Dynamic Indicators of Basic Early Literacy Skills

8th Edition

2018-19 Preliminary Goals

Technical Manual

University of Oregon (2018-2019). 8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS®): 2018-19 Preliminary Goals Technical Manual. Eugene, OR: University of Oregon. Available: https://dibels.uoregon.edu

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

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

David Larsen, M.A.
Senior Research Assistant II and DIBELS Data System Customer Support Manager
Center on Teaching & Learning
College of Education, University of Oregon

University of Oregon would also like to acknowledge and thank our partner Amplify Education for their work on copy editing and formatting.

Table of Contents

Introduction

This technical manual is for the DIBELS 8th Edition preliminary goals used in the 2018-2019 school year. Note that we also have published an updated technical manual that should be used for the 2019-2020 school year forward. The preliminary technical manual reflects the evidence for the 2018-2019 preliminary goals and is provided as historical documentation of the DIBELS 8 evidence base.

In this manual, you will find information about the DIBELS 8 approach to norming, as well as sections on the reliability and validity of DIBELS 8 as it stood for the 2018-2019 school year.

While the majority of the evidence is replicated in the 2019-2020 technical manual, the newer manual includes additional evidence for almost every type of evidence, new evidence regarding progress monitoring, and updated evidence for the finalized cut-scores regarding their classification accuracy. DIBELS users should reference the 2019-2020 technical manual for the 2019-2020 school year and beyond.

Chapter 1: DIBELS 8th Edition Preliminary Goals Normative Information

This chapter describes normative information regarding the DIBELS 8th Edition preliminary goals used during the 2018-19 school year, including sample recruitment and selection procedures used in DIBELS 8 research studies, and demographic characteristics of the research sample.

Sample Recruitment and Selection Procedures

We 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. We communicated information about the project, including participation requirements and incentives, to potential participating schools via a flyer, email, or by phone. Regardless of format, schools received a description of the study and participation options. We asked interested schools for contact and other basic information by one of two methods: using a revised Qualtrics survey, via a link in the email, on the website, or in the pop-up notice; or over the phone. We then sent an email confirmation of enrollment to the designated contact person.

We recruited schools until we met or exceed our recruitment goals, as determined by a minimum of 200 students per grade in each administration of the Iowa Assessments, and 100 struggling students per grade, 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 did exceed this amount for some grade levels in order 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 run K-5 and 6-8 or K-6 and 7-9. To achieve our minimum for transitional grade levels, we ended up yielding larger sample sizes in the other grade levels.

All children enrolled in participating schools, including students with disabilities and students who are English language learners were included, as long as they received English language reading instruction in general education classrooms and had the response capabilities to participate. At their discretion, schools could opt to not assess students with disabilities who require assessment modifications.

Description of the DIBELS 8 Preliminary Goals Research Sample

The DIBELS 8th Edition research sample consisted of 4,453 students in grades K – 8 from 29 schools (see Table 1.1 for distribution of students across grades). Schools ranged in size from very small (n = 7) to large (n = 666) and were located throughout the country. All four census regions and six of nine census divisions were represented (U.S. Department of Commerce, 1994). See Table 1.2 for breakdown of students and schools by census region and division.

Using the National Center for Education Statistics definitions, half of the sample lived in rural areas (22% fringe, 22% distant, and 6% remote); approximately a quarter lived in towns (13% fringe and 9% remote); and one fifth of the sample lived in large, suburban areas. The remainder of the sample lived in cities (3% large and 4% small) (Geverdt, 2015).

Demographic characteristics of the sample are displayed in Table 1.4: 48.1% of the sample was female, 50.9% was male. 18.1% of students were Hispanic. The sample included 0.6% Asian students, 14.3% Black/African American students, 0.4% Native Hawaiian or other Pacific Islander, 3.9% American Indian or Alaskan Native, 64.3% white, 3.2% two or more races, and 13.3% unknown or not reported. 6.3% of students were English Learners, and 13.9% were eligible for Special Education services. 57.4% of the students were eligible for the free or reduced lunch program.

Table 1.1 Number of Students by Grade

Grade	*n*
K	685
1	782
2	725
3	783
4	408
5	388
6	276
7	216
8	190

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

Region Division	Students *n*	Schools *n*
West		
Pacific	1771	15
Mountain	910	6
Midwest		
West North Central	156	1
East North Central	609	4
Northeast		
Middle Atlantic	846	2
New England	0	0
South		
West South Central	0	0
East South Central	0	0
South Atlantic	161	0

Table 1.3 Number of Students Schools by NCES Locale Classification

Locale Classification	Students *n*	Schools *n*
City		
Large	114	1
Midsize	0	0
Small	196	2
Suburb		
Large	913	4
Midsize	0	0
Small	0	0
Town		
Fringe	572	2
Distant	0	0
Remote	397	2
Rural		
Fringe	988	5
Distant	1002	7
Remote	271	2

Table 1.4 Demographic Characteristics of Sample

	n	Percent
Gender		
Female	2142	48.1
Male	2265	50.9
Unknown/not reported	46	1.0
Ethnicity		
Hispanic	804	18.1
Non-Hispanic	3649	81.9
Race		
Asian	26	0.6
Black/African American	636	14.3
Native Hawaiian/Pacific Islander	16	0.4
American Indian/Alaskan Native	173	3.9
White	2863	64.3
Two or more races	144	3.2
Unknown/not reported	594	13.3
Free/reduced lunch program eligible		
Yes	2557	57.4
No	1805	40.5
Unknown/not reported	92	2.1

Table 1.4 Demographic Characteristics of Sample

English-Lanuage Learners		
Yes	279	6.3
No	3979	89.4
Unknown/not reported	195	4.4
Special Education Eligibility		
Yes	619	13.9
No	3754	84.3
Unknown/not reported	81	1.8

Chapter 2: Reliability of DIBELS 8th Edition Preliminary Goals

Reliability refers to the extent to which a test score is a consistent and stable measure. Reliability is a necessary, but insufficient component of validity. Three forms of test reliability have been examined for DIBELS 8th Edition: concurrent alternate form reliability, delayed alternate form reliability, and test-retest reliability. All three forms can be thought of as estimates of the stability of scores, but both delayed alternate form and test-retest reliability address the stability of scores over time. We have provided individual coefficients in tables by subtest for every correlation calculated. We also provided median coefficients by subtest per grade in most instances, and median coefficients across grades by subtest are always provided. 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. First, concurrent and delayed alternate form reliability evidence is provided, followed by test-retest reliability evidence. We then report standard errors of measurement before providing a summary of 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. As illustrated in Table 2.1, results of concurrent reliability studies over two years reveal very strong correlations in all but one case, and the latter case still reveals moderately strong reliability. 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 and we examined each for concurrent alternate form reliability. As reported in Table 2.2 and 2.3, the results of all studies for both types of scores are highly reliable. For correct letter sounds on NWF, the median reliability was .89 or above in all grades, with an overall median reliability of .91. For words recoded correctly on NWF, the median reliability was .86 or above in all grades, and overall median reliability was .90.

