tablet with Internet connection to access Online Assessment and Data Management System	STEP trainer assigned to new school users; class coaching, online and in-person training available
GOLD by Teaching Strategies	Yes	English & Spanish	Computer/tablet with Internet connection for teacher to input scores, artifacts, and run reports	Includes access to online tutorials, access to help from Teaching Strategies staff, archived webinars, and tip sheets

In summary, REL Midwest compiled assessment information into two databases for quick access and analyses.



More than 20 early literacy assessments are currently included in the early literacy assessment databases.

Information on these assessments, including the skills assessed, the type of assessment, scores provided, training needed, and much more, was pulled from publisher manuals and technical reports.

A psychometric database was compiled to provide the following for each assessment:

- Reliability
- Validity
- Scores yielded
- Sensitivity
- Specificity

03 Putting it all together: Early literacy assessments

REL Midwest reviewed the literature and compiled all the information available on the early literacy assessments into databases to identify key assessments.

- Which of the early literacy assessments that REL Midwest reviewed target the research-recommended skills you are interested in?
- What is the reliability and validity evidence for those assessments?
- How do the assessment samples compare with the demographics in your district?

In this section:

- Characteristics of early literacy assessments
- Potential early literacy assessments
 - Identifying possible early literacy assessments
 - Discussion of sample characteristics

To identify assessments that meet research criteria, consider the information found in the literature and search the database to identify assessments for use.

For example, consider assessments that measure foundational skills and that are either screeners or progress monitoring for preK to grade 3 students.

The skills measured by each assessment are provided in the databases, making it easy to identify the assessments that target the foundational skills desired.



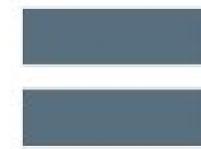
Literature

- Use screeners and progress monitoring assessments.
- Target foundational skills.
- Reliability and validity should be .60 and .70.



Database

- Assessment types
- Skills assessed
- Reliability
- Validity



Identify assessments

Remember, different early literacy assessments measure different skills depending on their type and purpose.

DIBELS, 8th ed.

- Letter Naming Fluency (LNF)
- Phonemic Segmentation Fluency (PSF)
- Nonsense Word Fluency (NWF)
- Word Reading Fluency (WRF)
- Oral Reading Fluency (ORF)
- Maze

FastBridge Curriculum-Based Measures of Early Reading

- Concepts of Print (Screening)
- Onset Sounds
- Letter Sounds
- Letter Names
- Nonsense Words
- Word Blending
- Sentence Reading
- Oral Language (Sentence Repetition)
- Word Rhyming
- Word Segmenting
- Decodable Words
- Sight Words

Acadience Reading

(previously known as DIBELS NEXT)

- Letter Naming Fluency (LNF)
- First Sound Fluency
- Phoneme Segmentation Fluency (PSF)
- Nonsense Word Fluency (NWF)
- Oral Reading Fluency and Retell Fluency (ORF)
- Retell Fluency (RTF)

NWEA MAP Reading Fluency

- Early Literacy Behaviors
- Phonemic Awareness
- Letter Naming
- Phonics and Word Awareness
- Vocabulary
- High Frequency Words
- Fluency
- Comprehension
- Writing About Reading (optional)

Fountas & Pinnell Benchmark Assessment Systems (BAS)

- Early Literacy Behaviors
- Phonemic Awareness
- Letter Naming
- Phonics and Word Awareness
- Vocabulary
- High Frequency Words
- Fluency
- Comprehension
- Writing About Reading (optional)

mCLASS DIBELS Indicadores Dinámicos del Éxito en la Lectura (IDEL)

- Letter Naming Fluency
- Phoneme Segmentation
- Word Identification

NWEA MAP Growth Reading (Screening, Skills Checklists, and Growth)

- Screening: Phonological awareness, letter identification, matching letters to sounds, concepts of print
- Checklists
 - Phonological Awareness
 - Letter Identification
 - Phonemic Awareness: Phoneme Identification
 - Phonemic Awareness: Manipulation of Sounds
 - Phonics: Matching Letters to Sounds
 - Syllable Types: Vowel, Digraphs/Diphthongs
 - Syllable Types: CVC, CVCe, R-Controlled
 - Decoding Consonant Blends/Digraphs
 - Decoding: Spelling Patterns/Word Families
 - Decoding: Multi-Syllable Words, Affixes, Open/C+le
- Growth
 - Foundational Skills
 - Language and Writing
 - Literature and Informational Text
 - Vocabulary Use and Functions

mCLASS Reading 3D TRC

- Retell/Recall
- Reading Record Accuracy
- Oral Comprehension

FastBridge Curriculum-Based Measures of Reading (CBM) Reading

- Oral Reading Fluency

Other Considerations When Choosing Early Literacy Assessments

Other mediating factors, like the language of the test, resources needed to administer the test, feasibility, and cost, determine whether an assessment is a good fit for your students.

