

A background image showing a smiling teacher and a young student high-fiving. The teacher is on the left, wearing a yellow top, and the student is on the right, wearing a striped shirt. They are both looking at each other and smiling. In the foreground, there is a wooden table with a piece of paper that has some red and blue markings on it.

Classworks Universal Screener

TABLE OF CONTENTS

ExecutiveSummary.....2

Test Design.....5

Item Development.....11

Test Validation.....13

National Center on Intensive Intervention.....18

Conclusion.....25

EXECUTIVE SUMMARY

Purpose

Classworks Universal Screeners are formal assessments used to measure readiness for grade level instruction, help identify baseline learning levels, and measure growth. The Universal Screeners were specifically designed for the purpose of screening students who may need additional intervention and can be used as part of the Response to Intervention (RtI) process.

In addition to reporting an overall scaled score based on the total test, Classworks provides nationally normed percentile ranks as well as student strengths and weaknesses for key strands. Key strands include a minimum of four test questions to provide a reasonable estimate of student strengths and weaknesses. This information, when used in conjunction with other data such as High Stakes Test results and classroom performance, can help provide a starting point for determining next steps.

Overview

Classworks Universal Screeners include multiple forms at each level for language arts and mathematics, grades K–10. The Universal Screeners are typically administered three times a year: at the beginning of the school year to assess readiness for instruction for all students, mid-year to measure progress for RtI tiers II and III, and end-of-year to measure overall growth for the year. Given that the test is primarily designed to identify readiness, the test includes multiple grade levels of content to allow sufficient reach for students who may be struggling.

The Universal Screeners are between 20 and 35 items in length depending on the grade level targeted. Although we recommend it is administered in a single sitting, the assessment can be bookmarked. Two forms of each Screener were developed; these forms are parallel and measure similar content. The kindergarten level assessments are an exception to this approach, with two different forms reflecting earlier and later kindergarten content given the rapid development at the kindergarten level.

Overall test results are reported as a scaled score. Scoring on a vertical scale provides a single point of reference to compare individual student gains from one test administration to the next, within and across school years. Measuring growth vertically serves a dual purpose: to track learning gains for individual students and to determine whether learning must be accelerated. Student grade level readiness and proficiency is reported as a Grade Level Equivalency or GLE for a student's overall performance and the performance within each tested domain. Based on this overall score, Classworks recommends a progress monitoring level for each student. The Classworks Universal Screener, then looks at those scores by grade level to assign norm referenced percentile scores for all grades and subjects in order to recommend levels of intervention.

Classworks Universal Screeners have been evaluated by the National Center on Intensive Intervention (NCII) receiving the highest reliability ranking.

UNIVERSAL SCREENER QUICK GUIDE

Item	Description
Purpose	Measure grade level readiness, help identify baseline, measure growth
Grades	K–10 Math, K–10 Reading
Levels of coverage per test	Test includes multiple grade levels of content to allow sufficient reach to help identify strugglers (exception: Kindergarten)
Audio	Audio support available for all grades
Length of test	Recommended to be taken in one sitting; 20–35 items depending on grade level/subject (Assessment will bookmark if needed)
Vertical scale?	Yes. All scores are vertically scaled from K–10 for longitudinal tracking.
Nationally Normed?	Yes. Test is nationally normed and includes norm-referenced percentile scores for all grades and subjects of the Universal Screener for Fall, Winter and Spring.
Output from test	Average readiness scaled score, recommended progress monitoring level and norm-referenced percentile rank of students by class, teacher, custom group, demographic, and/or grade level

TEST DESIGN

SEG Measurement (SEG) has been instrumental in the design, development, testing, and analysis of Classworks Universal Screeners. SEG is an assessment, measurement, and research firm that provides assessment design, development, and implementation services for K–12, higher education, and credentialing programs. They have delivered over 100 million assessments to tens of thousands of schools and colleges in all 50 states.

