WHITE PAPER

mCLASS Dyslexia Screening Measures Research Report

Amplify.

Table of Contents

Abstract

This paper presents initial reliability and validity evidence for three newly developed K-3 early literacy measures: Vocabulary, Spelling, and Rapid Automatized Naming (RAN). Performance data were collected for 584 students in grades K-3 at the beginning (BOY) and middle of year (MOY) benchmark periods in the 2017-2018 school year. mCLASS:DIBELS Next, the mCLASS:Early Literacy Measures (Vocabulary, Spelling, RAN), STAR Early Literacy (Grades K and 1) and STAR Reading (Grades 2 and 3) were administered to all students. Initial data collected at BOY and MOY indicate strong to very strong internal consistency reliability for the Vocabulary and Spelling measures. In addition, data collected at the end of the year indicate strong test-retest reliability for RAN. Similarly, strong concurrent validity correlations were demonstrated for Vocabulary and Spelling with the DIBELS Composite and STAR measures. The overall (K-3) correlation of mCLASS:RAN with PRO-ED RAN was strong, and moderate correlations of mCLASS:RAN with DIBELS Composite scores are explained by the differences in the two assessments. Research will continue through the end of the current school year and into the 2018-2019 school year.

Introduction to mCLASS:Early Literacy Measures

To better meet the growing needs of educators to screen students for reading difficulties, including dyslexia, Amplify has created three additional early literacy measures. These measures are designed to provide additional information on students' skills in the areas of vocabulary, spelling (encoding), and rapid automatized naming (RAN). These additional screeners will provide indication of risk in these domains, which are important for reading success, and may provide additional information related to risk for reading difficulties, including dyslexia. The early literacy measures (Vocabulary, Spelling, RAN) were created to complement the DIBELS Next assessment when screening students for reading difficulties. Educators may use the additional information collected from these measures to develop and implement targeted interventions for students at risk.

This Research Brief describes initial evidence for the reliability and validity of the mCLASS:Early Literacy Measures (i.e., Vocabulary, Spelling, and RAN) based on a study conducted during two benchmark periods of the 2017-2018 school year (Beginning of Year, BOY; and Middle of Year, MOY). In addition, test-retest data were collected during one benchmark period of the 2017-2018 school year (End of Year, EOY). The research will continue through the end of the current school year and into the 2018-2019 school year.

mCLASS:Vocabulary. The vocabulary measure provides additional information to help determine what level of knowledge a student has of grade specific words, whether the student has strategies for making meaning of words encountered in text, and whether the student is applying vocabulary knowledge to derive meaning from text. The tasks assess each student's depth of knowledge of grade level high utility (Tier 2) and content specific words. Words were selected from widely used core reading programs, lists of the most frequent and high utility words (Graves, Sales, & Ruda, 2008), and content specific words (Marzano & Sims, 2013) students should know.

mCLASS:Spelling. The spelling measure provides an indication of a student's level of general spelling skills. It is designed based on the principles of General Outcome Measurement and Curriculum-Based Measurement (CBM; Deno, 1992). Items on the spelling measure represent a random sample of grade-specific words. These words

are drawn from a pool of words covering the phoneme-grapheme correspondences that students at each grade level are expected to learn over the course of a year based upon the scopes and sequences of published reading and spelling curricula.

mCLASS:RAN. The rapid automatized naming (RAN) measure indicates how quickly students can name numeric symbols aloud. While RAN tasks can be completed with a variety of stimuli (numbers, colors, objects), alphanumeric RAN tasks are typically given to students from age 4 to adult. In a meta-analysis of Rapid Automatized Naming assessments, results showed that RAN tasks with letters or numbers showed higher correlations with reading than tasks with colors or pictures (Araujo, Reis, Magnus Petersson, & Faisca, 2015). RAN is considered a measure of phonological processing, specifically the retrieval of phonological information. Deficits in rapid automatized naming have been shown to be a robust indicator of risk for dyslexia in children (Gaab, 2017). Research has documented that students who have difficulty with both rapid automatized naming and phonological awareness experience more reading difficulty and are more likely at-risk for dyslexia than students who have difficulty in rapid automatized naming or phonological awareness alone (Wolf & Bowers, 2000; Ozernov-Palchik, Yu, Wang, & Gaab, 2016). While RAN data do not lead to specific instructional activities that would be incorporated into an intervention (i.e., it is not expected that students should practice rapidly naming objects if their RAN scores are low), RAN serves as an additional indicator of risk for reading difficulty including difficulty related to dyslexia.