Assessment	Factors to consider			
	Accommodations	Languages	Data systems used/needed	Types of professional development (PD) available
F&P BAS	Yes	English Spanish (Systema)	Computer/tablet with Internet connection for teacher to input scores and run reports	Onsite, offsite, and online PD available
mCLASS DIBELS, 8th ed.	Yes	English	Computer/tablet with Internet connection	Onsite, offsite, and online PD available
NWEA MAP Growth K–2	Yes	English/Spanish	Computer/tablet with Internet connection	1–4 hours of training required; online or in-person training available
NWEA MAP Reading Fluency	Yes	English/Spanish Progress monitoring English only	Computer/tablet with Internet connection	1–4 hours of training required; online or in-person training available
ACCESS (for ELs)	Yes	English	Computer/tablet with Internet connection	Onsite, offsite, and online PD available
Kindergarten Individual Development Survey (KIDS)	Yes	English	Access to KIDSTech an online data entry system to enter data	ISBE KIDS training required for all kindergarten teachers
mCLASS DIBELS/IDEL	Yes	Spanish	Computer/tablet with Internet connection	Onsite, offsite, and online PD available
mCLASS Reading 3D (Text Reading and Comprehension)	Yes	English & Spanish	Computer/tablet with Internet connection	4–8 hours of training required
Acadience (DIBELS Next)	Yes	Spanish Reading	Computer/tablet with Internet connection	4–8 hours of training required
STEP (Strategic Teaching and Evaluation of Progress) Assessment	No info	English K–5/Spanish K–3	Computer/tablet with Internet connection to access Online Assessment and Data Management System	STEP trainer assigned to new school users; class coaching, online and in-person training available
GOLD by Teaching Strategies	Yes	English & Spanish	Computer/tablet with Internet connection for teacher to input scores, artifacts, and run reports	Includes access to online tutorials, access to help from Teaching Strategies staff, archived webinars, and tip sheets



Selecting Early Literacy Assessments

- Consider the skills of interest when selecting an assessment. You can use the databases developed by REL Midwest to identify the early literacy assessments that include them.
- The databases will provide the reliability and validity associated with each assessment so that you can determine whether they are at or above the recommended thresholds.



A reliability of **.70** or higher



A predictive validity of **.60** or higher

- Then, use the database to identify other considerations that would affect how well the assessment could work in your district (for example, cost, technological requirements).

Another consideration: Are the students in the samples used to develop the assessments similar to students attending your schools?

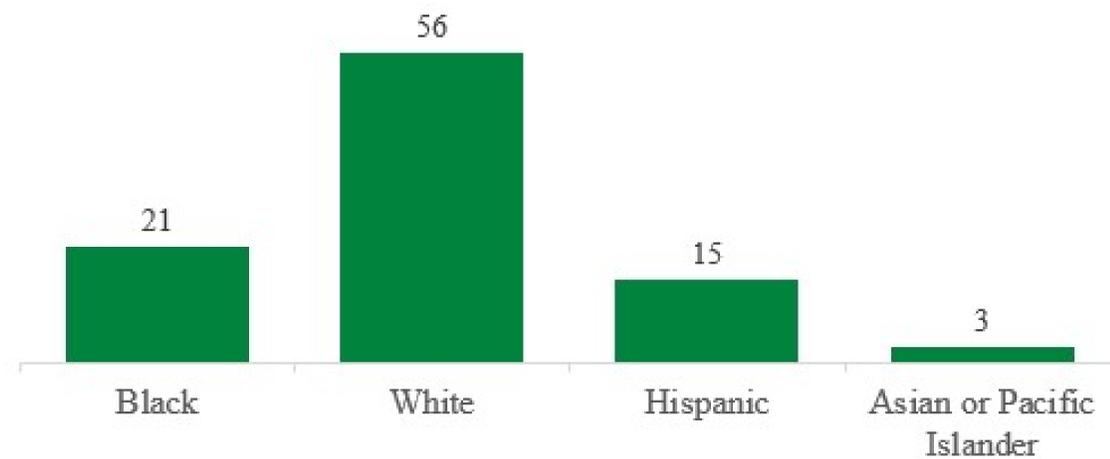


Districts can ask the publisher to provide statistics on the subgroups in their sample.