Classworks Universal Screeners were designed and built for the particular purpose they serve. For this reason, they meet all of the criteria that define quality screeners: the assessments are brief, reliable, valid, equated, and measured on a vertical scale.

SEG initially created the assessments by hand-selecting items for each level and form of the tests. Forms were then equated through field testing and calibration so that each measures the same sets of skills at the same level of difficulty. Individual test items and the assessments themselves were designed with diversity in mind: including populations of cultural and linguistically diverse students, and special needs students. Guiding principles for assessment design were integrated into the process, including ensuring all items are written in a clear, concise manner and free of age, gender, ethnic, religious, or disability bias.

There are two parallel forms for each test in grades K–10. For second grade and above, the test questions include content from the target grade level as well as from two grade levels below the target. Given that the test is primarily designed to identify readiness, the test includes multiple grade levels of content to allow sufficient reach and enough content coverage for students who may be struggling. The tests include approximately 50% of the content from the target grade, approximately 25% of the content from the grade below, and approximately 25% of the content from two grades below.

The first grade assessment contains content from both 1st grade and kindergarten. The kindergarten assessment contains content drawn only from kindergarten with two different forms reflecting earlier and later kindergarten content, given the rapid development at the kindergarten level.

Vertical Scale and Item Bank Calibration

The vertical scale was developed through a linked testing design such that all items could be calibrated together and placed on the same continuum. The field test data was used to calibrate the items and tests. Calibration is a process that places all tests and all test items on a common scale. This was used to create a single common scale from grade K to grade 10. In this way, scores from the tests are comparable across forms of the test and over time. A given score will have the same meaning regardless of which form is administered and regardless of when the student takes the test.

The assessments developed include sets of overlapping items across test forms at the same level and across adjacent grade levels. This facilitates the calibration of the item bank. SEG calibrated the items using IRT (one parameter Rasch model) to create a common vertical scale across grade levels.

The raw number of correct answers reflects a particular Rasch score (ranging from -4 to +4), which is then translated to the final scaled score for reporting purposes. When the student completes his/her screener, the scaled score and key strand level performance feedback are immediately available for reporting. The approach taken in the calibration and scoring process provides Rasch extrapolated norms.

As a further measure to ensure that the test questions and assessments are technically sound and are performing as expected, SEG analyzes the data from the Fall test takers each year.

Curriculum Advantage reviews the results from the Fall to make sure the tests are performing well. SEG examines the statistics for the tests as a whole (e.g., average scores, distribution of scores) and the statistics for

individual test items (e.g., question difficulty and the ability of the question to distinguish between different levels of student performance). Based on this analysis, Curriculum Advantage further refines the tests, revising and replacing questions as necessary.

During the 2014-2015 item analysis, Curriculum Advantage made the decision to update the Universal Screener. New items were created and field tested during the 2015-2016 school year and officially added to the assessment for the 2016-2017 school year.

Percentile Ranks

To create the Universal Screener Percentiles, we began by gathering the most recent Language Arts and Mathematics test scores from K to twelfth grade. However, due to massive school closures caused by Covid-19 there were not enough test scores from Spring 2020 for percentile creation thus scores from Fall 2019, Winter 2019, and Spring 2019 were used. The total sample for Mathematics was 25,743 for the Fall window, 17,536 for the Winter window, and 10,472 for the Spring window. For Language Arts the total sample for the Fall testing window was 22,834, 19,352 for Winter testing window, and 9,270 for the Spring window. Using these scores, thresholds were created at 10%, 20%, 25%, 40%, 50%, 70%, and 90% mark for K through twelfth grade.

Score Reporting

Score Reporting is designed to provide reliable information useful for understanding overall student readiness and estimated student strengths and weaknesses in specific strands measured by the test. Scores are based on scaled scores that allow all tests to be placed on a common scale regardless of which form is administered and at what grade level. Results are reported at the total test and key strand level. Strands assessed vary by grade level and subject of the assessment. This approach provides a reasonable balance between the need for information on student strengths and weaknesses and the need for sufficient score reliability.

