

Connecticut State Board of Education

Hartford

To Be Proposed:

May 7, 2025

Resolved, That the State Board of Education, pursuant to Connecticut General Statutes Section 10-14t:

1. Removes DIBELS 6th Edition and Amira Learning from the Approved Menu of Research-Based Grades K-3 Literacy Universal Screening Assessments;
2. Continues approval of the following early literacy universal screening assessments for a period of three years starting with the 2025-26 school year for use by all districts with all students in Grades K-3:
 - aimsWebPlus Early Literacy and Reading and aimswebPlus RAN;
 - easyCBM;
 - DIBELS 8th Edition paper;
 - mClass DIBELS 8th Edition on Amplify; and
 - Acadience Reading K-6 and Acadience RAN.
3. Approves a pilot of iReady Early Literacy Screening at the Stamford School District; and
4. Directs the CSDE to:
 - a. assist districts currently using DIBELS 6th Edition or Amira Learning to transition to one of the other fully approved assessments;
 - b. collect student benchmark performance data on the literacy universal screening assessments for all students in Grades K-3 from all districts, and report meaningful aggregate results publicly on EdSight; and
 - c. conduct studies to maximize the efficacy of these early literacy assessments to support instruction, inform additional diagnoses, and develop future reading comprehension.

Approved by a vote of _____, this seventh day of May, Two Thousand Twenty-Five.

Signed: _____

Charlene M. Russell-Tucker, Secretary
State Board of Education

Connecticut State Board of Education

Hartford

To: State Board of Education
From: Charlene M. Russell-Tucker, Commissioner of Education
Date: May 7, 2025
Subject: Revised Menu of Grades K-3 Literacy Universal Screening Assessments

Executive Summary

The purpose of this report is to provide the State Board of Education (SBE) with a recommended menu of research-based Grades K-3 literacy universal screening assessments for approval. The Approved Menu of Research-Based Grades K-3 Literacy Universal Screening Assessments shall be used by all districts for the purpose of universal screening for reading of all students in Grades K-3.

Purpose of Grades K-3 Literacy Assessments on the Approved Menu

The Standards for Educational and Psychological Testing developed jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) and published collaboratively by the three organizations since 1966, is considered the gold standard in guidance on testing. These standards define validity as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.” In other words, validity is intricately tied to how the results from the assessment are interpreted and used.

There are many reasons why students are assessed and how those results are used. A teacher may want to know if their students can isolate, blend, segment, add, delete, and substitute phonemes; a district/school leader may want to know who should be identified as gifted/talented; the Connecticut State Department of Education (CSDE), the SBE, legislators, and other community stakeholders may want to know which districts consistently reflect low student achievement statewide. **One single assessment cannot serve many purposes.**

Connecticut General Statutes Section 10-14t, in part, states that the literacy assessments in the approved menu for Grades K-3 are to be used “...to identify students in kindergarten to grade three, inclusive, who are below proficiency in reading.” The statute further requires that the assessments be brief, be evidence-based with proven psychometrics, measure a range of critical reading skills, provide opportunities for formative assessments at least three times a year, produce data that are useful for informing individual and classroom instruction, be compatible with best practices in reading instruction and research, and assist in identifying, in whole or in part, students at risk for dyslexia, or other reading-related learning disabilities.

In summary, the intended purpose of the state approved K-3 assessments according to state law is to screen all students in Grades K-3 three times a year to identify students who are below proficiency in reading and may be at risk for dyslexia or other reading disabilities. The purpose of these assessments is not, in and of themselves, to make a formal *diagnosis* for reading disabilities,

but to identify students *at-risk* for those reading disabilities. The students identified by these assessments as being *at-risk* may need additional assessments to pinpoint areas of targeted literacy development and reading instruction with frequent progress monitoring, and/or may need more *diagnostic* assessments to make a formal diagnosis.

The assessments on the K-3 approved menu are “universal screeners.” They are administered to the population i.e., universally. They are a “screeener” because they are designed to work like a sieve and identify a subset of the population who may have a greater likelihood of the condition for which the population is being screened i.e., reading disabilities. There are numerous examples of universal screeners in other disciplines like medicine. For example, the [pediatric screening schedule established by the American Academy of Pediatrics](#) incorporates assessments for body measurements, sensory screenings, and maternal screenings that are administered to all children and mothers at specified frequencies during those critical early years. A screener is not intended to make a particular diagnosis. It is, however, designed to point to areas that may need more in-depth assessment and possibly treatment. To make a formal diagnosis, *diagnostic* assessments will be necessary.

Re-Examining the Assessment Menu

In 2023-24, management responsibility for the Grades K-3 literacy universal screening assessments was transferred to the Performance Office within the CSDE. This change consolidated all assessment management under one office (the Performance Office), bringing the K-3 literacy screenings in line with how other assessments are handled and to ensure consistent practices across all grade levels and assessment types.

Historically, the Performance Office has handled all aspects of Connecticut’s student assessments. Staff in the Office have significant experience with assessment development, psychometrics, data collection, reporting, interpretation, and assessment evaluation. In recent years, the Office has also become involved in assessments used as part of teacher licensure. Over the past two decades, the Office has successfully shepherded numerous statewide student assessments through independent and rigorous peer reviews required by the U.S. Department of Education. Therefore, when the Office was assigned this new responsibility for K-3 literacy assessments, it recommended that though the CSDE had previously conducted thorough evaluations of all assessments on the menu, it would be best to commission an external and independent re-examination of the assessments by researchers with expertise in K-3 reading assessments and instruction. The Office decided to leverage the CSDE’s existing contract with the University of Connecticut that had partnerships with experts at Michigan State University, and Utah State University.

At the same time, the CSDE also conducted a district survey in April/May 2024 to learn about the assessments chosen by districts. The data revealed that most districts (73%; $n=129$) used DIBELS 8/mClass DIBELS 8. Smaller percentages used the other assessments: 11% ($n=20$) used aimswebPlus; 3% ($n=5$) used Acadience; 3% ($n=5$) used DIBELS 6th edition; 2% ($n=3$) used Amira Learning; 2% ($n=3$) used easyCBM, while approximately 7% ($n=13$) had not finalized their selection.

During the survey and in the subsequent months, the CSDE received requests from some districts to consider additional assessments for inclusion on the menu. Since the external review was underway, the CSDE asked the faculty to also review those additional assessments.

Findings And Recommendations

The full report submitted by the external reviewers is attached. Based on the findings from the external reviewers, and the purposes outlined in Connecticut General Statute Section 10-14t, and after considering the recommendations set forth in appendix g of the final report of the task force established pursuant to Special Act 19-8, the following changes to the menu are recommended for SBE approval.

Discontinue DIBELS 6th Edition

The DIBELS 6th Edition assessment was discontinued by the vendor many years ago. At the June 2022 meeting of the SBE, the CSDE recommended and the SBE approved the removal of this assessment. However, there was significant pressure from the advocacy community to reinstate the DIBELS 6th Edition assessment. Subsequently, in February 2023, the SBE approved the reintroduction of DIBELS 6th Edition into the menu.

The faculty experts conducting the independent review requested technical manuals and materials from vendors of all assessments including the DIBELS 6th Edition. The vendor indicated that the DIBELS 6th Edition had been discontinued, and no technical manuals were available. The vendor recommended that schools transition from DIBELS 6th Edition to DIBELS 8th Edition due to considerable updates including significantly higher classification accuracy, updated materials, updated benchmark data, reduced testing time, increased accuracy for identifying students' risk for dyslexia, greater equivalence across reading passages, and less error due to text difficulty.

Therefore, the DIBELS 6th Edition is being removed from the menu being presented for SBE approval.

Two Assessments Need Critical Evidence

According to the reviewers, one assessment on the prior menu, Amira Learning, demonstrates face validity but lacks critical evidence of concurrent validity. Without that evidence, the CSDE is unable to have the fullest confidence that this assessment will identify the right number of students and the right students who are below proficiency in reading and/or at risk for dyslexia or other reading disabilities. Therefore, Amira Learning is being removed from the approved menu being presented for SBE approval. The CSDE will work with the handful of districts implementing this assessment to transition to another fully approved assessment.

Another assessment, MAP Fluency, that is not currently on the menu but recommended by districts for external review, also demonstrates face validity but lacks critical evidence of concurrent validity. As with Amira Learning, without this evidence, the CSDE cannot recommend this assessment for inclusion.

Continue Approval of Five Assessments

Five assessments that were previously on the menu continue to demonstrate technical adequacy, are deemed to be fully approved, and are being recommended to remain on the menu for three

years. Districts can continue to use these assessments or switch to one of them. The five assessments are:

- aimsWebPlus Early Literacy and Reading and aimswebPlus RAN;
- easyCBM;
- DIBELS 8th Edition paper;
- mClass DIBELS 8th Edition on Amplify; and
- Acadience Reading K-6 and Acadience RAN.

The revised menu is attached. The individual subtests within each of the assessment suites may be amended periodically based on vendor recommendations, district feedback, and ongoing research. Note: The DIBELS 8th Edition paper version does not contain a vocabulary measure. Currently, districts using DIBELS 8th Edition paper need to separately procure the vocabulary measure that is part of mClass DIBELS 8th Edition on Amplify. The extent to which this is happening is unclear. Over the next three years, the CSDE will explore the feasibility of helping districts currently using the DIBELS 8th Edition paper assessment to transition to the mClass DIBELS 8th Edition on Amplify or one of the other approved assessments.

One New Assessment Recommended for a Limited Pilot

According to the reviewers, one assessment not on the current menu – iReady Early Literacy Screening – but requested by the Stamford School District demonstrates technical adequacy. However, given its newness in Connecticut, the CSDE does not recommend its full addition to the menu at this time. Instead, the CSDE recommends a three-year pilot of this assessment at the Stamford School District. This limited pilot will allow CSDE staff to gain a deeper knowledge of the specific subtests to be administered within this assessment suite, in different grades and at different times during the assessment window. A pilot will also enable the CSDE to evaluate the efficacy of this assessment and examine its comparability to other measures. In the third year of the pilot, the CSDE will be better positioned to make a recommendation regarding its full inclusion on the menu.

Opportunities for Improvement

An assessment is useful only in so far as the results from that assessment are used to bring the right interventions and supports to students. Over the past year, as the Performance Office has communicated about the assessments and interacted with districts, several opportunities for improvement in communication and local practices have been identified.

For example, some districts asked if the mandate applied to all districts or just Priority or Alliance Districts. Others asked if students should be tested on all the grade-appropriate subtests. Even others questioned if all students needed to be tested three times a year or if it was sufficient to only assess those who were at-risk in the Fall assessment.

Through the course of the past year, the CSDE has reiterated the expectation in writing and in numerous conversations with practitioners across the state that all students in Grades K-3 should be tested three times a year on all the grade-appropriate subtests.