Results for WRF are displayed in Table 2.4. Once again, concurrent alternate form reliability is very strong. The median reliability for WRF in all grades was greater than .94. Overall median reliability was .95.

Results for ORF and ORF-ACC are displayed in Table 2.5 and 2.6. As with NWF scores and WRF, concurrent alternate form reliability is very strong. The median reliability for ORF was .92 or above in all grades, and overall median reliability was .93. For accuracy on ORF, the median reliability ranged from .73 to .92 across grades, and overall median reliability was .79.

Maze concurrent alternate form reliability was somewhat weaker than other DIBELS subtests (see Table 2.7). Correlations ranged from .66 to .93. Medians were not calculated by grade due to the presence of a maximum of two studies per grade, but the overall median for Maze was .71. The 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
Kindergarten	2017–18	1:PM9	93	.80	.72–.87
		2:PM8	153	.88	.83–.91
		3:PM5	119	.80	.72–.85
		3:PM1	105	.86	.80–.90
	2018–19	1:PM7	221	.90	.87–.92
	Median			.86	.80–.90
First	2018–19	1:PM5	128	.81	.74–.86
		1:PM6	148	.53	.40–.64
		2:PM6	149	.81	.74–.86
	Median			.81	.74–.86
Median				.81	.74–.86

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:PM2	92	.95	.93-.97
		2:PM2	109	.92	.89-.95
		3:PM2	113	.93	.89-.95
		3:PM7	51	.97	.94-.98
		3:PM8	40	.90	.82-.95
	2018-19	1:PM10	63	.92	.88-.95
	Median			.92	.89-.95
First	2017-18	1:PM1	123	.95	.93-.96
		2:PM1	126	.89	.84-.92
		3:PM1	127	.85	.79-.89
	Median			.89	.84-.92
Second	2017-18	2:PM1	50	.94	.89-.97
		3:PM1	52	.93	.87-.96
	2018-19	1:PM9	73	.91	.87-.95
		1:PM10	227	.86	.82-.89
		1:PM12	138	.92	.89-.94
		2:PM3	135	.88	.84-.91
		2:PM8	69	.94	.90-.96
		2:PM13	255	.90	.88-.92
		2:PM14	81	.91	.86-.94
	Median			.91	.87-.95

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

Grade	Sample	Forms	N	r	CI
Third	2017-18	1:2	45	.75	.59-.86
		1:3	47	.92	.86-.95
	2018-19	1:PM10	161	.87	.83-.90
		1:PM13	66	.87	.79-.92
		1:PM14	205	.92	.89-.94
		2:PM4	164	.90	.87-.93
		2:PM6	69	.94	.91-.96
		2:PM9	225	.90	.87-.92
		2:PM15	68	.90	.85-.94
	Median			.90	.87-.92
Median				.91	.87-.95

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:PM2	87	.95	.92–.96
		2:PM2	107	.86	.80–.90
		3:PM2	112	.89	.84–.92
		3:PM7	51	.91	.84–.95
		3:PM8	40	.84	.72–.91
	2018-19	1:PM10	49	.83	.72–.90
	Median			.86	.80–.90
First	2017-18	1:PM1	123	.90	.86–.93
		2:PM1	126	.90	.85–.93
		3:PM1	127	.86	.81–.90
	Median			.90	.85–.93
Second	2017-18	2:PM1	50	.97	.95–.98
		3:PM1	52	.93	.88–.96
	2018-19	1:PM9	73	.90	.85–.94
		1:PM10	225	.91	.89–.93
		1:PM12	137	.92	.89–.94
		2:PM3	65	.88	.81–.92
	Median			.91	.89–.93

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

Grade	Sample	Forms	N	r	CI
Third	2017-18	2:1	45	.82	.70–.90
		3:1	47	.95	.90–.97
	2018-19	1:PM10	161	.88	.84–.91
		1:PM13	65	.87	.80–.92
		1:PM14	192	.90	.87–.93
		2:PM4	74	.90	.85–.94
		2:PM9	75	.86	.79–.91
	Median			.88	.84–.91
Median				.90	.85–.94

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	3:PM4	87	.97	.96–.98
	2018-19	1:PM6	127	.96	.95–.97
		2: PM7	50	.96	.94–.98
		2: PM11	179	.87	.83–.90
	Median			.96	.95–.97

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

Grade	Sample	Forms	N	r	CI
First	2017-18	2:PM3	59	.97	.96–.98
		3:PM3	59	.96	.93–.97
	2018-19	1: PM3	201	.96	.95–.97
		1: PM18	103	.95	.92–.96
		2: PM3	236	.97	.97–.98
		2: PM19	69	.97	.95–.98
	Median			.96	.95–.97
Second	2017-18	2:PM7	68	.95	.92–.97
		3:PM4	118	.92	.88–.94
	2018-19	1: PM7	139	.94	.92–.96
		1: PM13	71	.96	.94–.98
		1: PM15	230	.96	.94–.97
		2: PM12	69	.94	.91–.96
		2: PM16	82	.95	.92–.97
		2: PM17	136	.95	.93–.97
		2: PM20	241	.94	.93–.96
	Median			.95	.92–.97

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

Grade	Sample	Forms	N	r	CI
Third	2017-18	1:PM9	48	.93	.87–.96
		2:PM9	146	.96	.94–.97
		3:PM9	50	.95	.91–.97
	2018-19	1:PM9	367	.94	.92–.95
		1:PM19	66	.94	.91–.97
		2:PM12	216	.95	.94–.96
		2:PM14	68	.90	.85–.94
		2:PM17	163	.94	.92–.96
		2:PM20	71	.94	.91–.96
	Median			.94	.92–.96
Median				.95	.92–.97

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

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

Grade	Sample	Forms	N	r	CI
First	2017-18	1:PM5	128	.97	.96–.98
		2:PM7	184	.96	.95–.97
		3:PM8	186	.94	.92–.95
	Median			.96	.95–.97

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

Grade	Sample	Forms	N	r	CI
Second	2017-18	1: 3	118	.95	.92–.96
		1:PM5	109	.97	.96–.98
		1:PM7	108	.97	.95–.98
		2:PM2	159	.96	.94–.97
		2:PM6	159	.96	.95–.97
		3:PM3	118	.95	.92–.96
	2018-19	1:PM16	365	.95	.93–.96
		2:PM15	375	.96	.96–.97
	Median			.96	.94–.97
Third	2017-18	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
	2018-19	1:PM14	366	.94	.92–.95
		2:PM10	377	.94	.93–.95
	Median			.93	.91–.94

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

Grade	Sample	Forms	N	r	CI
	2017-18	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
Fourth		3:PM2	144	.92	.89–.94
		3:PM8	143	.88	.84–.91
	2018-19	1:PM12	440	.94	.93–.95
		1:PM20	339	.94	.93–.95
		2:PM18	356	.94	.93–.95
		2:PM20	506	.94	.93–.95
	Median			.93	.91–.95