Raw scores are calculated as the total number of items answered correctly on the screener. Performance on the assessments is reported as a scaled score on a vertical scale ranging from 200 to 800 spanning across grades K–10. Feedback is also provided at the key strand level. (see Vertical Scale and Item Bank Calibration above).

These strands were determined based on an analysis of over 31 state standards and then re-examined with the introduction of the Common Core State Standards.

Strands that are reported are required to include a minimum of four test questions to provide a reliable estimate of student strengths and weaknesses.

Math:

- » Algebra
- » Geometry
- » Mathematical Processes
- » Measurement
- » Numeration
- » Operations
- » Patterns
- » Statistics and Probability

Reading:

- » Grammar/Usage/Mechanics
- » Reading Comprehension
- » Study Skills
- » Word Analysis
- » Writing
- » Writing Process

Curriculum Advantage establishes score ranges that reflect levels of student readiness on the assessments. There are various approaches that can be used to identify appropriate cut points defining levels of readiness. Below details the method SEG recommended for creating appropriate cut points.

Establishing Cut Scores

The cut scores for Classworks Universal Screeners were established using a two-stage standard setting process. In the first stage, a BookMarking Procedure (Cizek and Bunch, 2007) was applied. This was followed by a second stage, in which the stage one potential cut scores were reviewed in light of student performance data and expectations for student performance.

The BookMarking Procedure is an item mapping approach to standard setting developed in the 1990's (Cizek and Bunch, 2007). The BookMarking Procedure as employed for Classworks involves the review of an ordered test booklet containing all the items for a given test arranged in order of difficulty from easiest to hardest (Mitzel, H.C., Lewis, D.M., Patz, R.J., and Green, D.R., 2001). The difficulty values for this procedure were obtained from the Rasch item calibrations obtained from the original development of the screeners. Based on the procedures suggested by Mitzel, et al (2001), content experts reviewed the ordered item booklet and were asked to identify ("bookmark") the item representing the first item for which the minimally proficient student would be unlikely to answer the item correctly (less than 50% probability). The difficulty of the item identified served as the potential cut score emerging from stage one of the standard setting.

In the second stage, the potential cut scores produced in stage one of the process were reviewed against the distribution of scores from operational testing to evaluate the number and percentage of students that would "pass" and the number and percentage of students that would "fail" the assessment based on the stage one potential cut scores. In some cases, the stage one potential cut score was raised or lowered based on the impact rates or expected performance for the students.

ITEM DEVELOPMENT

The Classworks assessment item bank was developed by a team of content experts from a third-party developer, a leader in the creation of high-stakes content for assessments produced by states and testing companies. The test items have been reviewed and refined through a multi-step process involving members of this test development team.

The Universal Screeners are composed of 100% four-response-option multiple-choice type questions. The items were specifically developed for the Universal Screener or were selected and modified from the existing Curriculum Advantage item bank.

Guiding Principles of Item Construction

In order to ensure item reliability and validity, guiding principles were used in the item construction process.

Item Construction:

- Items are written in clear, concise language at the appropriate grade level
- Items are written without age, gender, ethnic, religious, or disability bias
- Each item set measures both basic knowledge and higher-order thinking skills
- Items adhere to the objectives being assessed
- Items are constructed in a consistent manner
- Item content is current and relevant to audience
- Items are written in the form of questions, avoiding open ended or negative stems

Item Response Measurement:

- Items show consistency of student response
- Results can be generalized to the population
- Items are calibrated to ensure that scores have similar meaning over time

- After calibration, items are placed on a developmental/vertical scale to allow for the accurate comparison of students over time and across use of the items
- Student performance can be predicted from item response
- Target goals and norms can be developed from item response measures

Questions/Stems:

- Stems and reading passages will be at grade-level readability and must assess the skill being tested according to the level of Bloom's indicated
- Stems are free of age, gender, ethnic, religious, or disability stereotypes or bias
- Stems are written in question format and do not require sentence completion, true/false, and fill-in-the-blank
- Each stem has only one correct answer