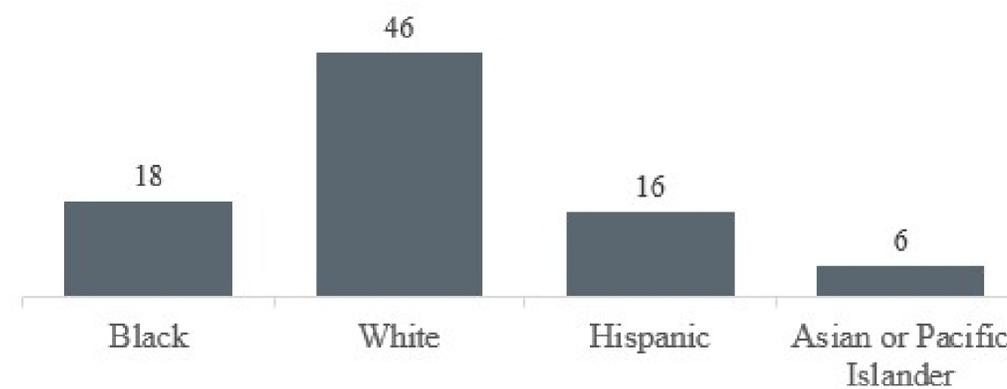
If the answer is no, interpret the findings with caution as they may not generalize to your students.

Here are examples of norming sample demographics used by test developers. Consider how the students of your district compare.

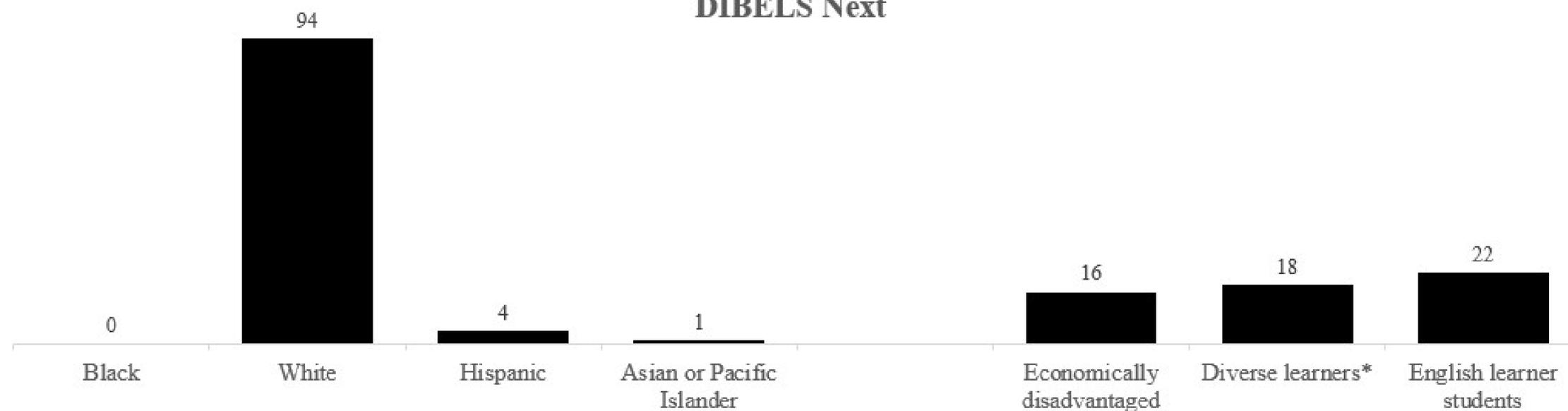
Text Reading and Comprehension



NWEA MAP Growth



DIBELS Next



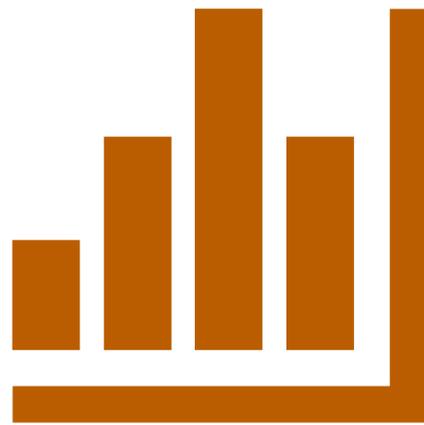
If an assessment's norming samples are not similar to the characteristics of your district's students, then you may want to request the validity and reliability information for subgroups of students.

