Dynamic Indicators of Basic Early Literacy Skills

8th Edition

Technical Manual

Contributing Authors

Gina Biancarosa, Ed. D.
Associate Professor and Ann Swindells Chair in Education
Department of Educational Methodology, Policy, and Leadership
College of Education, University of Oregon

Patrick C. Kennedy, Ph.D.
Research Associate and Director of Data Management and Analysis Group
Center on Teaching & Learning
College of Education, University of Oregon

Sunhi Park
Graduate Research Assistant Fellow
Center on Teaching & Learning
College of Education, University of Oregon

Janet Otterstedt, M.S.
Research Assistant and Project Coordinator
Center on Teaching & Learning
College of Education, University of Oregon

Brian Gearin, M.Ed.
Research Assistant and Project Coordinator
Center on Teaching & Learning
College of Education, University of Oregon

Christopher Ives
Graduate Research Assistant Fellow
Center on Teaching & Learning
College of Education, University of Oregon

HyeonJin Yoon, Ph.D.
Research Assistant and Project Coordinator
Center on Teaching & Learning
College of Education, University of Oregon

Contributing Editors

Maureen Warman, M.S.
Senior Research Assistant I and DIBELS Data System Manager
Center on Teaching & Learning
College of Education, University of Oregon

Table of Contents

Introduction

As part of our commitment to the continuous improvement of DIBELS, we proudly provide this technical manual for DIBELS 8th Edition. Note that we also have published a historical technical manual for those with DIBELS 8 data from the 2018-2019 school year. That technical manual reflects the evidence for the 2018-2019 preliminary goals. In this newer manual, we present information about the DIBELS 8 approach to norming, as well as sections on the reliability and validity evidence for DIBELS 8 as it stood for the 2019-2020 school year and onward. While much of the evidence presented here is the same as it was for the preliminary goals, this new 2019-2020 technical manual includes additional evidence in nearly every chapter including, new evidence regarding progress monitoring, and updated evidence regarding the classification accuracy of the finalized cut-scores. DIBELS users should reference this technical manual for the 2019-2020 school year and beyond.

Chapter 1: DIBELS 8th Edition Normative Information

This chapter describes normative information regarding DIBELS 8th Edition, including sample recruitment and selection procedures used in DIBELS 8 research studies, and demographic characteristics of the research sample. Four samples contribute to the technical adequacy evidence for DIBELS 8th Edition. The 2017-2018 Center on Teaching and Learning (CTL) research sample is dubbed Sample A for simplicity and contributes to reliability and validity evidence, as well as national norms. The 2018-2019 CTL research sample, or Sample B, contributes to classification accuracy analyses that yielded DIBELS 8 cut-scores, as well as to reliability and validity evidence and national norms. The 2018-2019 CTL dyslexia sample is called Sample C and represents a sub-sample of Sample B. Sample C took additional measures as part of validation of DIBELS 8 for use in dyslexia screening. The last sample, Sample D, was also collected in 2018-2019 by Amplify, and like Sample B, contributes to classification accuracy analyses that yielded DIBELS 8 cut-scores, as well as to reliability and validity evidence and national norms.

Sample Recruitment and Selection Procedures

CTL recruited elementary and middle schools from across the US to participate in DIBELS 8 research. Schools were recruited from the pool of DIBELS Data System users, through website postings and email contact, and via connections to colleagues of DIBELS 8 researchers, both within the University of Oregon and across the nation. Information about the project, including participation requirements and incentives, were communicated to potential participating schools via a flyer, email or phone. Regardless of format, schools received a description of the study and participation options. Interested schools were then asked for contact and other basic information by one of two methods: using a Qualtrics survey, via a link in the email, on the website or in the pop-up notice; or over the phone. An email confirmation of enrollment was then sent to the designated contact person.

We recruited schools until we met or exceeded our recruitment goals, or until it was no longer feasible for schools to assess students during the specified benchmark administration windows. Due to differences in school grade level configurations across the U.S., we expected to, and exceeded this amount for some grade levels to meet the goal for other grades. For example, for many states in the South, schools run K-4 and 5-8, while in the West and Northeast they more typically run K-5 and 6-8. Students were recruited from schools representing all these configurations, as well as from K-8 schools. To achieve our minimum for transitional grade levels like Grade 5, we ended up with larger sample sizes in the other grade levels.

All students enrolled in participating schools, including those with disabilities and students who were English language learners, were eligible for participation and were included based on local assessment procedures. Specifically, students who would normally be excluded from typical assessments in schools did not participate in DIBELS 8 assessments. At their discretion, schools could also opt not to assess students with disabilities who required assessment modifications.

These procedures apply to all three CTL samples (i.e., A, B, and C). Sample D was recruited by Amplify from the Amplify customer base. These users specifically took both DIBELS Next and DIBELS 8 benchmark measures. Due to human subjects limitations for this study, only scores and student-level demographic characteristics were shared with CTL, and students were not linked to schools.

Description of the DIBELS 8 Research Sample

The characteristics of the four research samples are summarized in Tables 1.1 to 1.4. In Table 1.1, we report the number of students by grade level. While Samples A and B both included Grades K to 8, Sample C was limited to kindergarten and first grade, and Sample D was conducted in Grades K to 3. Depending on the sample, schools ranged in size from very small (n = 7) to large (n = 790). In all but Sample C, which came exclusively from the Pacific division of the West region, schools were located throughout the country. All four Census regions and six of nine Census divisions were represented in Samples A and B (U.S. Department of Commerce, 1994; see Table 1.2). Sample D drew schools from three Census regions (i.e., Midwest, Northeast, and South) and four Census divisions: two states in the West South Central division, one state in the West North Central division, one state in the East North Central division, and one state in the Middle Atlantic division. Across the four samples, 17 states contributed to DIBELS 8 research (see Figure 1.1).

All four major locale Census designations (i.e., urban/city, suburban, town, and rural) are represented in Samples A and B, according to the National Center for Education Statistics (Geverdt, 2015; see Table 1.3). The Census locale definitions further classify locales into 12 distinct categories differentiated by size and proximity. Based on these classifications, about half of Sample A schools served rural areas (22% fringe, 22% distant, and 6% remote); approximately a quarter served towns (13% fringe and 9% remote); a fifth served large, suburban areas; and the remainder served cities (3% large and 4% small). In Sample B, almost two-fifths of the schools served rural areas (11 % fringe, 18 % distant, and 8 % remote); another two-fifths served urban areas (26 % large, 6% midsize and 5% small); and the remainder were evenly split between towns (6% fringe and 6% remote) and suburban areas (13% small). Sample C was almost evenly split between town (48% fringe) and rural (34% fringe and 18% remote) areas.

Individual demographic characteristics varied by sample (see Table 1.4). In all four samples, gender was evenly distributed between male and female. Across the four samples, the percentage of American Indian and Alaskan Native students ranged from 0.1% to 3.9%. Similarly, the percentage of Asian American students ranged from 0.1% to 2.5%, and the percentage of Black and African American students ranged from 0.1% to 17.2%. Hispanic and Latino students represented 9.7% to 25.1% of the samples. Native Hawaiian and Pacific Islands students represented 0% to 0.4% of the samples. Students identified as belonging to two or more races ranged from 1.2% to 7.0%, and White students ranged from 53% to 78.8%. Eligibility for special services also varied, with free and reduced meal eligibility ranging from 47.6% to 59.6%, English language support eligibility ranging from 1.1% to 7.3%, and special education eligibility ranging from 7.2% to 13.9%.

Importantly, all four samples contributed to DIBELS 8 percentile ranks. As a result, DIBELS 8 norms are more representative than ever before in terms of regions, states, locales, and student characteristics.

Table 1.1 Number of Students by Grade

	Sample			
Grade	A	B	C	D
K	685	678	213	429
1	782	694	195	369
2	725	853	--	221
3	783	710	--	256
4	408	683	--	--
5	388	645	--	--
6	276	422	--	--
7	216	314	--	--
8	190	260	--	--
Total	4453	5259	408	1275

Note. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study.

Table 1.2 Number of Students and Schools by Census Region and Division

| | | Sample | | | | | |
| | | A | | B | | C | |
Region	Division	Students	Schools	Students	Schools	Students	Schools
West		2681	21	3374	12	408	4
	Pacific	1771	15	2949	10	408	4
	Mountain	910	6	425	2	--	--
Midwest		765	5	741	4	--	--
	West North Central	156	1	162	1	--	--
	East North Central	609	4	579	3	--	--
Northeast		846	2	291	1	--	--
	Middle Atlantic	846	2	291	1	--	--
	New England	--	--	--	--	--	--
South		161	1	853	4	--	--
	West South Central	--	--	--	--	--	--
	East South Central	--	--	--	--	--	--
	South Atlantic	161	1	853	4	--	--

Note. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample.

Table 1.3 Number of Students and Schools by Census Locale Classification

| | | Sample | | | | | |
| | | A | | B | | C | |
Region	Division	Students	Schools	Students	Schools	Students	Schools
City		310	3	1,937	7	--	--
	Large	114	1	1,374	4	--	--
	Midsize	--	--	320	2	--	--
	Small	196	2	243	1	--	--
Suburb		913	4	704	2	--	--
	Large	913	4	--	--	--	--
	Midsize	--	--	--	--	--	--
	Small	--	--	704	2	--	--
Town		969	4	624	2	285	1
	Fringe	572	2	293	1	285	1
	Distant	--	--	--	--	--	--
	Remote	397	2	331	1	--	--
Rural		2,261	14	1,994	10	308	3
	Fringe	988	5	579	3	199	1
	Distant	1,002	7	966	6	109	2
	Remote	271	2	449	1	--	--

 Note. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample.

Table 1.4 Demographic Characteristics of Sample

	Sample							
	A		B		C		D	
Characteristic Category	n	%	n	%	n	%	n	%
Gender								
Female	2142	48.1	2570	48.9	290	48.9	603	47.3
Male	2265	50.9	2663	50.6	303	51.2	663	52.0
Race & ethnicity								
American Indian & Alaskan Native	173	3.9	77	1.5	23	3.9	1	0.1
Asian American	26	0.6	133	2.5	2	0.3	1	0.1
Black & African American	636	14.3	904	17.2	1	0.1	7	0.5
Hispanic & Latino	804	18.1	1101	20.9	57	9.7	320	25.1
Native Hawaiian & other Pacific Islander	16	0.4	23	0.4	2	0.3	--	--
Two or more races	144	3.2	215	4.1	41	7.0	15	1.2
White	2863	64.3	2787	53.0	467	78.8	920	72.2
Not reported	594	13.3	--	--	--	--	--	--
Free & reduced-price meals eligible	2557	57.4	2557	57.4	282	47.6	628	49.3
English learner designated	279	6.3	381	7.3	16	2.7	14	1.1
Special education eligible	619	13.9	730	13.9	106	17.9	92	7.2

Note. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study.
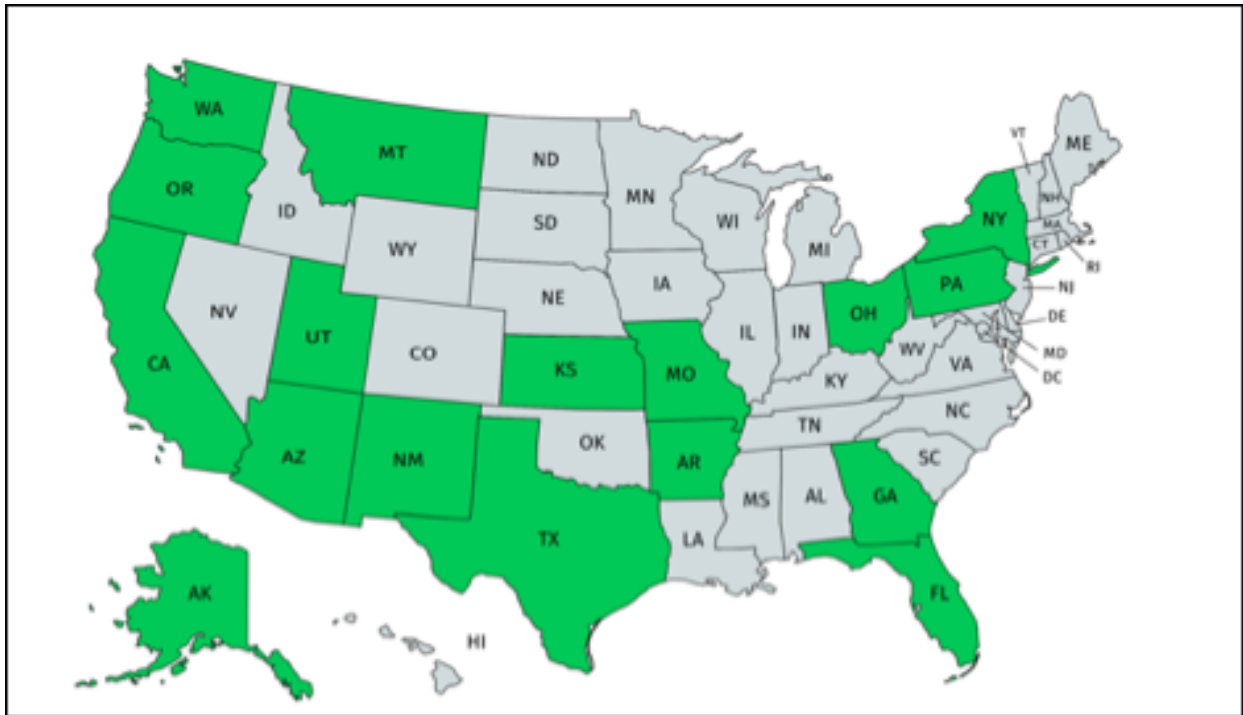
Figure 1. *Schools in states shaded green participated in DIBELS 8 studies and national norming sample. (Figure created with mapchart.net)*

Chapter 2: Reliability of DIBELS 8th Edition

We examined five forms of test reliability for DIBELS 8th Edition: concurrent alternate form reliability, delayed alternate form reliability, test-retest reliability, intercept and slope reliability for progress monitoring, and standard errors of measurement. Reliability refers to the extent to which a test score is a consistent and stable measure. All these forms of reliability can be thought of as estimates of the stability of scores. Delayed alternate form and test-retest reliability also address the stability of scores over time, while slope reliability addresses the stability of change estimates for progress monitoring. Reliability is a necessary, but insufficient component of validity. We present further validity evidence in the chapter entitled Validity of DIBELS 8th Edition.

We provide individual coefficients in tables by subtest and for the composite score for each type of reliability in turn. We also provide median coefficients by subtest per grade where more than one coefficient is present, which is most cases. In all cases, we provide median coefficients across grades as an overall estimate of each type of reliability for each subtest and the composite. In cases where the number of coefficients contributing to a median was even, the lower of the two coefficients is reported, thereby providing a conservative estimate of reliability. In cases where coefficients are repeated, we ranked these coefficients so that those with wider confidence intervals, and thus less precision, were treated as the lower coefficient. We only provide coefficients where sample sizes were 50 or greater to ensure a minimum of precision in the reliability estimates.

In the first sections, we present concurrent and delayed alternate form reliability evidence, followed by test-retest reliability evidence. Next, we present model-based estimates of intercept and slope reliability. Finally, we report standard errors of measurement and conclude with a summary of all DIBELS 8 reliability evidence.

Alternate Form Reliability

Alternate form reliability describes the relationship between scores produced with different versions of a test. In general, strong correlations are desirable because they imply that different versions of the test are capable of generating similar scores. To obtain excellent alternate form reliability, we used strict item writing and form generation guidelines. Nonetheless, reliability must be tested empirically to establish validity of a measure for almost any purpose.

To calculate alternate form reliability, different versions of each DIBELS subtest were administered at the beginning, middle, and end of each year in kindergarten through eighth grade. Because concurrent alternate form reliability measures are administered in the same sitting, the resulting correlations are expected to be quite strong. In contrast, delayed alternate form reliability measures are administered over longer periods of time and are expected to be more moderate in strength. These expectations are especially the case for measures like DIBELS, which targets precisely the skills that are the subject of instruction. Thus, if students are learning as intended, scores from delayed administrations should be less stable than those from concurrent administrations.

Concurrent alternate form reliability. We studied all DIBELS 8th Edition subtests except LNF for concurrent alternate form reliability. We do not report concurrent alternate form reliability for LNF because it is not used in progress monitoring, but other sources of reliability for LNF are reported in the next sections. As illustrated in Table 2.1, results of concurrent reliability studies over two years reveal very strong correlations (r > .90) for all subtests except PSF and Maze, and PSF and Maze still demonstrate adequate to strong reliability.

In every case, PSF alternate form reliability for forms administered concurrently was .80 or above. The median reliability of PSF in kindergarten was .86 and in first grade was .81. The overall median reliability of PSF was also .81.

NWF provides two scores (i.e., CLS and WRC), and we examined each for concurrent alternate form reliability. As reported in Tables 2.2 and 2.3, the results in all cases for both types of scores are highly reliable. That is, most coefficients are .90 or above. For NWF-CLS, the median reliability was .89 or above in all grades with an overall median reliability of .91. For NWF-WRC, the median reliability was .88 or above in all grades, and overall median reliability was .90.

Results for WRF are displayed in Table 2.4. Here, concurrent alternate form reliability is very strong and nearly at a maximum. All but four coefficients exceed .94, and the median reliability for WRF in all grades was .94 or above. Overall median reliability was .95.

Results for ORF and ORF-ACC are displayed in Tables 2.5 and 2.6. As with NWF scores and WRF, concurrent alternate form reliability is very strong for ORF. The median reliability for ORF was .92 or above in all grades with the strongest reliability seen in Grade 1 where the median reliability was .96. Overall median reliability for ORF words read correctly scores was .93. ORF-ACC is similar to PSF with reliability being strong, but not quite as strong as for ORF and the other subtests. ORF-ACC median reliability ranged from .75 to .89 across grades, and overall median reliability was .83.

Maze concurrent alternate form reliability was somewhat weaker than other DIBELS subtests (see Table 2.7). Medians by grade ranged from .66 to .81, and the overall median for Maze was

.72. It appears that passage differences make a bigger difference for Maze than they do for ORF and ORF-ACC. The general lower reliability of this subtest suggests it should not be used alone in making high-stakes decisions about students; however, good educational practice is never to rely on a single test for such decisions.

Table 2.1 Concurrent Alternate Form Reliability for DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	Sample	Forms	N	r	CI
	A	1:PM9	93	.80	.72–.87
		2:PM8	153	.88	.83–.91
		3:PM5	119	.80	.72–.85
Kindergarten		3:PM1	105	.86	.80–.90
	B	1:PM7	224	.90	.87–.92
		2:PM12	150	.92	.89–.94
	Median			.86	.80–.90
	B	1:PM5	132	.82	.76–.87
		1:PM6	150	.85	.80–.89
First		2:PM6	153	.81	.74–.85
		3:PM6	156	.84	.79–.88
	Median			.82	.76–.87
Median				.82	.76–.87

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.2 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
	A	1:PM2	92	.95	.93–.97
		2:PM2	109	.92	.89–.95
		3:PM2	113	.93	.89–.95
Kindergarten		3:PM7	51	.97	.94–.98
	B	1:PM10	179	.92	.89–.94
		3:PM8	164	.92	.89–.94
	Median			.92	.89–.94

Table 2.2 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
First	A	1:PM1	123	.95	.93–.96
		2:PM1	126	.89	.84–.92
		3:PM1	127	.85	.79–.89
	Median			.89	.84–.92
Second	A	2:PM1	50	.94	.89–.97
		3:PM1	52	.93	.87–.96
	B	1:PM9	71	.92	.87–.95
		1:PM10	228	.91	.88–.93
		1:PM12	138	.92	.89–.94
		2:PM3	136	.88	.84–.92
		2:PM8	69	.94	.90–.96
		2:PM13	258	.91	.88–.93
		2:PM14	83	.91	.86–.94
		3:PM5	66	.93	.89–.96
		3:PM15	85	.94	.91–.96
		3:PM17	201	.92	.90–.94
		3:PM20	135	.88	.84–.91
	Median			.92	.89–.94
Third	B	1:PM10	161	.87	.83–.90
		1:PM13	63	.86	.77–.91
		1:PM14	192	.90	.86–.92
		2:PM4	169	.90	.87–.93
		2:PM6	70	.94	.91–.97
		2:PM9	223	.92	.89–.94
		2:PM15	68	.90	.85–.94
		3:PM7	70	.89	.82–.93
		3:PM12	179	.89	.86–.92
		3:PM18	166	.95	.93–.96
		3:PM11	73	.96	.94–.97
	Median			.90	.87–.93
Median				.90	.87–.93

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.3 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Sample	Forms	N	r	CI
Kindergarten	A	1:PM2	87	.95	.92–.96
		2:PM2	107	.86	.80–.90
		3:PM2	112	.89	.84–.92
		3:PM7	51	.91	.84–.95
	B	1:PM10	179	.84	.79–.88
		3:PM8	164	.88	.85–.91
	Median			.88	.85–.91
First	A	1:PM1	123	.90	.86–.93
		2:PM1	126	.90	.85–.93
		3:PM1	127	.86	.81–.90
	Median			.90	.86–.93
Second	A	2:PM1	50	.97	.95–.98
		3:PM1	52	.93	.88–.96
	B	1:PM9	71	.91	.86–.94
		1:PM10	228	.91	.88–.93
		1:PM12	138	.92	.89–.94
		2:PM3	136	.90	.86–.93
		2:PM8	69	.95	.92–.97
		2:PM13	258	.91	.89–.93
		2:PM14	83	.93	.89–.95
		3:PM5	66	.95	.92–.97
		3:PM15	85	.94	.91–.96
		3:PM17	201	.94	.92–.95
		3:PM20	135	.89	.85–.92
	Median			.92	.89–.94