Answers/Distractors:

- Answers are presented in a multiple-choice format with 4 answer options
- Distractors are written in a logical order (alphabetical, chronological)
- Distractors are approximately the same length and must be grammatically parallel
- Distractors are plausible and should not contain grammatical clues
- Distractors address a variety of common errors rather than the same error
- Distractor rationale is provided for each answer choice

The test items are multiple-choice questions, offering an efficient and reliable way to assess students' knowledge and skills. All items have one single best answer and responses are scored as correct or incorrect. Multiple choice measures have advantages over other types of item response, in that they are capable of covering a large amount of content in a relatively short period of time. Moreover, they can achieve high levels of reliability, providing users with a consistent and stable measure of student knowledge and skills over time.

TEST VALIDATION

Following the creation of the tests, SEG conducted a second verification of the assessment items. The verification process consisted of a comprehensive alignment review to establish the validity of the assessment items and to determine if they were accurately aligned to the objectives they purport to measure.

Classworks continues to partner with SEG to ensure that the tests themselves, as well as assessment-related decisions, are psychometrically sound. This ongoing process includes further statistical analysis, item calibration, adjustments to the cut scores on the vertical scale, and overall evaluation of the quality of Classworks Universal Screeners.

Field Testing and Analysis

To ensure that the test items and assessments are psychometrically sound, SEG analyzed the item and test performance data based on the field test to be conducted by Curriculum Advantage in the Fall of 2009, the Fall of 2010 and the Fall of 2011. Subsequent field testing took place during the 2015-2016 school year resulting in an updated assessment for the 2016-2017 school year and beyond. Curriculum Advantage collected information from approximately 200–300 students per test form the first year, with exponential increases in each of the following years. SEG analyzes the results each year, providing both test and item level analyses including:

- Overall test and subtest statistics
 - » Mean
 - » Standard Deviation
 - » Reliability
 - » SEM (Standard Error of Measure)
 - » Overall Model Fit
 - » Frequency Distribution
- Item statistics
 - » P Value (percent correct)
 - » Point biserial correlation (measure of item discrimination)

- » Logit value from -3 to +3 (person and item independent measure of item difficulty)
- » Item Infit statistic
- » Item Outfit statistic

SEG reviews the item statistics, and any item that does not demonstrate suitable psychometric characteristics are recommended for replacement. These statistics help ensure on-going relevance and validity. Here are some of the statistics SEG calculates:

Total Test Statistics

- Average Score on the Assessment – SEG computes the average (mean) score achieved by students taking the assessment. This helps us determine if the assessment is properly targeted to the level of the students assessed.
- Variation and Distribution of Scores on the Assessment – SEG calculates the amount of variability (standard deviation) in the test scores achieved by students taking the assessment. This is another indicator of how well the test is targeted to the level of students assessed.
- Reliability – SEG computes the reliability of the test to ensure that the test is consistently measuring the knowledge and skills measured by the assessment across forms of the test and is stable over time.
- Score Accuracy – Any assessment score is subject to variation when a student takes the test multiple times. SEG estimates the amount of variation expected for a student score (Standard Error of Measure; SEM); this is an indicator of score accuracy.

Individual Question Statistics

- Question Difficulty – SEG computes the percentage of students who answer the questions correctly; this is an indicator of the difficulty of the question
- Question Differentiation – SEG computes the relationship between student performance on each individual question and the assessment as a whole; this is an indicator of how well the question differentiates between those students who have the knowledge and skills measured by the assessment and those who do not have the knowledge and skills.