mCLASS:Early Literacy Measures Research Overview

Purpose: The purpose of this study was to examine the reliability and validity of the vocabulary, spelling, and rapid automatized naming measures, and to inform improvements and changes to each measure based on item analysis.

Recruitment: Participants were recruited from existing mCLASS:DIBELS Next customers. The following criteria were used to screen for eligible participants: a) school must demonstrate a range of student reading proficiency levels, b) school must demonstrate a variety of demographic characteristics, c) school must have students in any or all grades kindergarten through three, and
d) schools must contribute a minimum of 15 students per grade.

Participants: The study was conducted during the 2017-2018 academic year, with performance data collected around two benchmark periods (BOY and MOY). In total, 584 students in grades kindergarten through three were assessed in three schools representing two districts.

All outcome measures including mCLASS:DIBELS Next data were collected by a combination of reading coaches and Amplify data collectors. The mCLASS:Early Literacy Measures were administered to students by reading coaches and Amplify data collectors in the school computer lab.

Demographic Information: Participants in the field study were educators and students from following geographic regions: Northeast and Midwest (US Census Bureau, n.d.). Demographic information including gender, ethnicity, Free and Reduced Price Lunch Eligibility (FRL; an indicator of socioeconomic status), English Language Learner status (ELL), and mCLASS:DIBELS Next performance is provided in Table 1.

Table 1: Sample Size and Demographics by Grade

	All	Kindergarten	Grade 1	Grade 2	Grade 3
Sample Size (n)					
Districts	2	2	2	2	2
Schools	3	3	3	3	3
Students	584	146 (25%)	165 (28%)	129 (22%)	144 (25%)
Gender (n)					
Female	302 (52%)	76 (52%)	87 (53%)	58 (45%)	81 (56%)
Male	282 (48%)	70 (48%)	78 (47%)	71 (55%)	63 (44%)
Ethnicity (n)					
White	548 (94%)	139 (96%)	153 (93%)	122 (96%)	134 (94%)
Black	11 (2%)	2 (1%)	3 (2%)	2 (1%)	3 (2%)
Asian	3 (0%)	0	1 (0%)	1 (1%)	2 (1%)
Multiracial	4 (1%)	0	2 (1%)	2 (1%)	2 (1%)
Ethnicity Not Specified	18 (3%)	5 (3%)	6 (4%)	2 (1%)	3 (2%)
Other Demographics (n)					
FRL Eligible	497 (85%)	127 (87%)	138 (84%)	112 (87%)	120 (83%)
FRL Not Eligible	87 (15%)	19 (13%)	27 (16%)	17 (13%)	24 (17%)

Research Procedures

All educators or data collectors administering the assessments attended a half-day, web-based training on standardized administration and scoring procedures for mCLASS:DIBELS Next and the mCLASS:Early Literacy Measures. A reliability check for the Early Literacy Measures and DIBELS Next was administered to attendees to confirm their understanding prior to assessing students.

Assessors administered the mCLASS:Early Literacy Measures to all students following standard procedures in a small group setting in kindergarten and Grade 1 and group administration in grades 2 and 3. The mCLASS:Early Literacy Measures were administered via paper and pencil at BOY and online at MOY within 2 weeks of the usual benchmark testing period and within 2 weeks of each other. Students were administered external outcome measures following standard procedures following their usual benchmark testing period. External outcome measures serve to demonstrate concurrent validity—that is, the degree to which the mCLASS:Early Literacy Measures compare with a measure that has been previously validated.