Table 2.3 Concurrent Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Sample	Forms	N	r	CI
	B	1:PM10	161	.88	.84−.91
		1:PM13	63	.85	.77−.91
		1:PM14	192	.90	.87−.93
		2:PM4	169	.92	.89−.94
		2:PM6	70	.95	.93−.97
Third		2:PM9	223	.93	.91−.95
		2:PM15	68	.93	.89−.96
		3:PM7	70	.89	.83−.93
		3:PM12	179	.92	.89−.94
		3:PM18	166	.95	.94−.96
		3:PM11	73	.97	.95−.98
	Median			.92	.89−.94
Median				.90	.86−.93

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.4 Concurrent Alternate Form Reliability for DIBELS 8th Edition Word Reading Fluency

Grade	Sample	Forms	N	r	CI
	A	3:PM4	87	.97	.96−.98
	B	1:PM6	123	.96	.95−.97
		2:PM7	51	.96	.94−.98
Kindergarten		2:PM11	193	.87	.83−.90
		3:PM20	140	.97	.95−.98
	Median			.96	.95−.97
	A	2:PM3	59	.97	.96−.98
		3:PM3	59	.96	.93−.97
	B	1:PM3	201	.96	.95−.97
		1:PM18	102	.95	.92−.96
First		2:PM3	236	.97	.97−.98
		2:PM19	71	.97	.96−.98
		3:PM3	193	.98	.97−.98
		3:PM20	60	.93	.89−.96
	Median			.96	.93−.97

Table 2.4 Concurrent Alternate Form Reliability for DIBELS 8th Edition Word Reading Fluency

Grade	Sample	Forms	N	r	CI
Second	A	2:PM7	68	.95	.92–.97
		3:PM4	118	.92	.88–.94
	B	1:PM7	139	.94	.92–.96
		1:PM13	69	.96	.94–.98
		1:PM15	230	.96	.94–.97
		2:PM12	69	.94	.91–.96
		2:PM16	84	.95	.92–.97
		2:PM17	136	.95	.93–.97
		2:PM20	259	.95	.93–.96
		3:PM10	66	.94	.91–.96
		3:PM11	84	.98	.97–.99
		3:PM14	136	.95	.93–.96
		3:PM18	202	.95	.94–.97
	Median			.95	.93–.97
Third	A	2:PM9	146	.96	.94–.97
		3:PM9	50	.95	.91–.97
	B	1:PM9	367	.94	.92–.95
		1:PM19	64	.96	.93–.97
		2:PM12	225	.95	.94–.96
		2:PM14	68	.90	.85–.94
		2:PM17	168	.94	.93–.96
		2:PM20	71	.94	.91–.96
	Median			.94	.93–.96
Median				.95	.93–.97

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	*N*	*r*	CI
First	A	1:PM5	128	.97	.96–.98
		1:PM1	75	.98	.96–.99
		2:PM7	184	.96	.95–.97
		3:PM8	186	.94	.92–.95
	Median			.96	.95–.97
Second	A	1:3	118	.95	.92–.96
		1:PM5	109	.97	.96–.98
		1:PM7	108	.97	.95–.98
		2:3	118	.95	.93–.97
		2:PM2	159	.96	.94–.97
		2:PM6	159	.96	.95–.97
		3:PM3	118	.95	.92–.96
	B	1:PM16	365	.95	.93–.96
		2:PM15	393	.96	.96–.97
		3:PM20	338	.95	.94–.96
	Median			.95	.93–.97
Third	A	1:PM3	114	.93	.90–.95
		2:PM1	196	.95	.94–.96
		2:PM2	196	.93	.91–.94
		3:PM5	180	.91	.88–.93
		3:PM12	180	.89	.86–.92
	B	1:PM14	366	.94	.92–.95
		2:PM10	391	.94	.93–.95
		3:PM19	349	.93	.91–.94
	Median			.93	.91–.94

Table 2.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	*N*	*r*	CI
Fourth	A	1:PM2	146	.93	.91–.95
		1:PM3	146	.94	.91–.95
		1:PM5	147	.94	.92–.96
		2:PM2	145	.81	.74–.86
		2:PM7	144	.87	.83–.91
		2:PM10	145	.85	.80–.89
		3:PM2	144	.92	.89–.94
		3:PM8	143	.88	.84–.91
	B	1:PM12	436	.94	.92–.95
		1:PM20	339	.94	.93–.95
		2:PM18	359	.94	.93–.95
		2:PM20	510	.94	.93–.95
		3:PM14	488	.95	.94–.95
		3:PM15	343	.94	.93–.95
	Median			.94	.93–.95
Fifth	A	1:2	123	.95	.92–.96
		1:3	131	.91	.87–.93
		1:PM5	133	.92	.89–.94
		1:PM1	132	.92	.89–.94
		1:PM3	133	.92	.89–.94
		2:PM2	181	.93	.91–.95
		2:PM20	58	.94	.89–.96
		3:PM7	131	.93	.90–.95
		3:PM8	131	.92	.89–.95
	B	1:PM14	327	.93	.91–.94
		1:PM15	387	.91	.89–.92
		2:PM16	399	.89	.87–.91
		2:PM19	519	.93	.91–.94
		3:PM17	349	.91	.88–.92
		3:PM20	467	.93	.92–.94
	Median			.92	.89–.94

Table 2.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Sixth	A	1:2	69	.94	.91–.96
		1:PM19	69	.95	.93–.97
		2:3	98	.93	.90–.95
		2:PM1	99	.89	.84–.93
		2:PM2	104	.93	.90–.96
		2:PM3	99	.94	.91–.96
		2:PM4	104	.93	.89–.95
		2:PM11	104	.93	.90–.95
		2:PM16	98	.95	.92–.96
		3:PM15	98	.94	.91–.96
	B	1:PM13	182	.92	.90–.94
		1:PM16	166	.92	.89–.94
		2:PM7	292	.94	.92–.95
		2:PM14	167	.91	.88–.93
		3:PM8	296	.92	.90–.93
		3:PM12	167	.94	.92–.96
	Median			.93	.90–.96
Seventh	A	1:2	136	.92	.89–.94
		1:3	123	.90	.86–.93
		1:PM8	83	.93	.89–.95
		1:PM9	83	.87	.80–.91
		1:PM11	83	.93	.89–.95
		2:PM1	142	.92	.88–.94
		2:PM2	141	.91	.88–.94
		3:PM3	123	.89	.84–.92
	B	1:PM13	63	.91	.86–.95
		1:PM19	77	.95	.92–.97
		2:PM12	201	.95	.93–.96
		2:PM18	64	.93	.89–.96
		3:PM7	196	.95	.94–.96
		3:PM20	64	.94	.90–.96
	Median			.92	.89–.94

Table 2.5 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency

Grade	Sample	Forms	N	r	CI
Eighth	A	1:2	110	.92	.89−.95
		1:3	102	.81	.73−.87
		1:PM7	69	.92	.87−.95
		1:PM11	70	.91	.86−.94
		1:PM12	70	.92	.86−.95
		2:PM2	114	.90	.86−.93
		2:PM4	114	.92	.89−.94
		3:PM1	102	.80	.72−.86
		3:PM3	102	.78	.69−.85
	B	1:PM13	66	.95	.92−.97
		1:PM16	74	.95	.92−.97
		2:PM6	137	.92	.89−.94
		2:PM10	64	.94	.91−.96
		3:PM3	144	.93	.90−.95
		3:PM19	65	.95	.92−.97
	Median			.92	.87−.95
Median				.93	.90−.96

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
First	A	1:PM5	128	.92	.89−.94
		2:PM7	184	.90	.87−.93
		2:PM1	75	.92	.86−.95
		3:PM8	186	.91	.88−.93
	Median			.91	.88−.93

Table 2.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Second	A	1:3	118	.77	.69–.84
		1:PM5	109	.91	.87–.94
		1:PM7	108	.89	.85–.93
		2:3	118	.83	.77–.88
		2:PM2	159	.90	.87–.93
		2:PM6	159	.87	.83–.90
		3:PM3	118	.78	.70–.84
	B	1:PM16	365	.85	.82–.88
		2:PM15	393	.92	.90–.93
		3:PM20	338	.92	.91–.94
	Median			.87	.83–.90
Third	A	1:2	114	.79	.71–.85
		1:PM3	114	.80	.90–.95
		2:PM1	196	.96	.94–.97
		2:PM2	196	.96	.95–.97
		3:PM5	180	.67	.58–.74
		3:PM12	180	.72	.64–.79
	B	1:PM14	366	.79	.75–.82
		2:PM10	391	.83	.80–.86
		3:PM19	349	.71	.65–.76
	Median			.79	.71–.85
Fourth	A	1:PM2	146	.74	.65–.80
		1:PM3	146	.75	.67–.81
		1:PM5	147	.78	.71–.84
		2:PM2	145	.76	.68–.82
		2:PM7	144	.60	.49–.70
		2:PM10	145	.61	.50–.71
		3:PM2	144	.86	.81–.90
		3:PM8	143	.67	.57–.75
	B	1:PM12	436	.88	.86–.90
		1:PM20	339	.87	.84–.89
		2:PM18	359	.87	.84–.89
		2:PM20	510	.87	.85–.89
		3:PM14	488	.90	.88–.91
		3:PM15	343	.77	.72–.81
	Median			.77	.72–.81

Table 2.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Fifth	A	1:2	123	.97	.95–.98
		1:3	131	.57	.45–.68
		1:PM1	132	.76	.68–.83
		1:PM3	133	.68	.58–.76
		1:PM5	133	.73	.64–.80
		2:PM2	181	.96	.95–.97
		2:PM20	58	.94	.91–.97
		3:PM7	131	.54	.41–.65
		3:PM8	131	.65	.53–.74
	B	1:PM14	327	.75	.70–.80
		1:PM15	387	.85	.82–.87
		2:PM16	399	.79	.75–.83
		2:PM19	519	.89	.87–.90
		3:PM17	349	.59	.52–.67
		3:PM20	467	.55	.48–.61
	Median			.75	.70–.80
Sixth	A	1:2	69	.77	.65–.85
		1:PM19	69	.84	.75–.90
		2:3	98	.78	.69–.85
		2:PM1	99	.68	.56–.78
		2:PM2	104	.90	.87–.94
		2:PM3	99	.64	.50–.74
		2:PM4	104	.83	.76–.88
		2:PM11	104	.77	.68–.84
		2:PM16	98	.76	.67–.84
		3:PM15	98	.81	.73–.87
	B	1:PM13	182	.97	.96–.98
		1:PM16	166	.96	.94–.97
		2:PM7	292	.93	.91–.94
		2:PM14	167	.92	.90–.94
		3:PM8	296	.98	.97–.98
		3:PM12	167	.94	.92–.96
	Median			.83	.76–.88

Table 2.6 Concurrent Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Seventh	A	1:2	136	.89	.85–.92
		1:3	123	.93	.90–.95
		1:PM8	83	.89	.83–.93
		1:PM9	83	.90	.85–.94
		1:PM11	83	.89	.83–.93
		2:PM1	142	.87	.82–.90
		2:PM2	141	.83	.78–.88
		3:PM3	123	.87	.81–.91
	B	1:PM13	63	.68	.51–.79
		1:PM19	77	.98	.97–.99
		2:PM12	201	.94	.92–.95
		2:PM18	64	.53	.32–.68
		3:PM7	196	.91	.88–.93
		3:PM20	64	.79	.68–.87
	Median			.89	.83–.93
Eighth	A	1:2	110	.85	.79–.89
		1:3	102	.82	.75–.88
		1:PM7	69	.84	.75–.90
		1:PM11	70	.86	.78–.91
		1:PM12	70	.73	.60–.83
		2:PM2	114	.84	.76–.89
		2:PM4	114	.89	.85–.92
		3:PM1	102	.76	.66–.83
		3:PM3	102	.85	.79–.90
	B	1:PM13	66	.90	.85–.94
		1:PM16	74	.88	.82–.92
		2:PM6	137	.76	.68–.83
		2:PM10	64	.91	.85–.94
		3:PM3	144	.80	.74–.86
		3:PM19	65	.78	.67–.86
	Median			.84	.75–.90
Median				.83	.76–.88

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.7 Concurrent Alternate Form Reliability Coefficients for DIBELS 8th Edition MAZE Adjusted Score

Grade	Forms	N	r	CI
Second	1:2	502	.70	.65−.74
	1:3	462	.72	.67−.76
	1:PM1	548	.70	.66−.74
	1:PM2	97	.74	.63−.82
	1:PM3	200	.73	.65−.79
	2:3	433	.66	.61−.71
	2:PM1	527	.72	.67−.76
	2:PM2	126	.71	.61−.78
	2:PM3	208	.62	.52−.69
	3:PM1	475	.73	.69−.77
	3:PM2	131	.78	.71−.84
	3:PM3	189	.64	.54−.71
	Median		.71	.61−.78
Third	1:2	553	.71	.66−.75
	1:3	465	.64	.58−.69
	1:PM1	575	.72	.67−.75
	1:PM2	150	.72	.63−.79
	1:PM3	180	.79	.73−.84
	2:3	477	.67	.62−.72
	2:PM1	572	.74	.70−.77
	2:PM2	160	.79	.72−.84
	2:PM3	207	.74	.67−.79
	3:PM1	490	.73	.68−.77
	3:PM2	163	.84	.79−.88
	3:PM3	191	.72	.64−.78
	Median		.72	.63−.79

Table 2.7 Concurrent Alternate Form Reliability Coefficients for DIBELS 8th Edition MAZE Adjusted Score

Grade	Forms	N	r	CI
Fourth	1:2	585	.72	.68–.76
	1:3	486	.66	.61–.71
	1:PM1	583	.65	.60–.70
	1:PM2	150	.73	.64–.80
	1:PM3	200	.76	.70–.81
	2:3	482	.76	.72–.79
	2:PM1	576	.68	.64–.73
	2:PM2	156	.86	.81–.90
	2:PM3	206	.75	.68–.80
	3:PM1	475	.78	.74–.81
	3:PM2	162	.87	.82–.90
	3:PM3	183	.74	.67–.80
	Median		.74	.67–.80
Fifth	1:2	521	.73	.68–.77
	1:3	426	.55	.49–.62
	1:PM1	485	.52	.46–.59
	1:PM2	146	.66	.56–.74
	1:PM3	162	.64	.54–.72
	2:3	429	.66	.61–.71
	2:PM1	508	.62	.57–.67
	2:PM2	152	.74	.66–.81
	2:PM3	181	.67	.59–.75
	3:PM1	398	.76	.71–.80
	3:PM2	160	.81	.74–.85
	3:PM3	152	.66	.56–.74
	Median		.66	.61–.71

Table 2.7 Concurrent Alternate Form Reliability Coefficients for DIBELS 8th Edition MAZE Adjusted Score

Grade	Forms	N	r	CI
Sixth	1:2	280	.73	.68–.78
	1:3	231	.81	.76–.85
	1:PM1	111	.69	.57–.77
	1:PM2	159	.73	.64–.79
	2:3	251	.82	.77–.86
	2:PM1	120	.83	.77–.88
	2:PM2	158	.86	.81–.89
	2:PM3	57	.74	.59–.84
	3:PM1	63	.83	.74–.90
	3:PM2	162	.84	.79–.88
	Median		.81	.76–.85
Seventh	1:2	268	.62	.54–.68
	1:3	172	.81	.75–.86
	1:PM1	125	.77	.69–.83
	1:PM2	64	.89	.83–.93
	1:PM3	62	.84	.75–.90
	2:3	164	.61	.50–.70
	2:PM1	146	.78	.71–.84
	2:PM2	64	.84	.75–.90
	2:PM3	89	.77	.66–.84
	3:PM1	65	.79	.68–.87
	3:PM2	64	.89	.82–.93
	Median		.79	.68–.87
Eighth	1:2	184	.48	.36–.59
	1:3	112	.75	.66–.82
	1:PM1	99	.75	.65–.83
	1:PM2	64	.72	.58–.82
	1:PM3	86	.72	.60–.81
	2:3	96	.40	.21–.55
	2:PM1	110	.62	.49–.72
	2:PM2	64	.77	.64–.85
	2:PM3	100	.59	.44–.70
	3:PM2	65	.84	.75–.90
	Median		.72	.58–.82
Median			.72	.63–.79

Note. Form numbers correspond to benchmark period forms. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Delayed alternate form reliability. We studied all DIBELS 8th Edition subtests for delayed alternate form reliability. We established delayed alternate form reliability by correlating benchmark forms; thus, the delay between one form and the next was approximately three months. As a result, these results were expected to be weaker than concurrent alternate form reliability because three months of instruction intervened between administration of alternate forms in the calculation of delayed alternate form correlations.

Median delayed alternate form reliability for LNF was .81 in kindergarten, .76 in first grade, and .76 overall (see Table 2.8). Median delayed alternate form reliability was lower for PSF (see Table 2.9). PSF median correlations were .49 in kindergarten, .54 in first grade, and .49 across the two grades. The lower correlations for PSF are expected given the strong growth that occurs in kindergarten in phonological awareness in general and phonemic awareness specifically.

For the NWF-CLS score, delayed alternate form reliability was .77 across grades and ranged from .71 in kindergarten to .83 in second grade (see Table 2.10). For the NWF-WRC score, delayed alternate form reliability was .72 across grades and ranged from .62 in kindergarten to .81 in second grade (see Table 2.11). Again, the lower correlations occurring in kindergarten are expected given the steep growth in early literacy skills seen in this grade. WRF coefficients (see Table 2.12) demonstrated strong delayed alternate form reliability of .87 overall and ranged from .85 in kindergarten to .91 in second grade.

ORF-WRC coefficients (see Table 2.13) also demonstrated strong delayed alternate form reliability ranging from .82 in fifth grade to .90 in first grade with an overall median reliability of .88. ORF-ACC scores demonstrated slightly weaker delayed alternate form reliability, ranging from .66 in third grade to .87 in sixth grade with an overall median reliability of .73 (see Table 2.14). Maze delayed alternate form reliability was similar to ORF-ACC in that medians ranged from .69 in second grade to .86 in seventh grade with an overall median of .77 (see Table 2.15).

We also calculated delayed alternate form reliability for the composite score, which was not possible for concurrent alternate form reliability. Since benchmark forms are defined for each time of year, correlating composite scores for two times of year generates an estimate of the aggregate stability of DIBELS 8 scores over time for different forms. These results were stronger than any individual subtest such that overall median delayed alternate form reliability of the composite was .89. Grade level medians ranged from a low of .80 in kindergarten, which is considered strong evidence of reliability, to a high of .94, which is considered exceptionally strong evidence of reliability.

Table 2.8 Delayed Alternate Form Reliability for DIBELS 8th Edition Letter Naming Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	A	1:2	150	.86	.82–.90
		1:3	100	.83	.76–.88
		2:3	137	.87	.81–.90
	B	1:2	420	.81	.77–.84
		1:3	531	.67	.62–.71
		2:3	375	.80	.76–.83
	Median			.81	.77–.84
First	A	1:2	153	.77	.70–.83
		1:3	140	.70	.61–.78
		2:3	183	.76	.70–.82
	B	1:2	413	.83	.80–.86
		1:3	511	.73	.68–.77
		2:3	413	.79	.75–.83
	Median			.76	.70–.82
Median				.76	.70–.82

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.9 Delayed Alternate Form Reliability for DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	A	1:2	94	.49	.32–.63
		1:3	72	.39	.18–.57
		2:3	115	.70	.59–.78
	B	1:2	391	.53	.46–.60
		1:3	505	.34	.26–.42
		2:3	361	.68	.62–.73
	Median			.49	.32–.63
First	A	1:2	148	.53	.40–.64
		1:3	138	.54	.41–.65
		2:3	181	.63	.53–.71
	B	1:2	412	.65	.59–.70
		1:3	509	.42	.35–.49
		2:3	413	.63	.57–.69
	Median			.54	.41–.65

Table 2.9 Delayed Alternate Form Reliability for DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	Sample	Forms	N	r	CI
Median				.49	.32–.63

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.10 Delayed Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Sample	Forms	N	r	CI
Kindergarten	A	1:2	191	.73	.65–.79
		1:3	180	.63	.53–.71
		2:3	224	.83	.78–.87
	B	1:2	343	.71	.65–.76
		1:3	447	.60	.54–.66
		2:3	353	.79	.75–.83
	Median			.71	.65–.76
First	A	1:2	268	.80	.75–.84
		1:3	256	.73	.67–.78
		2:3	308	.79	.75–.83
	B	1:2	409	.82	.79–.85
		1:3	503	.71	.67–.75
		2:3	408	.79	.75–.82
	Median			.79	.75–.83
Second	A	1:2	127	.83	.77–.88
		1:3	113	.84	.77–.88
		2:3	194	.84	.79–.87
	B	1:2	422	.74	.70–.78
		1:3	383	.78	.74–.82
		2:3	463	.86	.83–.88
	Median			.83	.77–.88
Third	A	2:3	218	.76	.70–.81
	B	1:2	401	.81	.77–.84
		1:3	367	.77	.72–.81
		2:3	468	.86	.84–.88
	Median			.77	.72–.81
Median				.77	.72–.81

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.11 Delayed Alternate Form Reliability for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Sample	Forms	N	r	CI
Kindergarten	A	1:2	171	.70	.62–.77
		1:3	161	.60	.49–.69
		2:3	219	.77	.71–.82
	B	1:2	543	.62	.55–.68
		1:3	447	.41	.33–.48
		2:3	353	.68	.63–.74
	Median			.62	.55–.68
First	A	1:2	268	.72	.66–.77
		1:3	256	.67	.59–.73
		2:3	308	.74	.68–.78
	B	1:2	409	.76	.72–.80
		1:3	503	.62	.56–.67
		2:3	408	.73	.67–.77
	Median			.72	.66–.77
Second	A	1:2	126	.83	.76–.87
		1:3	112	.81	.74–.87
		2:3	194	.83	.78–.87
	B	1:2	422	.75	.71–.79
		1:3	383	.77	.73–.81
		2:3	463	.86	.83–.88
	Median			.81	.74–.87
Third	A	2:3	218	.78	.72–.83
	B	1:2	401	.83	.79–.86
		1:3	367	.78	.74–.82
		2:3	468	.86	.84–.88
	Median			.78	.72–.83
Median				.72	.66–.77