ITEM DISTRIBUTION BY STRAND:

Reading

Grade	Grammar/ Usage/ Mechanics	Reading	Study Skills	Word Analysis	Writing	Writing Process	Not Covered - Listening/ Speaking/ Viewing	Grand Total
K		9	1	5				15
1	2	10		7		1		20
2	3	12	1	8		1		25
3	7	9	3	4	1	1		25
4	6	10	2	5	1	1		25
5	7	10	3	7	1	2		30
6	7	11	3	6	1	2		30
7	7	11	4	6		2		30
8	8	11	2	6		3		30
9	8	13	3	4		2		30
10	9	13	5	3				30

Mathematics

Grade	Algebra	Concepts of Calculus	Geometry	Mathematical Processes	Measurement	Numeration	Operations	Patterns	Statistics & Probability	Trig	Grand Total
K				2	3	5	2	2	1		15
1	1		5	4	4	4		1	1		20
2	2		4	3	5	4	2	2	3		25
3	2		2	2	6	1	5	2	5		25
4	1		5	1	3	4	4	1	6		25
5	5		8	4	4	1	2	1	5		30
6	6	1	8	3	1	3	3	1	4		30
7	6	2	6	2	4	2	2	1	5		30
8	8	1	8	1	4	1	2	1	4		30
9	8		7	5	2	1	1	1	4	1	30
10	8		6	5	4	1	1		4	1	30

ITEM DISTRIBUTION:

Form	Number of Scored Items	Number of Non-Scored Items	Number of linking items (additional non-scored linking items for field testing and calibration)	Total Test Length (scored and non-scored items)
Grade K Reading Screener				
A	20	5	3	21
B	20	5	3	21
Grade 1 Reading Screener				
A	25	5	3	26
B	25	5	2	26
Grade 2 Reading Screener				
A	30	5	3	33
B	30	5	3	33
Grade 3 Reading Screener				
A	30	5	3	35
B	30	5	3	35
Grade 4 Reading Screener				
A	30	5	4	35
B	30	5	4	35
Grade 5 Reading Screener				
A	35	5	2	39
B	35	5	2	39
Grade 6 Reading Screener				
A	35	5	3	39
B	35	5	2	39
Grade 7 Reading Screener				
A	35	5	2	39
B	35	5	2	39
Grade 8 Reading Screener				
A	35	5	3	39
B	35	5	2	39
Grade 9 Reading Screener				
A	35	5	2	39
B	35	5	2	39
Grade 10 Reading Screener				
A	35	5	2	39
B	35	5	2	39

Form	Number of Scored Items	Number of Non-Scored Items	Number of linking items (additional non-scored linking items for field testing and calibration)	Total Test Length (scored and non-scored items)
Grade K Math Screener				
A	20	5	3	21
B	20	5	3	21
Grade 1 Math Screener				
A	25	5	3	26
B	25	5	3	26
Grade 2 Math Screener				
A	30	5	3	33
B	30	5	3	33
Grade 3 Math Screener				
A	30	5	3	35
B	30	5	3	35
Grade 4 Math Screener				
A	30	5	4	35
B	30	5	4	35
Grade 5 Math Screener				
A	35	5	2	39
B	35	5	2	39
Grade 6 Math Screener				
A	35	5	3	39
B	35	5	3	39
Grade 7 Math Screener				
A	35	5	2	39
B	35	5	2	39
Grade 8 Math Screener				
A	35	5	3	39
B	35	5	2	39
Grade 9 Math Screener				
A	35	5	3	39
B	35	5	2	39
Grade 10 Math Screener				
A	35	5	2	39
B	35	5	2	39

NATIONAL CENTER ON INTENSIVE INTERVENTION

Classworks Universal Screeners have been validated by the National Center on Intensive Intervention (NCII), and they received the highest ratings for reliability and validity. Classworks Universal Screeners for Reading are formal assessments used to measure readiness for grade level instruction, help identify baseline learning levels, and measure growth. The Universal Screeners were specifically designed for the purpose of screening students who may need additional intervention and can be used as part of the RtI process.

In addition to reporting an overall scaled score based on the total test, Classworks provides student strengths and weaknesses for key strands. Key strands include a minimum of four test questions to provide a reasonable estimate of student strengths and weaknesses.