Students were administered mCLASS:DIBELS Next following standard procedures in a one-on- one setting during their usual benchmark testing period. mCLASS:DIBELS Next was selected as an outcome measure due to its established validity and reliability as a test of early literacy skills, which are related to the mCLASS:Early Literacy Measures. An additional external measure selected for all students in kindergarten and Grade 1 was STAR Early Literacy (Renaissance, 2014), and STAR Reading (Renaissance, 2015) served as the external measure for all students in Grade 2 and Grade 3 at BOY and MOY. A subset of students were administered the PRO-ED RAN (PRO-ED, 2005) measure as a measure of concurrent validity for mCLASS:RAN at MOY.

Measure Descriptions

mCLASS:DIBELS Next: DIBELS Next is a nationally recognized observational assessment of reading foundational skills including letter knowledge, phonological awareness, alphabetic principle, oral reading fluency, and comprehension. Skills assessed vary by grade level and time of year. Teachers score students' oral responses to determine risk level for reading failure, and scores for each individual measure as well as a composite score are reported.

Together, the measures administered at each benchmark period comprise a DIBELS Composite Score. The DIBELS Composite Score is a combination of multiple DIBELS Next scores and provides the best overall estimate of the student's reading proficiency and risk level (Good et al., 2013). The DIBELS Next Technical Manual reports strong reliability support for the measures and overall for the Composite Score. Specific evidence for the reliability and validity of one submeasure used in the analysis for mCLASS:Spelling and the Composite Score is presented below.

- **Correct Letter Sounds (CLS):** CLS is a submeasure within DIBELS Next Nonsense Word Fluency that represents the number of letter sounds produced correctly in one minute. This score provides important information to educators about a student's place in the progression from sounding out individual letter sounds to reading whole words. The authors report NWF- CLS alternate form reliability ranges from 0.71 to 0.94, inter-rater reliability ranges from 0.99 to 1.00, and test-retest ranges from 0.76 to 0.90 (Good et al., 2013). Predictive validity, as measured as the correlation with GRADE EOY, ranges from 0.43 to 0.56. This suggests that NWF-CLS has moderate validity and strong reliability evidence.
- **DIBELS Composite Score:** The authors report alternate form reliability ranges from 0.66 to 0.97, inter-rater reliability ranges from 0.81 to 0.94, and test-retest reliability ranges from 0.97 to 0.99 (Good et al., 2013). Predictive validity, as measured as the correlation with GRADE EOY, ranges from 0.50 to 0.80. This suggests strong validity and reliability evidence.

STAR Reading (SR): SR is a computer-adaptive assessment that tests students in five content domains: Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning, Analyzing Literary Text, Understanding Author's Craft, and

Analyzing Argument and Evaluating Text (Renaissance Learning, 2015). Students in grades 2 and 3 are administered 20 vocabulary- in-context (i.e., cloze) items and five authentic text passages with multiple-choice literal or inferential questions. Students respond to these questions using a mouse or keyboard. SR takes approximately 10 to 15 minutes for students to complete. It is intended to provide interim data on students' reading skills so educators can set goals, respond quickly to student needs, monitor progress, and maximize growth. Overall scale scores are reported. STAR Reading serves as an external assessment in grades 2 and 3 for this study.

The SR technical manual reports reliability coefficient ranges for grades 1 through 5 as follows: split-half reliability of 0.88 to 0.89; test-retest reliability of 0.82 to 0.89; and "generic" reliability (i.e., calculated from the conditional error variance of IRT ability estimates) of 0.89 to 0.91 (Renaissance Learning, 2011).

Concurrent validity coefficients for grades 1 through 4 ranged from 0.71 to 0.87 (i.e., students were administered SR and DIBELS Oral Reading Fluency (DORF) within a 2-week period in 32 schools across 9 states in 2007-2008). Predictive validity was also examined using test scores from a variety of other reading measures (e.g., Colorado Student Assessment Program, Illinois Standards Achievement Test). For grades 1 through 6 the validity coefficients ranged from 0.68 to 0.82 for STAR Reading scores predicting later performance on tests (Renaissance Learning, 2011).