Some districts asked about the timing of each of the three windows (i.e., Fall, Winter, and Spring). After internal and external consultations, the CSDE established and communicated a standard window for all districts. The CSDE has also created a website for posting all information related to the K-3 universal screening assessments.

Though the assessment administration requirements and policies are being formalized and disseminated, without student-level data collection and reporting, the CSDE cannot truly monitor and support district practice.

There is wide variability across districts in the extent to which results from the assessments are used effectively. While many districts have participated in the Science of Reading Masterclass or other professional learning and are engaged in the difficult work of helping their teachers to utilize the assessment results and make the necessary instructional shifts, others are continuing to use additional assessments that do not align to the science of reading and using results from unaligned assessments to inform their instructional approaches or for sharing results with parents. These anecdotes have highlighted the need for continued professional learning and policies from the CSDE in the effective use of results from the approved menu of screeners.

Data Collection, Reporting, and Research

Currently, there is no comprehensive database of Connecticut student performance on the assessments from the approved menu. The external reviewers had to rely on data and studies conducted by the assessment vendors that may or may not have included any data from Connecticut. Therefore, the CSDE cannot determine how well any of the approved assessments on the menu are performing with Connecticut students.

As permitted in Connecticut General Statutes Section 10-10a and explicitly authorized in subsection (d) of Section 10-14t, the CSDE in consultation with the external reviewers from the University of Connecticut and Michigan State University, will be collecting student benchmark performance data on the literacy universal screening assessments for all students in Grades K-3. The CSDE expects to report aggregate results through its EdSight data portal in 2026.

As suggested by the external reviewers, these data will enable the CSDE to examine the various profiles of students in districts across the state that are based on their performance on the universal screening benchmark assessments to better inform “the subtests and cut points that maximize classification at each time frame.” They will allow the CSDE to study the relation of these early literacy assessments to subsequent proficiency on the Smarter Balanced assessments, and “evaluate the effectiveness of different reading screening suites in the early grade levels for future reading comprehension.” These early literacy data will also inform the CSDE’s Early Indication Tool.

Conclusion

In summary, the attached SBE resolution:

1. Removes DIBELS 6th Edition and Amira Learning from the Approved Menu of Research-Based Grades K-3 Literacy Universal Screening Assessments;

2. Continues approval of the following early literacy universal screening assessments for a period of three years starting with 2025-26 for use by all districts with all students in Grades K-3:
 - aimsWebPlus Early Literacy and Reading and aimswebPlus RAN;
 - easyCBM;
 - DIBELS 8th Edition paper;
 - mClass DIBELS 8th Edition on Amplify; and
 - Acadience Reading K-6 and Acadience RAN.
3. Approves a pilot of iReady Early Literacy Screening at the Stamford School District; and
4. Directs the CSDE to:
 - a. assist districts using DIBELS 6th Edition or Amira Learning to transition to one of the approved assessments on the menu;
 - b. collect student benchmark performance data on the literacy universal screening assessments for all students in Grades K-3 from all districts, and report meaningful aggregate results publicly on EdSight; and
 - c. conduct studies to maximize the efficacy of these early literacy assessments to support instruction, inform additional diagnoses, and develop future reading comprehension.

Prepared by: _____
Katherine Seifert, Associate Education Consultant

Revised by: _____
Abe Krisst, Bureau Chief

Approved by: _____
Ajit Gopalakrishnan, Chief Performance Officer

Approved Menu of Research-based K-3 Literacy Universal Screening Assessments 2025-26



Pursuant to [Section 10-14t\(a\)](#) of the Connecticut General Statutes, all public school districts (e.g., local and regional boards of education, regional educational service centers, charter schools) serving students in Kindergarten to Grade 3, inclusive, must select and administer assessments from this *Approved Menu of Research-based Grades K-3 Literacy Universal Screening Assessments* to **all** K-3 students enrolled in the district. Administering a set of screening measures in the primary grades helps identify students who are below proficiency in reading, at risk for reading difficulties and require intervention, and assists in identifying, in whole or in part, students at risk for dyslexia, or other reading-related learning disabilities. The following five screening assessments demonstrate technical adequacy for the aforementioned purposes outlined in Section 10-14t, and after considering the recommendations set forth in appendix g of the final report of the task force established pursuant to Special Act 19-8, these assessments are recommended for continued listing on the menu. Districts can continue to use these assessments or switch to one of them.

- aimsWebPlus Early Literacy and Reading and aimswebPlus RAN;
- easyCBM;
- DIBELS 8th Edition paper;
- mClass DIBELS 8th Edition on Amplify; and
- Acadience Reading K-6 and Acadience RAN.

Based on where students are in their reading skills development, different aspects of critical reading skills become most appropriate to include when using a component-based approach to early reading assessment. Thus, some reading areas may not be appropriate to assess for all students (e.g., reading comprehension in the beginning of Kindergarten) as indicated by the shaded areas in the charts below. Each publisher assessment protocol indicates the grade-specific reading skills that are appropriate to assess in each grade.

Approved Kindergarten Menu

Reading Measure	aimswebPlus Early Literacy and Reading aimswebPlus RAN ¹	easyCBM	DIBELS 8th Edition (paper)	mClass DIBELS 8th Edition on Amplify	Acadience Reading K-6 ³ ; Acadience RAN ⁴
Phonological and Phonemic Awareness	<ul style="list-style-type: none"> • Phoneme Segmentation • Initial Sound Fluency 	<ul style="list-style-type: none"> • Phoneme Segmenting 	<ul style="list-style-type: none"> • Phoneme Segmentation Fluency 	<ul style="list-style-type: none"> • Phoneme Segmentation Fluency 	<ul style="list-style-type: none"> • First Sound Fluency • Phoneme Segmentation Fluency
Phonics	<ul style="list-style-type: none"> • Letter Word Sounds Fluency • Word Reading Fluency • Nonsense Word Fluency 	<ul style="list-style-type: none"> • Letter Sounds • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency
Fluency					
Vocabulary	<ul style="list-style-type: none"> • Auditory Vocabulary 			<ul style="list-style-type: none"> • Vocabulary² 	<ul style="list-style-type: none"> • Word Use Fluency - Revised⁵
Comprehension					
Rapid Automatic Name (RAN) or Letter Name Fluency	<ul style="list-style-type: none"> • RAN Objects <i>or</i> • RAN Colors <i>or</i> • RAN Shapes 	<ul style="list-style-type: none"> • Letter Names 	<ul style="list-style-type: none"> • Letter Naming Fluency 	<ul style="list-style-type: none"> • RAN Numbers² 	<ul style="list-style-type: none"> • RAN Objects, Letters and Numbers⁴

Please reference the [Footnote](#) on page 5.

Approved Grade 1 Menu

Reading Measure	aimswebPlus Early Literacy and Reading aimswebPlus RAN ¹	easyCBM	DIBELS 8th Edition (paper)	mCLASS DIBELS 8th Edition on Amplify	Acadience Reading K-6 ³ ; Acadience RAN ⁴
Phonological and Phonemic Awareness	<ul style="list-style-type: none"> • Phoneme Segmentation 	<ul style="list-style-type: none"> • Phoneme Segmenting 	<ul style="list-style-type: none"> • Phoneme Segmentation Fluency 	<ul style="list-style-type: none"> • Phoneme Segmentation Fluency 	<ul style="list-style-type: none"> • Phoneme Segmentation Fluency
Phonics	<ul style="list-style-type: none"> • Letter Word Sounds Fluency • Word Reading Fluency • Nonsense Word Fluency 	<ul style="list-style-type: none"> • Letter Sounds • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency
Fluency	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Passage Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency
Vocabulary	<ul style="list-style-type: none"> • Auditory Vocabulary 			<ul style="list-style-type: none"> • Vocabulary² 	<ul style="list-style-type: none"> • Word Use Fluency - Revised⁵
Comprehension					
Rapid Automatic Name (RAN) or Letter Name Fluency	<ul style="list-style-type: none"> • RAN Objects <i>or</i> • RAN Colors <i>or</i> • RAN Shapes 		<ul style="list-style-type: none"> • Letter Naming Fluency 	<ul style="list-style-type: none"> • RAN Numbers² 	<ul style="list-style-type: none"> • RAN Objects, Letters and Numbers⁴

Please reference the [Footnote](#) on page 5.

Approved Grade 2 Menu

Reading Measure	aimswebPlus Early Literacy and Reading aimswebPlus RAN ¹	easyCBM	DIBELS 8th Edition (paper)	mCLASS DIBELS 8th Edition on Amplify	Acadience Reading K-6 ³ ; Acadience RAN ⁴
Phonological and Phonemic Awareness					
Phonics			<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency
Fluency	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Passage Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency
Vocabulary	<ul style="list-style-type: none"> • Vocabulary 	<ul style="list-style-type: none"> • Vocabulary 		<ul style="list-style-type: none"> • Vocabulary² 	<ul style="list-style-type: none"> • Word Use Fluency - Revised⁵
Comprehension	<ul style="list-style-type: none"> • Reading Comprehension 	<ul style="list-style-type: none"> • Proficient Reading 	<ul style="list-style-type: none"> • Maze 	<ul style="list-style-type: none"> • Maze 	
Rapid Automatic Name (RAN) or Letter Name Fluency	<ul style="list-style-type: none"> • RAN Objects <i>or</i> • RAN Colors <i>or</i> • RAN Shapes 			<ul style="list-style-type: none"> • RAN Numbers² 	

Please reference the [Footnote](#) on page 5.

Approved Grade 3 Menu

Reading Measure	aimswebPlus Early Literacy and Reading aimswebPlus RAN ¹	easyCBM	DIBELS 8th Edition (paper)	mCLASS DIBELS 8th Edition on Amplify	Acadience Reading K-6 ³ ; Acadience RAN ⁴
Phonological and Phonemic Awareness					
Phonics			<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	<ul style="list-style-type: none"> • Nonsense Word Fluency • Word Reading Fluency 	
Fluency	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Passage Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency 	<ul style="list-style-type: none"> • Oral Reading Fluency
Vocabulary	<ul style="list-style-type: none"> • Vocabulary 	<ul style="list-style-type: none"> • Vocabulary 		<ul style="list-style-type: none"> • Vocabulary² 	<ul style="list-style-type: none"> • Word Use Fluency - Revised⁵
Comprehension	<ul style="list-style-type: none"> • Reading Comprehension 	<ul style="list-style-type: none"> • Proficient Reading 	<ul style="list-style-type: none"> • Maze 	<ul style="list-style-type: none"> • Maze 	<ul style="list-style-type: none"> • Maze
Rapid Automatic Name (RAN) or Letter Name Fluency	<ul style="list-style-type: none"> • RAN Objects <i>or</i> • RAN Colors <i>or</i> • RAN Shapes 			<ul style="list-style-type: none"> • RAN Numbers² 	

Footnotes

The individual subtests within each of the assessment suites may be amended periodically based on vendor recommendations, district feedback, and ongoing research.