DIBELS is Dynamic Indicators of Basic Early Literacy Skills. NWEA MAP is Northwest Evaluation Association Measures of Academic Progress.

Note: Diverse learners is a Chicago Public Schools term for students with disabilities or students with individualized education programs. The technical manuals for NWEA MAP growth and TRC did not report special population values.

Source: Technical manuals of NWEA MAP Growth, DIBELS Next, and Text Reading and Comprehension.

When applying publisher norms, be mindful of the extent to which the publisher's sample resembles your students.



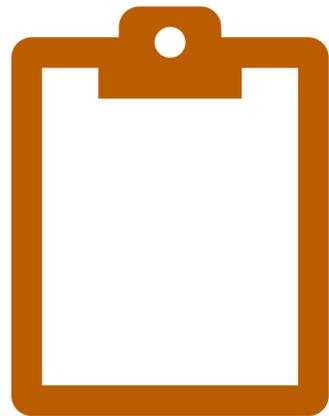
Districts can ask the publisher to provide statistics on the subgroups in their sample.

Districts can conduct analyses on their own assessment data to explore how students are progressing.

If the publisher's sample does not resemble your students:

- Ask the publisher for subgroup data.
- Conduct your own analyses.
- Consider how scores might vary.

In summary, the REL Midwest created Early Literacy Assessment databases to use to identify assessments that satisfy research criteria.



Early literacy skills identified by research-based review:

- Letter naming fluency
- Phoneme segmentation
- Nonsense word fluency
- Word identification
- Oral reading fluency



There are other considerations when choosing an assessment:

- Predictive validity
- Appropriateness for one's students
- Feasibility
- Cost

Consequently, more than one assessment measuring different skills might be warranted.



When your sample does not match the developers' sample:

- Ask for subgroup sample data.
- Look at validity evidence.
- Think about how your students might differ in terms of rate of development.

04 Conclusions and tools

REL Midwest reviewed the literature and compiled all the information available from the publishers on the early literacy assessments reviewed into databases. Next, REL Midwest identified assessments that met most of the research criteria. Different checklists and tools were developed during this process that can be used for assessments in other subject areas.

In this section:

- Conclusions from our research-based review of early literacy assessments
- Checklists and other tools

**A research-based assessment
review process**

**Building early literacy
assessment databases**

**Putting it all together:
Early literacy assessments**

Conclusions and tools

In summary

REL Midwest reviewed the literature to determine what skills should be assessed, by what type of assessment, and with which psychometrics.

Key research recommendations are as follows:

- **Screeners** should be administered in **the beginning and middle of the year** in kindergarten and grade 1.
- **Struggling readers should be monitored often**, ideally monthly, with progress monitoring assessments. These can be informal assessments.
- Screeners and progress monitoring assessments should have a **reliability of .70** and a **predictive validity of .60 or above**.

In summary, REL Midwest employed a research-based review process to identify potential assessments that meet research criteria.

To recap

REL Midwest reviewed the developer and publisher administration and technical manuals for over 20 early literacy assessments, entering all information about the assessments into two large databases.

The databases are useful because having all this information all in one place makes it easier to search on certain criteria, compare and contrast assessments on certain characteristics, and determine which assessments will work best in your district or school.

Information is provided on:

- ✓ **Assessment type**
- ✓ **Scores yielded**
- ✓ **Psychometric information**
- ✓ **Other considerations like cost, feasibility, training, and more!**

To recap

These tools allow you to search the early literacy databases to identify assessments that meet the criteria recommended by the research and the needs of your students.



Literature

- Use screeners and progress monitoring assessments.
- Target foundational skills.
- Reliability and validity should be .60 and .70.



Database

- Assessment types
- Skills assessed
- Reliability
- Validity



Identify assessments

REL Midwest developed tools to assist the review of assessments in any content domain.

REL Midwest employed a research-based review and selection process for early literacy assessments that could be used in any content domain. In identifying key assessments, several checklists were developed that could be used in future assessment reviews. Additionally, this chapter includes a list of resources for assessment information.