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

Grade	Sample	Forms	N	r	CI
Fifth	2017-18	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
	2018-19	1:PM14	327	.93	.91–.94
		1:PM15	387	.91	.89–.92
		2:PM16	393	.89	.87–.91
		2:PM19	510	.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	2017-18	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
	2018-19	1:PM13	182	.92	.90–.94
		1:PM16	166	.92	.89–.94
		2:PM7	290	.94	.92–.95
		2:PM14	164	.91	.88–.93
	Median			.93	.90–.95

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

Grade	Sample	Forms	N	r	CI
	2017-18	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
Seventh		2:PM2	141	.91	.88–.94
		3:PM3	123	.89	.84–.92
	2018-19	1:PM13	63	.91	.86–.95
		1:PM19	77	.95	.92–.97
		2:PM12	199	.95	.93–.96
		2:PM18	63	.93	.89–.95
	Median			.92	.88–.94

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

Grade	Sample	Forms	N	r	CI
Eighth	2017-18	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
	2018-19	1:PM13	66	.95	.92–.97
		1:PM16	74	.95	.92–.97
		2:PM6	134	.92	.89–.94
		2:PM10	61	.94	.91–.97
	Median			.92	.87–.95
Median				.93	.90–.95

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

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

Grade	Sample	Forms	N	r	CI
First	2017-18	1:PM5	128	.92	.89–.94
		2:PM7	184	.90	.87–.93
		2:PM1	75	.92	.86–.95
		3:PM1	44	.93	.87–.96
		3:PM8	186	.91	.88–.93
	Median			.92	.87–.94
Second	2017-18	1:3	118	.77	.69–.84
		1:PM5	109	.91	.87–.94
		1:PM7	108	.89	.85–.93
		2:3	118	.83	.77–.88
		3:PM3	118	.78	.70–.84
	2018-19	1:PM16	363	.85	.82–.88
		2:PM15	66	.96	.95–.98
	Median			.85	.77–.88

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

Grade	Sample	Forms	N	r	CI
Third	2017-18	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
	2018-19	1:PM14	366	.79	.75–.82
		2:PM10	148	.81	.74–.86
	Median			.79	.74–.85
Fourth	2017-18	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
	2018-19	1:PM12	440	.89	.87–.91
		2:PM18	203	.88	.84–.91
		2:PM20	247	.93	.89–.93
	Median			.76	.68–.82

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

Grade	Sample	Forms	N	r	CI
Fifth	2017-18	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
	2018-19	1:PM14	327	.75	.70–.80
		1:PM15	387	.85	.82–.87
		2:PM16	70	.57	.38–.71
		2:PM19	82	.69	.56–.79
	Median			.73	.64–.80

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

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	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
	2018-19	1:PM13	182	.97	.96–.98
		1:PM16	166	.96	.94–.97
		2:PM7	89	.66	.52–.76
	Median			.78	.69–.85

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

Grade	Sample	Forms	N	r	CI
Seventh	2017-18	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
	2018-19	1:PM13	63	.68	.51–.79
		1:PM19	77	.98	.97–.99
		2:PM12	92	.93	.90–.96
	Median			.89	.83–.93

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

Grade	Sample	Forms	N	r	CI
Eighth	2017-18	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
	2018-19	1:PM13	66	.90	.85–.94
		1:PM16	74	.88	.82–.92
		2:PM6	90	.81	.72–.87
	Median			.84	.76–.89
Median				.79	.74–.85

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers.

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

Grade	Forms	N	r	CI
Second	1:2	66	.82	.72-.89
	1:3	80	.80	.70-.87
Third	1:2	64	.71	.57-.82
	1:3	64	.66	.49-.77
Fourth	1:2	74	.71	.58-.81
	1:3	71	.71	.58-.81
Fifth	1:2	63	.82	.71-.89
	1:3	47	.71	.53-.83
Sixth	1:2	72	.83	.73-.89
Seventh	1:2	69	.93	.89-.96
Median			.71	.58-.81

Note. Form numbers correspond to benchmark period forms. All coefficients come from the 2018-2019 sample, and data was gathered within a six-week period in the fall of 2018.

Delayed alternate form reliability. We studied all DIBELS 8th Edition subtests except Maze for delayed alternate form reliability. Delayed alternate form reliability was established by correlating benchmark forms; thus, the delay was about 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 .83 in kindergarten, .76 in first grade, and .81 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 .54 across the two grades. For the NWF correct letter sounds score, delayed alternate form reliability was .76 across grades and was .71 in kindergarten, .79 in first grade, .83 in second grade, and .74 in third grade (see Table 2.10). For the NWF words recoded correctly score, delayed

alternate form reliability was .72 across grades and was .60 in kindergarten, .69 in first grade, .81 in second grade, and .78 in third grade (see Table 2.11). WRF coefficients are located in Table 2.12 and demonstrated strong delayed alternate form reliability of .90 overall and .85 in kindergarten, .88 in first grade, .91 in second grade, and .90 in third grade (see Table 2.12). ORF coefficients (see Table 2.13) demonstrated strong delayed alternate form reliability of .89 overall and .92 in first grade, .89 in second grade, .88 in third grade, .82 in fourth grade, .86 in fifth grade, .91 in sixth grade, .86 in seventh grade, and .74 in eighth grade. For the ORF words accuracy score, delayed alternate form reliability was .78 in first grade, .78 in second grade, .75 in third grade, .82 in fourth grade, .86 in fifth grade, .77 in sixth grade, .90 in seventh grade, and .73 in eighth grade. Across grades, median delayed alternate form reliability for ORF accuracy was .78 (see Table 2.14).

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017–18	1:2	150	.86	.82–.90
		1:3	100	.83	.76–.88
		2:3	137	.87	.81–.90
	2018–19	1:2	419	.81	.77–.84
	Median			.83	.76–.88
First	2017–18	1:2	153	.77	.70–.83
		1:3	140	.70	.61–.78
		2:3	183	.76	.70–.82
	2018–19	1:2	413	.83	.80–.86
	Median			.76	.70–.82
Median				.81	.77–.84

Note. Form numbers correspond to benchmark periods, and data was gathered in the periods indicated.