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.12 Delayed Alternate Form Reliability for DIBELS 8th Edition Word Reading Fluency

Grade	Sample	Forms	N	r	CI
Kindergarten	A	1:2	112	.88	.83–.92
		1:3	86	.85	.77–.90
		2:3	150	.90	.86–.92
	B	1:2	359	.81	.77–.84
		1:3	467	.66	.61–.71
		2:3	357	.89	.87–.91
	Median			.85	.77–.90
First	A	1:2	153	.91	.88–.93
		1:3	141	.82	.75–.86
		2:3	241	.88	.85–.91
	B	1:2	410	.92	.91–.93
		1:3	507	.80	.76–.83
		2:3	410	.87	.84–.89
	Median			.87	.84–.89
Second	A	1:2	163	.91	.88–.94
	B	1:2	424	.90	.88–.92
		1:3	384	.91	.89–.92
		2:3	465	.92	.90–.93
	Median			.91	.88–.94
Third	A	1:2	198	.90	.88–.93
		1:3	138	.90	.86–.93
		2:3	216	.90	.87–.92
	B	1:2	412	.89	.87–.91
		1:3	376	.85	.82–.88
		2:3	469	.90	.88–.91
	Median			.90	.86–.93
Median				.87	.84–.89

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.13 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Sample	Forms	N	r	CI
First	A	1:2	161	.94	.91–.95
		1:3	117	.87	.82–.91
		2:3	218	.92	.89–.94
	B	1:2	399	.94	.93–.95
		1:3	496	.84	.81–.86
		2:3	406	.90	.88–.92
	Median			.90	.88–.92
Second	A	1:2	225	.89	.86–.92
		1:3	116	.85	.79–.89
		2:3	116	.92	.89–.95
	B	1:2	418	.92	.90–.93
		1:3	379	.86	.84–.89
		2:3	464	.93	.91–.94
	Median			.89	.86–.92
Third	A	1:2	112	.91	.87–.94
		2:3	171	.84	.79–.88
	B	1:2	410	.88	.86–.91
		1:3	376	.85	.82–.88
		2:3	467	.88	.86–.90
	Median			.85	.82–.88
Fourth	A	1:2	142	.82	.76–.87
	B	1:2	391	.90	.88–.92
		1:3	369	.89	.87–.91
		2:3	460	.89	.87–.91
	Median			.89	.87–.91
Fifth	A	1:2	83	.89	.83–.92
		1:3	128	.86	.81–.90
	B	1:2	378	.81	.77–.84
		1:3	343	.82	.78–.85
		2:3	452	.90	.88–.91
	Median			.82	.78–.85

Table 2.13 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Sample	Forms	N	r	CI
Sixth	A	1:2	66	.86	.78–.91
		1:3	61	.91	.85–.94
		2:3	98	.91	.87–.94
	B	1:2	173	.89	.86–.92
		1:3	172	.89	.85–.92
		2:3	283	.89	.86–.91
	Median			.89	.85–.92
Seventh	A	1:2	79	.89	.83–.93
		1:3	65	.84	.74–.90
		2:3	127	.86	.81–.90
	B	1:2	75	.93	.89–.95
		1:3	71	.84	.81–.86
		2:3	190	.90	.88–.92
	Median			.86	.81–.90
Eighth	A	1:2	67	.92	.89–.95
		1:3	67	.73	.62–.81
		2:3	100	.74	.64–.82
	B	1:2	68	.94	.90–.96
		1:3	67	.94	.90–.96
		2:3	135	.88	.83–.91
	Median			.88	.83–.91
Median				.88	.83–.91

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.14 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
First	A	1:2	161	.89	.85–.92
		1:3	117	.71	.61–.79
		2:3	218	.78	.72–.82
	B	1:2	399	.84	.81–.87
		1:3	496	.65	.60–.70
		2:3	406	.78	.73–.81
	Median			.78	.72–.82
Second	A	1:2	225	.78	.73–.83
		1:3	116	.61	.48–.71
		2:3	116	.80	.72–.86
	B	1:2	418	.84	.81–.87
		1:3	379	.81	.78–.85
		2:3	464	.84	.81–.87
	Median			.80	.72–.86
Third	A	1:2	112	.75	.66–.82
		2:3	171	.65	.55–.73
	B	1:2	410	.79	.72–.82
		1:3	376	.66	.60–.72
		2:3	467	.78	.74–.81
	Median			.66	.60–.72
Fourth	A	1:2	142	.82	.76–.87
	B	1:2	391	.84	.80–.86
		1:3	369	.84	.81–.87
		2:3	460	.90	.88–.91
	Median			.84	.80–.86
Fifth	A	1:2	83	.89	.83–.92
		1:3	128	.86	.81–.90
	B	1:2	378	.81	.77–.84
		1:3	343	.62	.55–.69
		2:3	452	.71	.67–.76
	Median			.71	.67–.76

Table 2.14 Delayed Alternate Form Reliability for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Sample	Forms	N	r	CI
Sixth	A	1:2	66	.89	.83–.93
		1:3	61	.77	.64–.86
		2:3	98	.85	.78–.89
	B	1:2	173	.89	.86–.92
		1:3	172	.87	.83–.90
		2:3	283	.92	.91–.94
	Median			.87	.83–.90
Seventh	A	1:2	79	.90	.85–.93
		1:3	65	.96	.94–.98
		2:3	127	.83	.77–.88
	B	1:2	74	.67	.52–.78
		1:3	70	.49	.29–.65
		2:3	189	.49	.37–.59
	Median			.67	.52–.78
Eighth	A	1:2	67	.73	.59–.82
		1:3	67	.67	.51–.78
		2:3	100	.75	.65–.83
	B	1:2	68	.79	.68–.86
		1:3	67	.77	.64–.85
		2:3	135	.62	.60–.71
	Median			.73	.59–.82
Median				.73	.59–.82

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study.

Table 2.15 Delayed Alternate Form Reliability for DIBELS 8th Edition Maze

Grade	Forms	N	r	CI
Second	1:2	144	.69	.59–.77
	1:3	101	.73	.62–.82
	2:3	129	.68	.58–.77
	Median		.69	.59–.77
Third	1:2	219	.71	.64–.77
	1:3	153	.76	.69–.82
	2:3	163	.81	.75–.86
	Median		.76	.69–.82

Table 2.15 Delayed Alternate Form Reliability for DIBELS 8th Edition Maze

Grade	Forms	N	r	CI
Fourth	1:2	220	.77	.72–.82
	1:3	150	.72	.64–.79
	2:3	156	.82	.77–.87
	Median		.77	.72–.82
Fifth	1:2	184	.67	.58–.74
	1:3	146	.72	.54–.79
	2:3	152	.78	.71–.84
	Median		.72	.54–.79
Sixth	1:2	157	.78	.71–.83
	1:3	155	.75	.68–.82
	2:3	160	.85	.80–89
	Median		.78	.71–.83
Seventh	1:2	60	.86	.77–.91
	1:3	60	.85	.76–.91
	2:3	64	.90	.84–.94
	Median		.86	.77–.91
Eighth	1:2	57	.85	.75–.91
	1:3	57	.84	.75–.90
	2:3	64	.85	.77–.91
	Median		.85	.75–.91
Median			.77	.72–.82

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Table 2.16 Delayed Alternate Form Reliability for DIBELS 8th Edition Composite

Grade	Forms	N	r	CI
Kindergarten	1:2	330	.80	.76–.84
	1:3	430	.70	.65–.75
	2:3	350	.86	.83–.88
	Median		.80	.76–.84
First	1:2	396	.94	.92–.94
	1:3	489	.84	.81–.86
	2:3	401	.90	.88–.91
	Median		.90	.88–.91

Table 2.16 Delayed Alternate Form Reliability for DIBELS 8th Edition Composite

Grade	Forms	N	r	CI
Second	1:2	299	.89	.87–.91
	1:3	313	.88	.85–.90
	2:3	289	.93	.91–.94
	Median		.89	.87–.91
Third	1:2	354	.90	.88–.92
	1:3	345	.86	.83–.89
	2:3	442	.91	.90–.93
	Median		.90	.88–.92
Fourth	1:2	354	.90	.88–.92
	1:3	340	.89	.87–.91
	2:3	438	.90	.88–.92
	Median		.90	.88–.92
Fifth	1:2	321	.79	.74–.82
	1:3	269	.85	.81–.88
	2:3	359	.90	.87–.91
	Median		.85	.81–.88
Sixth	1:2	168	.89	.86–.92
	1:3	161	.90	.87–.93
	2:3	220	.88	.85–.91
	Median		.89	.86–.92
Seventh	1:2	68	.90	.85–.94
	1:3	67	.74	.60–.83
	2:3	126	.88	.83–.91
	Median		.88	.83–.91
Eighth	1:2	65	.94	.90–.96
	1:3	64	.94	.90–.96
	2:3	73	.92	.88–.95
	Median		.94	.90–.96
Median			.89	.87–.91

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Test-Retest Reliability

Test-retest reliability describes the correlation between scores on the same test administered at different points in time to the same test-takers. Test-retest reliability was evaluated by administering the same form in two different benchmark periods; thus, the delay

between test and retest was about three months. As a result, these results were expected to be weaker than concurrent alternate form reliability and more similar to delayed alternate form reliability because three months of instruction intervened between administrations of the same subtest form.

There are no universally accepted standards for judging the acceptability of test-retest reliability coefficients. The ideal degree of test-retest reliability depends on the purpose of the test, the construct it assesses, and the time between test administrations. In the case of DIBELS 8th Edition, we would emphasize that very high levels of reliability, especially for component skills like letter naming and phonemic segmentation, are undesirable because these skills develop quite rapidly in the grades in which they are assessed (Paris, 2005), especially over three months. Even when the time between test administrations is smaller, one should not expect levels of test-retest reliability to be as high as concurrent alternate form reliability when the skill measured develops rapidly and time between measurement occasions is sufficient for genuine growth to have occurred.

We studied test-retest reliability for all DIBELS 8th Edition subtests except PSF. For LNF, median test-retest reliability was .82 in kindergarten, .75 in first grade, and .77 overall (see Table 2.17).

For the NWF-CLS score, median test-retest reliability ranged from .75 in second grade to .81 in first grade and was .77 across grades; no median is reported for third grade because only one coefficient (.87) was available (see Table 2.18). For the NWF-WRC score, median test-retest reliability was .72 in kindergarten and second grade and .78 in first grade, with an overall median of .72 overall; again, no median is reported for third grade because only one coefficient (.84) was available (see Table 2.19).

For WRF, median test-retest reliability was strong at .92 in kindergarten and .90 in first grade, no median is reported in second and third grade because only one coefficient was available, but these were high at .95 and .88 respectively (see Table 2.20). Overall median reliability of WRF was .92.

For ORF, median test-retest reliability was also strong at .91 across grades with grade-level medians ranging from .86 in seventh grade to .94 in third grade (see Table 2.21). For ORF-ACC, median test-retest reliability was adequate at .75 across grades; median test-retest reliability by grade ranged from .74 in first and sixth grade to .90 in seventh grade (see Table 2.22). Maze test-retest reliability was more variable across grades ranging from .54 in fifth grade to .88 in seventh grade, but the median across grades was good at .82 (see Table 2.23).

Table 2.17 Test–Retest Reliability Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	122	.82	.76−.87
	1:3	123	.77	.69−.83
	2:3	121	.84	.78−.89
	Median		.82	.76−.87
First	1:2	124	.82	.75−.87
	1:3	123	.67	.56−.76
	2:3	128	.75	.66−.81
	Median		.75	.66−.81
Median			.77	.69−.83

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 2.18 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	92	.84	.77 - .89
	1:3	89	.79	.70 - .86
	2:3	107	.77	.67 - .83
	Median		.79	.70 - .86
First	1:2	120	.81	.74 - .87
	1:3	119	.75	.66 - .82
	2:3	126	.84	.78 - .84
	Median		.81	.74 - .87
Second	1:2	116	.75	.66 - .82
	1:3	115	.75	.65 - .82
	2:3	165	.88	.83 - .91
	Median		.75	.66 - .82
Third	2:3	158	.87	.82 - .90
Median			.79	.70 - .86

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 2.19 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	89	.79	.69–.86
	1:3	85	.72	.59–.81
	2:3	106	.69	.58–.78
	Median		.72	.59–.81
First	1:2	120	.78	.70–.84
	1:3	119	.68	.57–.77
	2:3	126	.80	.73–.85
	Median		.78	.70–.84
Second	1:2	116	.72	.62–.80
	1:3	115	.72	.62–.80
	2:3	165	.88	.84–.91
	Median		.72	.62–.80
Third	2:3	158	.84	.78–.88
Median			.72	.62–.80

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 2.20 Test-Retest Reliability Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Benchmark period	N	r	CI
Kindergarten	1:2	120	.92	.88–.94
	1:3	120	.88	.84–.92
	2:3	120	.93	.91–.95
	Median		.92	.88–.94
First	1:2	126	.90	.86–.93
	1:3	122	.82	.75–.87
	2:3	128	.92	.89–.94
	Median		.92	.89–.94
Second	2:3	82	.95	.93–.97
Third	2:3	90	.88	.82–.92
Median			.92	.88–.94

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 2.21 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Benchmark period	N	r	CI
First	1:2	156	.92	.89–.94
	1:3	123	.88	.83–.91
	2:3	164	.94	.92–.95
	Median		.92	.89–.94
Second	1:2	150	.87	.83–.91
	1:3	116	.85	.79–.90
	2:3	148	.93	.90–.95
	Median		.87	.83–.91
Third	1:2	159	.94	.91–.95
	1:3	110	.92	.89–.95
	2:3	156	.94	.91–.95
	Median		.94	.91–.95
Fourth	1:2	274	.91	.89–.93
	1:3	259	.90	.87–.92
	2:3	316	.88	.86–.90
	Median		.90	.87–.92
Fifth	1:2	229	.91	.89–.93
	1:3	221	.87	.83–.90
	2:3	298	.91	.89–.93
	Median		.91	.89–.93
Sixth	1:2	169	.91	.89–.94
	1:3	158	.91	.87–.93
	2:3	219	.93	.91–.95
	Median		.91	.89–.94
Seventh	1:2	79	.90	.84–.93
	1:3	65	.86	.78–.91
	2:3	121	.86	.81–.90
	Median		.86	.81–.90
Eighth	1:2	67	.91	.86–.94
	1:3	67	.92	.87–.95
	2:3	96	.93	.90–.95
	Median		.92	.87–.95
Median			.91	.89–.93

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 2.22 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Benchmark period	N	r	CI
First	1:2	156	.80	.74–.85
	1:3	123	.49	.34–.61
	2:3	164	.74	.67–.81
	Median		.74	.67–.81
Second	1:2	150	.75	.67–.81
	1:3	116	.75	.67–.81
	2:3	148	.83	.78–.88
	Median		.75	.67–.81
Third	1:2	159	.86	.81–.90
	1:3	110	.74	.64–.82
	2:3	156	.80	.73–.85
	Median		.80	.73–.85
Fourth	1:2	274	.83	.79–.87
	1:3	259	.75	.70–.80
	2:3	316	.75	.70–.79
	Median		.75	.70–.80
Fifth	1:2	229	.79	.73–.83
	1:3	221	.79	.73–.83
	2:3	298	.83	.80–.87
	Median		.79	.73–.83
Sixth	1:2	169	.76	.69–.82
	1:3	158	.60	.49–.69
	2:3	219	.74	.67–.79
	Median		.74	.67–.79
Seventh	1:2	79	.90	.84–.93
	1:3	65	.95	.93–.97
	2:3	121	.84	.78–.89
	Median		.90	.84–.93
Eighth	1:2	67	.84	.75–.90
	1:3	67	.79	.68–.86
	2:3	96	.83	.75–.88
	Median		.83	.75–.88
Median			.75	.70–.81

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 2.23 Test-Retest Reliability Coefficients for DIBELS 8th Edition Maze

Grade	Benchmark period	N	r	CI
Second	1:2	148	.90	.87–.93
	1:3	144	.48	.35–.60
	2:3	149	.57	.46–.67
	Median		.57	.46–.67
Third	1:2	127	.77	.69–.83
	1:3	128	.75	.66–.82
	2:3	131	.90	.86–.93
	Median		.77	.69–.83
Fourth	1:2	140	.84	.78–.88
	1:3	132	.73	.64–.80
	2:3	137	.82	.75–.87
	Median		.82	.75–.87
Fifth	1:2	102	.54	.38–.66
	1:3	105	.49	.33–.62
	2:3	98	.92	.88–.94
	Median		.54	.38–.66
Sixth	1:2	226	.83	.78–.87
	1:3	64	.85	.76–.90
	2:3	63	.85	.76–.91
	Median		.85	.76–.90
Seventh	1:2	126	.88	.84–.92
	1:3	63	.89	.82–.93
	2:3	115	.74	.65–.82
	Median		.88	.84–.92
Eighth	1:2	64	.83	.73–.89
	2:3	23	.88	.74–.95
	Median		.83	.73–.89
Median			.82	.75–.87

Note. The same form was administered in the benchmark periods indicated. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Intercept and Slope Reliability for Progress Monitoring

We also estimated model-based reliability estimates for the DIBELS 8 subtests recommended for use in frequent progress monitoring: PSF, NWF, WRF, and ORF. To calculate model-based reliability estimates for progress monitoring performance levels and

slopes, we created hierarchical linear models of student performance on each progress monitoring measure using data from Sample B (i.e., the 2018-2019 DDS study). In this study, students were progress-monitored up to 23 times over the course of the academic year with a minimum of 20 weeks elapsed. Test administrations typically took place every 2 weeks in kindergarten and first grade and every 3 weeks in all other grades. To be included in the analysis, students needed to have complete fall benchmark data on the word-reading measures and a minimum of three observed scores over the academic year. For the estimates of slope reliability, an additional requirement was that student needed to show evidence of need of intensive intervention. To obtain subsamples of students requiring intensive intervention, the highest performing students on each measure were removed until the sample mean for each measure fell below the 25th percentile according to national norms. Progress monitoring scores were used to create growth models represented by the general equations:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1jt_{ij}} + R_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{0j}$$

$$\beta_{0j} = \gamma_{10} + \gamma_{1j}$$

with

$$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N \begin{pmatrix} 0 & \tau^2_{00} & \tau_{01} \\ 0 & \tau_{01} & \tau^2_{00} \end{pmatrix}$$

and

$$R_{ij} \sim N\,(0,\,\sigma^2\,)$$

where Y equals a given DIBELS measure, β equals the number of days from the benchmark assessment, and γ represents across-group parameters. The group in these models is the student. Growth models were analyzed in R using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) with the maximum likelihood estimator and the Nelder Mead optimizer. For all Kindergarten measures and Grade 1 ORF and NWF-WRC, low variance in the effect of time led to singular fit warnings. These models were therefore rerun with blme package (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013), which uses an approximate Bayesian maximum a posteriori estimate to avoid singularity. Differences in fit indices and estimates between models were negligible. Results from the blme model are reported for these models.

The reliability of parameters was calculated using Raudenbush and Bryk's (2002) formulas for estimating parameter reliability, which are as follows:

Level 1

$$reliability \; \hat{\beta}_{qj} = \tau_{qq} / (\tau_{qq} + \upsilon_{qqj})$$

For each q = 0,....,Q.

Level 2

$$reliability \; \hat{\beta}_q = \frac{1}{J} \sum_{j=1}^{J} \tau_{qq} / (\tau_{qq} + \upsilon_{qqj})$$

For each q = 0,....,Q.

Intercept reliability estimates, which represent the reliability of the fall benchmark assessment within a growth model, were generally at or above .80, indicating high reliability performance level for DIBELS 8 (see Tables 2.24-2.28). The only exceptions to this occurred in Kindergarten, where reliability fell at .76 for PSF and .09 for NWF-WRC. The latter is unacceptably low, but importantly NWF-WRC (and also WRF) is not recommended as a progress-monitoring tool until later in Kindergarten. Both NWF-WRC and WRF have strong floor effects at the beginning of kindergarten (i.e., more scores of zero), which is one reason DIBELS 8 uses discontinue benchmark rules to limit the number of kindergarteners who takes these subtests at the beginning of the year. In other words, NWF-WRC (and WRF) would infrequently be used in practice to benchmark and progress monitor kindergarten students from the beginning of the year, making the model-based results for these scores less meaningful than results for PSF and NWF-CLS. It is also noteworthy that intercept reliability is consistently over .80 in all other grades for all subtests and that ORF reliability is over .90 consistently. These results indicate that DIBELS 8 is a highly reliable status indicator for progress monitoring.

Slope reliability estimates, which represent the reliability of change over time within a growth model, were also generally strong, falling at or above .33 for all scores and grades and far exceeding that threshold in most cases (see Tables 2.29-2.33). The lowest slope reliability estimates were found in third and eighth grade for ORF (see Table 2.33). Reliability estimates ranged as high as .90 and exceeded .70 for many grades and scores (e.g., kindergarten PSF, first grade WRF, seventh grade ORF). NWF-CLS, NWF-WRC, and WRF reliability was especially strong (see Tables 2.30-2.32). Based on these estimates, the suite of DIBELS 8th Edition

progress monitoring measures have acceptable to exceptional reliability for tracking change in student scores over time.