¹aimsweb Plus RAN – This is an add on assessment to the aimsweb Plus suite of assessments.

² mCLASS additional screeners that are included as part of mCLASS (Listed by mCLASS as optional screening measures).

³Acadience is a trademark of Acadience Learning Inc. The DIBELS Next copyrighted content is owned by Acadience Learning Inc. The DIBELS® and DIBELS Next registered trademarks were sold by Acadience Learning Inc. to the University of Oregon (UO) and are now owned by the UO.

⁴Use Acadience RAN (Print version available for download if using print version of Acadience Reading).

⁵Word Use Fluency-Revised (WUF-R) is available as an experimental measure.

Approved Spanish Assessments

Districts are required to administer a standard English K-3 Literacy Universal Screening Assessment to all students. Students who are enrolled in a dual language program or are receiving literacy instruction in their native language should also be assessed in their native language, if available. The rationale for administering a native language assessment in addition to the English assessment, is to identify students who may be at risk for reading difficulties, regardless of their language of instruction. Administering the assessments in both languages can help teams to determine if the potential deficit in literacy skills is an indication of reading difficulties or if it is due to the stage of the student's English language acquisition and development.

The following three Spanish K-3 Literacy Universal Screening Assessments demonstrate technical adequacy and may be used by districts:

- AimswebPlus – Spanish
- easyCBM
- mClass Lectura

Independent Researcher Review of Commercial Reading Screening Assessment Suites May 2025

Prepared for Connecticut State Department of Education,
Performance Office

PREPARED BY: Adrea Truckenmiller, PhD, Michigan State University

Michael Coyne, PhD, University of Connecticut

Katherine Valentine, MA, Michigan State University

Pamella Moura, MS, Michigan State University

Cherish Sarmiento, PhD, Utah State University

Authors' Note.

Adrea Truckenmiller, PhD, NCSP is an Associate Professor of Special Education at Michigan State University. She has co-authored more than 40 publications on reading and writing development and the use of assessment for guiding evidence-based instruction. She co-led the development of five research-based reading and writing assessments which were funded by the U.S. Department of Education and she is on the Florida Center for Reading Research Wall of Fame. She regularly partners with schools and states to transform assessment to instruction.

Michael Coyne, PhD is Professor and Department Head in the Neag School of Education at the University of Connecticut. He has directed seven federally-funded research grants focused on beginning reading assessment, instruction, and intervention and he is currently the principal investigator of the National Research Network on Multi-tiered Systems of Support (MTSS). Dr. Coyne has worked closely with the Connecticut State Department of Education over the last 12 years to support state-wide efforts to improve reading outcomes for all students.

Katherine Valentine, MA is a graduating PhD candidate at Michigan State University and an Assistant Professor of School Psychology at Florida State University. Her research, publications, and practice expertise include implementing, evaluating, and training educators to use literacy assessments as part of systematic decision making in schools.

Pamella Moura, MS is a doctoral student in the Special Education program at Michigan State University. She has extensive experience with multilingual learners as a teacher, administrator, and multilingual speaker. Her research focuses on using assessments to inform school intervention systems, particularly in determining special education programming for linguistically diverse students.

Cherish Sarmiento, PhD is an Assistant Professor in the Teacher and Educational Leadership program at Utah State University. Her research and practice expertise includes innovating vocabulary assessment, reading and writing intervention within science content, and formative assessment to support teachers' instructional decision making.

ATTESTATION. Authors of this report claim no conflict of interest with any of the vendors or products reviewed in this report. This report was commissioned as an independent review and solely reflects the expert opinions of the authors.

Independent Researcher Review of Commercial Reading Screening Assessment Suites

The **purpose of reading screening** is to identify which students would likely benefit from supplemental reading instruction to meet grade level standards by the end of the school year (Albers & Hall, 2023). Administering reading screening is a best practice when it leads to the provision of evidence-based supplemental instruction that is provided early enough (i.e., grades K through 3) to facilitate reading success and prevent students' later frustration in school and widening achievement gaps (e.g., Blachman et al. 2014; Connor, 2019).

A related purpose of reading screening assessment is to detect early signs of risk for dyslexia and other reading difficulties. This places a privilege on selecting assessments that measure components of reading (e.g., phonics, language comprehension) that are associated with reading difficulties. The two most common difficulties in reading are characteristics associated with dyslexia and developmental language disability (Adlof & Hogan, 2019; Catts et al., 2005). Reading skills associated with risk of dyslexia include phonemic awareness, phonics, fluency, and rapid automatic naming or letter name fluency. Assessments that use curriculum-based measurement in reading typically meet the criteria of being brief measures that are sensitive to earlier risk factors associated with dyslexia (Kilgus et al., 2014). Reading skills associated with developmental language disability (DLD) are vocabulary and oral language (Adlof & Hogan, 2019). Many students with or at risk for dyslexia also have co-occurring instructional needs associated with DLD. Therefore, assessment suites will likely be most effective to inform instructional planning when the assessment includes phonics and language comprehension (e.g., Connor, 2019).

Historically, the reading assessments used most in schools could detect some broad reading comprehension difficulties but did not pick up most reading difficulties until grade 3 or later (e.g., Lonigan & Burgess, 2017). Although current commercial reading screening suites provide more accurate and instructionally relevant information, they vary widely in how they were developed, the aspects of reading they focus on, the way they elicit responses from students, and the evidence collected in research on the assessment. Reading screening suites also provide a variety of other information, making it very challenging for schools to know which aspects of assessment to prioritize when choosing and using assessments.

This review was conducted to describe the research evidence for eight reading screening suites to help schools and districts in Connecticut select assessments that meet high standards of

rigor and also align with their unique needs and contexts. In this review, evidence was evaluated for three purposes of screening assessments.

1. The primary purpose of reading screening is to predict which students are likely to not meet end-of-year expectations for reading development. This requires evidence from research studies showing how well the screening score predicts end-of-year expectations (called classification accuracy).
2. Another purpose of reading screening is to detect the most common reading difficulties. Screening assessments need evidence that they measure students' development in the components of reading associated with dyslexia and DLD. In this review, we sought evidence for the degree to which reading screening suites measure the components identified in Connecticut statute (phonemic awareness, phonics, fluency, vocabulary, oral language, reading comprehension, and rapid automatized naming or letter naming fluency).
3. To supplement the screening process in biliteracy for multilingual learners, students can be assessed in their home language. In the current review, we describe the evidence available for Spanish versions of the reading screening suites.

Request for Review

In Connecticut, a list of approved reading screening assessment suites was released on July 1, 2023. In summer 2024, the Connecticut State Department of Education (CSDE), Performance Office commissioned an updated review of the assessments on the 2023 list and additional assessments requested by schools in Connecticut. This review was commissioned as an independent review by experts in reading screening assessment and research on reading development. We collected and evaluated evidence for each of the three purposes listed above.

Evidence Criteria for the Review

The evidence provided by each of vendors was evaluated with the criteria described below and defined in Appendix A.

Connecticut Review Criteria

The review for the updated 2024 list followed similar criteria to the previous 2023 review. The review included seven criteria from Connecticut General Statutes Sec. 10-14t Sec. 10-14t, which states: “Compilation of and guidance re reading assessments for students in kindergarten to grade three. Such reading assessments shall:

- (1) be brief,
- (2) be evidence-based, as defined in 20 USC 7801(21), with proven psychometrics for validity,

- (3) measure oral language, phonemic awareness, phonics, fluency, vocabulary, rapid automatic name or letter name fluency and reading comprehension,
- (4) provide opportunities for formative assessments at least three times, in the fall, winter and spring, during each school year,
- (5) produce data that is useful for informing individual and classroom instruction, including the grouping of students based on such data and the selection of instructional activities based on data of individual student response patterns during such progress monitoring,
- (6) be compatible with best practices in reading instruction and research, and
- (7) assist in identifying, in whole or in part, students at risk for dyslexia, as defined in section 10-3d, or other reading-related learning disabilities.”

National Center for Intensive Intervention Review Criteria

In the 2023 review, Connecticut used the technical standards from the National Center for Intensive Intervention (NCII) to evaluate assessment psychometrics defined in #2 above. The current review also employed the NCII technical adequacy criteria, with a few clarifications to ensure consistency in rating and one adaptation to fit the context of Connecticut. These clarifications include specifying the criterion that the screener predicts to, the cutscore on the criterion that the screener predicts to, and evidence of validity to the reading components that Connecticut specifies in their statute. The NCII criteria are detailed in Appendix A, with the additional clarifications and one adaptation highlighted.

Reviewing Spanish Supplemental Subtests

Connecticut schools value the linguistic strengths of multilingual learners in their schools. Assessment in a student’s home language can provide additional information to predict their later reading comprehension achievement (Mancilla-Martinez et al., 2010). It is important to note that the Spanish subtests may provide additional information but does not replace the use of English reading screening for multilingual learners because English word recognition remains the largest predictor of later reading comprehension for multilingual learners (e.g., Mancilla-Martinez et al., 2010; 2020). Accordingly, CSDE expressed interest in supporting schools to use high quality Spanish subtests as a supplemental piece of information, but the Spanish results should not supplant the English results for predicting instructional needs in English. In the current review, we sought evidence of reliability for Spanish subtests and how the assessment was normed. We used several of the same criteria from the NCII rubric.

Request for Information to Vendors

Requests for information were sent to the current vendors for the assessments on Connecticut’s 2023 approved list, which includes: Acadience, aimsweb, Amira, DIBELS 6th edition, DIBELS 8th edition/mClass, and EasyCBM. School districts requested additional reviews of FastBridge, i-Ready, and NWEA MAP Fluency. School districts also sought approval to use Spanish screening measures to supplement English screening measures. Therefore, additional requests for information were sent to the three new vendors and to seek technical

information about the Spanish measures associated with each screening suite. Vendors provided the review committee with technical manuals and example reports.

Results

The committee reviewed the materials provided by vendors using the rubric in Appendix A. Consensus scores for each of the reviewed assessments are included in Appendix B. To aid schools in their decisions and further their expertise in evaluating reading screening assessment, we also provide the exact evidence for each suite on the three purposes (screening, components of reading, supplemental Spanish assessment). The specific technical evidence for each of the strengths and weaknesses is listed in Appendix C and summarized in narrative text below.