STAR Early Literacy (SEL): STAR Early Literacy is a computer-adaptive assessment of student reading skills in seven domains: General Readiness, Graphophonemic Knowledge, Phonemic Awareness, Phonics, Comprehension, Structural Analysis, and Vocabulary (Renaissance Learning, 2014). Items within the comprehension and structural analysis domains are omitted in kindergarten administrations. Each administration consists of 25 items presented in multiple choice format (three answer choices per item). Each item includes a graphic display and is dictated by audio recordings. SEL takes approximately 10 to 15 minutes for students in kindergarten and Grade 1 to complete. Similar to STAR Reading, it is intended as an interim assessment to monitor student progress and instructional needs. STAR Early Literacy serves as an external assessment in kindergarten and Grade 1 for this study.

SEL's scaled score has generic reliability ranges from 0.78 to 0.86; split-half reliability ranges from 0.75 to 0.85; and alternate form reliability ranges from 0.63 to 0.78. The concurrent validity ranges from 0.50 to 0.88 (measured by the correlations with teachers' ratings of students' skills, Brigance scale, Developing Skills Checklist,

Metropolitan Early Childhood Assessment, Texas Primary Reading Inventory, and Test of Phonological Awareness); the concurrent validity ranges from 0.42 to 0.73 to predict STAR Reading scores (Renaissance Learning, 2014).

SEL total scaled scores were used in the present analyses rather than scores from the seven subscales within SEL because students may only see a limited number of items in some domains based on their item responses. Thus, scaled scores are considered the strongest estimate of a student's overall reading skills at a particular time (Renaissance Learning, 2014).

PRO-ED RAN: The PRO-ED Rapid Automatized Naming (RAN) measure indicates how quickly students can name common objects, colors, letters, or numeric symbols aloud. This measure is administered via paper and pencil, whereby a student says aloud the items presented on the stimulus card while the assessor records any errors and the student's total time.

The PRO-ED technical manual reports reliability coefficient ranges for kindergarten through Grade 5 as follows: test-retest reliability of 0.81 to 0.91; and interscorer reliability of 0.98 to 0.99 (PRO- ED, 2005). Criterion-prediction validity coefficients between PRO-ED RAN Number and Comprehensive Test of Phonological Processing (CTOPP) Rapid Digit Naming was .72 and between PRO-ED RAN Letters and CTOPP Rapid Letter Naming was .71. Concurrent validity was also examined using test scores from a variety of other reading measures (e.g., Woodcock Johnson Psycho-Educational Battery-Revised: Letter-Word Identification and Passage

Comprehension). The validity coefficients for PRO-ED RAN/RAS and Tests of Word Identification and Reading Comprehension ranged from 0.25 to 0.69 (PRO-ED, 2005).

mCLASS:Vocabulary: The vocabulary measure is administered on a computer or tablet and can be completed independently by the student. Each item and all answer options are spoken (by the computer) to the students who then select their answer choice. Students are first presented with tasks that require a deeper understanding of words. They are asked two questions about a word (a procedure based on Kearns and Biemiller's (2010) Two-Questions Vocabulary Measure) and asked to fill in the blank with the correct word. If they do not respond correctly, they are then asked to match words to their basic definitions. The time of year and grade level determine which of the three vocabulary tasks a student completes. The vocabulary measure is administered in kindergarten through Grade 3.

mCLASS:Spelling: The spelling measure is administered on a computer or tablet, and students complete the Spelling measure independently. The target word is spoken (by the computer) and the student uses letter tiles to spell the word. Like traditional spelling tests, the measure score provides the total number of words spelled correctly (WSC). To increase the sensitivity of the measure, the number of Correct Letter Sequences (CLS, number of letters correctly sequenced within a word) is also calculated to provide partial credit for words as students progress to becoming good spellers (Hosp, Hosp, & Howell, 2007). The spelling measure is administered from the middle of kindergarten through Grade 3.