Reliability

The Classworks assessment affords the means to screen students on multiple occasions (e.g., Fall, Winter, Spring, or weekly) during the school year. Thus, test-retest reliability is necessary, and we estimate test-retest reliability via the Pearson correlation between Classworks Screener scores of students taking tests in two terms within the school year (Fall/Winter, and Winter/Spring). The second reliability test is Cronbach's Alpha, a measure of internal consistency. This analysis was conducted on a sample of students who had posted scores for three sets of Classworks Screener questions, all of which aimed to measure a single construct - student's proficiency in mathematics.

Type of Reliability	Subscale	Age/Grade	n (sample/ examinees)	Median Coefficient	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Test-Retest (Fall/ Winter)	Reading	Grade 2	1283	0.61	0.57	0.64
Test-Retest (Winter/ Spring)	Reading	Grade 2	1038	0.72	0.68	0.74
Cronbach's Alpha (internal consistency)	Reading	Grade 2	927	0.85	0.83	0.86
Test-Retest (Fall/ Winter)	Reading	Grade 3	1202	0.74	0.71	0.76
Test-Retest (Winter/ Spring)	Reading	Grade 3	978	0.75	0.72	0.78
Cronbach's Alpha (internal consistency)	Reading	Grade 3	849	0.89	0.87	0.9
Test-Retest (Fall/ Winter)	Reading	Grade 4	1339	0.72	0.7	0.75
Test-Retest (Winter/ Spring)	Reading	Grade 4	1144	0.75	0.73	0.78
Cronbach's Alpha (internal consistency)	Reading	Grade 4	928	0.89	0.88	0.9
Test-Retest (Fall/ Winter)	Reading	Grade 5	1482	0.79	0.77	0.81
Test-Retest (Winter/ Spring)	Reading	Grade 5	1290	0.77	0.75	0.79
Cronbach's Alpha (internal consistency)	Reading	Grade 5	1126	0.91	0.9	0.92
Test-Retest (Fall/ Winter)	Reading	Grade 6	1733	0.76	0.74	0.78
Test-Retest (Winter/ Spring)	Reading	Grade 6	1484	0.76	0.74	0.79
Cronbach's Alpha (internal consistency)	Reading	Grade 6	1155	0.89	0.88	0.91
Test-Retest (Fall/ Winter)	Reading	Grade 7	1376	0.78	0.76	0.8
Test-Retest (Winter/ Spring)	Reading	Grade 7	1032	0.71	0.67	0.74
Cronbach's Alpha (internal consistency)	Reading	Grade 7	893	0.89	0.88	0.9

Type of Reliability	Subscale	Age/Grade	n (sample/ examinees)	Median Coefficient	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Test-Retest (Fall/ Winter)	Math	Grade 2	1237	0.6	0.56	0.63
Test-Retest (Winter/ Spring)	Math	Grade 2	967	0.7	0.67	0.74
Cronbach's Alpha (internal consistency)	Math	Grade 2	851	0.83	0.81	0.85
Test-Retest (Fall/ Winter)	Math	Grade 3	1212	0.67	0.63	0.7
Test-Retest (Winter/ Spring)	Math	Grade 3	908	0.76	0.73	0.79
Cronbach's Alpha (internal consistency)	Math	Grade 3	802	0.86	0.84	0.88
Test-Retest (Fall/ Winter)	Math	Grade 4	1350	0.65	0.62	0.68
Test-Retest (Winter/ Spring)	Math	Grade 4	1140	0.72	0.69	0.75
Cronbach's Alpha (internal consistency)	Math	Grade 4	916	0.86	0.85	0.88
Test-Retest (Fall/ Winter)	Math	Grade 5	1421	0.71	0.68	0.73
Test-Retest (Winter/ Spring)	Math	Grade 5	1245	0.72	0.69	0.75
Cronbach's Alpha (internal consistency)	Math	Grade 5	1072	0.88	0.87	0.89
Test-Retest (Fall/ Winter)	Math	Grade 6	1771	0.7	0.68	0.73
Test-Retest (Winter/ Spring)	Math	Grade 6	1264	0.71	0.68	0.74
Cronbach's Alpha (internal consistency)	Math	Grade 6	1016	0.88	0.86	0.89
Test-Retest (Fall/ Winter)	Math	Grade 7	1286	0.65	0.62	0.68
Test-Retest (Winter/ Spring)	Math	Grade 7	922	0.66	0.62	0.69
Cronbach's Alpha (internal consistency)	Math	Grade 7	791	0.84	0.82	0.86