Reading suites with minor weaknesses

1. **DIBELS 8th edition with mClass and Lectura** has many strengths that set it apart from other suites.
 - a. **Screening:** DIBELS 8th cut points match Connecticut’s reading proficiency rate (40th percentile) and has the highest rates of classification accuracy in K. The word reading fluency task further increases accuracy and reduces testing time. However, it is not clear how or if mClass vocabulary and oral language can be part of the composite score prediction to end-of-year outcomes.
 - b. **Components of reading:** A notable strength is that mClass includes measures of vocabulary, oral language, and spelling that all have adequate correlations with external measures.
 - c. **Supplemental Spanish assessment:** At this time, Lectura demonstrates greater evidence of technical adequacy and a more robust norming sample of Spanish speakers than other Spanish assessments.
 - d. Special note about DIBELS 6th Edition – Connecticut’s 2023 list included DIBELS 6th edition as an approved assessment suite. We requested technical manuals and materials for the DIBELS 6th Edition from the vendor. The vendor indicated that DIBELS 6th edition has been discontinued. Technical manuals were not available from the vendor nor publicly available online and therefore could not be reviewed. The vendor recommends schools transition from DIBELS 6th edition to DIBELS 8th edition due to [considerable updates](#). We concur with the recommendation, particularly regarding these updates:
 - i. Significantly higher classification accuracy, updated materials, and updated benchmark data that follows the latest reading research and statistical methodology.
 - ii. Reducing the amount of testing administration time by decreasing the number of oral reading fluency passages at each benchmark from three to one.

- iii. Increasing the accuracy for identifying students' risk for dyslexia by updating and validating subtests with the Comprehensive Test of Phonological Processing (CTOPP-2), which is the gold standard assessment for characteristics associated with dyslexia.
- iv. The passages for oral reading fluency and Maze have greater equivalence across passages and less error due to the text difficulty of the passage.

2. easyCBM

- a. **Screening:** The cut points in easyCBM are aligned with Connecticut proficiency levels and cut points predict to SBAC in grade 3. The assessment also integrates vocabulary. The classification accuracy statistics listed for easyCBM (see Appendix C) may be artificially higher than the other assessment suites because the outcome measure being predicted by the fall easyCBM score is the spring score on the same test (easyCBM score).
- b. **Components of reading:** Compared to other assessments, the vocabulary measure shows superior evidence of aligning with a gold standard measure of vocabulary. Weaknesses include poor evidence of concurrent validity for Word Reading Fluency and an absence of an oral language measure. Technical reports do not include reliability data or standard error of measurement (important for identifying quality of the forms and progress monitoring).
- c. **Supplemental Spanish assessment:** A small norming study was conducted on the Spanish measures. Additional evidence of technical adequacy is not yet available

3. Acadience

- a. **Screening:** Cut points are aligned with Connecticut proficiency levels (40th percentile) and a linking study was conducted with the SBAC. However, sensitivity rates suggest that it will miss many students in grade K.
- b. **Components of reading:** Acadience provides evidence of concurrent validity with gold standard measures of reading in early grade levels, including a measure of vocabulary (Word Use Fluency). Like other assessment suites, a weakness is that Acadience does not include an oral language measure in the screener. Acadience does have diagnostic assessments in Phonemic Awareness, Word Reading & Decoding, Comprehension, Fluency, Oral Language, and Spelling. However, information for these diagnostics were not submitted for review.
- c. **Supplemental Spanish assessment:** Evidence of technical adequacy and the norm sample is not yet available for the Spanish subtests of Acadience.

4. Aimsweb plus

- a. **Screening:** The inclusion of Word Reading Fluency and an optional spelling measure makes Aimsweb plus appealing for use in guiding instruction for students with and at risk for dyslexia. The current cut-scores are not set at Connecticut's base rate of proficiency, but Aimsweb plus allows local norms to be set. The use of local norms makes Aimsweb plus highly appealing.

- b. **Components of reading:** There is no vocabulary measure for grades K and 1. Aimsweb plus also needs additional evidence of validity data for the individual subtests.
- c. **Supplemental Spanish assessment:** Aimsweb plus Spanish subtests have many strengths including evidence of technical adequacy, inclusion of a listening comprehension measure, and a robust norming sample that included Spanish speakers.

Additional reading suites with minor weaknesses

5. FastBridge

- a. **Screening:** FastBridge has an extensive suite of CBM measures. They also have a CAT measure called aReading. Only the CBM measures were reviewed for this report. Screening accuracy is set to SBAC in grade 3 and align to Connecticut's base rate of proficiency. Screening accuracy rates are slightly lower than ideal rates (i.e., 0.80), but the rates are similar to other high quality CBM assessments.
 - b. **Components:** FastBridge has an extensive suite of CBM measures available which include multiple aspects of vocabulary and an oral language measure. Each show evidence of validity with gold standard measures.
 - c. **Supplemental Spanish assessment:** Evidence of technical adequacy and the norm sample is not yet available for the Spanish subtests of FastBridge
6. **i-Ready Early Literacy Screening** which includes the i-Ready Diagnostic for Reading + Fluency Measure
- a. **Screening:** Predicts well to the DIBELS 8th composite cut score of some risk (yellow to green cut point). Information was also provided for predicting to the at-risk cut point for DIBELS 8th (red to yellow cut point), however, those classification accuracy statistics were too low. We recommend using the scores that predict to the "some risk" cut point. The correlations of the i-Ready scores in grades K through 2 to later performance on the SBAC in grades 3 and 4 is promising. It also outperforms other suites in terms of sensitivity rates.
 - b. **Components of reading:** Strengths include spelling items and a reading comprehension score, particularly for informational text. Additional evidence of external validity is needed for vocabulary and comprehension. Similar to other measures, there is not a separate oral language score.
 - c. **Supplemental Spanish Assessment:** Evidence of technical adequacy and the norm sample is not yet available for the Spanish version of i-Ready.

Reading suites that require additional evidence

7. Amira

- a. **Screening:** In grades K through 2, the cut points of Amira and its reported classification accuracy statistics are calibrated for predicting MAP Growth, which is likely not capturing important aspects of early reading development. Researchers were heavily involved in the creation of the assessment and it contains items that have face validity, so Amira may show evidence of accuracy in the future. To accumulate that evidence, Amira needs a study on the K, 1, and 2 measures predicting to a gold standard assessment. We also could not evaluate the norming sample, as it was not adequately described.
 - b. **Components of reading:** Amira includes vocabulary, spelling, and listening comprehension, although evidence of validity for these individual measures are not provided. A unique strength of Amira is the study that shows its high classification accuracy for screening for specific language impairment.
 - c. **Supplemental Spanish Assessment:** The Spanish assessment has many strengths including evidence of technical adequacy and a robust norming sample of bilingual speakers.
8. **NWEA MAP Fluency** is a suite of CBM subtests that is completely separate from the MAP Growth suite which is a CAT. The current review is of the MAP Fluency suite and not the MAP Growth suite.
- a. **Screening:** In grades K through 2, the cut points of MAP Fluency and its reported classification accuracy statistics are calibrated for predicting MAP Growth, which is likely not capturing important aspects of early reading development. It is also calibrated to the 10th percentile of performance on MAP Growth, which is likely to miss many students who need support. The items in MAP Fluency have face validity, so MAP Fluency may show evidence of accuracy in the future. To accumulate that evidence, MAP Fluency needs a study on the K, 1, and 2 measures predicting to a gold standard criterion.
 - b. **Components of reading:** Evidence of external validity for the individual measures is not provided.
 - c. **Supplemental Spanish assessment:** The Spanish assessment parallels the English assessment but has uniquely created items. The norming study has adequate numbers and demographics to match the diversity of the United States. However, it is unclear if the norms for the Spanish items were normed on Spanish speakers.

A Significant Difference in the Current Review Criteria

Other reviews of reading screeners (e.g., NCII) do not specify which measures of reading that vendors should aim to predict, especially in grades K through 2. Therefore, NCII reviews accept any end-of-year criterion measure that the vendor chooses to predict to. However, the accuracy of a reading screener depends entirely on what end-of-year criterion assessment the screening assessment predicts to and the variations of reading skills the criterion measures (e.g.,

Keenan & Meenan, 2014; Klingbeil et al., 2018). For example, in grades 3 through 8, the screening assessment typically predicts to proficiency on a state test. Vendors conduct studies to recommend a cut score (also called a benchmark score) on the screening assessment that optimizes the accuracy of the screening assessment predicting to that state test. All accuracy statistics reported by vendors (and evaluated by independent reviewers) are based on the cut score that the vendors calculate in their norming studies. Two assessments could have adequate accuracy statistics but are very different because they predict to different outcomes. Therefore, it matters what end-of-year test is being used and that education agencies define their valued outcome measure.

For grades 3 and up, the current review sought accuracy statistics that predict to the Smarter Balanced Assessment Consortium (SBAC) but also accepted statistics if they predicted to multiple states' tests. There is widespread agreement that state tests are appropriate and valid end-of-year reading outcomes to use for calculating classification accuracy statistics for reading screeners in grades 3 and above. For grades K through 2, however, there is ongoing debate in the research community about what measures can be considered a "gold standard" reading outcomes at the ends of grades K through 2 to use for calculating classification accuracy statistics. Given converging evidence that most measures of overall reading in grades K-2 (i.e., reading comprehension) are not sensitive enough to early signs of dyslexia and word decoding difficulties, reading researchers recommend that end-of-year outcome measures in grade K through 2 should assess word decoding (e.g., Foorman et al., 2015; Petscher et al., 2019). Vendors' selection of end-of-year outcome measures to use for calculating classification accuracy statistics for K-2 reading screeners is highly variable, with most selecting assessments that have evidence of measuring decoding. However, there is one outcome measure that does not have evidence of measuring word decoding – NWEA MAP Growth. Therefore, we excluded the NWEA MAP Growth measure as an acceptable external validity measure for calculating classification accuracy statistics for reading in Grades K and 1. We excluded the NWEA MAP Growth assessment for three reasons.

1. Although some items in MAP Growth have face validity evidence that they measure phonemic awareness and phonics in grades K-2, we were unable to locate any data to confirm that the items are weighted enough to show evidence of concurrent validity with external measures. The technical manual for MAP Growth does not provide correlations to any external measures of phonemic awareness, phonics, or fluency.

2. The MAP Growth scale is based on its robust evidence of validity for later grade levels. The RIT scaled score for MAP Growth has excellent evidence of external validity and classification accuracy for most students in Grades 3 and up predicting to SBAC. For Grades K through 2, NWEA extended the RIT scaled score down so that schools could monitor growth on the same scale throughout schooling. **This would be an ideal feature of MAP Growth except that the downward extension of the scale into grade K through 2 is not supported by evidence.** Research consistently shows that there are problems with this approach because

assessments need to measure different components of reading in grades K through 2 than in grades 3 and above (e.g., Foorman et al., 2015; 2017). Reading is commonly multidimensional in the early grade levels and unidimensional in later grade levels (e.g., Foorman et al., 2018; Lonigan & Burgess, 2017). Therefore, separate types of reading measures should be used in the early grades compared to later grades. The grade at which this switch occurs is debatable as the range of development in grades 3 through 5 is wide.