Table 2.24 Mean Reliability Coefficients for the Intercept of DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	N	Total Observations	Average Observations	Maximum Observations	Rel.	CI
Kindergarten	386	2,084	7.18	12	.76	.61−.91
First	430	2,780	9.17	16	.87	.76−.98

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Table 2.25 Mean Reliability Coefficients for the Intercept of DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	N	Total Observations	Average Observations	Maximum Observations	Rel.	CI
Kindergarten	393	2,475	8.33	15	.82	.71−.93
First	330	1,980	7.68	23	.92	.85−.99
Second	432	3,112	8.44	14	.93	.90−.96
Third	408	2,893	8.16	14	.93	.91−.96

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Table 2.26 Mean Reliability Coefficients for the Intercept of DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	N	Total Observations	Average Observations	Maximum Observations	Rel.	CI
Kindergarten	393	2,475	8.33	15	.09	.03−.16
First	330	1,980	7.68	23	.90	.81−.99
Second	432	3,112	8.44	14	.93	.90−.96
Third	408	2,893	8.16	14	.94	.91−.96

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Table 2.27 Mean Reliability Coefficients for the Intercept of DIBELS 8th Edition Word Reading Fluency

Grade	N	Total Observations	Average Observations	Maximum Observations	Rel.	CI
Kindergarten	442	2,151	5.89	14	.85	.79–.92
First	430	2,873	8.67	19	.96	.93–.98
Second	432	3,105	8.40	14	.97	.96–.98
Third	408	2,898	8.17	14	.97	.95–.98

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Table 2.28 Mean Reliability Coefficients for the Intercept of DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	N	Total Observations	Average Observations	Maximum Observations	Rel.	CI
First	333	1,976	7.60	13	.95	.92–.97
Second	421	2,822	8.02	14	.96	.93–.99
Third	408	2,665	7.69	14	.94	.89–.98
Fourth	345	4,589	15.69	23	.96	.90–.99
Fifth	294	4,107	15.86	23	.96	.91–.99
Sixth	137	2,242	17.84	23	.97	.93–.99
Seventh	61	944	17.11	23	.98	.96–.99
Eighth	57	925	18.16	23	.97	.93–.99

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study).

Table 2.29 Mean Reliability Coefficients for the Slope of DIBELS 8th Edition Phonemic Segmentation Fluency

Grade	N	Total Observations	Maximum Observations	Average Observations	Rel.	CI
Kindergarten	182	1,052	12	7.70	.71	.50–.93
First	356	2,203	16	8.82	.60	.36–.83

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study). In Kindergarten, the progress monitoring period ranged from .69 to 8.44 months, with an average of 7.11. In First Grade, the progress monitoring period ranged from 1.38 to 8.51 months, with an average of 7.18.

Table 2.30 Median Reliability Coefficients for the Slope of DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	N	Total Observations	Maximum Observations	Average Observations	Rel.	CI
Kindergarten	206	1,351	15	8.75	.85	.55–.99
First	263	1,758	23	8.31	.78	.61–.95
Second	286	2,233	14	9.27	.76	.47–.99
Third	232	1,818	14	9.24	.64	.33–.95

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study). In Kindergarten, the progress monitoring period ranged from .69 to 8.44 months, with an average of 7.12. In First Grade, the progress monitoring period ranged from 1.38 to 8.51 months, with an average of 7.46. In Second Grade, the progress monitoring period ranged from 1.57 to 8.54 months, with an average of 6.90. In Third Grade, the progress monitoring period ranged from 1.15 to 8.54 months, with an average of 7.03.

Table 2.31 Mean Reliability Coefficients for the Slope of DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly

Grade	N	Total Observations	Maximum Observations	Average Observations	Rel.	CI
Kindergarten	363	2,333	15	8.54	.84	.53–.99
First	212	1,541	23	8.86	.82	.64–.99
Second	318	2,428	14	9.07	.71	.41–.99
Third	282	2,081	14	8.87	.64	.35–.92

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study). In Kindergarten, the progress monitoring period ranged from .69 to 8.48 months, with an average of 7.19. In First Grade, the progress monitoring period ranged from 1.38 to 8.51 months, with an average of 7.44. In Second Grade, the progress monitoring period ranged from 1.58 to 8.56 months, with an average of 6.86. In Third Grade, the progress monitoring period ranged from 1.15 to 8.54 months, with an average of 7.02.

Table 2.32 Mean Reliability Coefficients for the Slope of DIBELS 8th Edition Word Reading Fluency

Grade	N	Total Observations	Maximum Observations	Average Observations	Rel.	CI
Kindergarten	391	1,923	15	8.54	.88	.72–.99

Table 2.32 Mean Reliability Coefficients for the Slope of DIBELS 8th Edition Word Reading Fluency

Grade	N	Total Observations	Maximum Observations	Average Observations	Rel.	CI
First	315	2,321	19	9.43	.90	.76–.99
Second	271	2,160	14	9.46	.68	.37–.99
Third	288	2,189	14	8.89	.62	.32–.91

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study). In Kindergarten, the progress monitoring period ranged from .69 to 8.48 months, with an average of 7.18. In First Grade, the progress monitoring period ranged from 2.27 to 8.51 months, with an average of 6.87. In Second Grade, the progress monitoring period ranged from 1.58 to 8.56 months, with an average of 6.87. In Third Grade, the progress monitoring period ranged from 1.15 to 8.54 months, with an average of 7.06.

Table 2.33 Mean Reliability Coefficients for the Slope of DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	N	Total Observations	Maximum Observations	Average Observations	Rel.	CI
First	191	1,374	13	8.58	.87	.77–.97
Second	277	1,958	14	9.05	.73	.45–.99
Third	311	2,125	14	8.21	.44	.14–.73
Fourth	277	3,701	23	16.44	.60	.25–.96
Fifth	234	3,330	23	16.94	.59	.32–.85
Sixth	80	1,579	23	20.98	.60	.37–.83
Seventh	41	670	23	19.11	.85	.53–.99
Eighth	56	941	23	19.25	.33	.12–.54

Note. Rel. = reliability. The minimum number of observations was 3 for all grades. All coefficients come from Sample B (i.e., 2018-2019 CTL norming study). In First Grade, the progress monitoring period ranged from 3.65 to 8.51 months, with an average of 7.47. In Second Grade, the progress monitoring period ranged from 1.58 to 8.56 months, with an average of 7.02. In Third Grade, the progress monitoring period ranged from 1.15 to 8.54 months, with an average of 7.04. In Fourth Grade, the progress monitoring period ranged from 0 months to 8.51 months, with an average of 6.65. In Fifth Grade, the progress monitoring period ranged from 0 months to 8.58 months, with an average of 6.71. In Sixth Grade, the progress monitoring period ranged from 2.23 months to 8.54 months, with an average of 7.94. In Seventh Grade, the progress monitoring period ranged from 2.49 months to 8.58 months, with an average of 7.42. In Eighth Grade, the progress monitoring period ranged from 0 months to 8.48 months, with an average of 7.24.

Standard Error of Measurement

We also estimated the standard error of measurement (SEM) using a classical test theory approach, which multiplies the standard deviation for a measure by the square root of one minus the reliability of the measure. The SEM for each score DIBELS 8 produces in each grade and benchmark period is reported in Table 2.34. In all cases except for LNF and the composite, we used the median concurrent alternate form reliability for a grade drawn from both Samples A and B and the standard deviation (SD) for each benchmark period, again drawn from both Samples A and B, in these calculations. Because concurrent alternate form reliability was not available for LNF or the composite, we used delayed alternate form reliability in these two cases.

By definition, scores with the best reliability have the smallest SEMs, and SEMs are also affected by the variability and range of typical scores across students, as reflected in their SDs. That is, as variability in students' scores increases, SEMs will naturally become larger.

LNF has an SEM of 8-9 letters per minute in kindergarten and 9-11 in first grade. PSF has an SEM of 5-8 phonemes per minute in kindergarten and 7 in first grade. NWF-CLS, which has uniformly good reliability, demonstrates clearly the effect of increasing range of scores with its SEM increasing from 5-7 in kindergarten, to 10-14 in first grade, to 12-14 in second grade, and finally to 14-17 in third grade. Similarly, NWF-WRC SEMs increase from 1-3 in kindergarten, to 3-5 in first grade, and ultimately to 4-5 in the second and third grades. Showing the same developmental pattern, WRF SEMs are 1-3 in kindergarten, 3-5 in first grade, 5-6 in second grade, and 5-7 in third grade. ORF SEMs are 5-8 in first grade, 9-11 in second grade, 10-11 in third grade, 9-13 in fourth grade, 10-11 in fifth grade, 10-12 in sixth grade, 10-13 in seventh grade, and 9-10 in eighth grade. Expressed in percentage points, ORF-ACC SEMs demonstrate a trend of decreasing SEMs due to the increasing accuracy and decreasing variability of students' reading across grades. Thus, ORF-ACC SEMs are 5-10 in first grade, 4-7 in second grade, 2-5 in third grade, 3-4 in fourth grade, 1-4 in fifth grade, 3-5 in sixth grade, 2 in seventh grade, and 1-2 in eighth grade. Maze SEMs are relatively stable across grades; they are 3-4 in second grade, 4-5 in third, fourth, and sixth grade, 5-6 in fifth and seventh grade, and 5-7 in eighth grade. The composite score SEMs are also relatively stable across grades and are 17-25 in kindergarten, 9-14 in first grade, 10-13 in second grade, 11-13 in third grade, 10-14 in fourth grade, 14-15 in fifth grade, 12-13 in sixth grade, 11-14 in seventh grade, and 9-10 in eighth grade.

Table 2.34 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Subtest	Beginning	Middle	End
Kindergarten	LNF	8.07	9.30	8.39
	PSF	5.22	7.74	7.18
	NWF-CLS	4.76	6.50	7.37
	NWF-WRC	1.73	2.79	3.31
	WRF	1.23	2.14	2.81
	Composite	24.69	16.96	18.42
First	LNF	9.03	10.65	10.32
	PSF	6.71	7.24	7.52
	NWF-CLS	9.58	10.60	13.62
	NWF-WRC	3.42	3.96	4.91
	WRF	2.94	4.06	4.95
	ORF	5.31	6.36	8.11
	ORF-ACC	9.53	8.72	5.36
	Composite	8.78	10.49	13.60
Second	NWF-CLS	11.68	13.10	13.81
	NWF-WRC	3.97	4.62	4.62
	WRF	5.12	5.58	6.16
	ORF	8.54	9.92	10.59
	ORF-ACC	6.54	5.51	4.01
	Maze	3.56	3.43	4.22
	Composite	10.18	12.45	13.30
Third	NWF-CLS	14.26	17.30	17.49
	NWF-WRC	3.88	4.74	4.77
	WRF	5.14	5.52	7.40
	ORF	10.46	10.67	10.28
	ORF-ACC	4.64	3.72	2.20
	Maze	3.56	4.71	4.54
	Composite	10.85	12.81	12.70
Fourth	ORF	9.33	12.74	9.50
	ORF-ACC	4.08	2.93	3.03
	Maze	4.27	4.24	5.36
	Composite	10.10	13.75	10.50
Fifth	ORF	11.23	10.42	11.09
	ORF-ACC	3.51	3.43	1.45
	Maze	5.57	4.58	5.22
	Composite	14.63	13.52	15.10

Table 2.34 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Subtest	Beginning	Middle	End
Sixth	ORF	10.84	10.23	12.34
	ORF-ACC	4.56	3.02	3.76
	Maze	3.69	4.53	3.70
	Composite	12.78	12.47	13.30
Seventh	ORF	10.32	11.65	12.79
	ORF-ACC	1.54	1.87	1.66
	Maze	5.30	6.30	5.87
	Composite	10.67	12.53	14.30
Eighth	ORF	9.37	9.55	10.23
	ORF-ACC	2.08	1.93	1.33
	Maze	5.03	5.45	6.71
	Composite	8.94	9.99	10.20

Note. SEMs were calculated using median concurrent alternate form reliability for a grade and the standard deviation (SD) for each benchmark period. Medians and SDs were drawn from Samples A and B.

Inter-rater Reliability

Finally, we also estimated inter-rater reliability for all measures except LNF and the composite. Although DIBELS 8th Edition is a low inference measure, some human judgment is required in order to produce scores. For example, test administrators must decide whether student pronunciation of a word or sound is correct or incorrect. It is therefore useful to consider evidence of inter-rater reliability. To estimate inter-rater reliability, we calculated intra-class coefficients (ICCs) for a subsample of students from Sample 1. We calculated ICCs using the one-way random effects analysis in SPSS with consistency as the criterion. We used a one-way random effects analysis because students were rated by different sets of raters, and we used consistency as the criterion because scores on DIBELS can range from 0 to over 100, making absolute agreement unlikely and less consequential.

Table 2.35 Inter-rater Reliability of DIBELS 8th Edition Phoneme Segmentation Fluency

Grade	Measure	*N*	ICC	CI
Kindergarten	Average	28	.992	.982–.996
	Single	28	.984	.966–.992
First	Average	8	.985	.933–.997
	Single	8	.971	.874–.994

Table 2.36 Inter-rater Reliability of DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Measure	N	ICC	CI
Kindergarten	Average	16	.998	.994–.999
	Single	16	.996	.989–.999
First	Average	5	.990	.927–.999
	Single	5	.980	.864–.998
Second	Average	6	.999	.997–.999
	Single	6	.999	.995–.999
Third	Average	15	.987	.961–.995
	Single	15	.973	.925-.991

Table 2.37 Inter-rater Reliability of DIBELS 8th Edition Nonsense Word Fluency-Words Recoded

Grade	Measure	N	ICC	CI
Kindergarten	Average	16	.977	.936–.992
	Single	16	.955	.880–.984
First	Average	5	.788	-.565–.977
	Single	5	.650	-.220–.956
Second	Average	6	.939	.633–.991
	Single	6	.884	.463–.983
Third	Average	15	.953	.864–.984
	Single	15	.910	.761–.969

Table 2.38 Inter-rater Reliability of DIBELS 8th Oral Reading Fluency

Grade	Measure	N	ICC	CI
First	Average	60	.997	.995–.998
	Single	60	.994	.990–.996
Second	Average	18	.996	.989–.998
	Single	18	.992	.989–.998
Third	Average	45	.998	.996–.999
	Single	45	.996	.992–.998

Table 2.39 Inter-rater Reliability of DIBELS 8th Edition Word Reading Fluency

Grade	Measure	N	ICC	CI
Kindergarten	Average	17	.996	.990–.999
	Single	17	.992	.979–.997
First	Average	15	.999	.998–.999
	Single	15	.999	.997–.999
Second	Average	6	.999	.998–.999
	Single	6	.999	.992–.999
Third	Average	15	.998	.995–.999
	Single	15	.996	.989–.999

Summary

Taken together, the reliability evidence for DIBELS 8 is strong. The strongest evidence regards concurrent alternate form reliability and SEMs. Research into the reliability of DIBELS 8 scores is ongoing, and regular addendums to this manual will continue to build the validity argument for DIBELS 8.

Chapter 3: Validity of DIBELS 8th Edition

Validity is an argument that hinges on the desired inferences to be made about an individual (Messick, 1995). As such, assuming an adequate reliability of scores, different forms of validity can serve as evidence for different claims. In the sections below, we present and discuss concurrent and predictive validity evidence for the use of DIBELS 8 measures for screening and progress monitoring purposes.

Concurrent validity is generally seen as a means of validating that the intended construct is being captured by a measure. Concurrent validity for DIBELS 8 was evaluated by correlating each DIBELS 8 subtest with the corresponding subtest from the previous version of DIBELS (i.e., DIBELS Next), with related but external criterion measures, and with the other DIBELS 8 subtests. Correlations are only reported when the study sample size was 50 or greater to ensure a minimum threshold of precision in the correlation estimate.

Predictive validity can also be seen as a means of validating that the intended construct has been captured, but in addition, it serves as a means of validating the use of a measure for predicting performance at a later period (e.g., often the end of a grade). Predictive validity traditionally includes correlations, but when intended uses of measure include identification of subgroups of students, then an evaluation of screening accuracy provides critical evidence that a measure is functioning as intended (Jenkins, Hudson, & Johnson, 2007). In addition, when progress monitoring is an intended use, then evidence that change on a measure, which is typically expressed as a slope, predicts a criterion measure. All three are evaluated for DIBELS 8.

Concurrent Validity

The correlation between two measures of the same construct should be higher than the correlation between two measures of distinct, but related constructs. Thus, when correlating like measures (e.g., DIBELS Next and DIBELS 8 NWF), correlations should be quite strong. However, when correlating component skills like PSF and LNF with reading achievement scores, correlations ought to be lower. As a result, in the sections that follow, relationships are expected to be strongest between the same subtest for the previous and current editions of DIBELS and weakest for subtests like PSF and LNF with reading achievement measures.

Correlations with DIBELS Next. The equivalency of the previous and current editions of DIBELS was evaluated by correlating like subtests for students who took both DIBELS 8 and DIBELS Next in a given benchmark period. Correlations are reported for the same subtest when it would traditionally be administered on DIBELS Next only.

As shown in Table 3.1, DIBELS Next and DIBELS 8th Edition LNF are strongly related. Similarly strong are the relations between DIBELS Next and DIBELS 8th Edition PSF, although these relations are somewhat more variable, which may be due to the different approaches to item ordering in the two editions (see Table 3.2). Of particular note are the correlations between NWF scores on the two editions of DIBELS. Despite substantial changes in the items used and in form construction, correlations for NWF-CLS are quite strong at .73 or above (see Table 3.3). More interesting is that despite the additional difference in scoring for blending on the two editions (i.e., words recoded correctly in DIBELS 8th Edition and whole words read in DIBELS Next), the correlations are again very strong, ranging from .66 to .90 (see Table 3.4). These results suggest that DIBELS Next WWR and DIBELS 8th Edition WRC scores rank students similarly, despite differences in scoring method and form composition. Finally, both ORF and ORF-ACC scores in DIBELS 8th Edition also demonstrate strong correlations with the corresponding score in DIBELS Next, but ORF is particularly good, with nearly every correlation being .90 or above (see Tables 3.5 and 3.6 respectively). Taken together, this evidence suggests the equivalence of DIBELS Next and DIBELS 8th Edition.

Table 3.1 Correlations between DIBELS 8th Edition and DIBELS Next Letter Naming Fluency

Grade	Period	r	N	CI
Kindergarten	1	.75	153	.67–.81
	2	.87	181	.84–.90
	3	.88	128	.83–.91
First	1	.72	157	.63–.79

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 3.2 Correlations between DIBELS 8th Edition and DIBELS Next Phoneme Segmentation Fluency

Grade	Period	r	N	CI
Kindergarten	2	.83	28	.65–.92
	3	.75	125	.66–.82
First	1	.49	154	.35–.60
	2	.84	16	.60–.94
	3	.96	16	.88–.99

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 3.3 Correlations between DIBELS 8th Edition and DIBELS Next Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	r	N	CI
Kindergarten	2	.84	163	.78–.88
	3	.87	139	.82–.91
First	1	.73	184	.66–.79
	2	.81	223	.76–.85
	3	.87	211	.84–.90
Second	1	.83	129	.77–.88

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 3.4 Correlations between DIBELS 8th Edition and DIBELS Next Nonsense Word Fluency-Words Recoded Correctly

Grade	Period	r	N	CI
Kindergarten	2	.88	130	.84–.92
	3	.84	108	.78–.89
First	1	.66	163	.56–.74
	2	.78	202	.72–.83
	3	.90	190	.86–.92
Second	1	.82	118	.75–.87

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 3.5 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Period	r	N	CI
First	2	.93	227	.91–.95
	3	.94	188	.92–.96
Second	1	.83	173	.77–.87
	2	.95	192	.93–.96
	3	.97	25	.92–.98
Third	1	.92	100	.89–.95
	2	.90	209	.87–.92
	3	.90	176	.87–.93
Fourth	1	.90	35	.82–.95
	2	.90	37	.81–.95
	3	.92	58	.87–.95
Fifth	1	.95	51	.91–.97
	2	.93	84	.89–.95
	3	.98	25	.94–.99
Sixth	2	.93	40	.88–.96

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Table 3.6 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	r	N	CI
First	2	.89	196	.86–.89
	3	.84	163	.78–.88
Second	1	.79	131	.71–.84
	2	.90	171	.87–.92
	3	.98	24	.95–.99
Third	1	.89	73	.83–.93
	2	.76	186	.70–.82
	3	.69	154	.60–.77
Fourth	1	.85	29	.70–.93
	2	.89	26	.77–.95
	3	.88	46	.78–.93
Fifth	1	.90	41	.81–.94
	2	.99	23	.97–.99

Table 3.6 Concurrent Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	r	N	CI
Sixth	2	.91	30	.82–.96

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year. All coefficients come from Sample A (i.e., 2017-2018 CTL norming study).

Correlations with external criterion measures. The concurrent validity of DIBELS 8 was also evaluated by correlating its subtests with external criterion measures given in the same benchmark period. These measures included DIBELS Next composite scores, the Comprehensive Test of Phonological Processing-2nd Edition (CTOPP-2) composite scores, and Iowa Assessment Total Reading and Word Analysis raw scores. All four study samples described in this manual contributed to these correlations: Sample A (2017-2018 CTL norming study), Sample B (2018-2019 CTL norming study), Sample C (2018-2019 CTL CTOPP subsample), and Sample D (2018-2019 Amplify study). Correlations for each DIBELS 8 subtest and the DIBELS 8 composite score with each available criterion measure in a given grade are reported in Tables 3.7-3.15.