mCLASS:RAN: The rapid automatized naming measure is administered on a computer or tablet, with a shared teacher-student interface. Students are presented with 50 stimulus items (five rows of 10) consisting of five randomly alternated numbers (i.e., the numbers 7, 9, 4, 1, and 6 are repeated in varying random orders 10 times). The educator administering the assessment sits with the student and enters each response as correct or incorrect when the student says the number aloud while working through each item in the row. When the student reaches the end of the row, he or she must move to the next row until all 50 items have been administered. The measure score is the total time in seconds that the student took to complete all 50 items. RAN is administered in kindergarten through Grade 3.

Research Design: During BOY and MOY benchmark periods, students were administered five assessments: mCLASS:DIBELS Next (Dynamic Measurement Group, 2010), mCLASS:Vocabulary, mCLASS:Spelling, mCLASS:RAN, STAR Reading (in grades 2 and 3) (SR: Renaissance Learning, 2011), and STAR Early Literacy (in kindergarten and Grade 1) (SEL; Renaissance Learning, 2014). One school was administered six assessments to include the PRO- ED version of the RAN assessment at MOY.

Descriptive Statistics: Descriptive information for mCLASS:DIBELS Next, mCLASS:Vocabulary, mCLASS:Spelling, mCLASS:RAN, and STAR Scaled Scores are provided in Tables 2, 3, 4, 5, and 6, respectively.

Table 2 shows percentages of students at each performance level on mCLASS:DIBELS Next. Percentages of students at each benchmark category are similar to those in a national sample of students; however, the differences in the research sample represent a more equal distribution of students in each benchmark category to ensure students with all skills are represented.

Table 2: Percentages of students at each benchmark performance level on mCLASS:DIBELS Next by Grade

	Beginning of Year					Middle of Year				
Measure by Grade	**Well Below**	**Below**	**At**	**Above**	**Overall CS**	**Well Below**	**Below**	**At**	**Above**	**Overall CS**
Kindergarten										
DIBELS Composite Score	40 (31%)	32 (25%)	19 (15%)	37 (29%)	128	43 (34%)	26 (21%)	20 (16%)	37 (29%)	126
National sample	26%	18%	15%	41%	364271	19%	19%	24%	39%	365662
Grade 1										
DIBELS Composite Score	43 (28%)	25 (16%)	29 (23%)	56 (37%)	158	45 (28%)	24 (15%)	34 (21%)	56 (35%)	159
National Sample	26%	14%	16%	43%	367313	24%	11%	18%	46%	363660
Grade 2										
DIBELS Composite Score	24 (21%)	18 (16%)	31 (27%)	41 (36%)	114	25 (22%)	16 (14%)	33 (29%)	40 (35%)	114
National Sample	20%	9%	26%	45%	357430	20%	8%	23%	48%	355253
Grade 3										
DIBELS Composite Score	45 (30%)	11 (7%)	33 (22%)	61 (41%)	150	39 (30%)	10 (8%)	32 (25%)	49 (38%)	130
National Sample	23%	9%	23%	44%	295858	22%	12%	23%	43%	292618

Table 3 shows STAR Scaled Scores by grade; the grade-level expectations for SEL in kindergarten and Grade 1 and SR in grades 2 and 3 are also listed for reference (Renaissance Learning, 2011 & 2015 respectively). Student mean performance increases across benchmark periods within a grade. In most cases, mean student performance was above grade-level expectations. The only exception was in Grade 3 at MOY.