3. Likely due to the issue described above, the MAP Growth assessment has very poor sensitivity in Grade K and Grade 1. The [National Center for Intensive Intervention tools chart](#) shows that MAP Growth detects only 33% of students who perform in the lowest 10th percentile at the end of grade K and detects only 77% of students who perform in the lowest 10th percentile in Grade 1. In other words, MAP Growth misses many of the students with or at-risk of dyslexia, which is a key demographic the legislation was designed to address.

The implication of using the MAP Growth assessment as an outcome in vendors' studies of classification accuracy is that cut scores on screening measures are based on predicting to an outcome (MAP Growth) that does not appropriately assess word decoding and therefore may miss early signs of word decoding difficulties. Therefore, the classification accuracy statistics reported by vendors that predict to MAP Growth are not reliable. This is not to say that screening assessments that predict to MAP Growth will never be accurate. Rather, vendors need to collect additional evidence by conducting studies that predict to a different end-of-year outcome that is more appropriate for students in grades K-2.

Considerations and Recommendations

1. **Curriculum-based measurement (CBM) compared with computer adaptive testing (CAT).** There is a common misconception that CAT style assessments (e.g., iReady, iStation, NWEA MAP Growth, Star) are inferior to CBM style measures (e.g., Acadience, Aimsweb, DIBELS, EasyCBM). In the context of early reading screening, however, CBMs and CATs have different strengths. A CBM will often be more accurate for measuring fluency and a CAT will often be better (and more efficient) for measuring vocabulary and later, reading comprehension. CBMs have the advantage of teachers being able to hear students' strengths and weaknesses as students read aloud, while CATs have the advantage of using Item Response Theory to more appropriately weight certain items. Furthermore, most assessment suites are beginning to use both CAT and CBM to leverage the strengths of each. The purpose of reading screening is to ensure that students receive effective intervention. To accomplish this, we encourage vendors and state education agencies to attend to research on early reading assessment-to-instruction practices (see Connor, 2019; Connor & Morrison, 2016) to evaluate the best combination of scores (e.g., decoding and language comprehension) that both predict student outcomes and inform effective instruction.

2. **Screening versus diagnosis.** It should also be noted that CBMs and CATs for reading screening will never be as reliable as individually administered diagnostic assessments used for the identification of learning disabilities (e.g., CTOPP-2, CELF-5, Woodcock-Johnson). If a student scores below benchmark on the screening assessment, that is not a diagnosis of dyslexia or developmental language disorder or a qualification for special education. Screening is designed to be brief so that it can be administered to all students. The brief nature sacrifices some accuracy and coverage of all the complexities of reading. No assessment is 100% accurate 100% of the time. Screening measures must balance specificity and sensitivity. When specificity rates are lower, schools should plan for overidentifying 10% to 20% of students who score below the benchmark. This means that, if a student is receiving decoding intervention but they show they have mastered decoding, the school should have a process for exiting those students from the intervention. When sensitivity rates (% of students who needed support and were correctly identified by the screener) of an assessment are below 90%, schools will also under identify students who would benefit from instruction. This is a harder issue to address and, arguably, more important (Fletcher et al., 2020). It can be addressed by selecting assessments that maximize sensitivity rates and by potentially raising the cut point on the screener. Maximizing sensitivity rates often comes at a cost to specificity. Luckily, screeners with lower specificity rates (% of students who did not need support but were incorrectly identified by the screener) can be addressed by schools having a decision rule for exiting a student from intervention early if it is clear they do not need the intervention (see Compton et al., 2010).
3. **Improving Classification Accuracy.** The actual classification accuracy of screening in local school contexts may vary widely from what vendors report in their norming studies. The state education agency can aid schools in evaluating the accuracy of their current assessment systems by calculating the percentage of students correctly identified as at-risk on the screener compared with proficiency on the SBAC state test. It is possible that different cut points and different configurations of score types will increase the accuracy of screening (Barrett et al., 2014; Klingbeil et al., 2018; Lemke et al., 2023). Very few vendors used Smarter Balanced as the outcome measure for determining cut scores on screening measures and calculating classification accuracy. With its expertise in large scale assessment and literacy, the CSDE is in the unique position to help determine cut score(s) and subtests that are more sensitive for schools and districts across Connecticut (i.e., schools that have high proficiency rates and schools that have lower proficiency rates). With a longitudinal database of the same students across grade levels K through 5, the CSDE can also evaluate the effectiveness of different reading screening suites in the early grade levels for predicting future reading comprehension.
4. **A note about selecting outcome measures.** In the evaluation criteria, we recommended that screening assessments should be identifying the same percentage of students as needing intervention as the Connecticut base rate of proficiency on the SBAC in grade 3.

There is ongoing conversation about what percentage of students should be identified by screeners as well as conversation about the content of the end-of-year assessment in grades K through 2. For example, if a school has 40% of their students performing below proficient on the grade 3 SBAC, it may not actually be desirable to have their screening assessment identify 40% of students in K as performing below benchmark on the screener. Skills develop rapidly in the early grade levels, and it is possible that using the 40th percentile will overidentify students (Fuchs & Fuchs, 2017). More research is needed to explore the intended and unintended results of the use of screening assessments and the nuances of screening decisions that school employ to adjust reading instruction. Research practice partnerships between state agencies and researchers are an important way for accelerating what is known about reading development and scaling effective practices in assessment, instruction, and intervention (e.g., Foorman et al., 2010).

5. **Updating assessment practices.** Research on reading development continues to evolve and screening assessments continue to become incrementally more accurate. Therefore, we recommend using assessments that continue to make updates based on evidence. For example, the DIBELS 8th edition is intended to replace the DIBELS 6th edition. Although DIBELS 6th edition had adequate psychometric properties as previously reviewed by CSDE and NCII, there are enhancements in DIBELS 8th edition that significantly improve accuracy and include subtests that align to the most recent research on reading development and assessment practices. For example, Initial Sound Fluency in grade K had poor sensitivity rates and was removed from DIBELS 8th edition. Therefore, we recommend that the CSDE assist schools using the DIBELS 6th edition to transition to the DIBELS 8th edition. Regardless of what assessment schools use, schools will need periodic support in updating their assessment practices. This will require technical resources and funding for training. For an analysis of time, costs, and procedures for implementation of screening processes, see Barrett et al. (2024).
6. **Increasing expertise in language comprehension.** The evidence for measuring oral language and vocabulary varies widely across the assessment suites (see Appendix C), with most assessments not providing evidence of concurrent validity with gold standard measures of vocabulary. Only two assessment suites have **oral language measures**, and only Amira has studied the validity of their oral language measure. All assessment suites could make improvements in this area. **Vocabulary** is an important component of early literacy that is highly predictive of reading at all grade levels (Foorman et al., 2015; 2018). It is important to measure vocabulary, in addition to decoding, for guidance in selecting appropriate interventions in grades K through 3 (Adlof & Hogan, 2019; Connor, 2019; Truckenmiller et al., 2020). Only one assessment suite (easyCBM) has a vocabulary measure that has been evaluated for concurrent validity and is completely integrated into the benchmark assessment administration and reports for grades K through 3. Other assessments (Acadience, DIBELS 8th edition, Lectura) have vocabulary measures that have evidence of validity, but the vocabulary subtest is provided as an

option. In the case of DIBELS 8th edition, schools must elect to purchase and administer the mClass vocabulary. It is currently unclear how much schools use these assessments.

- a. We recommend that CSDE assist schools in funding, training, administering, and interpreting vocabulary measures for those schools that have not previously administered these measures. Although this may increase the amount of time spent in assessment for those schools, we believe that it will provide important useful information. See recommendation #4 for ways to mitigate excessive intrusion into instruction time.
 - b. Vocabulary and oral language are key predictors of reading difficulties, especially for students with developmental language disabilities. Instruction in vocabulary and oral language are also needed for a large number of students with dyslexia (Adlof & Hogan, 2019). Vendors should consider conducting new research on classification accuracy for both decoding and language comprehension (oral language, vocabulary). Vendors should also consider providing practical support to interpret both students' decoding and vocabulary/oral language needs in research-supported ways to balance instruction (Connor, 2019). CSDE might also conduct their own independent research to assist schools with interpreting profiles of readers in both decoding and language comprehension to support more effective intervention.
 - c. Research continues to evolve on the best ways to measure vocabulary and oral language in ways that inform effective instruction. For example, each assessment suite includes measures that assess vocabulary using a different format. CSDE may consider running pilot trials to evaluate the accuracy, assessment time, and instructional utility of the measures included in the assessment suites. CSDE might also consider running pilot trials of other screeners of oral language that show evidence of validity. See Bao, Komesidou, & Hogan, 2024 for a list of high-quality vocabulary/oral language screeners.
7. **Reducing testing time.** Assessing all the constructs listed in CT legislation three times per year could extend testing time beyond what is generally recommended for screening measures (Fletcher et al., 2020). Current research suggests that reading screeners administered in K to 2 assess three somewhat separable factors: decoding, vocabulary/language comprehension, and fluency (Connor & Morrison, 2016; Foorman et al., 2015; 2017; Lonigan & Burgess, 2017; Petscher et al., 2021). These three dimensions are also associated with different implications for choosing interventions (Truckenmiller et al., 2020). Therefore, it may be most efficient and effective to administer just the subtests that have the highest sensitivity/AUC at a specific time points and cover the three factors. However, additional studies with the SmarterBalanced assessment and Connecticut's base rate of proficiency will be needed to support that recommendation. CSDE and/or vendors may consider conducting profile studies (e.g., Foorman et al., 2017) or Classification and Regression Tree analyses (e.g., Koon et al.,

2014) to choose the subtests and cut points that maximize classification accuracy at each time frame.