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017–18	1:2	94	.49	.32–.63
		1:3	72	.39	.18–.57
		2:3	115	.70	.59–.78
	2018–19	1:2	347	.56	.48–.63
	Median			.49	.32–.63
First	2017–18	1:2	148	.53	.40–.64
		1:3	138	.54	.41–.65
		2:3	181	.63	.53–.71
	2018–19	1:2	411	.64	.58–.69
	Median			.54	.41–.65
Median				.54	.41–.65

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

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

Grade	Sample	Forms	N	r	CI
	2017-18	1:2	191	.73	.65-.79
		1:3	180	.63	.53-.71
Kindergarten		2:3	224	.83	.78-.87
	2018-19	1:2	192	.71	.63-.78
	Median			.71	.63-.78
	2017-18	1:2	268	.80	.75-.84
		1:3	256	.73	.67-.78
First		2:3	308	.79	.75-.83
	2018-19	1:2	404	.82	.79-.85
	Median			.79	.75-.83
	2017-18	1:2	127	.83	.77-.88
		1:3	113	.84	.77-.88
Second		2:3	194	.84	.79-.87
	2018-19	1:2	420	.74	.70-.78
	Median			.83	.77-.88
	2017-18	2:3	218	.76	.70-.81
Third	2018-19	1:2	412	.74	.69-.78
	Median			.74	.69-.78
Median				.76	.70-.81

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:2	171	.70	.62–.77
		1:3	161	.60	.49–.69
		2:3	219	.77	.71–.82
	2018-19	1:2	77	.45	.25–.61
	Median			.60	.49–.69
First	2017-18	1:2	268	.72	.66–.77
		1:3	256	.67	.59–.73
		2:3	308	.74	.68–.78
	2018-19	1:2	233	.69	.62–.75
	Median			.69	.62–.75
Second	2017-18	1:2	126	.83	.76–.87
		1:3	112	.81	.74–.87
		2:3	194	.83	.78–.87
	2018-19	1:2	105	.72	.61–.80
	Median			.81	.74–.87
Third	2017-18	2:3	218	.78	.72–.83
	2018-19	1:2	173	.83	.78–.87
	Median			.78	.72–.83
Median				.72	.66–.77

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

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

Grade	Sample	Forms	N	r	CI
Kindergarten	2017-18	1:2	112	.88	.83–.92
		1:3	86	.85	.77–.90
		2:3	150	.90	.86–.92
	2018-19	1:2	357	.81	.77–.84
	Median			.85	.77–.90
First	2017-18	1:2	153	.91	.88–.93
		1:3	141	.82	.75–.86
		2:3	241	.88	.85–.91
	2018-19	1:2	410	.92	.91–.93
	Median			.88	.85–.91
Second	2017-18	1:2	163	.91	.88–.94
	2018-19	1:2	407	.92	.88–.95
	Median			.91	.88–.94
Third	2017-18	1:2	198	.90	.88–.93
		1:3	138	.90	.86–.93
		2:3	216	.90	.87–.92
	2018-19	1:2	405	.89	.87–.91
	Median			.90	.88–.93
Median				.90	.88–.93

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

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

Grade	Sample	Forms	N	r	CI
First	2017-18	1:2	161	.94	.91–.95
		1:3	117	.87	.82–.91
		2:3	218	.92	.89–.94
	2018-19	1:2	399	.94	.93–.95
	Median			.92	.89–.94
Second	2017-18	1:2	225	.89	.86–.92
		1:3	116	.85	.79–.89
		2:3	116	.92	.89–.95
	2018-19	1:2	401	.92	.90–.93
	Median			.89	.86–.92
Third	2017-18	1:2	112	.91	.87–.94
		2:3	171	.84	.79–.88
	2018-19	1:2	403	.88	.86–.90
	Median			.88	.86–.90
Fourth	2017-18	1:2	142	.82	.76–.87
	2018-19	1:2	391	.90	.88–.92
	Median			.82	.76–.87
Fifth	2017-18	1:2	83	.89	.83–.92
		1:3	128	.86	.81–.90
	2018-19	1:2	378	.81	.77–.84
	Median			.86	.81–.90

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

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	1:2	66	.86	.78–.91
		1:3	61	.91	.85–.94
		2:3	98	.91	.87–.94
	2018-19	1:2	173	.89	.86–.92
	Median			.91	.85–.94
Seventh	2017-18	1:2	79	.89	.83–.93
		1:3	65	.84	.74–.90
		2:3	127	.86	.81–.90
	2018-19	1:2	75	.93	.89–.95
	Median			.86	.81–.90
Eighth	2017-18	1:2	67	.92	.89–.95
		1:3	67	.73	.62–.81
		2:3	100	.74	.64–.82
	2018-19	1:2	68	.94	.90–.96
	Median			.74	.64–.82
Median				.89	.83–.92

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

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

Grade	Sample	Forms	N	r	CI
First	2017-18	1:2	161	.89	.85–.92
		1:3	117	.71	.61–.79
		2:3	218	.78	.72–.82
	2018-19	1:2	227	.84	.80–.88
	Median			.78	.72–.82
Second	2017-18	1:2	225	.78	.73–.83
		1:3	116	.61	.48–.71
		2:3	116	.80	.72–.86
	2018-19	1:2	100	.83	.76–.88
	Median			.78	.72–.83
Third	2017-18	1:2	112	.75	.66–.82
		2:3	171	.65	.55–.73
	2018-19	1:2	176	.81	.76–.86
	Median			.75	.66–.82
Fourth	2017-18	1:2	142	.82	.76–.87
	2018-19	1:2	240	.82	.77–.86
	Median			.82	.76–.86
Fifth	2017-18	1:2	83	.89	.83–.92
		1:3	128	.86	.81–.90
	2018-19	1:2	76	.51	.32–.66
	Median			.86	.81–.90

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

Grade	Sample	Forms	N	r	CI
Sixth	2017-18	1:2	66	.89	.83–.93
		1:3	61	.77	.64–.86
		2:3	98	.85	.78–.89
	Median			.77	.78–.89
Seventh	2017-18	1:2	79	.90	.85–.93
		1:3	65	.96	.94–.98
		2:3	127	.83	.77–.88
	Median			.90	.85–.93
Eighth	2017-18	1:2	67	.73	.59–.82
		1:3	67	.67	.51–.78
		2:3	100	.75	.65–.83
	Median			.73	.59–.82
Median				.78	.72–.83

Note. Form numbers without a prefix correspond to benchmark periods, while prefixed numbers refer to progress monitoring passage numbers. Benchmark forms were administered during typical benchmark periods.

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. 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 like to 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). More generally, one should not expect levels of test-retest reliability to be as high as alternate form reliability when the skill measured develops rapidly and time between measurement occasions is sufficient for genuine growth to have occurred.

We studied all DIBELS 8th Edition subtests except PSF and Maze for test-retest reliability. Test-retest reliability was evaluated by administering the same form in two different benchmark periods. Thus, as with delayed alternate form reliability, about three months passed between administrations. As a result of the long delay between administrations, test-retest reliability coefficients are again expected to be lower than concurrent alternate form reliability coefficients due to instructional effects.