Validity

The validity evidence for the Classworks assessments comes from the relationships of Classworks test scores to NWEA MAP Growth test scores.

These relationships include:

- The concurrent performance of students on Classworks tests with their performance on MAP Growth tests.
- The predictive relationship between students' performance on Classworks tests with their performance, two testing terms later, on MAP Growth tests.

The Measures of Academic Progress (MAP) is used as the outcome measure. Published by the NWEA the MAP Growth is regarded as a highly valid and reliable measure of broad reading ability. The NWEA website states, "Our tools are trusted by educators in 140 countries and more than half the schools in the US" which indicates it can be considered an excellent outcome measure for classification studies. For the validity analysis conducted, we used concurrent and predictive validity. Concurrent validity was estimated as the Pearson correlation coefficient between student scores from Fall 2017 and the same students' total scale score on the Map Growth assessment (also administered in Fall 2017). Predictive validity was estimated as the Pearson correlation coefficient between student scores from a given term (Fall 2017) and the same students' total scale score on the MAP Growth assessment administered in Winter 2017-2018. 3 Concurrent and predictive validity coefficients, for each grade and each time of year, were consistently in the mid to high 0.60s to 0.70s. This validity evidence demonstrates a strong relationship between the Classworks assessment and the MAP Growth assessments across the grades and times of year reported.

Type of Validity	Subscale	Age/Grade	Test or Criterion	n (sample/ examinees)	Median Coefficient	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Concurrent (Fall/ Fall)	Reading	Grade 2	Map Growth	150	0.69	0.6	0.77
Predictive (Fall/ Winter)	Reading	Grade 2	Map Growth	150	0.76	0.68	0.82
Concurrent (Fall/ Fall)	Reading	Grade 3	Map Growth	139	0.72	0.62	0.79
Predictive (Fall/ Winter)	Reading	Grade 3	Map Growth	138	0.72	0.63	0.79
Concurrent (Fall/ Fall)	Reading	Grade 4	Map Growth	189	0.79	0.73	0.84
Predictive (Fall/ Winter)	Reading	Grade 4	Map Growth	190	0.72	0.64	0.78
Concurrent (Fall/ Fall)	Reading	Grade 5	Map Growth	198	0.74	0.66	0.79
Predictive (Fall/ Winter)	Reading	Grade 5	Map Growth	196	0.71	0.63	0.77
Concurrent (Fall/ Fall)	Reading	Grade 6	Map Growth	153	0.72	0.64	0.79
Predictive (Fall/ Winter)	Reading	Grade 6	Map Growth	153	0.7	0.61	0.77
Concurrent (Fall/ Fall)	Reading	Grade 7	Map Growth	57	0.84	0.74	0.9
Predictive (Fall/ Winter)	Reading	Grade 7	Map Growth	57	0.8	0.68	0.88