Table 3: Descriptive Statistics for STAR Scale Score by Grade

	Beginning of Year				Middle of Year			
Grade	N	Mean	SD	Grade-Level Expectation	N	Mean	SD	Grade-Level Expectation
K	136	538.47	97.77	499	103	577.26	89.14	570
1	153	658.23	104.63	611	147	705.55	96.93	681
2	121	221.93	107.90	182	113	256.96	125.92	230
3	133	329.38	142.07	323	133	353.73	145.75	360

Table 4 shows student performance on mCLASS:Vocabulary. Depending on a student's grade level, he or she receives different types of vocabulary questions. In the kindergarten BOY assessment, students complete one item type for a total of 20 questions. In kindergarten MOY, students complete two item types for a total of 30 questions. In the Grade 1 BOY assessment, students complete two item types for a total of 30 questions. In Grade 1 MOY, students complete three item types for a total of up to 50 questions. In grades 2 and 3, students complete three item types for a total of up to 50 questions for each benchmark assessment during the year.

Thus, mean scores in BOY to MOY in kindergarten and Grade 1 reflect a different number of items. For grades 2 and 3, means increase from BOY to MOY. The means for Grade 3 at BOY is lower than for Grade 2, which likely reflects the increased difficulty of the words and items included for Grade 3. The means for MOY scores for grades 1 through 3 (the grades and times of year where the item types and number of possible items is consistent) increase slightly.

Table 4: Descriptive Statistics for mCLASS:Vocabulary by Grade

	Beginning of Year			Middle of Year		
Grade	N	Mean	SD	N	Mean	SD
K	115	13.91	2.68	115	18.92	5.31
1	128	19.21	3.74	137	32.66	8.72
2	106	31.62	7.21	114	35.67	7.26
3	111	22.18	3.90	119	36.18	7.02

Table 5 shows student performance on mCLASS:Spelling by grade level. The first kindergarten administration of the spelling measure was at MOY. While the student mean performance on spelling decreases from BOY to MOY, it's important to note that the format of the measure changed from paper and pencil at BOY to online at MOY.

Table 5: Descriptive Statistics for mCLASS:Spelling by Grade

	Beginning of Year			Middle of Year		
Grade	N	Mean	SD	N	Mean	SD
K	n/a	n/a	n/a	138	13.21	11.62
1	153	34.48	15.42	159	25.75	13.48
2	123	52.07	14.96	124	38.56	13.57
3	132	91.52	28.83	140	78.14	24.65

Table 6 shows student performance on mCLASS:RAN. RAN scores represent the total time in seconds a student takes to complete the measure. As such, the lower the mean the more quickly the student completed the measure. In most cases, student mean performance in seconds is consistent between BOY and MOY.

Table 6: Descriptive Statistics for mCLASS:RAN by Grade

	Beginning of Year					Middle of Year				
Grade	N	Mean	Min	Max	SD	N	Mean	Min	Max	SD
K	139	88.75	32	208	37.78	132	80.43	31.8	214.6	36.38
1	154	48.19	25	173	17.25	156	51.04	28.0	96.8	13.94
2	117	36.80	22	68	9.12	118	39.93	23.4	94.5	10.20
3	142	30.01	17	66	7.49	133	35.53	22.2	71.7	9.32

Reliability

Internal consistency reliability estimates (Cronbach's alpha) are shown in Table 7 for mCLASS:Vocabulary and Table 8 for mCLASS:Spelling. Overall, For the vocabulary measure, internal consistency reliability ranged from 0.51 to 0.85 in all but one instance (Kindergarten BOY) representing consistency above 0.60, which is characterized as adequate reliability (Salvia, Ysseldyke, & Bolt, 2013).

Table 7: Internal Consistency of mCLASS:Vocabulary

	Beginning of Year			Middle of Year		
Grade	N	Alpha	Number of items included in analysis	N	Alph	Number of items included in analysis
K	115	0.51	20	115	0.69	32
1	128	0.65	30	137	0.80	53
2	106	0.83	50	114	0.83	53
3	111	0.85	50	119	0.82	53

For spelling, internal consistency reliability ranged from 0.76 to 0.93, demonstrating strong to very strong internal consistency.