For LNF, median test-retest reliability was .82 in kindergarten, .75 in first grade, and .77 overall (see Table 2.15). For the NWF correct letter sounds score, median test-retest reliability was .79 in kindergarten, .81 in first grade, .75 in second grade, and .77 overall; no median is reported for third grade because only one coefficient (.87) was available (see Table 2.16). For the NWF words recoded correctly score, median test-retest reliability was .72 in kindergarten, .78 in first grade, and .72 in second grade and overall; no median is reported for third grade because only one coefficient (.84) was available (see Table 2.17). For WRF, median test-retest reliability was strong, at .92 in kindergarten, .90 in first grade, and .92 overall; no median is reported for second or third grade because only two and one coefficients were available respectively, although all were .88 or above (see Table 2.18). For the ORF words read correctly score, median test-retest reliability was also strong, at .91 across grades; median test-retest reliability by grade was .92 for first, .87 for second, .94 for third, .90 for fourth, .91 for fifth, .91 for sixth, .86 for seventh, and .92 for eighth (see Table 2.19). For the ORF accuracy, median test-retest reliability was adequate at .75 across grades; median test-retest reliability by grade was .74 for first, .75 for second, .80 for third, .75 for fourth, .79 for fifth, .74 for sixth, .90 for seventh, and .83 for eighth (see Table 2.20).

Table 2.15 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

Table 2.15 Test–Retest Reliability Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Benchmark period	N	r	CI
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. All data comes from the 2017–18 sample and was gathered in the benchmark periods indicated.

Table 2.16 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

Table 2.16 Test-Retest Reliability Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Benchmark period	N	r	CI
Third	2:3	158	.87	.82 - .90
Median			.77	.67 - .83

Note. The same form was given in the benchmark periods indicated.

Table 2.17 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 given in the benchmark periods indicated.

Table 2.18 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		.90	.86–.93
Second	2:3	82	.95	.93–.97
Third	1:2	48	.93	.88–.96
	2:3	90	.88	.82–.92
Median			.92	.88–.94

Note. The same form was given in the benchmark periods indicated.

Table 2.19 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

Table 2.19 Test-Retest Reliability Coefficients for DIBELS 8th Edition Oral Reading Fluency-Words Read Correctly

Grade	Benchmark period	N	r	CI
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 given in the benchmark periods indicated.

Table 2.20 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

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

Grade	Benchmark period	N	r	CI
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 given in the benchmark periods indicated.

Standard Error of Measurement Finally, 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 measure in each grade and benchmark period is reported in Table 2.21. In all cases except for LNF, we used the median concurrent alternate form reliability for a grade and the standard deviation (SD) for each benchmark period in these calculations. Because concurrent alternate form reliability was not available for LNF, we used delayed alternate form reliability in this case. By definition, measures with the best reliability have the smallest SEMs relative to each measure's SD.

Table 2.21 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Subtest	Beginning	Middle	End
Kindergarten	LNF	8.42	9.06	8.56
	PSF	6.01	6.47	6.29
	NWF-CLS	5.88	6.60	7.40
	NWF-WRC	2.59	3.06	3.52
	WRF	1.58	2.31	2.89
First	LNF	8.64	9.22	8.86
	PSF	7.58	6.26	6.76
	NWF-CLS	10.55	11.57	14.33
	NWF-WRC	3.96	4.64	5.28
	WRF	3.75	4.52	5.06
	ORF	6.06	6.88	8.43
	ORF-ACC	0.10	0.08	0.06
Second	NWF-CLS	12.96	14.51	13.59
	NWF-WRC	4.29	5.27	4.60
	WRF	5.10	5.48	5.90
	ORF	7.84	8.85	9.12
	ORF-ACC	0.05	0.06	0.02
	Maze	2.88	TBD	TBD

Table 2.21 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Subtest	Beginning	Middle	End
Third	NWF-CLS	11.68	15.72	17.53
	NWF-WRC	3.96	5.62	5.89
	WRF	5.24	5.69	5.58
	ORF	10.39	10.24	9.59
	ORF-ACC	0.05	0.04	0.02
	Maze	3.74	TBD	TBD
Fourth	ORF	9.63	12.86	10.08
	ORF-ACC	0.03	0.01	0.03
	Maze	4.49	TBD	TBD
Fifth	ORF	11.31	11.01	11.23
	ORF-ACC	0.04	0.05	0.02
	Maze	4.46	TBD	TBD
Sixth	ORF	11.00	9.82	12.87
	ORF-ACC	0.02	0.02	0.02
	Maze	3.46	TBD	TBD
Seventh	ORF	10.16	10.06	9.83
	ORF-ACC	0.02	0.02	0.02
	Maze	3.06	TBD	TBD

Table 2.21 Standard Errors of Measurement for DIBELS 8th Edition by Grade, Subtest, and Time of Year

Grade	Subtest	Beginning	Middle	End
Eighth	ORF	8.08	9.16	10.23
	ORF-ACC	0.02	0.02	0.01
	Maze	TBD	TBD	TBD

Note. SEMs for Maze middle and end of year benchmarks will be available in summer 2019.

Summary

Taken together, the reliability evidence for DIBELS 8 is strong. The strongest evidence regards concurrent alternate form reliability. Research into the reliability of DIBELS 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 Preliminary Goals

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 discuss concurrent and predictive validity.

Concurrent validity is generally seen as a means of validating that the intended construct is being captured by a measure. Concurrent validity is evaluated by correlating each DIBELS 8 subtest with like DIBELS Next subtests, with related external criterion measures, and with the other DIBELS 8 subtests.

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 screening accuracy is the more valuable evidence that a measure is functioning as intended. Both are evaluated for DIBELS 8.

Concurrent Validity

The more similar the two measures given, the higher the correlation between the scores should be. Thus, when correlating like measures (e.g., DIBELS Next and DIBELS 8 NWF), correlations should be quite strong (i.e., .8 or above). 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 Next and DIBELS 8 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 more variable, which may be due to the different approaches to item ordering of 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 (see Table 3.4). The strongest correlations are in kindergarten and second grade. The lowest is still quite strong at .66 in fall of first grade, and correlations strengthen over the course of the first grade year. 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 subtest in DIBELS Next, but ORF is particularly good, with nearly every correlation being .90 or above (see Table 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.

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.

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.

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.

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.

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
Sixth	2	.91	30	.82–.96

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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 predominant subtests for the grade and period of the assessment (so long as they were not covered in the previous section), DIBELS Next composite scores, various raw scores from the Comprehensive Test of Phonological Processing (CTOPP), and Total Reading and Word Analysis scores from the Iowa Assessment. Correlations with other DIBELS 8 subtests are reported in the following section.

Correlations for each DIBELS 8 subtest are reported in Table 3.7-3.14. Where available, concurrent validity with the DIBELS Next composite scores was quite good. Correlations among unrelated DIBELS 8 and DIBELS Next subtests varied in strength as would be predicted based on the similarity of the component skills assessed. Correlations with some of the CTOPP are negative, but negative correlations are sensible for RAN measures because CTOPP scores are times while DIBELS scores are rates (i.e., items per minute), and time and rate have a naturally inverse relationship with each other. CTOPP composite scores were not correlated with DIBELS 8 scores because CTOPP composite scores require using age-level norms, which eliminates some variability in scores and obscures the relationship between performance on specific component skills measured by both CTOPP and DIBELS.