Correlations among DIBELS 8 subtests with DIBELS Next composite scores are generally quite strong, suggesting that DIBELS 8th Edition is a similar measure compared to DIBELS Next. LNF correlations range from .49 to .89 depending on grade, time of year, and study, and are strongest in kindergarten (see Table 3.7). Not surprisingly, given the weak contributions of PSF to DIBELS Next composite scores in first grade, PSF correlations range from .47 to .62 in kindergarten but only from .14 to .27 in first grade (see Table 3.8). Also as expected, NWF-CLS correlations with the DIBELS Next composite are lowest at the beginning of kindergarten ($r = .36$), but are otherwise very strong, regardless of grade, time of year, and study (see Table 3.9). Interestingly, NWF-WRC correlations are uniformly strong (.48 or above) even in kindergarten, regardless of grade, time of year, and study (see Table 3.10). Another interesting result is the strong correlations between the new DIBELS 8 WRF subtest and DIBELS Next composite scores, with some correlations exceeding .90 (see Table 3.11). More predictably, correlations between DIBELS 8 ORF-WRC and DIBELS Next composite scores are .74 or above (see Table 3.12), although ORF-ACC correlations are more varied (see Table 3.13). DIBELS Next composite score correlate with DIBELS 8 Maze between .62 to .67 (see Table 3.14) and with DIBELS 8 composite scores between .70 to .87 (see Table 3.15). Given this pattern of relationships across measures, we conclude that DIBELS 8 is a highly similar measure to DIBELS Next and that the weaker relationships that were observed were predictable given intended differences between editions.

Similarly, correlations between DIBELS 8 subtests and the CTOPP-2 are generally as would be expected, given the nature of the measures. Of particular note are the moderate to strong correlations between CTOPP-2 rapid naming composite scores and LNF, depending on the grade and time-of-year DIBELS 8 was administered. As might be expected, relations are strongest with the symbolic naming composite, which incorporates a letter naming subtest (r = .31 and .60 for the beginning and end of kindergarten respectively, and r = .59 and .70 for the beginning and end of first grade respectively). Relations are weaker for non-symbolic composite scores, but still moderate to strong in magnitude, and as with symbolic naming, they get stronger from beginning to end of year and from grade to grade. These results suggest that LNF can operate as a decent screener for processing speed, especially from the end of kindergarten onward (see Table 3.7).

In contrast, correlations between PSF and the CTOPP-2 phonological awareness composite show a different pattern, depending on the time of year and age of the student assessed. While relations are relatively strong in kindergarten (r = .53 and .43 for the beginning and end of year respectively) and at the beginning of first grade (r = .51), they are weak by the end of first grade (see Table 3.8). In contrast to the pattern of increasing relationships seen with the naming speed composite scores for LNF, correlations of PSF with the phonological awareness composite get weaker from beginning to end of year and more dramatically so in first grade. These results again suggest the validity of DIBELS as a screener for phonological awareness difficulties but suggest that it only operates well in this manner from the beginning of kindergarten to the beginning of first grade.

Finally, concurrent relations with the Iowa Assessment scores vary predictably by Iowa score type and by DIBELS 8 subtest. As expected, given the discrete skills assessed by PSF, it showed the weakest relationships with the Iowa Total Reading and Word Analysis scores in both kindergarten and first grade. LNF was most weakly related to Iowa Total Reading in kindergarten, but otherwise was moderately strongly correlated with Iowa scores. The remaining DIBELS 8 subtests displayed moderate to very strong relations with the Iowa external criterion measures regardless of grade, although relationships were consistently stronger with the Iowa Total Reading score than with the Word Analysis score. Iowa Total Reading correlations ranged from a low of .50 to a high of .83 depending on subtest, grade, and time of year. The sole exceptions to these strong correlations was the beginning of fifth grade when ORF-WRC and the composite score correlated at .41 and .42 with the Iowa Total Reading score. However, these exceptions do not represent a trend of decreasing relations between DIBELS 8 and Iowa Total Reading, because by the end of the year, the correlations are quite strong again and remain strong in sixth and seventh grade (see Tables 3.12 and 3.15). In fact, the strength of relations between DIBELS 8 and Iowa Total Reading in the upper elementary and middle grades is generally quite strong. For example, DIBELS Maze

correlations range from .52 to .78 in Grades 4 to 8, depending on subtest, grade, and time of year, with seven out of ten correlations above .70. For the DIBELS 8 composite, the same correlations range from .42 to .74. These results suggest that DIBELS 8 is a very good to excellent indicator of reading proficiency as measured by a widely used reading achievement test not only in kindergarten through third grade, but also in fourth through eighth grade.

Table 3.7 Concurrent Validity Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	128	.70	.60–.78
			2	156	.80	.74–.85
			3	98	.89	.84–.93
		D	3	321	.74	.68–.78
	Iowa Total Reading	B	3	135	.51	.38–.63
	Iowa Word Analysis	B	3	136	.36	.20–.50
	CTOPP-2 Non-symbolic composite	C	1	153	.27	.12–.42
			3	179	.39	.26–.51
	CTOPP-2 Symbolic composite	C	1	63	.31	.06–.51
			3	177	.60	.50–.69
First	DIBELS Next composite	A	1	144	.65	.54–.74
			2	163	.70	.61–.77
			3	163	.63	.53–.72
		D	3	302	.49	.40-.57
	Iowa Total Reading	A	3	117	.54	.40–.66
		B	3	128	.40	.25–.54
	Iowa Word Analysis	A	3	119	.46	.31–.61
		B	3	103	.28	.09–.45
	CTOPP-2 Non-symbolic composite	C	1	136	.51	.37–.63
			3	62	.52	.30–.68
	CTOPP-2 Symbolic composite	C	1	164	.59	.48–.68
			3	190	.70	.62–.77

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.8 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Phoneme Segmentation Fluency

Grade	Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	75	.47	.27–.63
			3	95	.62	.48- .73
		D	3	321	.49	.41-.57
	Iowa Total Reading	B	3	132	.18	.01–.34
	Iowa Word Analysis	B	3	133	.18	.10–.34
	CTOPP-2 Phonological awareness composite	C	1	187	.53	.42–.63
			3	207	.43	.31–.54
First	DIBELS Next composite	A	1	141	.27	.11–.41
			2	161	.17	.01–.31
			3	163	.14	-.01–.29
		D	3	301	.14	.03–.25
	Iowa Total Reading	A	3	117	.12	-.06–.30
		B	3	128	.26	.09–.42
	Iowa Word Analysis	A	3	119	.14	-.05–.31
	CTOPP-2 Phonological awareness composite	C	1	148	.51	.38–.62
			3	68[a]	.20	-.04–.42
			3	124[a]	.12	-.06–.12

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composite is derived from measures of sound matching, blending, elision, and phoneme isolation depending on student age at time of testing. Students younger than 7 years old take the first three subtests, and students 7 years old and older take the last three subtests.

[a] First-grade phonological awareness composite has two samples because of CTOPP-2's separate age-based norms. The smaller subsample is compared to the younger norming group and the larger to the older based on each child's age at time of CTOPP-2 testing.

Table 3.9 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	92	.36	.26–.52
			2	133	.68	.57–.76
			3	109	.82	.75–.87
		D	3	321	.75	.70-.80
	Iowa Total Reading	A	3	113	.65	.53–.75
		B	3	129	.59	.47–.69
	Iowa Word Analysis	A	3	96	.43	.26–.58
		B	3	130	.27	.10–.42
First	DIBELS Next composite	A	1	166	.71	.62–.78
			2	185	.80	.74–.85
			3	186	.85	.81–.89
		D	3	302	.71	.65-.76
	Iowa Total Reading	A	3	198	.65	.56–.73
		B	3	124	.57	.44-.68
	Iowa Word Analysis	A	3	200	.54	.43–.63
		B	3	99	.34	.15-.50
Second	DIBELS Next composite	A	1	115	.80	.72–.85
			2	107	.62	.49–.72
			3	112	.74	.65–.82
		D	3	197	.44	.32–.55
	Iowa Total Reading	A	3	84	.62	.47–.74
		B	1	83	.68	.54–.78
			3	190	.67	.59–.75
	Iowa Word Analysis	A	3	51	.60	.39–.75
		B	3	158	.70	.61–.77
Third	DIBELS Next composite	A	2	109	.71	.61–.80
			3	99	.69	.57–.78
		D	3	216	.58	.49–.66
	Iowa Total Reading	A	3	90	.50	.33–.64
		B	1	93	.66	.53–.76
			3	150	.68	.58–.76
	Iowa Word Analysis	A	3	72	.34	.12–.53
		B	1	51	.48	.24–.69
			3	115	.65	.53–.75

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study.

Table 3.10 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly

Grade	Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	76	.60	.43–.73
			2	130	.66	.55–.75
			3	108	.74	.65–.82
		D	3	321	.74	.68–.78
	Iowa Total Reading	A	3	112	.65	.53–.74
		B	3	129	.61	.49–.71
	Iowa Word Analysis	A	3	95	.35	.16–.52
		B	3	130	.26	.09–.41
First	DIBELS Next composite	A	1	163	.61	.50–.70
			2	185	.79	.73–.84
			3	186	.86	.81–.89
		D	3	302	.66	.59–.72
	Iowa Total Reading	A	3	198	.63	.54–.71
		B	3	124	.52	.38–.64
	Iowa Word Analysis	A	3	200	.56	.45–.65
		B	3	99	.26	.06–.64
Second	DIBELS Next composite	A	1	115	.79	.71–.85
			2	107	.70	.58–.78
			3	112	.74	.64–.82
		D	3	197	.48	.36–.58
	Iowa Total Reading	A	3	84	.60	.45–.72
		B	1	83	.63	.47–.74
			3	190	.70	.62–.77
	Iowa Word Analysis	A	3	51	.62	.42–.76
		B	3	158	.73	.65–.79
Third	DIBELS Next composite	A	2	109	.74	.65–.82
			3	97	.73	.62–.81
		D	3	216	.59	.50–.67
	Iowa Total Reading	A	3	90	.51	.34–.65
		B	1	93	.71	.59–.80
			3	150	.69	.59–.76
	Iowa Word Analysis	A	3	72	.36	.14–.54
		B	1	51	.58	.36–.74
			3	115	.67	.56–.76

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study.

Table 3.11 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	89	.57	.41–.70
			2	124	.63	.52–.73
			3	92	.75	.64–.82
		D	3	321	.73	.68–.78
	Iowa Total Reading	B	3	128	.61	.49–.71
	Iowa Word Analysis	B	3	129	.26	.09–.41
First	DIBELS Next composite	A	1	144	.69	.60–.77
			2	163	.88	.85–.91
			3	163	.88	.85–.91
		D	3	302	.63	.51–.72
	Iowa Total Reading	A	3	117	.79	.71–.85
		B	3	124	.51	.35–.64
	Iowa Word Analysis	A	3	119	.67	.56–.76
		B	3	99	.84	.80–.87
Second	DIBELS Next composite	A	1	151	.91	.88–.94
			2	138	.87	.83–.91
		D	3	197	.77	.70–.82
	Iowa Total Reading	A	3	87	.62	.47–.74
		B	1	83	.83	.75–.89
			3	190	.77	.70–.83
	Iowa Word Analysis	A	3	89	.60	.44–.72
		B	3	158	.82	.76–.86
Third	DIBELS Next composite	A	1	138	.85	.79–.89
			2	154	.85	.80–.89
			3	97	.84	.77–.89
		D	3	216	.35	.23–.47
	Iowa Total Reading	A	3	90	.56	.40–.69
		B	1	93	.71	.59–.80
			3	150	.70	.61–.78
	Iowa Word Analysis	A	3	72	.32	.09–.51
		B	1	51	.52	.28–.69
			3	115	.68	.57–.77

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study.

Table 3.12 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Criterion	Sample	Period	N	r	CI
First	DIBELS Next composite	A	1	154	.75	.67−.81
			2	196	.91	.88−.93
			3	163	.91	.88−.94
		D	3	302	.88	.85−.90
	Iowa Total Reading	A	3	116	.82	.75−.87
		B	3	126	.71	.61−.79
	Iowa Word Analysis	A	3	118	.67	.55−.76
		B	3	102	.51	.35−.64
Second	DIBELS Next composite	A	1	131	.84	.78−.88
			2	130	.87	.82−.90
		D	3	197	.84	.79−.88
	Iowa Total Reading	A	3	87	.71	.59−.80
		B	1	83	.77	.66−.84
			3	190	.80	.74−.84
	Iowa Word Analysis	A	3	89	.60	.45−.72
		B	3	158	.78	.71−.84
Third	DIBELS Next composite	A	1	73	.89	.83−.93
			2	128	.83	.77−.88
			3	96	.83	.75−.88
		D	3	216	.74	.67−.80
	Iowa Total Reading	A	3	90	.58	.42−.70
		B	1	93	.72	.61−.81
			3	150	.73	.64−.80
	Iowa Word Analysis	A	3	72	.24	.01−.45
		B	1	51	.48	.23−.67
			3	115	.70	.60−.78
Fourth	Iowa Total Reading	A	3	91	.61	.47−.73
		B	1	96	.71	.60−.80
			3	165	.74	.66−.80
Fifth	Iowa Total Reading	A	3	59	.65	.48−.78
		B	1	109	.41	.24−.55
			3	148	.70	.60−.77
Sixth	Iowa Total Reading	A	3	82	.67	.52−.77
		B	3	152	.64	.53−.72
Seventh	Iowa Total Reading	A	3	91	.54	.38−.67
		B	3	150	.70	.61−.78

Table 3.12 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Criterion	Sample	Period	N	r	CI
Eighth	Iowa Total Reading	A	3	77	.59	.42–.72
		B	3	106	.60	.46–.71

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study.

Table 3.13 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Criterion	Sample	Period	N	r	CI
First	DIBELS Next composite	A	1	154	.76	.69–.82
			2	196	.78	.72–.83
			3	163	.76	.68–.82
		D	3	302	.82	.77–.85
	DIBELS Next NWF-CLS	A	1	166	.67	.58–.75
	Iowa Total Reading	A	3	116	.61	.48–.71
		B	3	126	.49	.34–.61
	Iowa Word Analysis	A	3	118	.60	.47–.71
		B	3	102	.33	.15–.50
Second	DIBELS Next composite	A	1	131	.63	.52–.73
			2	130	.68	.57–.76
		D	3	197	.65	.56–.73
	Iowa Total Reading	A	3	87	.48	.30–.62
		B	1	83	.76	.65–.84
			3	190	.58	.47–.66
	Iowa Word Analysis	A	3	87	.33	.13–.50
		B	3	158	.63	.52–.71
Third	DIBELS Next composite	A	1	73	.68	.53–.78
			2	128	.68	.58–.76
			3	96	.55	.39–.67
		D	3	216	.37	.25–.48
	Iowa Total Reading	A	3	90	.36	.17–.53
		B	1	93	.63	.48–.74
			3	150	.38	.22–.50
	Iowa Word Analysis	A	3	72	.14	-.10–.36
		B	1	51	.37	.11–.59
			3	115	.27	.09–.43

Table 3.13 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Criterion	Sample	Period	*N*	*r*	CI
Fourth	Iowa Total Reading	A	3	91	.37	.17–.53
		B	1	96	.50	.34–.64
			3	165	.45	.32–.57
Fifth	Iowa Total Reading	A	3	59	.22	-.04–.45
		B	1	109	.53	.46–.60
			3	148	.45	.31–.57
Sixth	Iowa Total Reading	A	3	82	.49	.31–.64
		B	3	152	.39	.24–.52
Seventh	Iowa Total Reading	A	3	91	.29	.09–.47
		B	3	150	.45	.31–.57
Eighth	Iowa Total Reading	A	3	77	.43	.22–.59
		B	3	106	.52	.37–.65

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study.

Table 3.14 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Maze

Grade	Criterion	Sample	Period	Form	*N*	*r*	CI
Second	DIBELS Next composite	D	3	3	195	.67	.58–.74
	Iowa Total Reading	B	1	1	121	.72	.63–.80
				2	88	.64	.49–.75
				3	61	.67	.51–.79
	Iowa Word Analysis	B	1	1	102	.67	.55–.77
				2	93	.60	.45–.72
Third	DIBELS Next composite	D	3	3	199	.62	.53–.70
	Iowa Total Reading	B	1	1	125	.66	.55–.75
				2	100	.69	.58–.78
				3	65	.73	.59–.82
	Iowa Word Analysis	B	1	1	107	.62	.49–.72
				2	106	.61	.48–.72
Fourth	Iowa Total Reading	B	1	1	134	.78	.70–.84
				2	97	.77	.67–.84
				3	64	.70	.54–.80

Table 3.14 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Maze

Grade	Criterion	Sample	Period	Form	*N*	*r*	CI
Fifth	Iowa Total Reading	B	1	1	137	.78	.71–.84
				2	95	.73	.62–.81
				3	68	.52	.32–.68
Sixth	Iowa Total Reading	B	1	1	59	.55	.35–.71
				2	57	.69	.53–.81
Seventh	Iowa Total Reading	B	1	1	70	.76	.65–.85
				2	55	.75	.61–.85

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study. Period indicates period administered, and form indicates benchmark form correlated with the criterion in the indicated period.

Table 3.15 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Composite

Grade	Criterion	Sample	Period	*N*	*r*	CI
Kindergarten	DIBELS Next composite	D	3	321	.85	.81–.88
	Iowa Total Reading	B	3	127	.61	.48–.71
	Iowa Word Analysis	B	3	128	.32	.16–.47
First	DIBELS Next composite	D	3	302	.87	.85–.90
	Iowa Total Reading	B	3	122	.68	.58–.77
	Iowa Word Analysis	B	3	98	.45	.27–.59
Second	DIBELS Next composite	D	3	193	.78	.72–.83
	Iowa Total Reading	B	1	80	.79	.69–.86
			3	179	.76	.69–.81
	Iowa Word Analysis	B	3	153	.78	.71–.83
Third	DIBELS Next composite	D	3	194	.70	.62–.76
	Iowa Total Reading	B	1	91	.73	.61–.81
			3	143	.74	.66–.81
	Iowa Word Analysis	B	1	50	.51	.27–.69
			3	115	.71	.61–.79
Fourth	Iowa Total Reading	B	1	95	.71	.60–.80
			3	157	.74	.66–.80

Table 3.15 Concurrent Criterion Validity Coefficients for DIBELS 8th Edition Composite

Grade	Criterion	Sample	Period	N	r	CI
Fifth	Iowa Total Reading	B	1	107	.42	.25–.57
			3	109	.67	.55–.76
Sixth	Iowa Total Reading	B	3	94	.52	.35–.65
Seventh	Iowa Total Reading	B	3	93	.71	.59–.80

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. D = 2018-2019 Amplify study.

Correlations among DIBELS 8 subtests. One way of establishing that subtests are measuring the same underlying construct, which here is reading, is to correlate them. However, interpreting correlations can be somewhat complicated. Generally, moderate to strong positive relationships among subtests are desirable when the constructs they assess are similar (e.g., WRF and ORF). Correlations that are too strong or nearly perfect (i.e., 1.0) likely indicate substantial redundancy. However, near perfect correlations are not a problem when the two measures predict different abilities or aspects of risk. In such cases, there is added-value in using both measures for screening purposes.

Correlations among DIBELS 8 subtests are reported by grade in Tables 3.16 to 3.21. Beginning of year correlations are above the diagonal; end of year correlations are below it. In kindergarten, the correlations are all positive, with the two NWF scores showing the strongest relationship with each other and with WRF (see Table 3.16). LNF is most strongly related to NWF-CLS at both the beginning and end of year, and moderately related to the other subtests. PSF consistently has the weakest relations to the other subtests, as might be expected given that it is administered differently and taps a component skill that does not directly involve reading. In first grade, all DIBELS 8 subtests are strongly correlated except for PSF (see Table 3.17). Again, PSF shows moderate to weak correlations with the other subtests and the weakest correlations with ORF. From second grade onward, DIBELS 8 subtests administered in these grades are all strongly related, both at the beginning and at the end of year (see Tables 3.18-3.21). In these later grades, ORF Accuracy shows the weakest relations, but the correlations are still strong. The two NWF scores demonstrate an almost perfect relationship throughout second and third grade, most likely due to the fact that the two scores are derived from the same subtest and that students in these grades increasingly read words, even those without meaning, without sounding them out. ORF and WRF are also strongly related throughout second and third grade. Maze and ORF Accuracy have moderate to strong relationships with ORF, but relatively weaker relations with each other. Taken together

these results suggest that each DIBELS 8 subtest offers unique value in the measurement of reading.

Table 3.16 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Kindergarten

	LNF	NWF-CLS	NWF-WRC	PSF	WRF
LNF					
N		542	542	606	571
r		.63	.28	.42	.39
CI		.57–.67	.20–.36	.35–.48	.32–.39
NWF-CLS					
N	559		543	541	532
r	.71		.75	.44	.70
CI	.67–.75		.71–.78	.37–.51	.65–.74
NWF-WRC					
N	559	559		541	532
r	.59	.86		.18	.76
CI	.53–.64	.84–.88		.10–.26	.72–.79
PSF					
N	567	558	558		570
r	.45	.42	.37		.26
CI	.38–.51	.35–.48	.30–.44		.17–.33
WRF					
N	553	554	554	556	
r	.61	.81	.78	.29	
CI	.55–.66	.75–.81	.75–.81	.21–.36	

Note. Correlations above the diagonal are for beginning-of-year benchmark forms, and those below the diagonal are for end-of-year benchmark forms. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study).

Table 3.17 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of First Grade

	LNF	NWF-CLS	NWF-WRC	ORF	ORF-ACC	PSF	WRF
LNF							
N		603	603	591	591	607	603
r		.67	.59	.53	.61	.33	.61
CI		.63–.72	.53–.64	.47–.59	.55–.66	.26–.40	.55–.65
NWF-CLS							
N	583		603	591	591	604	602
r	.63		.90	.77	.70	.26	.83
CI	.58–.68		.88–.91	.74–.80	.66–.74	.18–.33	.80–.85
NWF-WRC							
N	583	587		591	591	604	602
r	.57	.94		.71	.62	.23	.79
CI	.51–.62	.93–.95		.67–.75	.57–.67	.15–.30	.75–.82
ORF							
N	583	583	583		593	592	593
r	.64	.81	.76		.77	.09	.91
CI	.59–.68	.78–.84	.72–.79		.73–.80	.01–.17	.90–.93
ORF–ACC							
N	583	583	583	587		592	593
r	.63	.56	.54	.68		.26	.75
CI	.58–.68	.50–.61	.48–.60	.63–.72		.19–.34	.71–.78
PSF							
N	583	587	587	587	587		604
r	.26	.24	.27	.12	.28		.18
CI	.18–.33	.16–.31	.19–.34	.04–.20	.21–.36		.10–.25
WRF							
N	583	586	586	587	587	591	
r	.62	.83	.77	.94	.63	.12	
CI	.57–.67	.81–.86	.74–.80	.93–.95	.58–.68	.04–.19	

Note. Correlations above the diagonal are for beginning-of-year benchmark forms, and those below the diagonal are for end-of-year benchmark forms. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study).