- a. Early literacy skills develop rapidly in early elementary grade levels, so some measures have very limited developmental time frames when they are useful.
 - b. For example, assessment of overall reading comprehension is most useful in grades 3 and up. However, in Grades K through 2, assessment of reading comprehension is not sensitive enough to identify students who need supplemental instruction. Therefore, the opportunity cost (time and resources that could be reallocated to instruction) to administer a reading comprehension measure to all students in grades K to 3 may be too high.
 - c. As another example, measures of RAN are most useful for students very early in their reading development (i.e., students who know very few letter sounds). For students who know letter sounds, any measure that includes fluency (e.g., nonsense word fluency, word reading fluency, oral reading fluency) will better capture the same cognitive process as RAN. Measures of phonemic awareness are similar; they are very sensitive in PreK and Kindergarten. However, when students know most of their letter sounds, phonemic awareness is captured well by measures of word reading fluency.
8. **Schools do not need to use the same suite of assessments across all grade levels.** Schools often choose an assessment suite because it does particularly well in one grade band. For example, many schools use NWEA MAP Growth to monitor yearly growth on one consistent scale from grades Kindergarten to 10. Because reading development is so different in early grades compared with later grades, schools should feel empowered to use different assessment suites in the early grades than the later grades. In other words, there is no advantage to selecting DIBELS 8th for all students in grades K through 8 because it does better at the early grade levels and no advantage of selecting NWEA for all grades because it is most accurate in later grades.
9. **Selecting Interventions and Designing Supplemental Instruction.** Because the purpose of screening is to provide information that leads to more effective instruction many vendors offer interventions or online practice activities as part of the assessment package or for an additional purchase. We recommend that schools make decisions about purchasing interventions separately from decisions about purchasing screening assessments. Although assessments evaluated in this review have evidence of validity for screening, instructional components or interventions associated with these assessments may or may not have evidence of efficacy. Therefore, we caution schools who already use effective evidence-based interventions from replacing those interventions with potentially less effective practices just because they are associated with an assessment suite.
10. **Multilingual learners.** It is important for schools to assemble a team of multi-lingual experts and multi-lingual school and community partners to fully understand the effective use, and potential misuse, of English screening assessment scores for improving

instruction and outcomes for multilingual learners. To more acutely represent students' vocabulary knowledge, for example, conceptual scoring (i.e., counting students' responses as correct when the correct word is supplied in English or Spanish) can be used (Hwang et al., 2019). New assessments (e.g., [Multitudes](#)) are being developed that elicit conceptual scoring and also measure the important aspects of word reading.

Appendix A.

Technical Review Standards from the National Center for Intensive Intervention (2023). Updates for the 2024 review are in tracked changes.

Reliability in Scoring:		
Evidence of test reliability and internal consistency reliability	<p>Results of reliability studies are reported for each grade assessment.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> studies are appropriate given the purpose of the measure; and for each grade-level, studies provide evidence of: <ul style="list-style-type: none"> split-half reliability, coefficient alpha, test-retest reliability; classification consistency 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS --most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. Correlations demonstrate ranges of .7 or higher. (2)</p>
	<p>Standard error of measurement (SEM) or standard estimate of error is reported.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> SEM estimates reported for score ranges and cut-scores; and SEM estimates are reported for score ranges and cut-scores for each assessment (grade-level, form, subtest). 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS --Information and data provided suggests acceptable or strong evidence. (2)</p>
	<p>Inter-rater reliability studies have been conducted. The group of raters used to establish inter-rater reliability is representative of test administrators.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> inter-rater reliability studies have been conducted for each grade level and are based on a representative 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS --most information for the criterion is</p>

	<p>sample of educators who will administer and score the assessment; and</p> <ul style="list-style-type: none"> inter-rater reliability coefficients that exceed .7. <p>Studies have been conducted to establish reliability with all subcategories of students who will take the assessment.</p> <p>Evidence Includes:</p> <ul style="list-style-type: none"> reliability established from scoring samples of students, i.e., non-English learners (ELs) with and without reading deficiencies and ELs with and without reading deficiencies (gender, English learner status, special needs status, socioeconomic status, and race). 	<p>provided. Information and data provided suggests acceptable or strong evidence. (2)</p> <p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
<p>Alternative forms available for multiple assessments with demonstrated equivalence or comparability</p>	<p>If alternative forms are provided, all forms have demonstrated evidence of equivalence or comparability. Technical reviews indicate all forms for each grade level have demonstrated evidence of comparability and content specifications.</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> sufficient forms are provided to allow for progress monitoring between interim assessments; and split-half reliability and coefficient alpha reliability. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence correlations demonstrate ranges of .7 or higher. (2)</p>
CONTENT AND CONSTRUCT VALIDITY:		
<p>Evidence of content and construct validity</p>	<p>Evidence reported to demonstrate the assessment helps correctly identify students with “<i>significant reading deficiencies</i>” so that successful remediation and intervention can be provided. Studies have been conducted with similar assessments to show that the assessment measures reading ability, not other irrelevant criteria.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests</p>

Standard	Description	Rating
	<p>Evidence includes the provision of:</p> <ul style="list-style-type: none"> • a clear description is provided that demonstrates the purpose of the assessment is to screen students for reading concerns; and • content specifications for each grade-level, including a complete description of the test content, purpose(s), and intended use(s), and assessment blueprint as appropriate. 	acceptable or strong evidence. (2)
	<p>There are studies of construct validity, such as convergent and discriminant analysis, demonstrating significant indicators of relationship (i.e., correlations of .7 or above). Update: Vendors should provide evidence that their measure correlates with other measures of each of the subskills specified in CT Gen Stat § 10-14t. (2023): oral language, phonemic awareness, phonics, fluency, vocabulary, rapid automatic name or letter naming fluency, and reading comprehension.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
Evidence of criterion/predictive validity accurately identifying students with “significant reading deficiency”	<p>Evidence reported to demonstrate that the assessment has established criterion and/or predictive validity to correctly identify students with and without a “significant reading deficiency.”</p> <p>Evidence includes:</p> <ul style="list-style-type: none"> • a clear definition of the criterion or measure that were used to establish concurrent validity; and • studies with similar assessments that demonstrate the assessment measures reading ability, not other irrelevant criteria; and • predictive validity correlations above .7. 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>

Standard	Description	Rating
Determination of cut-scores based upon well-designed pilot study and standard-setting process	<p>The assessment has established cut-scores for decision making about students' "<i>significant reading deficiency</i>" using adequate demographics (e.g., ELs, free and reduced-price meals), appropriate criterion assessment, adequate sample size, and appropriate statistics.</p> <p>Evidence indicates:</p> <ul style="list-style-type: none"> • a description of the process used to establish the cut points; Point of clarification – vendors should define the cut score on the screener and the associated cut-score on the outcome measure • a full description of the norming sample; and • the norming sample is a large representative national sample of students at the same grade level and is representative of the testing population according to gender, EL status, special needs status, socioeconomic status, and race. • Update: The outcome that the screening predicts to needs to measure word reading. One of the outcome cut scores should be aligned with Connecticut's base rate of proficiency in grade 3 (i.e., 40th to 50th percentile).¹ 	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
	<p>Studies of classification accuracy analysis provide evidence that the measure appropriately identifies students as indicated in the description of purpose of the assessment, demonstrating values that exceed .8 or higher.</p>	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p>

¹ Rationale: Several studies have shown that the accuracy of a reading screening cut score depends on the base rate of proficiency in the location where the assessment is used (Albers & Hall, 2023; Grapin et al., 2017; Kilgus et al., 2014; Klingbeil et al., 2018; Lemke et al., 2023).

		<p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
	Acceptable, recognized procedures are followed for setting cut-scores.	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence.(0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
	There is guidance for cut-score for score interpretation. Point of clarification: The report should include some designation or categorization of scores (e.g., some risk/40 th percentile, at-risk/20 th percentile) AND defines the decision that can be made for that category (e.g., students in the some-risk category receive strategic/supplemental/Tier 2 intervention; students in the at-risk category receive more intensive/Tier 3 intervention).	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
ADMINISTRATION & SCORING		
Standardization of materials and procedures for administration	Administration protocol is scripted and provides precise guidelines; administration windows are clearly identified; materials are provided, or clear guidelines are provided if materials are to be created; includes both electronic and hard copy administration manual that is clear and concise.	<p>DOES NOT MEET-evidence was not provided for this criteria or information does not demonstrate evidence. (0)</p> <p>PARTIALLY MEETS-partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1)</p> <p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
Efficiency of	The amount of time needed to administer the assessment	DOES NOT MEET -evidence was not provided for this criteria

Standard	Description	Rating
administration	is reasonable and balanced to the information provided.	or information does not demonstrate evidence. (0) PARTIALLY MEETS -partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)
Efficiency of scoring	The amount of time needed to score the assessment is reasonable and balanced to the information provided; computer-assisted scoring is available; procedures for calculating scores are clear; scores can be stored and reported electronically.	DOES NOT MEET -evidence was not provided for this criteria or information does not demonstrate evidence.(0) PARTIALLY MEETS -partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)
Accommodations clearly stated and described for Second Language Learners	The accommodations directly address the linguistic needs of the student. Evidence includes: <ul style="list-style-type: none"> • approved accommodation that do not compromise the interpretation or purpose of the test; • specific administration guidelines provided for implementing any accommodations; • how to address accommodations, and is specifically addressed in the training; and • suggested accommodations are research or evidence-based. 	DOES NOT MEET -evidence was not provided for this criteria or information does not demonstrate evidence. (0) PARTIALLY MEETS -partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence. (1) MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)
Accommodations clearly stated and described for students with disabilities and students with special needs (504, etc.)	The differing needs of students with disabilities are specifically addressed. Evidence includes: <ul style="list-style-type: none"> • approved accommodation that do not compromise the interpretation or purpose of the test; 	DOES NOT MEET -evidence was not provided for this criteria or information does not demonstrate evidence.(0) PARTIALLY MEETS -partial evidence was provided related to the criterion and/ or data provided demonstrates weak evidence (1)

	<ul style="list-style-type: none"> • the provision of specific administration guidelines for implementing any accommodations; • information about how to address accommodations specifically addressed in the training materials or program; and • suggested accommodations are research or evidence-based. 	<p>MEETS OR EXCEEDS –most information for the criterion is provided. Information and data provided suggests acceptable or strong evidence. (2)</p>
--	--	---

<p>STUDENT DATA PRIVACY COMPLIANCE</p>
<p>Connecticut General Statutes §§ 10-234aa through 10-234dd</p>
<p>Educators and school leaders should review and understand their obligations under Connecticut’s student data privacy law (Connecticut General Statutes §§ 10-234aa through 10-234dd). As a key element of compliance, districts must enter into contracts with providers of educational technology whenever such providers capture or have access to personal student information, records, or data. For purposes of this assessment review, districts will need to communicate with such companies in advance of submitting the assessment proposal to ensure adherence to the privacy and security assurances outlined in the law. Subsequent non-compliance with Connecticut’s student data privacy law may void any previous CSDE approval decisions.</p>

Appendix B.

Consensus rubric scores for each assessment.