Of particular interest are the moderate to strong correlations between CTOPP rapid naming measures and LNF depending on the CTOPP raw score examined. As might be expected, relations are strongest for rapid letter naming ($r = -.56$ and $-.65$ in kindergarten and first grade respectively). Relations are weaker for other naming scores, suggesting that while LNF can operate as a decent screener for processing speed, it cannot replace more diagnostic measures like CTOPP.

Similarly, PSF is moderately to strongly related to CTOPP phonological awareness raw scores. Relations here are more uniform. These results again suggest the validity of DIBELS as a screener for phonological awareness difficulties, but suggest that it cannot replace more diagnostic measures like CTOPP.

Finally, concurrent relations with the Iowa Assessment scores vary predictably by subtest. PSF showed the weakest relationships with the Iowa total reading and word analysis scores in 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 rest of the DIBELS 8 subtests displayed moderate to strong relations with external criterion measures regardless of grade.

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

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next Composite	.70	128	.60–.78
		CTOPP2 Rapid Digit Naming Raw Score	-.52	124	-.64–.38
		CTOPP2 Rapid Letter Naming Raw Score	-.56	64	-.71–.36
		CTOPP2 Rapid Color Naming Raw Score	-.27	160	-.41–.12
		CTOPP2 Rapid Object Naming Raw Score	-.31	170	-.44–.17
	2	DIBELS Next NWF-CLS	.72	181	.64–.78
		DIBELS Next Composite	.80	156	.74–.85
	3	DIBELS Next NWF-CLS	.78	128	.70–.84
		DIBELS Next Composite	.89	98	. 84–.93

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

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF-CLS	.49	157	.36–.60
		DIBELS Next Composite	.65	144	.54–.74
		CTOPP2 Rapid Digit Naming Raw Score	-.62	174	-.71–.52
		CTOPP2 Rapid Letter Naming Raw Score	-.65	165	-.73–.56
		CTOPP2 Rapid Color Naming Raw Score	-.47	171	-.58–.34
		CTOPP2 Rapid Object Naming Raw Score	-.48	171	-.59–.36
	2	DIBELS Next ORF-WRC	.61	196	.52–.69
		DIBELS Next Composite	.70	163	.61–.77
	3	DIBELS Next ORF-WRC	.64	188	.55–.72
		DIBELS Next Composite	.63	163	.53–.72
		Iowa Total Reading	.54	117	.40–.66
		Iowa Word Analysis	.46	119	.31–.61

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.19	99	-.01–.37
		DIBELS Next Composite	.47	75	.27–.63
		CTOPP2 Elision Raw Score	.58	189	.48–.67
		CTOPP2 Blending Raw Score	.45	188	.32–.55
		CTOPP2 Sound Matching Raw Score	.53	187	.41–.62
	3	DIBELS Next NWF-CLS	.32	125	.16–.47
		DIBELS Next Composite	.62	95	.48- .73
First	1	DIBELS Next NWF-CLS	.09	154	-.07–.25
		DIBELS Next Composite	.27	141	.11–.41
		CTOPP2 Elision Raw Score	.40	179	.27–.52
		CTOPP2 Blending Raw Score	.35	179	.21–.47
		CTOPP2 Sound Matching Raw Score	.39	179	.26–.51
	2	DIBELS Next ORF-WRC	.15	194	.01–.28
		DIBELS Next Composite	.17	161	.01–.31
	3	DIBELS Next ORF-WRC	.15	183	.01–.29
		DIBELS Next Composite	.14	163	-.01–.29
		Iowa Total Reading	.12	117	-.06–.30
		Iowa Word Analysis	.14	119	-.05–.31

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.34	116	.16–.49
		DIBELS Next Composite	.36	92	.26–.52
	2	DIBELS Next Composite	.68	133	.57–.76
	3	DIBELS Next Composite	.82	109	.75–.87
		Iowa Total Reading	.65	113	.53–.75
		Iowa Word Analysis	.43	96	.26–.58
First	1	DIBELS Next Composite	.71	166	.62–.78
	2	DIBELS Next ORF	.79	223	.74–.84
		DIBELS Next Composite	.80	185	.74–.85
	3	DIBELS Next ORF	.78	212	.72–.82
		DIBELS Next Composite	.85	186	.81–.89
		Iowa Total Reading	.65	198	.56–.73
		Iowa Word Analysis	.54	200	.43–.63
Second	1	DIBELS Next ORF	.80	129	.72–.85
		DIBELS Next Composite	.80	115	.72–.85
	2	DIBELS Next ORF	.75	211	.69–.81
		DIBELS Next Composite	.62	107	.49–.72
	3	DIBELS Next ORF	.76	201	.70–.82
		DIBELS Next Composite	.74	112	.65–.82
		Iowa Total Reading	.62	84	.47–.74
		Iowa Word Analysis	.60	51	.39–.75

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

Grade	Period	Criterion	r	N	CI
Third	2	DIBELS Next ORF	.72	229	.65–.78
		DIBELS Next Composite	.71	109	.61–.80
	3	DIBELS Next ORF	.70	224	.63–.76
		DIBELS Next Composite	.69	99	.57–.78
		Iowa Total Reading	.50	90	.33–.64
		Iowa Word Analysis	.34	72	.12–.53

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.60	100	.46–.72
		DIBELS Next Composite	.60	76	.43–.73
	2	DIBELS Next NWF-CLS	.80	160	.53–.85
		DIBELS Next Composite	.66	130	.55–.75
	3	DIBELS Next NWF–CLS	.80	138	.73–.85
		DIBELS Next Composite	.74	108	.65–.82
		Iowa Total Reading	.65	112	.53–.74
		Iowa Word Analysis	.35	95	.16–.52

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

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF–CLS	.66	180	.57–.73
		DIBELS Next Composite	.61	163	.50–.70
	2	DIBELS Next NWF–CLS	.78	223	.72–.82
		DIBELS Next ORF	.76	223	.70–.81
		DIBELS Next Composite	.79	185	.73–.84
	3	DIBELS Next NWF–CLS	.87	206	.84–.90
		DIBELS Next ORF	.78	207	.72–.83
		DIBELS Next Composite	.86	186	.81–.89
		Iowa Total Reading	.63	198	.54–.71
		Iowa Word Analysis	.56	200	.45–.65
Second	1	DIBELS Next NWF–CLS	.81	128	.74–.86
		DIBELS Next ORF	.77	128	.69–.84
		DIBELS Next Composite	.79	115	.71–.85
	2	DIBELS Next ORF	.76	211	.70–.81
		DIBELS Next Composite	.70	107	.58–.78
	3	DIBELS Next ORF	.80	198	.74–.84
		DIBELS Next Composite	.74	112	.64–.82
		Iowa Total Reading	.60	84	.45–.72
		Iowa Word Analysis	.62	51	.42–.76