Table 3.18 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Second Grade

	NWF-CLS	NWF-WRC	ORF	ORF-ACC	WRF	Maze
NWF–CLS						
N		469	463	463	469	201
r		.95	.77	.54	.77	.70
CI		.94–.96	.73–.80	.47–.60	.73–.81	.62–.77
NWF–WRC						
N	639		463	463	469	201
r	.98		.74	.53	.75	.69
CI	.97–.98		.70–.78	.46–.59	.71–.79	.61–.76
ORF						
N	638	638		461	463	196
r	.80	.79		.69	.92	.78
CI	.76–.82	.76–.82		.64–.74	.91–.93	.72–.83
ORF–ACC						
N	638	638	640		463	196
r	.49	.50	.64		.70	.48
CI	.43–.55	.44–.56	.59–.68		.65–.74	.37–.58
WRF						
N	639	639	640	640		201
r	.84	.83	.90	.57		.71
CI	.82–.86	.81–.86	.88–.91	.52–.62		.67–.76
Maze						
N	132	132	133	133	133	
r	.62	.63	.72	.45	.73	
CI	.50–.72	.51–.72	.63–.80	.30–.57	.64–.80	

Note. Correlations above the diagonal are for beginning-of-year benchmark forms, and those below the diagonal are for end-of-year benchmark forms. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study).

Table 3.19 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Third Grade

	NWF-CLS	NWF-WRC	ORF	ORF-ACC	WRF	Maze
NWF–CLS						
N		438	437	437	438	233
r		.97	.77	.53	.79	.60
CI		.96–.97	.72–.80	.46–.60	.75–.82	.51–.67
NWF–WRC						
N	491		437	437	438	233
r	.98		.75	.56	.78	.58
CI	.98–.98		.71–.79	.49–.62	.75–.82	.49–.66
ORF						
N	490	490		448	448	233
r	.79	.78		.67	.90	.76
CI	.75–.82	.74–.81		.61–.71	.88–.91	.70–.81
ORF–ACC						
N	490	490	492		448	233
r	.37	.37	.50		.70	.72
CI	.29–.44	.30–.45	.44–.57		.65–.74	.65–.78
WRF						
N	490	490	492	492		233
r	.85	.84	.89	.47		.68
CI	.82–.87	.81–.86	.86–.90	.39–.53		.63–.73
Maze						
N	166	166	166	166	166	
r	.69	.70	.79	.46	.75	
CI	.60–.76	.62–.77	.73–.84	.33–.57	.67–.81	

Note. Correlations above the diagonal are for beginning-of-year benchmark forms, and those below the diagonal are for end-of-year benchmark forms. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study).

Table 3.20 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Fourth, Fifth, and Sixth Grades

Grade		ORF	ORF-ACC	Maze
Fourth	**ORF**			
	N		439	246
	r		.69	.74
	CI		.64–.74	.68–.79
	ORF–ACC			
	N	488		246
	r	.61		.55
	CI	.55–.66		.45–.63
	Maze			
	N	162	162	
	r	.72	.37	
	CI	.63–.78	.23–.49	
Fifth	**ORF**			
	N		391	211
	r		.53	.40
	CI		.46-.60	.28–.51
	ORF–ACC			
	N	467		211
	r	.47		.33
	CI	.40–.54		.20-.44
	Maze			
	N	160	160	
	r	.99	.40	
	CI	.99–.99+	.26–.52	

Table 3.20 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Fourth, Fifth, and Sixth Grades

Grade		ORF	ORF-ACC	Maze
	ORF			
	N		197	191
	r		.53	.67
	CI		.42-.63	.58–.74
	ORF–ACC			
Sixth	N	297		191
	r	.54		.30
	CI	.46–.62		.16–.43
	Maze			
	N	164	164	
	r	.76	.46	
	CI	.69–.82	.33–.57	

Note. Correlations above the diagonal are for beginning-of-year benchmark forms, and those below the diagonal are for end-of-year benchmark forms. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study).

Table 3.21 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Seventh and Eighth Grades

Grade		ORF	ORF-ACC	Maze
	ORF			
	N		77	73
	r		.50	.72
	CI		.31–.65	.58–.81
	ORF–ACC			
Seventh	N	197		73
	r	.53		.33
	CI	.43–.63		.11–.52
	Maze			
	N	63	63	
	r	.76	.33	
	CI	.63–.85	.09–.54	

Table 3.21 Correlations among DIBELS 8th Edition Subtests at the Beginning and End of Seventh and Eighth Grades

Grade		ORF	ORF-ACC	Maze
Eighth	**ORF**			
	N		74	68
	r		.75	.81
	CI		.63−.84	.71-.87
	ORF−ACC			
	N	145		68
	r	.72		.47
	CI	.63−.79		.26−.64
	Maze			
	N	64	64	
	r	.85	.53	
	CI	.76−.91	.32−.68	

Note. Correlations above the diagonal are for beginning-of-year benchmark forms, and those below the diagonal are for end-of-year benchmark forms. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study).

Predictive Validity

Another way of establishing the validity of a test is by examining its ability to predict scores on criterion measures taken at a later time. Given the use of DIBELS as a screening and risk prediction tool, this type of validity evidence is arguably the most important. Predictive validity can be evaluated using correlations or through receiver operating characteristic (ROC) curve analyses. In the context of progress monitoring, correlations are between estimates of change (i.e., slopes) and a criterion measure at the end of the year. DIBELS 8 was evaluated using each of these methods and results are presented in this section.

Predictive correlations. Depending on the grade, DIBELS 8 measures were correlated with end of year administrations of the DIBELS Next Composite, the Total Reading and Word Analysis scores from the Iowa Assessment, and the CTOPP-2 symbolic and non-symbolic composite scores. Results are presented in Tables 3.22 through 3.30 by subtest, grade, sample, and the benchmark period in which the DIBELS 8 measure was administered.

Similar to external criterion concurrent validity, predictive validity with the DIBELS Next composite scores, where available, was quite good, with the exception of PSF in first grade, and NWF-CLS and WRC at the beginning of kindergarten. LNF correlations ranged from .53 - .82. Correlations for PSF ranged from .44 to .65 in kindergarten, but only .10 to .23 in

first grade. The correlation for NWF-CLS in the beginning of kindergarten was only .43, but otherwise, NWF-CLS and NWF-WRC correlations were strong across grades and times of year, ranging from .54 to .79. Similarly, the correlation for WRF in the beginning of kindergarten for sample D was only .48, but otherwise ranged from .65 to .87. ORF in first through third grades was consistently strongly predictive of the DIBELS Next Composite, with correlations across grades and benchmark periods ranging from .75 to .93, as was ORF Accuracy. Maze was only moderately predictive of the DIBELS Next Composite in second and third grades, but the predictive correlations for the DIBELS 8 were strong, ranging from .68 to .85.

Predictive relations with the Iowa Assessment scores varied in strength, as would be predicted based on the similarity of the component skills assessed. For example, PSF showed the weakest relationships with the Iowa Total Reading and Word Analysis scores in first grade. LNF was somewhat weakly related to the two Iowa scores in kindergarten, but they were moderately to strongly correlated in first grade. Similarly, the remaining DIBELS 8 subtests and the DIBELS 8 Composite consistently displayed moderate to strong relations with the Iowa Total Reading, regardless of grade and benchmark period. In contrast, relationships between DIBELS 8 subtests and the Iowa Word Analysis were more varied, with relatively wide ranges in correlations depending on the subtest, grade, and benchmark period.

Finally, predictive validity correlations between beginning of year LNF and PSF varied in strength from moderate to strong based on subtest, grade level, and composite. Specifically, LNF demonstrated moderate correlations with both the CTOPP-2 non-symbolic and symbolic composite scores and strong correlations for the same scores in first grade. In addition, PSF demonstrated a moderately strong correlation with the CTOPP-2 phonological awareness composite in kindergarten. In first grade, the correlation with the phonological composite was strong for the students who took the CTOPP-2 subtests for younger students but moderate for the students who took the CTOPP-2 subtests for older students. The difference between the two first grade correlations here is most likely due to the changing nature of subtests contributing to the CTOPP-2 phonological awareness composite.

Taken together, the predictive correlations with DIBELS Next and Iowa Total Reading suggest that DIBELS 8 is a very good to excellent predictor of end of year reading proficiency as measured by two widely used measures of reading, throughout kindergarten through eighth grade. In addition, DIBELS 8 LNF and PSF are moderately to strongly predictive of performance on the CTOPP-2 measures of naming speed and phonological awareness. The latter results offer support for the use of DIBELS 8 subtests in dyslexia screening.

Table 3.22 Predictive Validity Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	123	.78	.71–.84
			2	154	.82	.76–.86
		D	1	306	.70	.63–.75
			2	314	.69	.63–.74
	CTOPP-2 Non-symbolic composite	C	1	151	.31	.16–.45
	CTOPP-2 Symbolic composite	C	1	151	.35	.21–.51
	Iowa Total Reading	B	1	126	.52	.38–.64
			2	128	.49	.35–.61
	Iowa Word Analysis	B	1	127	.19	.02–.36
			2	129	.25	.08–.40
First	DIBELS Next composite	A	1	132	.64	.53–.73
			2	170	.70	.62–.77
		D	1	298	.66	.59–.72
			2	295	.53	.45–.61
	CTOPP-2 Non-symbolic composite	C	1	54	.56	.34–.72
	CTOPP-2 Symbolic composite	C	1	157	.50	.37–.61
	Iowa Total Reading	A	1	80	.57	.40–.70
			2	115	.57	.43–.68
		B	1	119	.43	.27–.57
			2	132	.42	.27–.56
	Iowa Word Analysis	A	1	80	.57	.40–.70
			2	117	.52	.37–.64
		B	1	100	.32	.13–.49
			2	106	.33	.15–.49

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

.

Table 3.23 Predictive Validity Coefficients for DIBELS 8th Edition Phoneme Segmentation Fluency

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	72	.44	.23–.61
		D	1	306	.51	.43–.59
			2	309	.65	.58–.71
	CTOPP-2 Phonological awareness composite	C	1	177	.45	.13–.40
	Iowa Total Reading	B	1	101	.30	.11–.47
			2	114	.30	.12–.46
	Iowa Word Analysis	B	1	102	.24	.04–.41
			2	115	.24	.06–.40
First	DIBELS Next composite	A	1	129	.11	-.02–.23
			2	168	.23	.05–.31
		D	1	294	.20	.09–.31
			2	287	.14	.02–.25
	CTOPP-2 Phonological awareness composite	C	1	60[a]	.54	.33–.70
			1	100[a]	.37	.18–.53
	Iowa Total Reading	A	1	79	.12	-.05–.28
			2	113	.29	.12–.45
		B	1	118	.27	.09–.43
			2	132	.33	.17–.47
	Iowa Word Analysis	A	1	79	.02	-.14–.19
			2	115	.23	.08–.36
		B	1	100	.13	-.07–.32
			2	106	.05	-.14–.24

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composite is derived from measures of sound matching, blending, elision, and phoneme isolation depending on student age at time of testing. Students younger than 7 years old take the first three subtests, and students 7 years old and older take the last three subtests.

[a] The first-grade phonological awareness composite has two samples because the CTOPP-2 provides separate age-based norms. The smaller subsample is compared to the younger norming group and the larger to the older based on each child's age at time of CTOPP-2 testing.

Table 3.24 Predictive Validity Coefficients of DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	89	.43	.24–.59
			2	130	.72	.63–.80
		D	1	306	.55	.47–.62
			2	309	.79	.74–.83
	Iowa Total Reading	B	1	87	.57	.41–.70
			2	114	.47	.31–.60
	Iowa Word Analysis	B	1	89	.25	.04–.44
			2	115	.24	.06–.40
First	DIBELS Next composite	A	1	154	.76	.69–.82
			2	194	.76	.70–.82
		D	1	294	.62	.55–.69
			2	287	.64	.57–.70
	Iowa Total Reading	A	1	153	.55	.43–.65
			2	197	.60	.50–.68
		B	1	118	.50	.35–.63
			2	132	.55	.42–.66
	Iowa Word Analysis	A	1	153	.51	.38–.62
			2	199	.49	.38–.59
		B	1	99	.32	.13–.49
			2	106	.35	.17–.70
Second	DIBELS Next composite	A	1	97	.72	.61–.80
			2	117	.67	.55–.76
		D	1	201	.64	.56–.72
			2	195	.55	.44–.64
	Iowa Total Reading	A	1	49	.66	.47–.79
			2	76	.56	.38–.70
		B	1	132	.57	.45–.68
			2	186	.64	.55–.72
	Iowa Word Analysis	A	2	49	.47	.22–.66
		B	1	102	.60	.46–.71
			2	152	.66	.56–.74

Table 3.24 Predictive Validity Coefficients of DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
	DIBELS Next composite	A	2	107	.66	.54–.76
		D	1	211	.64	.55–.71
			2	214	.54	.43–.62
	Iowa Total Reading	A	2	89	.39	.19–.55
Third		B	1	86	.44	.25–.60
			2	143	.66	.56–.75
	Iowa Word Analysis	A	2	72	.27	.05–.48
		B	1	54	.60	.39–.75
			2	111	.63	.51–.73

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.25 Predictive Validity Coefficients of DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
	DIBELS Next composite	A	1	73	.71	.58–.81
			2	126	.71	.61–.79
		D	1	306	.55	.47–.62
			2	309	.74	.68–.79
Kindergarten	Iowa Total Reading	B	1	87	.29	.09–.47
			2	114	.44	.28–.58
	Iowa Word Analysis	A	2	49	.56	.33–.73
		B	1	89	-.01	-.22–.20
			2	115	.29	.11–.45

Table 3.25 Predictive Validity Coefficients of DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
First	DIBELS Next composite	A	1	163	.70	.61–.77
			2	185	.75	.68–.80
		D	1	294	.72	.66–.77
			2	287	.65	.58–.71
	Iowa Total Reading	A	1	153	.51	.39–.62
			2	197	.58	.47–.66
		B	1	118	.46	.30–.59
			2	132	.56	.43–.66
	Iowa Word Analysis	A	1	153	.50	.37–61
			2	199	.51	.40–.61
		B	1	99	.41	.23–.56
			2	106	.40	.23–.55
Second	DIBELS Next composite	A	1	97	.71	.60–.85
			2	117	.70	.58–.78
		D	1	201	.65	.56–.72
			2	195	.59	.49–.67
	Iowa Total Reading	A	1	49	.64	.44–.78
			2	76	.60	.43–.73
		B	1	132	.52	.38–.64
			2	186	.67	.58–.74
	Iowa Word Analysis	A	1	49	.56	.33–.73
		B	1	102	.58	.44–.70
			2	152	.70	.61–.77
Third	DIBELS Next composite	A	2	107	.69	.65–.82
		D	1	211	.68	.61–.75
			2	214	.56	.45–.64
	Iowa Total Reading	A	2	89	.41	.22–.57
		B	1	86	.47	.29–.62
			2	143	.67	.57–.75
	Iowa Word Analysis	A	2	72	.31	.09–.51
		B	1	54	.54	.32–.70
			2	111	.66	.54–.75
			2	111	.66	.54–.75

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.26 Predictive Validity Coefficients of DIBELS 8th Edition Word Reading Fluency

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
Kindergarten	DIBELS Next composite	A	1	86	.65	.50–.76
			2	121	.67	.56–.76
		D	1	306	.48	.39–.56
			2	309	.70	.63–.75
	Iowa Total Reading	B	1	88	.28	.08–.47
			2	110	.42	.25–.56
	Iowa Word Analysis		1	89	.08	-.13–.29
			2	111	.14	-.05–.32
First	DIBELS Next composite	A	1	132	.78	.71–.84
			2	170	.85	.80–.89
		D	1	294	.74	.69–.79
			2	287	.72	.66–.77
	Iowa Total Reading	A	1	80	.65	.50–.76
			2	115	.74	.64–.81
		B	1	116	.58	.44–.69
			2	132	.60	.47–.70
	Iowa Word Analysis	A	1	80	.59	.43–.72
			2	117	.63	.50–.73
		B	1	97	.45	.27–.60
			2	106	.41	.24–.56
Second	DIBELS Next composite	A	1	130	.87	.82–.91
			2	150	.87	.82–.90
		D	1	201	.80	.74–.84
			2	195	.78	.71–.83
	Iowa Total Reading	A	1	51	.78	.64–.87
			2	77	.72	.59–.81
		B	1	132	.70	.61–.78
			2	186	.76	.69–.81
	Iowa Word Analysis	A	1	49	.69	.50–.81
		B	1	102	.69	.57–.78
			2	153	.75	.67–.81

Table 3.26 Predictive Validity Coefficients of DIBELS 8th Edition Word Reading Fluency

Grade	End-of-Year Criterion	Sample	Period	N	r	CI
Third	DIBELS Next composite	A	1	131	.84	.78–.88
			2	154	.82	.77–.87
		D	1	211	.78	.72–.83
			2	214	.77	.71–.82
	Iowa Total Reading	A	1	71	.61	.44–.74
			2	89	.59	.43–.71
		B	1	86	.54	.37–.67
			2	143	.71	.61–.78
	Iowa Word Analysis	A	1	53	.47	.23–.66
			2	72	.46	.25–.62
		B	1	54	.62	.37–.67
			2	111	.68	.57–.77

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.27 Predictive Validity Coefficients of DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Criterion	Sample	Period	N	r	CI
First	DIBELS Next composite	A	1	141	.82	.76–.87
			2	203	.88	.84–.91
		D	1	293	.75	.70–.80
			2	287	.76	.70–.80
	Iowa Total Reading	A	1	59	.73	.58–.83
			2	115	.79	.71–.85
		B	1	109	.62	.48–.72
			2	130	.68	.57–.76
	Iowa Word Analysis	A	1	59	.60	.40–.74
			2	117	.69	.58–.77
		B	1	90	.37	.18–.54
			2	104	.43	.25–.57

Table 3.27 Predictive Validity Coefficients of DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Criterion	Sample	Period	N	r	CI
Second	DIBELS Next composite	A	1	121	.79	.71–.85
			2	140	.89	.85–.92
		D	1	201	.80	.74–.84
			2	195	.79	.74–.84
	Iowa Total Reading	A	1	137	.63	.51–.72
			2	163	.74	.66–.80
		B	1	127	.68	.58–.77
			2	185	.77	.70–.82
	Iowa Word Analysis	A	1	129	.55	.41–.66
			2	137	.64	.53–.73
		B	1	97	.62	.48–.73
			2	152	.73	.65–.80
Third	DIBELS Next composite	A	1	81	.93	.89–.95
			2	126	.86	.80–.90
		D	1	211	.82	.77–.86
			2	214	.85	.81–.88
	Iowa Total Reading	A	1	61	.74	.61–.84
			2	132	.69	.59–.77
		B	1	86	.56	.39–.69
			2	142	.74	.66–.81
	Iowa Word Analysis	A	2	96	.50	.33–.64
		B	1	54	.62	.43–.76
			2	110	.69	.57–.78
Fourth	Iowa Total Reading	A	1	55	.53	.30–.70
			2	55	.67	.49–.79
		B	1	110	.69	.57–.77
			2	164	.80	.73–.85
Fifth	Iowa Total Reading	A	1	128	.63	.52–.73
			2	99	.69	.57–.78
		B	1	95	.46	.29–.61
			2	144	.73	.65–.80
Sixth	Iowa Total Reading	A	1	49	.50	.25–.68
			2	86	.65	.50–.75
		B	1	44	.69	.50–.82
			2	149	.75	.67–.81

Table 3.27 Predictive Validity Coefficients of DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Criterion	Sample	Period	N	r	CI
Seventh	Iowa Total Reading	A	1	59	.52	.31–.69
			2	101	.57	.42–.69
		B	2	149	.74	.65–.80
Eighth	Iowa Total Reading	A	1	46	.69	.50–.82
			2	85	.48	.30–.63
		B	1	43	.77	.60–.87
			2	99	.63	.49–.73

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.28 Predictive Validity Coefficients of DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Criterion	Sample	Period	N	r	CI
First	DIBELS Next composite	A	1	141	.87	.82–.90
			2	203	.83	.78–.87
		D	1	293	.77	.72–.82
			2	285	.77	.71–.81
	Iowa Total Reading	A	1	59	.77	.71–.89
			2	115	.78	.70–.84
		B	1	109	.65	.53–.75
			2	130	.62	.50–.71
	Iowa Word Analysis	A	1	59	.72	.56–.82
			2	117	.72	.62–.80
		B	1	90	.50	.32–.64
			2	104	.42	.25–.57