Standard	Description	Acadience	Aimsweb	Amira	DIBELS 8th	Easy CBM	Fastbridge	iReady	NWEA MAP Fluency
1. Evidence of test reliability and internal consistency reliability	Reported for each grade	1	1	2	1	2	1	2	2
2. Evidence of test reliability and internal consistency reliability	SEM is reported	1	0	0	1	0	1	2	1
3. Evidence of test reliability and internal consistency reliability	Inter-rater reliability studies conducted	2	0	1	2	0	1	NA	2
4. Evidence of test reliability and internal consistency reliability	Reliability studies for all subcategories of students who will take assessment	1	2	1	0	1	1	2	2
5. Alternate forms available	All alternate forms have evidence of equivalence	2	2	1	2	2	2	2	2
6. Evidence of content and construct validity	Evidence that assessment measures reading	2	2	2	2	2	2	2	1
7. Evidence of content and construct validity	Studies of construct validity demonstrating relationship (correlations above .70)	1	2	0	2	1	1	1	0
8. Evidence of criterion/predictive validity	Predicts to an outcome that appropriately weighs word decoding, especially in earlier grade levels.	2	2	0	2	1	2	2	0
9. Determination of cut-scores	Established cut-scores for decision making that also align with Connecticut's	2	1	1	2	2	2	1	1

	base rate (~40 th to 50 th percentile)								
10. Determination of cut-scores	Studies of classification accuracy show measure appropriately identifies students (sensitivity & AUC >.80)	1	2	2	2	2	1	2	1
11. Determination of cut-scores	Acceptable procedures for setting cut-scores	2	1	2	2	1	2	2	1
12. Determination of cut-scores	Guidance for cut-score interpretation	2	2	2	2	1	2	2	2
13. Standardization of materials and procedures for administration	Scripted protocol with precise guidelines	2	2	2	2	2	2	2	2
14. Efficiency of administration	Administration time is reasonable	2	2	2	2	2	2	1	2
15. Efficiency of scoring	Scoring time is reasonable	2	2	2	2	2	2	2	2
16. Accommodations clearly stated for Second Language Learners	Accommodations directly address linguistic needs of the student	0	1	2	2	2	0	2	2
17. Accommodations clearly stated for students with disabilities and special needs	Differing needs of students with disabilities are specifically addressed	2	2	2	2	2	1	2	2

Appendix C

Evidence for Screening Decisions, Construct Validity for Components of Reading, and Evidence for Supplemental Spanish Subtests

Note. For screening, AUC and sensitivity should be above 0.80. For components of reading, correlations (r) should be close to 0.60, but can be marginally acceptable at 0.30.

Reading suites with minor weaknesses		
Screening	Acadience	aimswebPlus
Predicts to	40th percentile on the GRADE	25th percentile on WRF, ORF, ITBS, and NWEA; local cut-scores can also be set
Accuracy in grade K	First Sound Fluency 0.34 sensitivity; 0.72 AUC	Composite (predicting to Word Reading Fluency) 1.00 sensitivity; 0.99 AUC
Accuracy in grade 1	Reading Composite 0.70 sensitivity; 0.90AUC	Oral Reading Fluency (predicting to ITBS) 1.00 sensitivity; 0.85 AUC
Accuracy in grade 2	Oral Reading Fluency 0.75 sensitivity; 0.89 AUC	Reading Composite 0.88 sensitivity; 0.88 AUC
Accuracy in grade 3	Reading Composite 0.74 sensitivity; 0.90 AUC	Reading Composite 0.85 sensitivity; 0.92 AUC
Components	Acadience	aimswebPlus
Phonological & phonemic awareness	First sound fluency $r = .52$ with GRADE; Phonemic segmentation fluency $r = .34$ with GRADE	Initial Sounds, Phoneme Segmentation, subtests not individually evaluated
Phonics	Nonsense word fluency .47 to .52 with GRADE	Nonsense word fluency, Word Reading Fluency, Spelling, subtests not individually evaluated
Fluency	Oral Reading Fluency $r = .64$ to .76 with GRADE	Oral Reading Fluency, Silent Reading Fluency, subtests not individually evaluated
Vocabulary	Word Use Fluency-Revised $r = .22$ to .72 with PPVT & Test of Language Development	Vocabulary (starts in grade 2), subtest not individually evaluated
Comprehension	Maze $r = .65$ with GRADE; Composite score $r = .48$ to .78 with GRADE	Reading Comprehension, subtest not individually evaluated
RAN or letter naming fluency	RAN objects, letters, numbers $r = .44$ to -.68 with Reading Composite score	Fall Letter Naming Fluency $r = .89$ with Winter Early Literacy Composite (LWSF+PS); RAN available separately
Oral language	NA	Listening Comprehension, not reported

Notes	WUF format is saying a word in a sentence. Linking study conducted with Smarter Balanced with $r = .65$ to $.75$.	Vocabulary is multiple choice. Additionally sells the Shaywitz Dyslexia Screener. We did not evaluate that assessment.
Spanish	Acadience Reading Español	Aimsweb Plus
Reliability	NR	Alternate form reliability = $.82$ to $.91$
Subtests	Same subtests as the English version.	Letter naming fluency, oral reading fluency, letter sounds fluency, spelling, syllable reading fluency, syllable segmentation, listening comprehension. Correlations NR
Administration time	NR	22 to 30 min per student
Requires Spanish-speaking administrator?	NR	Yes
Norm sample	NR	The norming sample included 16,399 students in grades K-5 during the 2010-2012 school years who go to school in the U.S. The students were all Spanish speakers as their primary language who qualified for ESL services. No Spanish variation or newcomer status were reported but 81-94% of students were reported as Hispanic.

Reading suites with minor weaknesses		
Screening	DIBELS 8th edition/mClass	easyCBM
Predicts to	Cut points at 20th and 40th percentile predicting to DIBELS Next Composite & IOWA Assessment Total Reading Score	Kindergarten: 50th percentile on the Spring Letter Sounds measure; Grade 1 measures: 50th percentile on the Spring Word Reading Fluency measure; Grade 2 measures: 50th percentile on the Spring Passage Reading Fluency measure; Grade 3: 20th percentile on Smarter Balanced
Accuracy in grade K	Phoneme segmentation fluency (with DIBELS Next composite) .83 sensitivity, .78 AUC	Phoneme Segmenting 1.00 sensitivity; 0.37 AUC
Accuracy in grade 1	Word Reading Fluency (with Iowa Test) .95 sensitivity, .77 AUC; Oral Reading Fluency .94 sensitivity, .81 AUC	Word reading fluency 0.84 sensitivity; 0.95 AUC
Accuracy in grade 2	Word Reading Fluency (with Iowa Test) .82 sensitivity, .83 AUC; Oral Reading Fluency .78 sensitivity, .80 AUC; Maze .97 sensitivity, .87 AUC	Vocabulary 0.85 sensitivity; 0.85 AUC
Accuracy in grade 3	Maze (with Iowa Test) 1.00 sensitivity, .86 AUC; Oral Reading Fluency .77 sensitivity, .70 AUC	Vocabulary 0.85 sensitivity; 0.87 AUC Passage reading fluency correlates with SBAC $r = .64$ Vocabulary correlates with SBAC $r = .68$
Components	DIBELS 8th mClass	easyCBM
Phonological & phonemic awareness	Phoneme segmentation fluency $r = .37$ to $.54$ with CTOPP-2	Phoneme segmentation $r = .85$ with DIBELS Phonemic Segmentation
Phonics	Nonsense word fluency $r = .50$ to $.66$ with IOWA total reading; Word Reading Fluency $r = .54$ to $.78$ with IOWA total reading mClass Spelling $r = .32$ to $.78$ with Iowa Word Analysis	Word Reading Fluency $r = .15$ with SAT-10, Syllable Segmenting measured but correlations not reported
Fluency	Oral reading fluency $r = .62$ to $.80$ with Iowa Total Reading	Word Reading Fluency $r = .15$ with SAT-10, Passage reading fluency $r = .19$ to $.22$ with SAT-10 and $r = .88$ to $.95$ with DIBELS ORF
Vocabulary	mClass vocabulary $r = .35$ to $.71$ with Star	Vocabulary $r = .39$ to $.76$ with Gates-MacGinitie Word Knowledge

Comprehension	Composite $r=.54$ to $.80$ with IOWA total reading; Maze $r=.67$ to $.70$ with IOWA total reading	Multiple Choice Reading Comprehension $r=.41$ to $.70$ with Gates-MacGinitie reading comprehension tests
RAN or letter naming fluency	Letter naming fluency $r=.31$ to $.50$ with CTOPP-2 symbolic composite	Letter names $.86$ with DIBELS , RAN measured but correlations not reported
Oral language	mClass Oral Language $r= .45$ to $.62$ with TOLD-P:4 SI	NA
Notes	Vocabulary is fill-in-the-blank.	Vocabulary is multiple choice.
Spanish	Lectura	easyCBM
Reliability	Internal consistency = $.75$ to $.90$ grades K to 3; Interrater reliability $>.94$	NR
Subtests	Letter names, phonics, syllable awareness, beginning decoding, advanced decoding, oral reading fluency, Maze. Shows adequate correlations with Woodcock Munoz and SELS.	Letter sounds, syllable sounds, phoneme segmenting, syllable segmenting, word reading fluency, sentence reading fluency. Correlations with external measures NR.
Administration time	5 to 6 minutes per student, plus 5 min for whole class administration	1 min per subtest; <10 min per student.
Requires Spanish-speaking administrator?	Yes	Yes
Norm sample	The norming sample included 32,933 students in grades K through 5, who lived and go to school in the U.S. 9,883 students were considered to have English Learner status and 863 students spoke Spanish as their home language. Country of origin and newcomer status NR.	The norming sample included 170 students in grades K through 2 who were native English speakers in Spanish language only instruction and native Spanish speakers. Country of origin and newcomer status NR.

Additional reading suites with minor weaknesses		
Screening	FastBridge earlyReading FastBridge CBMreading	i-Ready Diagnostic + Fluency Measure
Predicts to	earlyReading predicts to 40 th percentile on the GRADE CBMreading predicts to 20 th percentile on MAP Growth	DIBELS 8 th composite some risk cut point

Accuracy in grade K	earlyReading (with GRADE) .75 sensitivity, .95 AUC	Sensitivity = 0.81 to 0.93, AUC > 0.90
Accuracy in grade 1	earlyReading (with GRADE) .89 sensitivity, .99 AUC CBMreading (with AIMSweb) .75 sensitivity, .81 AUC	Sensitivity = 0.84 to 0.96, AUC > 0.90
Accuracy in grade 2	AUTOreading vocabulary (with aReading ²) .76 sensitivity, .85 AUC CBMreading (with AIMSweb) .88 sensitivity, .93 AUC	Sensitivity = 0.81 to 0.93, AUC > 0.90
Accuracy in grade 3	earlyReading (with SBAC) NR sensitivity, .89 AUC CBMreading (with AIMSweb) .84 sensitivity, .89 AUC	Sensitivity = 0.77 to 0.92, AUC > 0.90
Components	FastBridge earlyReading, AUTOreading, CBMreading	iReady Diagnostic + Fluency Measure
Phonological & phonemic awareness	Onset sounds, word rhyming, word blending, word segmenting (with GRADE) $r = .38$ to $.58$	Phonemic awareness $r = 0.36$ to 0.57 with DIBELS 8 th Phoneme Segmentation, $r = 0.58$ to 0.74 with Nonsense Word Fluency
Phonics	Letter sounds, letter names, decodable words, nonsense words, syllable reading encoding, decoding (with GRADE) $r = .19$ to $.66$	Phonics $r = 0.60$ to 0.70 with DIBELS 8 th Encoding and spelling items NR
Fluency	Sentence reading, sight word reading, oral reading fluency (with TOSREC) $r = .47$ to $.87$	Fluency measure $r > 0.90$ with DIBELS 8 th ORF
Vocabulary	vocabulary, matching synonyms, morphology (CBMreading) $r = .51$ to $.67$	Vocabulary NR
Comprehension	earlyreading composite (with GRADE) $r = .68$ to $.81$	Comprehension: Informational Text NR Comprehension: Literature NR
RAN or letter naming fluency	RAN Rapid Naming (with CTOPP-2) .50 to .81 sensitivity, >.88 AUC	RAN, Letter Naming Fluency >0.90 with DIBELS 8 th
Oral language	Oral repetition	NR
Notes	CBMreading, earlyReading, and RAN are all paper/pencil one-to-one administration. AUTOreading is administered digitally in a group. Progress monitoring also available.	Scores in K, 1, and 2 correlate with grade 3 SBAC $r = .61$ to $.82$ i-Ready also has additional subtests to create a dyslexia risk indicator. We did not evaluate these additional subtests.