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

Grade	Period	Criterion	r	N	CI
Third	2	DIBELS Next ORF	.71	239	.64–.77
		DIBELS Next Composite	.74	109	.65–.82
	3	DIBELS Next ORF	.71	216	.64–.77
		DIBELS Next Composite	.73	97	.62–.81
		Iowa Total Reading	.51	90	.34–.65
		Iowa Word Analysis	.36	72	.14–.54

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.63	113	.51–.73
		DIBELS Next Composite	.57	89	.41–.70
	2	DIBELS Next NWF-CLS	.75	181	.68–.81
		DIBELS Next Composite	.63	124	.52–.73
	3	DIBELS Next NWF-CLS	.80	150	.74–.85
		DIBELS Next Composite	.75	92	.64–.82

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

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF-CLS	.74	157	.66–.80
		DIBELS Next Composite	.69	144	.60–.77
	2	DIBELS Next ORF-WRC	.91	249	.89–.93
		DIBELS Next Composite	.88	163	.85–.91
	3	DIBELS Next ORF-WRC	.90	239	.87–.92
		DIBELS Next Composite	.88	163	.85–.91
		Iowa Total Reading	.79	117	.71–.85
		Iowa Word Analysis	.67	119	.56–.76
Second	1	DIBELS Next ORF-WRC	.91	166	.87–.93
		DIBELS Next Composite	.91	151	.88–.94
	2	DIBELS Next ORF-WRC	.92	192	.89–.94
		DIBELS Next Composite	.87	138	.83–.91
	3	Iowa Total Reading	.62	87	.47–.74
		Iowa Word Analysis	.60	89	.44–.72

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

Grade	Period	Criterion	r	N	CI
Third	1	DIBELS Next ORF-WRC	.83	201	.78–.87
		DIBELS Next Composite	.85	138	.79–.89
	2	DIBELS Next ORF-WRC	.85	228	.81–.88
		DIBELS Next Composite	.85	154	.80–.89
	3	DIBELS Next ORF-WRC	.83	177	.78–.87
		DIBELS Next Composite	.84	97	.77–.89
		Iowa Total Reading	.56	90	.40–.69
		Iowa Word Analysis	.32	72	.09–.51

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF	.78	166	.72–.84
		DIBELS Next Composite	.75	154	.67–.81
	2	DIBELS Next Composite	.91	196	.88–.93
	3	DIBELS Next Composite	.91	163	.88–.94
		Iowa Total Reading	.82	116	.75–.87
		Iowa Word Analysis	.67	118	.55–.76

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

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next Composite	.84	131	.78–.88
	2	DIBELS Next Composite	.87	130	.82–.90
	3	Iowa Total Reading	.71	87	.59–.80
		Iowa Word Analysis	.60	89	.45–.72
Third	1	DIBELS Next Composite	.89	73	.83–.93
	2	DIBELS Next Composite	.83	128	.77–.88
	3	DIBELS Next Composite	.83	96	.75–.88
		Iowa Total Reading	.58	90	.42–.70
		Iowa Word Analysis	.24	72	.01–.45
Fourth	3	Iowa Total Reading	.61	91	.47–.73
Fifth	1	DIBELS Next Composite	.94	41	.90–.97
	3	Iowa Total Reading	.65	59	.48–.78
Sixth	3	Iowa Total Reading	.67	82	.52–.77
Seventh	3	Iowa Total Reading	.54	91	.38–.67
Eighth	3	Iowa Total Reading	.59	77	.42–.72

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next NWF	.67	166	.58–.75
		DIBELS Next Composite	.76	154	.69–.82
	2	DIBELS Next ORF	.69	227	.61–.75
		DIBELS Next Composite	.78	196	.72–.83
	3	DIBELS Next ORF	.66	188	.58–.74
		DIBELS Next Composite	.76	163	.68–.82
		Iowa Total Reading	.61	116	.48–.71
		Iowa Word Analysis	.60	118	.47–.71
Second	1	DIBELS Next ORF	.62	173	.51–.71
		DIBELS Next Composite	.63	131	.52–.73
	2	DIBELS Next ORF	.62	192	.52–.70
		DIBELS Next Composite	.68	130	.57–.76
	3	Iowa Total Reading	.48	87	.30–.62
Third	1	DIBELS Next ORF	.65	100	.52–.75
		DIBELS Next Composite	.68	73	.53–.78
	2	DIBELS Next ORF	.58	209	.48–.66
		DIBELS Next Composite	.68	128	.58–.76
	3	DIBELS Next ORF	.39	176	.25–.51
		DIBELS Next Composite	.55	96	.39–.67
		Iowa Total Reading	.36	90	.17–.53

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

Grade	Period	Criterion	r	N	CI
Fourth	3	DIBELS Next ORF	.57	58	.37–.72
		Iowa Total Reading	.37	91	.17–.53
Fifth	1	DIBELS Next ORF	.69	51	.52–.81
		DIBELS Next Composite	.73	41	.54–.85
	2	DIBELS Next ORF	.55	84	.38–.68
	3	Iowa Total Reading	.22	59	-.04–.45
Sixth	2	DIBELS Next ORF	.74	40	.56–.85
		Iowa Total Reading	.53	86	.36–.67
	3	Iowa Total Reading	.49	82	.31–.64
Seventh	3	Iowa Total Reading	.29	91	.09–.47
Eighth	3	Iowa Total Reading	.43	77	.22–.59

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.14 Concurrent Validity Coefficients for DIBELS 8th Edition Maze with Fall Administration of Iowa Assessment Total Reading

Grade	Benchmark form	r	N	CI
Second	1	.77	126	.69–.83
	2	.74	94	.63–.82
	3	.69	60	.53–.80
Third	1	.59	123	.46–.70
	2	.66	100	.53–.76
	3	.66	65	.49–.78
Fourth	1	.76	135	.67–.82
	2	.77	97	.67–.84
	3	.67	65	.51–.79
Fifth	1	.84	137	.72–.84
	2	.81	95	.62–.81
	3	.66	69	.31–.66
Sixth	1	.54	59	.33–.70
	2	.65	57	.47–.78
Seventh	1	.77	70	.65–.85
	2	.75	55	.61–.85

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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. Relationships should generally be positive and at least moderate, with more closely related subtests (e.g., WRF and ORF) exhibiting the strongest relationships. At the same time, correlations should not be perfect (i.e., 1.0), as that would indicate redundancy between the subtests. Nonetheless, correlations look at the relationships among all test-takers and can obscure the added value for screening purposes of two subtests that are highly related on average. Thus, even though some subtests may be correlated at .9 or above, differences in predictive validity and screening results can suggest apparent redundancy is not problematic for screening purposes.