Table 8: Internal Consistency of Spelling

		Beginning of Year		Middle of Year	
Grade	Total number of items	N	Alpha	N	Alpha
K	12	n/a	n/a	138	0.91
1	12	153	0.91	159	0.88
2	12	123	0.86	124	0.76
3	17	132	0.93	140	0.88

These results demonstrate that the vocabulary and spelling measures are reliable assessments for making educational decisions. Note internal consistency reliability was not conducted for mCLASS:RAN. More appropriate measures of reliability for RAN measures are test-retest and interscorer reliability; analyses of internal consistency are not appropriate as RAN is based on speed with repeated presentation of the same items. RAN was administered to a subset of students approximately two weeks after the first administration to examine test-retest reliability. A reliability coefficient of 0.88 demonstrates a high degree of test-retest reliability for mCLASS:RAN.

Table 9: Test Reliability of mCLASS:RAN

	Middle of Year
Grade	**Test-Retest**
K-3 overall	0.88

Validity

Concurrent validity results for the vocabulary, spelling, and RAN measures with respect to mCLASS:DIBELS Next and the STAR Scaled Score are presented in Tables 10 through 13. For vocabulary and spelling comparison, correlations are presented between each of the mCLASS early literacy measures, the mCLASS:DIBELS Next Composite Score, and the STAR Scaled Score. The spelling measure was also correlated with Nonsense Word Fluency - Correct Letter Sounds (NWF-CLS), a measure of knowledge of letter-sound correspondence, for grades and times of year when NWF is administered as part of typical benchmarking assessment. For RAN comparisons, correlations are presented between RAN and the mCLASS:DIBELS Next Composite Score and PRO-ED RAN Number. All students who completed each pair of measures are included in the analyses. Correlations between RAN and PRO-ED RAN Number were not available for BOY as PRO-ED RAN was not administered.

mCLASS:Vocabulary: The relationship between mCLASS:Vocabulary and the DIBELS Next Composite Score was examined at BOY and MOY. The strongest correlation was demonstrated between vocabulary and Grade 3 DIBELS Next Composite Scores (0.57) at MOY. Similarly, the strongest correlation between vocabulary and the STAR Scaled Scores was demonstrated in Grade 3 (0.60) at MOY with correlations in kindergarten, Grade 1, and Grade 2 at or above 0.47 (0.50 , 0.53, and 0.47 respectively). It is important to note that mCLASS:Vocabulary is an assessment of vocabulary specifically, while the DIBELS Next Composite and STAR Scaled Scores are measures of overall reading skill.

Table 10: Concurrent Validity of mCLASS:Vocabulary with mCLASS:DIBELS Next and STAR Scaled Scores

	Beginning of Year		Middle of Year	
Grade	mCLASS DIBELS Next Composite Score	STAR Scaled Score	mCLASS DIBELS Next Composite Score	STAR Scaled Score
K	0.54	0.58	0.43	0.50
1	0.44	0.40	0.47	0.53
2	0.51	0.63	0.29	0.47
3	0.46	0.62	0.57	0.60

mCLASS:Spelling: Strong, positive relationships were found at BOY between spelling and CLS in grades 1 and 2; between spelling and DIBELS Next Composite Scores in grades 1, 2, and 3; and between spelling and STAR Scaled Scores in grades 1, 2, and 3. Validity coefficients ranged from 0.60 to 0.73. At MOY, moderate to strong correlations (at or above 0.50) were determined between spelling and NWF-CLS in grades 1, 2, and 3; spelling and DIBELS Next Composite Scores in kindergarten, Grade 1, Grade 2, and Grade 3; and between spelling and STAR Scaled Scores in kindergarten, Grade 1, and Grade 3. All correlations are as predicted; as expected, mCLASS:Spelling measures skills that are highly related, but not the same as, the skills assessed by DIBELS Next and STAR.