² aReading and AUTOreading are FastBridge end-of-year measures. No external criterion measure was provided.

Spanish	FastBridge earlyReading Spanish	i-Ready Assessment of Spanish Reading
Reliability	NR	NR
Subtests	Concepts of print, Onset Sounds, Letter names, Letter sounds, Rhyming, Word blending, Word segmenting, Sight words, Decodable words, syllable reading, sentence reading, oral repetition. Correlations not reported	Same as the English subtests
Administration time	1-4 min per subtest; ~4-16 min/student	NR
Requires Spanish-speaking administrator?	Yes	Computer administered
Norm sample	NR	NR

Reading suites that require additional evidence		
Screening	NWEA MAP Fluency	Amira
Predicts to	10 th percentile on NWEA MAP Growth	30 th percentile on NWEA MAP Growth, Star, & iReady; also provides 10 th , 25 th , and 40 th benchmarks
Accuracy in grade K	Foundational Skills Domain Score, sensitivity=.72 to .84, AUC=.80 to .90	Reading Composite (with MAP Growth): Sensitivity=.74, AUC=.73
Accuracy in grade 1	Foundational Skills Domain Score, sensitivity=.81, AUC=.90	Reading Composite (with NWEA): Sensitivity=.82, AUC=.88
Accuracy in grade 2	Foundational Skills Domain Score, sensitivity=.77 to .86, AUC=.87 to .91	Reading Composite (with MAP Growth): Sensitivity=.84, AUC=.88; Reading Composite (with STAR): Sensitivity=.81, AUC=.81; Reading Composite (with iReady): Sensitivity=.83, AUC=.84
Accuracy in grade 3	Foundational Skills Domain Score, sensitivity=.80 to .82, AUC=.88 to .91	Reading Composite (with MAP Growth): Sensitivity=.86, AUC=.88; Reading Composite (with iReady): Sensitivity=.83, AUC=.84

Components	NWEA MAP Fluency	Amira
Phonological & phonemic awareness	Phonological awareness domain not individually evaluated	Phonological awareness subtest not individually evaluated
Phonics	Phonics & Word Recognition domain not individually evaluated	Nonsense word fluency, Word Identification, & Spelling subtests not individually evaluated
Fluency	Sentence Reading Fluency subtest not individually evaluated	Oral reading fluency subtests not individually evaluated
Vocabulary	Picture Vocabulary subtest not individually evaluated	Vocabulary measured but correlations not reported
Comprehension	Language comprehension domain not individually evaluated	Composite score $r=.70$ to $.77$ with NWEA MAP, Star Reading, & iReady; Reading Comprehension subtest also available K to 6
RAN or letter naming fluency	RAN	RAN subtest not individually evaluated
Oral language	Listening comprehension subtest not individually evaluated	Listening Comprehension subtest not individually evaluated
Notes	NWEA MAP Growth is a separate product from MAP Fluency and was not evaluated as part of this review.	Amira studied with Grammaggio (an assessment of specific language impairment). Amira risk flag had .85 sensitivity and .91 AUC predicting to Grammaggio.
Spanish	NWEA MAP Fluency	Amira
Reliability	Computer scored. Human-computer scoring agreement is very high.	Computer scored. Human-computer scoring agreement is .88 to .95 for grades K to 3.
Subtests	All subtests offered in both English and Spanish.	RAN, phonemic awareness, phonics, nonsense word fluency, word identification fluency, oral reading fluency. Correlations with Star and iReady $> .70$
Administration time	1-5 min per subtest; 20-30 min per student.	15 min per student.
Requires Spanish-speaking administrator?	No	Computer scored
Norm sample	The user sample for NWEA has the greatest reach of any assessment covering 49 states and closely matches the demographics of the U.S.,	Sample is Spanish-English bilingual population ($n = 52,280$) who take Amira's screener in grades K-3. Language status, countries of origin, and newcomer status NR.

	including Hispanic students and Spanish speakers. Language status, countries of origin, and newcomer status NR.	
--	---	--

References

- Adlof, S. M., & Hogan, T. P. (2019). If we don't look, we won't see: Measuring language development to inform literacy instruction. *Policy Insights from the Behavioral and Brain Sciences*, 6(2), 210-217. <https://doi.org/10.1177/2372732219839075>
- Albers, C. A., & Hall, G. J. (2023). Best practices in universal screening. In Harrison, Proctor, & Thomas (Eds.) *Best Practices in School Psychology, 7th Edition* (pp. 91-104). National Association of School Psychologists.
- Bao, X., Komesidou, R., & Hogan, T. P. (2024). A Review of Screeners to Identify Risk of Developmental Language Disorder. *American Journal of Speech-Language Pathology*, 33(3), 1548-1571. https://pubs.asha.org/doi/pdf/10.1044/2023_AJSLP-23-00286
- Barrett, C. A., Johnson, L. J., Truckenmiller, A. J., & VanDerHeyden, A. M. (2024). Comparing the cost-accuracy ratio of multiple approaches to reading screening in elementary schools. *Remedial and Special Education*, 45, 71-84. Preprint: <https://psyarxiv.com/2q7sm/>
- Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech Language and Hearing Research*, 48, 1387-1396.
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102(2), 327-340. <https://doi.org/10.1037/a0018448>
- Connor, C.M. (2019). Using Technology and Assessment to Personalize Instruction: Preventing Reading Problems. *Prevention Science*, 20, 89-99. <https://doi.org/10.1007/s11121-017-0842-9>
- Connor, C. M., & Morrison, F. J. (2016). Individualizing Student Instruction in Reading. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 54-61. <https://doi.org/10.1177/2372732215624931>
- Fletcher, J. M., Francis, D. J., Foorman, B. R., & Schatschneider, C. (2020). Early Detection of Dyslexia Risk: Development of Brief, Teacher-Administered Screens. *Learning Disability Quarterly*, 44(3), 145-157. <https://doi.org/10.1177/0731948720931870>
- Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. J. (2015). The structure of oral language and its relation to reading at various levels of comprehension in kindergarten through grade 2. *Reading and Writing*, 28(5), 655-681. doi: 10.1007/s11145-015-9544-5
- Foorman, B. R., Petscher, Y., & Herrera, S. (2018). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1-10. *Learning and Individual Differences*, 63, 12-23. <https://doi.org/10.1016/j.lindif.2018.02.011>

- Foorman, B. R., Petscher, Y., Lefsky, E. B., Toste, J. R. (2010). Reading First in Florida: Five years of improvement. *Journal of Literacy Research*, 42, 71-93. DOI: 10.1080/10862960903583202
- Foorman, B. R., Petscher, Y., Stanley, C., & Truckenmiller, A. J. (2017). Latent profiles of reading and language and their association with standardized reading outcomes in kindergarten through tenth grade. *Journal of Research on Educational Effectiveness*, 10, 619-645. doi: 10.1080/19345747.2016.1237597
- Fuchs, D., & Fuchs, L. S. (2017). Critique of the National Evaluation of Response to Intervention: A Case for Simpler Frameworks. *Exceptional Children*, 83(3), 255–268. <https://doi.org/10.1177/0014402917693580>
- Grapin, S. L., Kranzler, J. H., Waldron, N., Joyce-Beaulieu, D., & Algina, J. (2017). Developing local oral reading fluency cut scores for predicting high-stakes test performance. *Psychology in the Schools*, 54(9), 932-946. <https://doi.org/10.1002/pits.22035>
- Hwang, J. K., Mancilla-Martinez, J., McClain, J. B., Oh, M. H., & Flores, I. (2019). Spanish-speaking English learners' English language and literacy skills: The predictive role of conceptually scored vocabulary. *Applied Psycholinguistics*, 41(1), 1–24. <https://doi.org/10.1017/s0142716419000365>
- Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, 47(2), 125-135.
- Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. *Journal of School Psychology*, 52(4), 377-405. <https://doi.org/10.1016/j.jsp.2014.06.002>
- Koon, S., Petscher, Y., & Foorman, B.R. (2014). Using evidence-based decision trees instead of formulas to identify at-risk readers (REL 2014–036). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Klingbeil, D. A., Van Norman, E. R., Nelson, P. M., & Birr, C. (2018). Evaluating screening procedures across changes to the statewide achievement test. *Assessment for Effective Intervention*, 44(1), 17-31. <https://doi.org/10.1177/1534508417747390>
- Lemke, M., Murphy, D., Chow, A. S. P., Spencer, H., & Zhang, A. (2023). A first look at early literacy performance in Massachusetts: Results of initial analysis based on state grantee literacy screening assessments. *WestEd*.

- Lonigan, C. J., & Burgess, S. R. (2017). Dimensionality of reading skills with elementary-school-age children. *Scientific Studies of Reading, 21*(3), 239-253.
<https://doi.org/10.1080/10888438.2017.1285918>
- Mancilla-Martinez, J., Hwang, J. K., Oh, M. H., & McClain, J. B. (2020). Early elementary grade dual language learners from Spanish-speaking homes struggling with English reading comprehension: The dormant role of language skills. *Journal of Educational Psychology, 112*(5), 880.
- Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology, 102*(3), 701.
- National Center on Intensive Intervention (2023). *Academic Screening Tools Chart Rating Rubrics*. Author.
- Petscher, Y., Fien, H., Stanley, C., Gearin, B., Gaab, N., Fletcher, J.M., & Johnson, E. (2019). Screening for Dyslexia. Retrieved from improvingliteracy.org
- Truckenmiller, A. J., Yohannan, J., & Cho, E. (2020). Linking reading assessment data to instructional planning: A component-skills approach. *Communiq  , 48*, 15-18.