Table 3.28 Predictive Validity Coefficients of DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Criterion	Sample	Period	*N*	*r*	CI
Second	DIBELS Next composite	A	1	121	.70	.60–.78
			2	140	.69	.59–.77
		D	1	198	.64	.55–.72
			2	195	.70	.63–.77
	Iowa Total Reading	A	1	137	.61	.49–.70
			2	163	.51	.38–.61
		B	1	127	.67	.56–.76
			2	185	.64	.54–.71
	Iowa Word Analysis	A	1	129	.46	.31–.59
			2	137	.54	.41–.65
		B	1	97	.65	.52–.76
			2	152	.63	.52–.72
Third	DIBELS Next composite	A	1	81	.76	.65–.84
			2	126	.70	.70–.78
		D	1	210	.64	.55–.71
			2	214	.57	.47–.66
	Iowa Total Reading	A	1	61	.67	.51–.79
			2	132	.53	.53–.64
		B	1	86	.49	.31–.63
			2	142	.54	.41–.65
	Iowa Word Analysis	A	2	96	.44	.44–.59
		B	1	54	.40	.15–.60
			2	110	.51	.36–.64
Fourth	Iowa Total Reading	A	1	55	.44	.20–.63
			2	55	.37	.11–.58
		B	1	110	.55	.41–.67
			2	164	.48	.36–.59
Fifth	Iowa Total Reading	A	1	128	.46	.31–.59
			2	99	.41	.23–.56
		B	1	95	.49	.32–.63
			2	144	.51	.38–.62
Sixth	Iowa Total Reading	A	1	49	.46	.20–.65
			2	86	.53	.36–.67
		B	1	44	.48	.21–.68
			2	149	.61	.50–.70

Table 3.28 Predictive Validity Coefficients of DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Criterion	Sample	Period	N	r	CI
Seventh	Iowa Total Reading	A	1	59	.29	.04–.51
			2	101	.36	.17–.52
		B	2	149	.34	.19–.47
Eighth	Iowa Total Reading	A	1	46	.44	.17–.65
			2	85	.23	.02–.42
		B	1	43	.47	.19–.67
			2	99	.47	.30–.61

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.29 Predictive Validity Coefficients of DIBELS 8th Edition Maze

Grade	Criterion	Sample	Period	N	r	CI
Second	DIBELS Next Composite	D	2	193	.59	.49–.68
	Iowa Total Reading	B	1	168	.67	.58–.75
			2	67	.69	.54–.80
	Iowa Word Analysis	B	1	132	.71	.61–.78
			2	54	.64	.45–.77
Third	DIBELS Next Composite	D	2	214	.36	.24–.47
	Iowa Total Reading	B	1	122	.70	.59–.78
			2	61	.38	.38–.72
	Iowa Word Analysis	B	1	87	.49	.31–.64
			2	25	.64	.33–.83
Fourth	Iowa Total Reading	B	1	143	.76	.69–.82
			2	78	.68	.60–.81
Fifth	Iowa Total Reading	B	1	122	.40	.24–.54
			2	57	.49	.27–.67
Sixth	Iowa Total Reading	B	1	99	.68	.56–.77
			2	33	.79	.61–.89
Seventh	Iowa Total Reading	B	1	91	.73	.62–.81
			2	28	.71	.46–.86
Eighth	Iowa Total Reading	B	1	44	.74	.57–.85
			2	39	.63	.40–.79

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Table 3.30 Predictive Validity Coefficients of DIBELS 8th Edition Composite

Grade	Criterion	Sample	Period	*N*	*r*	CI
Kindergarten	DIBELS Next composite	D	1	306	.68	.65–.76
			2	309	.85	.81–.88
	Iowa Total Reading	B	1	82	.59	.43–.72
			2	110	.52	.37–.64
	Iowa Word Analysis	B	1	84	.24	.02–.43
			2	111	.29	.11–.45
First	DIBELS Next Composite	B	1	293	.80	.75–.84
			2	285	.78	.73–.82
	Iowa Total Reading	B	1	108	.63	.50–.73
			2	130	.66	.56–.75
	Iowa Word Analysis	B	1	90	.39	.20–.55
			2	104	.42	.25–.57
Second	DIBELS Next Composite	D	2	190	.77	.71–.82
	Iowa Total Reading	B	1	122	.68	.58–.77
			2	144	.72	.63–.79
	Iowa Word Analysis	B	1	91	.70	.58–.79
			2	135	.71	.61–.78
Third	DIBELS Next Composite	D	2	211	.74	.67–.79
	Iowa Total Reading	B	1	82	.54	.36–.68
			2	138	.75	.67–.81
	Iowa Word Analysis	B	1	51	.65	.46–.79
			2	107	.69	.58–.78
Fourth	Iowa Total Reading	B	1	105	.70	.59–.79
			2	159	.80	.74–.85
Fifth	Iowa Total Reading	B	1	94	.47	.30–.62
			2	133	.73	.64–.80
Sixth	Iowa Total Reading	B	1	42	.69	.49–.82
			2	100	.66	.54–.76
Seventh	Iowa Total Reading	B	1	36	.78	.61–.88
			2	91	.77	.67–.84
Eighth	Iowa Total Reading	B	1	43	.77	.61–.87
			2	46	.74	.58–.85

Note. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. A = 2017-2018 CTL norming study. B = 2018-2019 CTL norming study. C = 2018-2019 CTL dyslexia subsample. D = 2018-2019 Amplify study. CTOPP-2 composites are derived from measures of rapid naming of colors and objects (non-symbolic) and digits and letters (symbolic).

Screening accuracy. One of the uses of DIBELS 8th Edition is to identify students who are not on track for meeting reading proficiency standards and those who are at pronounced risk for reading difficulties. To support this use, we have provided two types of cut-scores for classifying students. The first score, called the risk cut-score, can be used to classify students who are at risk for reading difficulties, including dyslexia. The second score, called the benchmark goal, can be used to classify students who are at some risk for not meeting proficiency goals versus those who are on track for meeting proficiency goals.

The cut-scores were calculated using ROC curve analyses, which describe the relation between true positive rates (i.e., scores that correctly identify students who were not on track for attaining proficiency) and false positive rates (i.e., scores that indicate a student was not on-track when they really were). In this case, the ROC results characterize the extent to which DIBELS 8 scores correctly predicted performance on a criterion measure of reading: the DIBELS Next composite score percentile ranks in kindergarten and the Iowa Assessment Total Reading percentile ranks in all other grades. ROC analyses yield an area under the curve (AUC) estimate, which summarizes a test's classification accuracy. An AUC of .5 indicates the test predicts no better than chance, whereas an AUC of 1.0 indicates that a test is perfectly predictive (Habibzadeh, Habibzadeh, & Yadollahie, 2016).

In addition to the AUC, ROC analyses provide information about the sensitivity and specificity of a screener. Sensitivity scores summarize how well a subtest's cut-score identifies students who have not met a criterion goal. It is expressed as a proportion, ranging from 0 to 1. The sensitivity value represents the proportion of "truly" at-risk students who are correctly identified by the screener as being at risk. Specificity, which is also expressed as a proportion, is the counterpart to sensitivity. Specificity represents the proportion of "truly healthy" readers who are accurately identified as not at risk by the screener (i.e., identified as "okay"). Sensitivity can also be interpreted as the probability (likelihood) that a student who meets the criterion goal has been identified as such by the screener.

Although sensitivity and specificity are stable indicators of screening effectiveness regardless of the prevalence of reading difficulties in the population (Pepe, 2003), an important determinant of sensitivity and specificity that does not affect the AUC is how the cut-score for a test is set. DIBELS 8 sets cut scores that balance sensitivity and specificity, given their complementary roles in a prevention model in education. Specifically, balancing both statistics results in maximizing the proportion of students correctly identified for intervention without under-identifying students correctly identified as not in need of intervention. Thus, wherever possible, recommended cut points for DIBELS 8th Edition subtests were set to maximize sensitivity while maintaining specificity at or above .80. More specifically, for

each benchmark, the cut was set at the score with the highest sensitivity among scores with a specificity at or above .80. In cases where the maximum sensitivity value exceeded .90, the cut point selected was the score that minimized the difference between sensitivity and specificity among scores with specificity at or above .80; in other words, maximizing both statistics. For the few measures and periods with no cut scores that met the minimum threshold for specificity, the cut point represents the score that best balances the goals of providing additional instruction where needed while keeping demands on teachers reasonable.

AUC, sensitivity, and specificity results are reported by grade and within grade by subtest and benchmark period for two cuts on a criterion measure (i.e., 20th percentile rank and 40th percentile rank). For kindergarten, the DIBELS Next composite score at the end of the year served as the criterion measure, and in all other grades the criterion measure was the Iowa Assessment Total Reading Score at the end of the year. The Iowa Assessment is a published, group-administered, multiple-choice, norm-referenced measure of reading achievement (Welch & Dunbar, 2012).

Regardless of criterion measure, the 20th percentile rank cut is intended for use in identifying students who are well below benchmark, at risk for not meeting end of year learning goals, and in need of intensive intervention. For the LNF, PSF, and NWF subtests, students falling below this cut may also be at risk for reading disabilities, including dyslexia. The 40th percentile cut is intended for use in identifying students who are below benchmark, and thus at some risk of not meeting end of year learning goals and in need of strategic support. Note that in some grades, statistics are not reported for the 40th percentile cut-score (e.g., at the beginning of kindergarten for NWF-WRC and WRF) because ROC analyses indicated that a single cut worked best at the given time of year and grade. In other words, there was little to no distinction between recommended cuts for risk and the benchmark goal, suggesting that the strategic support designation was not meaningful for that score in that grade at that time of year.

In kindergarten, AUCs are uniformly high, with the majority falling at .8 or above and sensitivity and specificity are also routinely high (see Table 3.31). The exception is for NWF-WRC and WRF at the beginning of the year. Due to substantial floor effects, these scores are not ideal for predicting risk at the beginning of kindergarten and is one reason for the discontinue benchmark rules used in DIBELS 8. As a result, LNF, PSF, NWF-CLS, and the Composite score are the strongest indicators for the beginning of kindergarten, while all scores are highly predictive at the middle and end of the kindergarten year.

In first grade, both NWF scores, the WRF score, both ORF scores, and the Composite score have uniformly high AUCs with the majority falling at .8 or above, and sensitivity and

specificity are also routinely high (see Table 3.32). These statistics are somewhat lower for LNF and PSF, which is partly due to the change in criterion measure from DIBELS Next Composite Score to the Iowa Assessment Total Reading Score. Whereas DIBELS Next includes letter naming and phonemic awareness component skills in the composite score, the Iowa Total Reading Score does not assess these same component skills, making it a more distal criterion measure. The strongest predictors in first grade are NWF-CLS, WRF, ORF, and the Composite score.

In second and third grade, AUCs are again uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Tables 3.33 and 3.34). The strongest predictors are ORF, Maze, and the Composite score, although NWF and WRF remain very good predictors. However, ORF-ACC shows a declining trend in its predictive power. While NWF is a robust predictor, the strongest predictors in second and third grade are WRF, ORF, and the Composite score.

In fourth through eighth grade, AUCs, sensitivity, and specificity for ORF, Maze, and the Composite score remain quite high (see Tables 3.35 through 3.39). ORF Accuracy shows the same declining trend in predictive power. Interestingly, Maze is an incredibly strong predictor of risk in Grades 5 to 8. Taken together, the results suggest that from kindergarten to eighth grade, DIBELS 8 subtests and the Composite score predict end of year reading achievement very well.

Table 3.31 ROC Curve Results for DIBELS 8 Kindergarten Subtests Predicting DIBELS Next Composite

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
LNF	20th	1	306	0.85	0.79-0.91	0.745	0.847
		2	314	0.81	0.74-0.89	0.750	0.779
		3	321	0.87	0.81-0.93	0.818	0.789
	40th	1	306	0.83	0.78-0.88	0.726	0.755
		2	314	0.84	0.79-0.89	0.784	0.744
		3	321	0.88	0.84-0.92	0.852	0.801
PSF	20th	1	306	0.79	0.73-0.85	0.686	0.729
		2	309	0.88	0.84-0.92	0.745	0.853
		3	321	0.86	0.81-0.91	0.764	0.816
	40th	1	306	0.78	0.72-0.83	0.830	0.575
		2	309	0.83	0.78-0.87	0.716	0.780
		3	321	0.77	0.72-0.83	0.739	0.680

Table 3.31 ROC Curve Results for DIBELS 8 Kindergarten Subtests Predicting DIBELS Next Composite

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1	306	0.82	0.76-0.87	0.667	0.780
		2	309	0.89	0.84-0.93	0.745	0.837
		3	321	0.91	0.87-0.96	0.782	0.887
	40th	1	306	0.81	0.76-0.86	0.802	0.710
		2	309	0.87	0.83-0.91	0.817	0.770
		3	321	0.90	0.86-0.93	0.783	0.845
NWF-WRC	20th	1	306	0.27	0.23-0.30	1.000	0.400
		2	309	0.86	0.82-0.90	0.863	0.771
		3	321	0.90	0.86-0.94	0.745	0.857
	40th	1	NA	NA	NA	NA	NA
		2	309	0.84	0.80-0.89	0.789	0.775
		3	321	0.90	0.87-0.93	0.800	0.820
WRF	20th	1	306	0.33	0.28-0.39	0.863	0.431
		2	309	0.87	0.82-0.92	0.745	0.857
		3	321	0.89	0.84-0.93	0.782	0.865
	40th	1	NA	NA	NA	NA	NA
		2	309	0.86	0.82-0.90	0.771	0.785
		3	321	0.90	0.87-0.94	0.791	0.820
Composite	20th	1	306	0.86	0.81-0.92	0.71	0.88
		2	309	0.92	0.88-0.95	0.78	0.87
		3	321	0.94	0.91-0.97	0.80	0.93
	40th	1	306	0.84	0.80-0.89	0.65	0.83
		2	309	0.91	0.88-0.94	0.80	0.85
		3	321	0.94	0.92-0.96	0.88	0.84

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Table 3.32 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
LNF	20th	1	123	0.71	0.61-0.82	0.679	0.621
		2	137	0.73	0.63-0.82	0.714	0.637
		3	134	0.73	0.64-0.83	0.719	0.676
	40th	1	123	0.67	0.58-0.77	0.800	0.413
		2	137	0.70	0.62-0.79	0.704	0.561
		3	134	0.69	0.60-0.78	0.672	0.627
PSF	20th	1	122	0.67	0.57-0.77	0.643	0.628
		2	137	0.68	0.58-0.78	0.600	0.588
		3	134	0.69	0.60-0.79	0.688	0.637
	40th	1	122	0.62	0.52-0.72	0.783	0.484
		2	137	0.64	0.54-0.73	0.718	0.439
		3	134	0.62	0.53-0.72	0.597	0.597
NWF-CLS	20th	1	122	0.79	0.71-0.88	0.714	0.702
		2	137	0.79	0.71-0.86	0.771	0.686
		3	130	0.82	0.73-0.90	0.800	0.690
	40th	1	122	0.72	0.63-0.81	0.900	0.371
		2	137	0.74	0.66-0.83	0.859	0.439
		3	130	0.75	0.66-0.83	0.766	0.591
NWF-WRC	20th	1	122	0.74	0.66-0.82	0.750	0.596
		2	137	0.74	0.66-0.82	0.629	0.696
		3	130	0.75	0.66-0.84	0.733	0.680
	40th	1	122	0.71	0.62-0.80	0.950	0.403
		2	137	0.75	0.66-0.83	0.831	0.636
		3	130	0.73	0.64-0.81	0.766	0.545
WRF	20th	1	120	0.78	0.70-0.86	0.778	0.667
		2	137	0.82	0.76-0.90	0.829	0.725
		3	134	0.82	0.74-0.90	0.750	0.716
	40th	1	120	0.77	0.68-0.86	0.948	0.403
		2	137	0.78	0.70-0.86	0.817	0.652
		3	134	0.77	0.69-0.86	0.791	0.716
ORF	20th	1	113	0.82	0.75-0.90	0.769	0.724
		2	135	0.82	0.75-0.89	0.800	0.730
		3	132	0.84	0.77-0.91	0.750	0.810
	40th	1	113	0.81	0.73-0.89	0.944	0.407
		2	135	0.78	0.70-0.86	0.943	0.431
		3	132	0.80	0.72-0.88	0.776	0.738

Table 3.32 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF-ACC	20th	1	113	0.81	0.73-0.89	0.846	0.701
		2	135	0.80	0.72-0.88	0.714	0.770
		3	132	0.81	0.73-0.88	0.750	0.700
	40th	1	113	0.81	0.73-0.89	0.926	0.525
		2	135	0.76	0.68-0.85	0.929	0.400
		3	132	0.80	0.72-0.88	0.791	0.723
Composite	20th	1	112	0.86	0.79-0.93	0.808	0.791
		2	135	0.85	0.78-0.91	0.771	0.790
		3	128	0.87	0.80-0.94	0.767	0.878
	40th	1	112	0.79	0.71-0.88	0.778	0.690
		2	135	0.81	0.73-0.89	0.829	0.677
		3	128	0.81	0.73-0.89	0.797	0.781

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.33 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1	140	0.83	0.77-0.90	0.739	0.766
		2	193	0.89	0.85-0.94	0.804	0.803
		3	198	0.89	0.85-0.94	0.784	0.850
	40th	1	140	0.77	0.69-0.85	0.809	0.597
		2	193	0.87	0.82-0.92	0.803	0.761
		3	198	0.87	0.81-0.92	0.833	0.817
NWF-WRC	20th	1	140	0.82	0.75-0.89	0.804	0.734
		2	193	0.87	0.82-0.92	0.863	0.761
		3	198	0.90	0.85-0.94	0.863	0.803
	40th	1	140	0.76	0.68-0.84	0.853	0.500
		2	193	0.86	0.80-0.91	0.868	0.684
		3	198	0.87	0.82-0.92	0.846	0.775

Table 3.33 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
WRF	20th	1	140	0.87	0.81-0.93	0.826	0.809
		2	194	0.92	0.88-0.96	0.902	0.811
		3	198	0.93	0.89-0.96	0.902	0.830
	40th	1	140	0.83	0.76-0.90	0.824	0.681
		2	194	0.88	0.82-0.93	0.829	0.797
		3	198	0.87	0.82-0.93	0.859	0.808
ORF	20th	1	135	0.87	0.80-0.93	0.786	0.817
		2	193	0.92	0.88-0.96	0.860	0.818
		3	198	0.93	0.89-0.96	0.882	0.884
	40th	1	135	0.80	0.72-0.87	0.778	0.708
		2	193	0.87	0.82-0.92	0.827	0.788
		3	198	0.87	0.82-0.92	0.808	0.842
ORF-ACC	20th	1	135	0.85	0.78-0.92	0.738	0.860
		2	193	0.90	0.86-0.95	0.760	0.888
		3	198	0.86	0.79-0.92	0.863	0.707
	40th	1	135	0.79	0.71-0.87	0.746	0.708
		2	193	0.85	0.80-0.91	0.880	0.686
		3	NA	NA	NA	NA	NA
Maze	20th	1	189	0.88	0.83-0.93	0.940	0.755
		2	153	0.89	0.83-0.94	0.933	0.683
		3	190	0.89	0.84-0.94	0.936	0.664
	40th	1	189	0.87	0.82-0.93	0.973	0.405
		2	153	0.84	0.78-0.91	0.959	0.375
		3	190	0.85	0.79-0.91	0.944	0.364
Composite	20th	1	127	0.90	0.84-0.96	0.821	0.864
		2	149	0.93	0.89-0.97	0.815	0.885
		3	187	0.92	0.88-0.96	0.756	0.901
	40th	1	127	0.82	0.75-0.90	0.780	0.765
		2	149	0.88	0.82-0.93	0.783	0.854
		3	187	0.87	0.82-0.93	0.814	0.846

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Table 3.34 ROC Curve Results for DIBELS 8 Third Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	*N*	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1	114	0.66	0.50-0.82	0.650	0.596
		2	172	0.75	0.64-0.86	0.682	0.693
		3	179	0.81	0.72-0.90	0.720	0.701
	40th	1	114	0.67	0.56-0.77	0.698	0.656
		2	172	0.81	0.75-0.88	0.789	0.739
		3	179	0.81	0.74-0.87	0.770	0.712
NWF-WRC	20th	1	114	0.64	0.51-0.78	0.650	0.617
		2	172	0.73	0.62-0.83	0.591	0.693
		3	179	0.81	0.73-0.88	0.680	0.695
	40th	1	114	0.68	0.58-0.78	0.792	0.574
		2	172	0.81	0.74-0.88	0.807	0.713
		3	179	0.81	0.74-0.87	0.738	0.737
WRF	20th	1	114	0.72	0.59-0.85	0.700	0.723
		2	172	0.81	0.72-0.90	0.727	0.733
		3	179	0.83	0.74-0.91	0.760	0.727
	40th	1	114	0.71	0.62-0.81	0.755	0.574
		2	172	0.82	0.76-0.88	0.789	0.678
		3	179	0.80	0.74-0.87	0.787	0.737
ORF	20th	1	114	0.75	0.64-0.87	0.700	0.702
		2	171	0.84	0.77-0.92	0.773	0.785
		3	179	0.85	0.78-0.93	0.720	0.838
	40th	1	114	0.70	0.60-0.79	0.774	0.508
		2	171	0.81	0.75-0.88	0.772	0.693
		3	179	0.82	0.76-0.88	0.787	0.763
ORF-ACC	20th	1	114	0.73	0.61-0.85	0.800	0.511
		2	171	0.78	0.66-0.89	0.864	0.483
		3	179	0.68	0.55-0.82	0.720	0.383
	40th	1	114	0.75	0.66-0.84	0.981	0.164
		2	NA	NA	NA	NA	NA
		3	NA	NA	NA	NA	NA
Maze	20th	1	166	0.87	0.80-0.93	0.905	0.607
		2	168	0.87	0.80-0.94	0.952	0.605
		3	171	0.89	0.83-0.94	0.957	0.561
	40th	1	166	0.86	0.80-0.92	1.000	0.375
		2	168	0.84	0.78-0.90	1.000	0.277
		3	171	0.84	0.78-0.90	0.982	0.426