Table 11: Concurrent Validity of mCLASS:Spelling with mCLASS:DIBELS Next and STAR Scaled Scores

	Beginning of Year			Middle of Year		
Grade	NWF-CLS	mCLASS DIBELS Next Composite Score	STAR Scaled Score	NWF-CLS	mCLASS DIBELS Next Composite Score	STAR Scaled Score
K	n/a	n/a	n/a	0.66	0.73	0.60
1	0.61	0.68	0.73	0.61	0.68	0.58
2	0.60	0.70	0.60	0.56	0.57	0.48
3	n/a	0.73	0.71	n/a	0.70	0.65

mCLASS:RAN: Concurrent validity of mCLASS:RAN with PRO-ED RAN was examined at MOY. The relationship between mCLASS:RAN with PRO-ED RAN overall (for all students in kindergarten through Grade 3) was strong at 0.84.

Table 12: Concurrent Validity of mCLASS:RAN with PRO-ED RAN Number

	Middle of Year
Grade	PRO-ED RAN Number*
K-3 overall	0.84

Concurrent validity of mCLASS:RAN with DIBELS Next Composite Scores demonstrated moderate correlations at both BOY and MOY, which may be explained by the differences in the two assessments. The DIBELS Next Composite Score is a measure of overall reading whereas mCLASS:RAN measures only one aspect of reading. These findings are similar to those described above for concurrent correlations of PRO-ED RAN with measures of overall reading. The validity coefficients for PRO-ED RAN/RAS and Tests of Word Identification and Reading Comprehension ranged from 0.25 to 0.69 (PRO-ED, 2005).

Table 13: Concurrent Validity of mCLASS:RAN with mCLASS:DIBELS Next Composite Scores

	Beginning of Year	Middle of Year
Grade	**mCLASS DIBELS Next Composite Score**	**mCLASS DIBELS Next Composite Score**
K-3 overall	0.61	0.56

References

Araujo, S., Reis, A., Petersson, K. M., & Faisca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. **Journal of Educational Psychology, 107(3)**, 868-883.

Dynamic Measurement Group. (2010). Dynamic Indicators of Basic Early Literacy Skills (7th ed.). Eugene, OR: Author.

Dynamic Measurement Group. (2016). DIBELS Benchmark Goals and Composite Score report. Eugene, OR: Author.

Good, R.H., Kaminski, R.A., Dewey, E.N., Wallin, J., Powell-Smith, K.A., & Latimer, R.J. (2013). DIBELS Next Technical Manual. Eugene, OR: Dynamic Measurement Group.

Graves, M. F., Sales, G. C., & Ruda, M. (2008). **The first 4,000 words**. Minneapolis, MN: Seward. Hosp, M.K.,

Hosp, J.L., & Howell, K.W. (2007). **The ABCs of CBM: A Practical Guide to Curriculum-Based Measurement.** New York, NY: Guilford.

Kearns, G., & Biemiller, A. (2010/2011). Two-questions vocabulary assessment: Developing a new method for group testing in kindergarten through second grade. **Journal of Education, 190**, 31-41.

Marzano, R.J., & Simms, J.A. (2013). **Vocabulary for the Common Core**. Centennial, CO: Marzano Research.

PRO-ED. (2005). RAN/RAS: Rapid Automatized Naming and Rapid Alternating Stimulus Tests. Austin, TX: Author.

Renaissance Learning. (2011). **STAR Reading**. Wisconsin Rapids, WI: Author.

Renaissance Learning. (2014). **STAR Early Literacy technical manual**. Wisconsin Rapids, WI: Author. Available by request to research@renaissance.com

Renaissance Learning. (2015). **STAR Reading technical manual**. Wisconsin Rapids, WI: Author. Available by request to research@renaissance.com

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2013). **Assessment in special and inclusive education**. Belmont, CA: Wadsworth/Cengage Learning.

Woodcock, R.W., & Johnson, M.B. (1989). **Woodcock-Johnson Psycho-Educational Battery- Revised**. Itasca, IL: Riverside.

For more information visit amplify.com

Corporate:
55 Washington Street
Suite 900
Brooklyn, NY 11201-1071
(212) 796-2200

Sales Inquiries:
(866) 212-8688 • amplify.com

Amplify.