

Literacy Assessments Meeting the Requirements of 279.68/ELI

for universal screening and progress monitoring
2017-2018 school year

June 2017



Overview

This document contains information about tools approved for use as universal screening and progress monitoring measures to meet the requirements of 279.68/ELI.

Background

In the winter of the 2012-2013 school year the Department requested and reviewed information and proposals for a statewide universal screener (US) and progress monitoring (PM) assessment system for preschool through 6th grade literacy. There were two outcomes from this process. First, the Department identified tools to be provided and supported statewide for schools interested in participating in a literacy early warning system to support implementation of a Multi-Tiered System of Supports (MTSS). Secondly, the results of the reviews were published to inform users of the relevant technical qualities of additional literacy measures that may be used in local schools as they made selections for their own universal screening and progress monitoring measures.

Legislation funded at the end of the 2013 legislative session caused the Department of Education to revisit these earlier reviews in 2014 for the purposes of setting minimum standards for assessments approved for use in universal screening and progress monitoring of K-3 literacy as required by Iowa Code section 279.68. Annually, the Department requests information from vendors to update the list of assessments approved that meet requirements of Iowa Code section 279.68.

The approved list used through the end of the 2016-2017 academic year was cleared of *all* measures and established anew for the 2017-2018 year. This document represents the updated approved list for the 2017-2018 academic year. The technical aspects considered in the evaluation are provided following the table. The Department will continue annual reviews to identify assessment to satisfy the screening and monitoring requirements of Iowa Code section 279.68.

Each submitted assessment was evaluated separately at each grade level and at for each purpose (i.e., universal screening and progress monitoring) for which information was submitted. To be approved at a grade level, the measure needed to meet all criteria for the respective assessment purpose with the information provided through the submission.

In the table that follows, checks indicate an assessment met the minimum requirements for the indicated purpose and grade. In some situations letters are reported to indicate the measure was developed for and approved for specific seasons at that grade (e.g., “W,S” indicates approved for Winter and Spring of that grade). Footnotes and technical assistance are provided following the table for clarity and considerations in using universal screening and progress monitoring measures. Approximate administration times in minutes are listed. Please note reported approximate administration are listed independent of administration format (i.e., individually or group-administered).

Selecting an assessment

Please attend to the relevant footnotes, and also see page 3 for definitions and page 4 for an expanded discussion of the details in the table and important considerations for selecting screening and progress monitoring assessments.

2017-2018 State Supported Assessments	Universal Screening							Progress Monitoring						Admin. Mins.	
	K	1	2	3	4	5	6	K	1	2	3	4	5		6
FAST: CBMreading (CBMR)		W,S ^{3*}	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	3
FAST: Adaptive Reading (aReading)	✓ ²	✓	✓	✓	✓	✓	✓								10-20
FAST: earlyReading Composite**	✓	✓													5
Concepts of Print	F*														2
Letter Naming	F*							✓							1
Onset Sounds	F,W*							✓ ¹							2
Letter Sounds	✓*							✓							1
Word Blending	✓*							✓							2
Word Segmenting	W,S*	F*						✓	✓ ¹						2
Nonsense Words	W,S*	✓*						✓	✓						1
Decodable Real Words		✓*							✓						1
Sight Words 150		✓*							✓						1
Additional Approved Assessments	Universal Screening							Progress Monitoring							
AIMSwebPlus: Early Lit. K-1 Composite**	✓	✓													25-30
AIMSwebPlus: 2-6 Composite**			✓	✓	✓	✓	✓								20-40
Letter Naming	✓*														1
Oral Reading Fluency		✓*													2
AIMSweb: Nonsense Words								✓	✓						1
AIMSweb: Reading CBM		✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	3
AIMSweb: Maze											✓	✓		✓	3
DIBELS Next: Composite**		✓	✓	✓	✓	✓	✓								3-9
First Sound	F,W*							✓							1
Phoneme Segmentation	W*							✓ ¹							1
Letter Naming	✓*														1
Nonsense Words-CLS	W,S*	✓*	F					✓	✓	✓					1
Nonsense Words-WWR	S	✓*	F*					✓	✓	✓					1
Oral Reading Fluency		W,S*	✓*	✓*	✓*	✓*	✓*		✓	✓					3-6
DAZE				✓*	✓*	✓*	✓*				✓	✓			3
DIBELS 6th ed.: Letter Naming	✓	F													1
DIBELS 6th ed.: Phoneme Segmentation	W							✓ ¹							1
DIBELS 6th ed.: Nonsense Word	W,S	✓							✓						1
DIBELS 6th ed.: Oral Reading Fluency		W,S	✓	✓											3
Measures of Academic Progress(MAP)-Primary	✓ ⁴	✓ ⁴	✓ ⁴												40
Measures of Academic Progress(MAP)			✓ ⁴	✓	✓	✓	✓								40
STAR Early Literacy	✓ ⁵	✓ ⁵	✓ ^{5,6}	✓ ^{5,6}				✓ ⁵	✓ ⁵	✓ ^{5,6}	✓ ^{5,6}				6-10
STAR Reading		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	7-15

* A composite measure is available for this grade and season. Use of the broader composite measure is **strongly recommended** as the primary indicator of risk, rather than individual subtests which likely measure a narrower sample of reading skills. Composites within this list are indicated with a double asterisk (**). Using subtests in isolation for screening purposes is discouraged.