REL Midwest uncovered the following important considerations to include when doing an assessment review:



Research literature

What does the research indicate are important skills to assess for your outcome of interest? What does the research indicate are important assessment types for your outcome of interest?



Purpose and type of assessment

What do you want to know?
What is the purpose of doing the assessment?
What type of assessment is best to get the data you need?



Psychometric properties

Does the assessment have convincing or promising evidence of reliability and validity?
Does the norming sample match your population?



Feasibility and cost

Can you use the assessment, and does it fit your students?
Is it affordable?
Do you have the data systems needed?

REL Midwest developed a more detailed checklist to use for future assessment reviews

- What does the research state are important skills to measure?
- What do you want to know? What is the purpose of doing the assessment?
- What type of assessment is best to get the data you need?
- Does the assessment have promising or convincing reliability and validity?
- Is the assessment feasible to use in terms of administration time, professional development, and cost?
- Do you need more information?
- Does the norming sample match your population?

If You Need More Information

- **For the early literacy assessments, review the databases provided by AIR.**
- *Review other key sources:*
 - Refer to the [National Center for Intensive Intervention](#)
 - See summary information from [Annenberg](#)
 - Look at [What Works Clearinghouse](#)

Ask the developer for more information.



In conclusion, REL Midwest developed a research-based process to review and identify early literacy assessments that adhere to research guidelines. This process generated tools to be used with assessments in any domain.



REL Midwest reviewed the literature for information about what skills should be assessed at what grades and which types of assessments are developmentally appropriate.

REL Midwest entered the assessment information into databases. The databases can be searched using research criteria to select assessments that are reliable and valid and that meet the needs of your students.

REL Midwest developed checklists and other tools to help others review assessment in other subject areas.

References

- Council of Chief State School Officers (CCSSO). (2018). *Revising the definition of formative assessment*. <https://ccsso.org/sites/default/files/2018-06/Revising%20the%20Definition%20of%20Formative%20Assessment.pdf>
- Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394.
- Foorman, B. R., Kershaw, S., & Petscher, Y. (2013). *Evaluating the screening accuracy of the Florida Assessments for Instruction in Reading (FAIR)* (REL 2013–008). U.S. Department of Education, Institute of Education Sciences.
- Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade: Educator's practice guide* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/practiceGuide/wwc_foundationalreading_040717.pdf
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*(1), 7–21.
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide*. (NCEE 2009-4045). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuide/3>
- Jhangiani, R. S., Chiang, I. A., & Price, P. C. (2015). Chapter 5: Psychological measurement. In *Research methods in psychology (2nd Canadian ed.)*. BC Campus. <https://opentextbc.ca/researchmethods/chapter/reliability-and-validity-of-measurement/#navigation>
- McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research & Practice, 16*(4), 230–239.

References (continued)

- Najarian, M., Tourangeau, K., Nord, C., & Wallner-Allen, K. (2018). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), Kindergarten Psychometric Report* (NCES 2018-182). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. <https://nces.ed.gov/pubs2018/2018182.pdf>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.
- O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3(2), 159–197.
- Regional Educational Laboratory Southeast. (2017). *Assessment terms used in reading*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/edlabs/infographics/pdf/REL_SE_Assessment_Terms_Used_in_Reading.pdf
- Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, 48(1), 115–136.
- Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuide/14>
- Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education*, 36(4), 223–233.
- What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/handbooks>

Resources

- Amplify. (2018). *mCLASS DIBELS Next vs DIBELS 8 comparison guide*. <https://amplify-com-mktg.imgix.net/app/uploads/2018/10/18215916/mCLASS-DN-vs-D8-comparison-guide.pdf>
- Chicago Public Schools (CPS). (2017). *ACCESS fact sheet*. <https://www.cps.edu/globalassets/cps-pages/academics/student-assessments/accessfactsheet.pdf>
- Fountas & Pinnell. (2012). *Research executive summary*. https://www.fountasandpinnell.com/shared/resources/FP_BAS_2ED_Research_Executive-Summary_v2012-08.pdf
- National Center on Intensive Intervention. (2020). *Academic screening tools chart rating rubrics*. American Institutes for Research. https://intensiveintervention.org/sites/default/files/NCII_AcademicScreening_RatingRubric_2020-06-30.pdf

Thank you!

Contacts

Jodi Jacobson Chernoff
jchernoff@air.org

Jennie Jiang
jjiang@air.org

Arielle Lentz
alentz@air.org