Table 3.34 ROC Curve Results for DIBELS 8 Third Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
Composite	20th	1	109	0.72	0.58-0.87	0.632	0.778
		2	165	0.82	0.72-0.91	0.650	0.779
		3	171	0.84	0.77-0.92	0.696	0.804
	40th	1	109	0.71	0.61-0.81	0.706	0.690
		2	165	0.85	0.79-0.91	0.800	0.773
		3	171	0.85	0.79-0.90	0.804	0.765

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Table 3.35 ROC Curve Results for DIBELS 8 Fourth Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	134	0.88	0.83-0.94	0.697	0.812
		2	187	0.92	0.87-0.98	0.857	0.882
		3	189	0.89	0.84-0.94	0.750	0.837
	40th	1	134	0.74	0.65-0.82	0.662	0.759
		2	187	0.86	0.81-0.92	0.706	0.902
		3	189	0.84	0.79-0.90	0.671	0.798
ORF-ACC	20th	1	134	0.85	0.78-0.92	0.788	0.733
		2	187	0.80	0.71-0.90	0.886	0.368
		3	189	0.76	0.66-0.86	0.778	0.477
	40th	1	134	0.74	0.65-0.82	0.962	0.111
		2	NA	NA	NA	NA	NA
		3	NA	NA	NA	NA	NA
Maze	20th	1	182	0.92	0.88-0.96	0.969	0.720
		2	184	0.92	0.88-0.97	0.914	0.718
		3	181	0.90	0.84-0.96	0.939	0.669
	40th	1	182	0.82	0.76-0.88	0.988	0.431
		2	184	0.85	0.80-0.90	0.953	0.495
		3	181	0.84	0.78-0.90	0.900	0.485

Table 3.35 ROC Curve Results for DIBELS 8 Fourth Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
Composite	20th	1	129	0.89	0.83-0.94	0.710	0.796
		2	182	0.92	0.87-0.98	0.853	0.878
		3	180	0.90	0.84-0.95	0.727	0.857
	40th	1	129	0.75	0.66-0.84	0.737	0.642
		2	182	0.86	0.81-0.92	0.798	0.745
		3	180	0.85	0.79-0.90	0.911	0.614

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Table 3.36 ROC Curve Results for DIBELS 8 Fifth Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	96	0.77	0.65-0.89	0.708	0.764
		2	145	0.87	0.80-0.93	0.769	0.807
		3	149	0.86	0.79-0.94	0.704	0.828
	40th	1	96	0.77	0.68-0.87	0.778	0.647
		2	145	0.88	0.82-0.94	0.800	0.811
		3	149	0.86	0.80-0.92	0.788	0.784
ORF-ACC	20th	1	96	0.75	0.63-0.87	0.667	0.736
		2	145	0.86	0.79-0.93	1.000	0.403
		3	149	0.69	0.57-0.81	0.778	0.402
	40th	1	96	0.72	0.62-0.83	0.911	0.255
		2	NA	NA	NA	NA	NA
		3	NA	NA	NA	NA	NA
Maze	20th	1	142	0.80	0.70-0.91	0.840	0.504
		2	133	0.88	0.80-0.96	1.000	0.330
		3	109	0.88	0.77-0.99	0.929	0.421
	40th	1	142	0.77	0.68-0.86	0.917	0.287
		2	133	0.85	0.78-0.92	1.000	0.022
		3	109	0.89	0.82-0.95	1.000	0.222

Table 3.36 ROC Curve Results for DIBELS 8 Fifth Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
Composite	20th	1	95	0.78	0.66-0.90	0.708	0.789
		2	133	0.88	0.81-0.95	0.762	0.857
		3	109	0.86	0.76-0.97	0.714	0.832
	40th	1	95	0.78	0.68-0.87	0.822	0.600
		2	133	0.89	0.83-0.94	0.864	0.742
		3	109	0.88	0.82-0.95	0.821	0.741

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Table 3.37 ROC Curve Results for DIBELS 8 Sixth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	46	0.84	0.71-0.97	0.714	0.821
		2	153	0.90	0.84-0.95	0.789	0.843
		3	156	0.83	0.75-0.91	0.684	0.873
	40th	1	46	0.71	0.56-0.86	0.650	0.654
		2	153	0.85	0.79-0.91	0.698	0.856
		3	156	0.78	0.71-0.85	0.556	0.763
ORF-ACC	20th	1	46	0.90	0.79-1.00	0.857	0.718
		2	153	0.88	0.83-0.94	1.000	0.409
		3	156	0.70	0.60-0.79	0.711	0.602
	40th	1	46	0.68	0.52-0.84	0.950	0.231
		2	NA	NA	NA	NA	NA
		3	NA	NA	NA	NA	NA
Maze	20th	1	101	0.94	0.86-1.00	1.000	0.617
		2	103	0.97	0.94-1.00	1.000	0.723
		3	104	0.93	0.85-1.00	0.889	0.674
	40th	1	101	0.85	0.77-0.94	1.000	0.329
		2	103	0.87	0.79-0.95	0.963	0.592
		3	104	0.88	0.80-0.96	0.963	0.597

Table 3.37 ROC Curve Results for DIBELS 8 Sixth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
Composite	20th	1	44	0.84	0.69-0.99	0.833	0.684
		2	102	0.89	0.79-0.99	0.778	0.817
		3	96	0.83	0.70-0.96	0.625	0.739
	40th	1	44	0.69	0.53-0.86	0.842	0.520
		2	102	0.84	0.76-0.92	0.926	0.667
		3	96	0.77	0.67-0.86	0.833	0.667

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this period. *Criterion measure was administered concurrently at the beginning of the year instead of at end of year where indicated.

Table 3.38 ROC Curve Results for DIBELS 8 Seventh Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	37	0.92	0.81-1.00	0.909	0.962
		2	155	0.87	0.81-0.93	0.791	0.795
		3	155	0.88	0.81-0.94	0.738	0.876
	40th	1	37	0.88	0.77-0.99	0.789	0.778
		2	155	0.87	0.81-0.92	0.836	0.723
		3	155	0.85	0.79-0.91	0.738	0.787
ORF-ACC	20th	1	37	0.88	0.76-1.00	1.000	0.346
		2	155	0.82	0.74-0.89	0.930	0.491
		3	155	0.75	0.66-0.84	0.762	0.593
	40th	1	NA	NA	NA	NA	NA
		2	NA	NA	NA	NA	NA
		3	NA	NA	NA	NA	NA
Maze	20th	1	93	0.93	0.87-0.99	1.000	0.807
		2	93	0.97	0.93-1.00	1.000	0.741
		3	95	0.90	0.83-0.98	0.923	0.768
	40th	1	93	0.89	0.82-0.96	1.000	0.500
		2	93	0.92	0.87-0.98	1.000	0.629
		3	95	0.89	0.81-0.96	1.000	0.352

Table 3.38 ROC Curve Results for DIBELS 8 Seventh Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	*N*	AUC	AUC CI	Sensitivity	Specificity
Composite	20th	1	36	0.93	0.82-1.00	0.900	0.962
		2	92	0.92	0.85-0.99	0.750	0.850
		3	94	0.90	0.78-1.00	0.769	0.864
	40th	1	36	0.88	0.77-0.99	0.778	0.833
		2	92	0.92	0.86-0.98	0.957	0.797
		3	94	0.87	0.79-0.96	0.792	0.829

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Table 3.39 ROC Curve Results for DIBELS 8 Eighth Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	*N*	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	45	0.89	0.79-0.98	0.810	0.792
		2	105	0.83	0.75-0.91	0.864	0.738
		3	112	0.78	0.69-0.87	0.750	0.625
	40th	1	45	0.83	0.69-0.98	0.781	0.769
		2	105	0.76	0.67-0.86	0.776	0.632
		3	112	0.70	0.60-0.81	0.750	0.550
ORF-ACC	20th	1	45	0.69	0.52-0.85	0.905	0.083
		2	105	0.75	0.66-0.85	0.955	0.098
		3	112	0.73	0.64-0.83	0.958	0.125
	40th	1	NA	NA	NA	NA	NA
		2	NA	NA	NA	NA	NA
		3	NA	NA	NA	NA	NA
Maze	20th	1	46	0.84	0.71-0.96	0.909	0.625
		2	50	0.85	0.75-0.96	0.917	0.654
		3	52	0.84	0.73-0.94	0.833	0.643
	40th	1	46	0.77	0.61-0.93	0.909	0.231
		2	50	0.75	0.58-0.91	0.972	0.143
		3	52	0.77	0.63-0.92	0.973	0.267

Table 3.39 ROC Curve Results for DIBELS 8 Eighth Grade Subtests Predicting Iowa Total Reading

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
Composite	20th	1	45	0.89	0.79-0.98	0.810	0.833
		2	49	0.86	0.75-0.96	0.696	0.808
		3	52	0.82	0.71-0.94	0.708	0.786
	40th	1	45	0.83	0.69-0.98	0.781	0.769
		2	49	0.81	0.67-0.95	0.771	0.714
		3	52	0.69	0.52-0.85	0.676	0.600

Note. Criteria were percentile ranks on end-of-year administration of criterion measure. Data is drawn from Sample B (i.e., the 2018-2019 CTL norming study). 1 = Beginning of year. 2 = Middle of year. 3 = End of year. NA = No cut score provided for this subtest in this period.

Slope validity for progress monitoring. We also estimated model-based validity estimates for the DIBELS 8 subtests recommended for use in frequent progress monitoring: PSF, NWF, WRF, and ORF. To calculate model-based validity estimates for progress monitoring performance levels and slopes, we created hierarchical linear models of student performance on each progress monitoring measure using data from Sample B (i.e., the 2018-2019 DDS study).

In this study, students were progress-monitored up to 23 times over the course of the academic year with a minimum of 20 weeks elapsed. Test administrations typically took place every 2 weeks in kindergarten and first grade and every 3 weeks in all other grades. To be included in the analysis, students needed to have complete fall benchmark data on the word-reading measures and a minimum of three observed scores over the academic year. For the estimates of slope validity, an additional requirement was that student needed to show evidence of need of intensive intervention. To obtain subsamples of students requiring intensive intervention, the highest performing students on each measure were removed until the sample mean for each measure fell below the 25th percentile according to national norms. Progress monitoring scores were used to create growth models represented by the general equations:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1jt ij} + R_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{0j}$$

$$\beta_{0j} = \gamma_{10} + \gamma_{1j}$$

with

$$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N \begin{pmatrix} 0 & \tau^2_{00} & \tau_{01} \\ 0 & \tau_{01} & \tau^2_{00} \end{pmatrix}$$

and

$$R_{ij} \sim N(0, \sigma^2)$$

where Y equals a given DIBELS measure, β equals the number of days from the benchmark assessment, and γ represents across-group parameters. The group in these models is the student. Growth models were analyzed in R using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) with the maximum likelihood estimator and the Nelder Mead optimizer. For all Kindergarten measures and Grade 1 ORF and NWF-WRC, low variance in the effect of time led to singular fit warnings. These models were therefore rerun with blme package (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013), which uses an approximate Bayesian maximum a posteriori estimate to avoid singularity. Differences in fit indices and estimates between models were negligible. Results from the blme model are reported for these models.

We then correlated the slopes with Iowa Total Reading and Iowa Word Analysis scores at the end of the year. For PSF, we also correlated slopes with end of year DIBELS 8 NWF, WRF, and when possible, ORF-WRC scores and concurrent slopes. For NWF, we correlated CLS and WRC scores with end of year WRF and ORF-WRC scores and concurrent slopes.

One critical detail in the study's design is that students were progress monitored with subtests that were not necessarily aligned to the ones on which they demonstrated the most risk, nor necessarily to the intervention they received. In fact, students who were progress monitored may not have received intervention at all. These design features ensured a standardized approach to collecting data, but do not represent good practice well. As such, the validity coefficients reported here represent a lower bound to slope validity.

Slope validity estimates represent how well change over time within a growth model predicts student performance on the indicated criterion measures administered at the end of the year. Results for PSF were best in kindergarten and inadequate in first grade (see Table 3.40). Given that students in kindergarten are much more likely to need phonological awareness intervention than are those in first grade, and that the demands of the end of year first grade assessments are not well-aligned to phonological awareness, this pattern of results is not surprising. Results for NWF-CLS are inadequate in kindergarten, but quite good in first through third grade (see Table 3.41). For NWF-WRC, however, the results are stronger in kindergarten and remain quite good in Grades 1 to 3 (see Table 3.42). One hypothesis for the

discrepancy between the NWF CLS and WRC slope validity in kindergarten is that NWF-WRC is better aligned to how the Iowa assessment measures reading and word analysis skills. Results for WRF and ORF-WRC are both quite good (see Tables 3.43 and 3.44 respectively), indicating moderate to strong relationships between progress as measured by these subtests and end of year performance. In sum, student progress as measured using the suite of DIBELS 8 measures demonstrates moderate to good validity for predicting student performance at the end of the year.

Table 3.40 Predictive Validity for the Slope of DIBELS 8th Edition Phoneme Segmentation Fluency with End of Year Measures

Grade	Criterion	*N*	*r*	CI
Kindergarten	Iowa Total Reading	32	.05	-.31–.39
	Iowa Word Analysis	32	.14	-.22–.46
	DIBELS 8th NWF-CLS	70	.45	.25–.62
	DIBELS 8th NWF-WRC	70	.48	.27–.64
	DIBELS 8th WRF	70	.43	.22–.60
First	Iowa Total Reading	79	.04	-.19–.26
	Iowa Word Analysis	62	-.25	-.47–.00
	DIBELS 8th NWF-CLS	150	.02	-.14–.18
	DIBELS 8th NWF-WRC	150	.01	-.15–.17
	DIBELS 8th WRF	151	.03	-.14–.18
	DIBELS 8th ORF-WRC	151	.03	-.13–.19

Table 3.41 Predictive Validity for the Slope of DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds with End of Year Measures

Grade	Criterion	N	r	CI
Kindergarten	Iowa Total Reading	33	.22	-.14–.52
	Iowa Word Analysis	34	.24	-.11–.53
	DIBELS 8th PSF	127	.25	.08–.40
	DIBELS 8th NWF-WRC	127	.47	.32–.60
	DIBELS 8th WRF	127	.66	.54–.74
First	Iowa Total Reading	87	.42	.21–.59
	Iowa Word Analysis	73	.42	.20–.59
	DIBELS 8th PSF	162	.20	.05–.35
	DIBELS 8th NWF-WRC	161	.53	.41–.63
	DIBELS 8th WRF	162	.69	.60–.76
	DIBELS 8th ORF-WRC	162	.68	.59–.76
Second	Iowa Total Reading	93	.48	.31–.62
	Iowa Word Analysis	70	.52	.33–.67
	DIBELS 8th NWF-WRC	240	.71	.64–.76
	DIBELS 8th WRF	241	.69	.62–.75
	DIBELS 8th ORF-WRC	241	.62	.53–.69
	DIBELS 8th MAZE	177	.50	.38–.60
Third	Iowa Total Reading	52	.29	.02–.52
	Iowa Word Analysis	38	.10	-.22–.41
	DIBELS 8th NWF-WRC	199	.55	.44–.64
	DIBELS 8th WRF	200	.56	.46–.65
	DIBELS 8th ORF-WRC	200	.51	.40–.61
	DIBELS 8th MAZE	147	.25	.09–.40

Table 3.42 Predictive Validity for the Slope of DIBELS 8th Edition Nonsense Word Fluency-Words Recoded Correctly with End of Year Measures

Grade	Criterion	N	r	CI
Kindergarten	Iowa Total Reading	73	.53	.35–.68
	Iowa Word Analysis	75	.24	.02–.44
	DIBELS 8th PSF	206	.37	.25–.49
	DIBELS 8th NWF-CLS	206	.57	.47–.66
	DIBELS 8th WRF	206	.58	.54–.70
First	Iowa Total Reading	73	.42	.21–.60
	Iowa Word Analysis	61	.32	.08–.53
	DIBELS 8th PSF	145	.31	.15–.45
	DIBELS 8th NWF-CLS	144	.60	.49–.70
	DIBELS 8th WRF	145	.59	.47–.68
	DIBELS 8th ORF-WRC	162	.55	.42–.65
	Iowa Total Reading	98	.55	.40–.68
	Iowa Word Analysis	74	.55	.36–.69
Second	DIBELS 8th NWF-CLS	266	.71	.64–.76
	DIBELS 8th WRF	267	.68	.61–.74
	DIBELS 8th ORF-WRC	267	.62	.54–.69
	DIBELS 8th MAZE	197	.51	.40–.61
	Iowa Total Reading	67	.26	.02–.47
	Iowa Word Analysis	44	.11	-.19–.39
Third	DIBELS 8th NWF-CLS	231	.55	.44–.64
	DIBELS 8th WRF	247	.56	.46–.65
	DIBELS 8th ORF-WRC	200	.51	.40–.61
	DIBELS 8th MAZE	181	.25	.09–.40
	DIBELS 8th ORF-WRC	200	.51	.40–.61
	DIBELS 8th MAZE	147	.25	.09–.40

Table 3.43 Predictive Validity for the Slope of DIBELS 8th Edition Word Reading Fluency with End of Year Measures

Grade	Criterion	N	r	CI
Kindergarten	Iowa Total Reading	69	.66	.50–.77
	Iowa Word Analysis	71	.32	.10–.52
	DIBELS 8th PSF	226	.35	.23–.46
	DIBELS 8th NWF-CLS	226	.71	.64–.77
	DIBELS 8th NWF-WRC	226	.67	.60–.74
First	Iowa Total Reading	81	.62	.46–.74
	Iowa Word Analysis	70	.43	.22–.61
	DIBELS 8th PSF	224	.24	.11–.36
	DIBELS 8th NWF-CLS	220	.69	.61–.75
	DIBELS 8th NWF-WRC	220	.65	.56–.72
	DIBELS 8th ORF-WRC	223	.82	.78–.86
Second	Iowa Total Reading	85	.65	.51–.76
	Iowa Word Analysis	66	.59	.41–.73
	DIBELS 8th NWF-CLS	224	.54	.45–.64
	DIBELS 8th NWF-WRC	224	.52	.41–.61
	DIBELS 8th ORF-WRC	225	.74	.67–.79
	DIBELS 8th MAZE	166	.52	.40–.62
Third	Iowa Total Reading	66	.43	.21–.61
	Iowa Word Analysis	46	.29	.00–.54
	DIBELS 8th NWF-CLS	213	.64	.56–.71
	DIBELS 8th NWF-WRC	247	.62	.54–.69
	DIBELS 8th ORF-WRC	210	.65	.58–.72
	DIBELS 8th MAZE	191	.48	.37–.59

Table 3.44 Predictive Validity for the Slope of DIBELS 8th Edition Oral Reading Fluency with End of Year Measures

Grade	Criterion	N	r	CI
First	Iowa Total Reading	56	.54	.33–.71
	Iowa Word Analysis	53	.45	.20–.64
	DIBELS 8th PSF	129	.39	.23–.52
	DIBELS 8th NWF-CLS	128	.72	.63–.80
	DIBELS 8th NWF-WRC	128	.66	.55–.75
	DIBELS 8th WRF	129	.88	.83–.91

Table 3.44 Predictive Validity for the Slope of DIBELS 8th Edition Oral Reading Fluency with End of Year Measures

Grade	Criterion	N	r	CI
Second	Iowa Total Reading	82	.74	.62–.82
	Iowa Word Analysis	63	.61	.43–.75
	DIBELS 8th NWF-CLS	187	.69	.60–.76
	DIBELS 8th NWF-WRC	187	.69	.61–.76
	DIBELS 8th WRF	188	.83	.78–.87
	DIBELS 8th MAZE	186	.67	.58–74
Third	Iowa Total Reading	68	.38	.16–.57
	Iowa Word Analysis	46	.20	-.09–.47
	DIBELS 8th NWF-CLS	174	.71	.64–.77
	DIBELS 8th NWF-WRC	200	.71	.63–.77
	DIBELS 8th WRF	210	.88	.85–.91
	DIBELS 8th MAZE	204	.67	.59–.74
Fourth	Iowa Total Reading	77	.27	.04–.46
	DIBELS 8th MAZE	235	.72	.65–.78
Fifth	Iowa Total Reading	67	.56	.37–.70
	DIBELS 8th MAZE	194	.83	.78–.87
Sixth	Iowa Total Reading	21	-.39	-.70– .05
	DIBELS 8th MAZE	77	.29	.07–.48
Seventh	Iowa Total Reading	24	-.19	-.55–.24
	DIBELS 8th MAZE	38	.57	.30–.75
Eighth	Iowa Total Reading	37	-.04	-.36–.29
	DIBELS 8th MAZE	51	-.07	-.34–.21

Summary

This chapter presents and summarizes validity evidence for DIBELS 8, including concurrent and predictive validity comparing DIBELS 8 subtests to each other, and to a range of external criterion measures of reading achievement and phonological processing. The chapter also includes evidence of screening accuracy and slope validity for progress monitoring. Taken together, the validity evidence for DIBELS 8 is strong. The strongest evidence regards its primary use, which is as a screener for students at risk for reading difficulties, including reading disabilities like dyslexia. Research into the valid interpretations and uses of DIBELS scores is ongoing, and regular addendums to this manual will continue to build the validity argument for DIBELS 8.

References

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2015.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. Psychometrika, 78(4), 685-709.

Geverdt, D. (2015). Education Demographic and Geographic Estimates Program (EDGE): Locale Boundaries User's Manual (NCES 2016-012). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved January 11, 2019 from http://nces.ed.gov/pubsearch.

Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: The case of tests with continuous results. Biochemia Medica, 26(3), 297–307. doi: 10.11613/BM.2016.034

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. School Psychology Review, 36(4), 582-600.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741–749. doi: 10.1037/0003-066X.50.9.741

Paris, S. G. (2005). Reinterpreting the development of reading skills. Reading Research Quarterly, 40, 184–202.

Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. New York: Oxford.

Welch, C., & Dunbar, S. (2012). Iowa assessments. Boston, MA: Houghton Mifflin Harcourt.