- These measures, while meeting minimum technical requirements for progress monitoring at this grade, have established benchmarks for fewer than three windows. The benchmarks *may not* represent desired end of year performance, therefore care should be used with these measures.
- Fast Bridge Learning (FBL) makes aReading available in the fall of kindergarten and it meets the minimum technical requirements. FBL, however, recommends the earlyReading composite for Kdg. If schools consider using aReading in kindergarten the developer states "We encourage that educators are confident that students are exposed to the use of computers prior to the first aReading administration. This promotes consistent familiarity with the interface and ecology of a computer adaptive assessment."
- CBM-R is administered in the form of Sentence Reading in the fall of first grade within the earlyReading Composite.
- MAP-Primary and MAP in 2nd met minimum criteria using a study which linked performance to 3rd grade outcomes (i.e., a 1-3 year time lag) rather than using the standard same-year prediction approach to identify technical adequacy characteristics.
- STAR Early Literacy technical characteristics are based upon predicting to STAR Reading. Comparisons predicting to an external measure are preferred, but not required.
- It is reported that STAR Early Literacy, while developed for early elementary grades, can be used grades 2 and 3 for students with less developed literacy skills. Benchmark setting at these grades is based on an upward extension (in normative percentile rank) of those established in the early elementary grades, and should be used carefully.

Definitions

Reliability. Reliability is a measure of measurement consistency, or stability. The higher the reliability, the more confidently one can interpret an obtained score being due to students' skills rather than day-to-day or week-to-week variability.

Validity. Validity refers to how well a tool measures the skills (e.g., literacy skills) it intends to. The higher the validity, the more confidently one can believe a student's performance on the measure accurately represents reading development and/or key foundational reading skills.

Area Under the Curve (AUC). Area Under the Curve is a calculation that represents the relative value of a test for accurately classifying outcomes (e.g., above or below proficiency on high-stakes assessments).

Sensitivity/Specificity. Sensitivity and Specificity are statistics that represent the ability of the test to correctly identify students. That is, the ability of the tool to correctly identify students predicted to be on track for success and/or correctly identify students predicted not on track for success.

Administrations per Year. For the purposes of universal screening it is ideal that a screenings be completed at least three times a year. This will support 279.68/ELI requirements and enable schools to periodically check to ensure students are on track for later reading success.

Benchmarks Availability. Benchmarks are established to identify a level of performance which identifies students to be on track to be successful readers. Assessments should have well-established benchmarks.

Standard Setting. Seasonal benchmarks can be set in a number of different ways including normative and criterion-referenced studies, expert opinions, and alignment studies to other instruments. It is important that developers have sound standard setting methodology.

Equivalent Forms. When using an assessment to monitor progress it is important multiple forms of similar difficulty exist. This helps minimize the possibility of increases in scores being due to test familiarity rather than true student skill gains.

Reliability of Slope. Reliability of slope is a statistic that represents the ability of the test to produce a consistent estimate of student growth over time with the least amount of "bounce" caused by the test.

Administration Time. Because screening and progress monitoring assessment takes some time away from instruction, it is important to find tests that are as efficient as possible.

For the Technical Definition of each term, please see also: [ELI Guidance](#)

Technical Assistance for Selecting Universal Screening and Progress Monitoring Measures

When reviewing assessments to select screening and progress monitoring tools consider all variables. Combinations of tests not designed to work together will create potentially inconsistent, fragmented data, complicated training and potential for confusing administration and scoring practices. Balance the time spent administering tests against the quality of data that come from the assessment. Also consider training time, as well as time spent scoring the test and/or waiting for results to be returned.

Universal screening measures should be selected because of their ability to efficiently and accurately identify students who are at risk for reading difficulty. The Area Under the Curve and sensitivity statistics are used to estimate the predictiveness of the tests. We want to use tests that make the best predictions. Progress monitoring measures are selected based on their ability to measure and show improvement with repeated use. Reliability of slope is the statistic that helps pinpoint the measures that reliably show improvement. We want to use progress monitoring measures that do the best job of helping us know whether students are making enough progress quickly enough. Benchmarks are important to both screening and progress monitoring because they define targets for student outcomes. We want to use well established targets that represent meaningful outcomes for children.

Some test developers submitted composite measures as well as individual subtests for review. In some cases, measures used in composites might also have met the minimum criteria to be used as a standalone screener. It is strongly recommended that composite scores be used as the broader (i.e., more comprehensive) indicator of risk rather than individual, possibly narrow, subtests. Composites are likely to more accurately and consistently identify student need over the course of the school year because they measure multiple skills that contribute to overall literacy development. Individual tests may produce fragmented and potentially inconsistent data, especially when the tests and/or benchmarks are not available for the entire year.

In some cases measures met minimum technical requirements for progress monitoring, yet did not have established benchmarks for use through the entire school year (all three windows). Users should be aware of the available benchmarks for each season and grade and adjust decision making as appropriate. Using these measures for monitoring outside of the season or year intended (i.e., off level) may provide a false sense of progress because the available benchmark provides a lower target than appropriate for the student at the current grade/season.

One of the most important pieces of evidence for evaluating a universal screener is the ability of the test to predict to future outcomes. Typically this kind of prediction uses an external measure of the broad skill of reading, often an end of year outcome measure. In some cases, test developers used data prediction studies that compared screening data against outcome measures that was collected up to three years later, rather than predictions to outcomes during the same year. This practice may be less desirable because of the extended time lag between tests. In other cases, developers predicted to results of other tests in the same developer's test battery. This practice may be less desirable because the predictions are not to external measures. There is a risk of bias in the analysis because of a narrowed outcome comparison.

The goal is to increase student outcomes. This can be done, in part, by efficiently screening all students to accurately identify those at risk for difficulties, purposefully intervening then monitoring the progress. This list helps systems select measures best suited for screening and monitoring. Be thoughtful about creating a coherent testing plan. Select tests to maximize instructional time while getting the assessment data necessary to make sound decisions. Then focus on improving reading outcomes for all.