Type of Validity	Subscale	Age/Grade	Test or Criterion	n (sample/ examinees)	Median Coefficient	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Concurrent (Fall/ Fall)	Math	Grade 2	Map Growth	262	0.63	0.55	0.7
Predictive (Fall/ Winter)	Math	Grade 2	Map Growth	263	0.66	0.59	0.72
Concurrent (Fall/ Fall)	Math	Grade 3	Map Growth	273	0.77	0.71	0.81
Predictive (Fall/ Winter)	Math	Grade 3	Map Growth	269	0.78	0.72	0.82
Concurrent (Fall/ Fall)	Math	Grade 4	Map Growth	310	0.67	0.6	0.72
Predictive (Fall/ Winter)	Math	Grade 4	Map Growth	309	0.64	0.57	0.71
Concurrent (Fall/ Fall)	Math	Grade 5	Map Growth	295	0.74	0.69	0.79
Predictive (Fall/ Winter)	Math	Grade 5	Map Growth	292	0.73	0.67	0.78
Concurrent (Fall/ Fall)	Math	Grade 6	Map Growth	192	0.71	0.63	0.77
Predictive (Fall/ Winter)	Math	Grade 6	Map Growth	191	0.71	0.63	0.77
Concurrent (Fall/ Fall)	Math	Grade 7	Map Growth	95	0.66	0.53	0.76
Predictive (Fall/ Winter)	Math	Grade 7	Map Growth	94	0.68	0.56	0.78

Classification Analyses

In addition to the reliability and validity of the measures, the Universal Screeners were also evaluated with regard to the accuracy of classifying students as at-risk in comparison to an independent measure. It is important that the screeners are able to appropriately identify students who are at-risk and those who are not at-risk. In particular, it is critical that at-risk students are properly identified as being at-risk to get the instructional help that they need.

In order to evaluate the classification accuracy, Classworks Universal Screeners classifications were compared to the classifications determined by performance on NWEA MAP Growth assessments in reading and math. The comparisons provided a classification of students into one of four cells in a “confusion matrix.” Students could be classified as at-risk or not at-risk based on the passing status for each of the two assessments as Pass-Pass, Pass-Fail, Fail-Pass, or Fail-Fail. The classification analyses were performed by evaluating sensitivity and specificity.

Negative predictive power is a measure that estimates the accuracy of classifying students as “not at-risk.” A useful screening tool should have very high negative predictive power such that at-risk students are not misidentified as not being at-risk. Using test data for more than 11,300 students, the Universal Screeners were found to have 93% and 97% negative predictive power for math and reading, respectively.

Dyslexia

The Classworks Universal Screeners measure student performance with key domains that are indicative of future reading performance: phonological awareness, letter-sound correspondence, decoding, and reading comprehension. Each of these strands has been identified as early predictors for further screening for learning disabilities, specifically dyslexia.

In addition to measuring performance of key domains, Classworks Universal Screeners are recommended by NCII because of:

The breadth of content assessed:

- The fact that Individualized Learning is generated that meets the child at their instructional level
- The automated connection to Progress Monitoring probes that will continually fine-tune lessons included in each child's intervention.
- This continuation of assessment follows the Data-Based Intervention (DBI) model and is easily acted upon by educators.

Research indicates that universal screeners for kindergarten to second grade students are most effective for indicating when further Dyslexia testing is appropriate when they include:

- Phonological processing
- Spelling, and
- Decoding tasks

Academic screeners measure some, but not all, of the discrete skills that should be considered when determining whether a learning disability is present. The Classworks K-2 Reading screeners indicate whether further testing for dyslexia is needed

CONCLUSION

Classworks Universal Screeners have been evaluated by the National Center on Intensive Intervention (NCII) receiving the highest reliability rating. It is a vertically scaled assessment using percentile rankings to pinpoint student growth and performance. The Universal Screeners were specifically designed for the purpose of screening students who may need additional intervention and can be used as part of the Response to Intervention (RtI) and Multi-Tiered System of Supports (MTSS) process. This information, when used in conjunction with other data such as High Stakes Test results and classroom performance can help provide a starting point for determining the next steps in the intervention process.

ABOUT CLASSWORKS

Since 2003, millions of students have used Classworks to close learning gaps, keep pace, and grow! Our mission is to deliver the most compelling individualized learning experience that helps teachers create an equitable learning environment for students of all abilities and ethnic and social backgrounds.



For more information:

Contact Classworks at hello@classworks.com

Call 770-325-5555 or visit www.classworks.com