Correlations among DIBELS 8 subtests are reported by grade. In kindergarten, correlations are all positive, with the two NWF scores showing the strongest relationship with each other and WRF (see Table 3.15). LNF is most strongly related to NWF-CLS and moderately related to the other subtests. PSF 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 involve reading. In first grade, all DIBELS 8 subtests are strongly correlated except for PSF (see Table 3.16). 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 (see Table 3.17-3.19). In these later grades, ORF Accuracy shows the weakest relations, but the correlations are still strong. The two NWF scores demonstrate the strongest relationship in second and third grade, most likely due to the fact that the scores come from the same subtest and that students increasingly read without sounding out in these grades. ORF and WRF are also strongly related in second and third grade. Maze and ORF Accuracy have moderate to strong relationships with ORF but relatively weaker relations with each other.

Table 3.15 Correlations among DIBELS 8th Edition Subtests in Fall for Kindergarten

	NWF-CLS	NWF-WRC	PSF	WRF
LNF				
N	304	268	569	575
r	.61	.29	.42	.40
CI	.53–.68	.18–.40	.35–.48	.33–.47

Table 3.15 Correlations among DIBELS 8th Edition Subtests in Fall for Kindergarten

	NWF-CLS	NWF-WRC	PSF	WRF
NWF-CLS				
N		268	280	297
r		.79	.38	.71
CI		.74–.83	.27–.48	.65–.76
NWF-WRC				
N			253	263
r			.15	.75
CI			.03–.27	.69–.80
PSF				
N				536
r				.26
CI				.17–.33

Table 3.16 Correlations among DIBELS 8th Edition Subtests in Fall for First Grade

	NWF-CLS	NWF-WRC	ORF	ORF-ACC	PSF	WRF
LNF						
N	594	591	592	592	605	604
r	.67	.59	.53	.61	.32	.61
CI	.63–.72	.53–.64	.47–.59	.55–.65	.25–.39	.55–.65

Table 3.16 Correlations among DIBELS 8th Edition Subtests in Fall for First Grade

	NWF-CLS	NWF-WRC	ORF	ORF-ACC	PSF	WRF
NWF-CLS						
N		591	581	581	591	592
r		.90	.77	.69	.23	.83
CI		.88–.91	.74–.80	.65–.73	.15–.30	.80–.85
NWF-WRC						
N			579	579	588	589
r			.71	.61	.21	.78
CI			.67–.75	.56–.66	.13–.19	.75–.81
ORF						
N				592	589	592
r				.77	.08	.91
CI				.73–.80	.00–.16	.90–.93
ORF–ACC						
N					589	592
r					.27	.75
CI					.19–.34	.71–.78
PSF						
N						601
r						.18
CI						.10–.25

Table 3.17 Correlations among DIBELS 8th Edition Subtests in Fall for Second Grade

	NWF-WRC	ORF	ORF-ACC	WRF	Maze
NWF–CLS					
N	467	463	461	469	469
r	.94	.75	.54	.77	.62
CI	.93–.95	.73–.80	.47–.60	.73–.81	.56–.67
NWF–WRC					
N		461	460	467	468
r		.75	.53	.76	.59
CI		.71–.79	.46–.59	.71–.79	.53–.65
ORF					
N			461	463	463
r			.69	.92	.73
CI			.64–.79	.91–.94	.68–.77
ORF–ACC					
N				461	461
r				.70	.48
CI				.65–.74	.40–.54
WRF					
N					469
r					.71
CI					.67–.76

Table 3.18 Correlations among DIBELS 8th Edition Subtests in Fall for Third Grade

	NWF-WRC	ORF	ORF-ACC	WRF	Maze
NWF–CLS					
N	440	450	450	451	495
r	.96	.73	.49	.72	.50
CI	.95-.96	.68-.77	.41-.56	.67-.76	.43-.56
NWF–WRC					
N		439	439	440	484
r		.76	.56	.78	.56
CI		.71-.79	.49-.62	.75-.82	.49-.62
ORF					
N			450	450	494
r			.67	.89	.69
CI			.61-.72	.87-.91	.64-.74
ORF–ACC					
N				450	494
r				.70	.54
CI				.65-.74	.47-.60
WRF					
N					495
r					.68
CI					.63-.73

Table 3.19 Correlations among DIBELS 8th Edition Subtests in Fall for Fourth, Fifth, Sixth, Seventh, and Eighth Grade

Grade		ORF-ACC	Maze
Fourth	**ORF**		
	N	532	506
	r	.70	.69
	CI	.65-.74	.65-.74
	ORF–ACC		
	N		506
	r		.50
	CI		.43-.56
Fifth	**ORF**		
	N	447	419
	r	.54	.44
	CI	.47-.60	.36-.51
	ORF–ACC		
	N		419
	r		.39
	CI		.30-.47

Table 3.19 Correlations among DIBELS 8th Edition Subtests in Fall for Fourth, Fifth, Sixth, Seventh, and Eighth Grade

Grade		ORF-ACC	Maze
Sixth	**ORF**		
	N	197	191
	r	.59	.69
	CI	.49-.67	.61-.76
	ORF–ACC		
	N		191
	r		.35
	CI		.22-.47
Seventh	**ORF**		
	N	89	85
	r	.54	.75
	CI	.38-.68	.64-.83
	ORF–ACC		
	N		85
	r		.37
	CI		.17-.54

Table 3.19 Correlations among DIBELS 8th Edition Subtests in Fall for Fourth, Fifth, Sixth, Seventh, and Eighth Grade

Grade		ORF-ACC	Maze
Eighth	**ORF**		
	N	82	76
	r	.78	.81
	CI	.67-.85	.71-.87
	ORF–ACC		
	N		76
	r		.53
	CI		.35-.68

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. DIBELS 8 was evaluated through both methods and results are presented in this section.

Predictive correlations. DIBELS 8th Edition subtests were correlated with end of year administrations of DIBELS Next and the Iowa Assessment. The only exceptions are for NWF-CLS and NWF-WRC in the beginning of third grade and Maze in all grades. DIBELS Next related and predominant measure scores and the DIBELS Next composite score were used as criterion measures, as were Iowa Total Reading and Word Analysis scores. Results are presented in tables by subtest, grade, and the benchmark period in which DIBELS 8 was administered.

Correlations for each DIBELS 8 subtest are reported in Table 3.20-3.26 and are only reported when the study sample size was 40 or greater to ensure a minimum threshold of precision in the correlation estimate. As with external criterion concurrent validity, where available, predictive validity with the DIBELS Next composite scores was quite good.

Correlations among unrelated DIBELS 8 and DIBELS Next subtests varied in strength as would be predicted based on the similarity of the component skills assessed. Finally, concurrent relations with the Iowa Assessment scores also vary predictably by subtest. PSF showed the weakest relationships with the Iowa total reading and word analysis scores in 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 rest of the DIBELS 8 subtests displayed moderate to strong relations with external criterion measures regardless of grade.

Table 3.20 Predictive Validity Coefficients for DIBELS 8th Edition Letter Naming Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.79	149	.72–.84
		DIBELS Next NWF-CLS	.71	149	.62-.78
		DIBELS Next Composite	.78	123	.71–.84
	2	DIBELS Next LNF	.84	183	.79–.88
		DIBELS Next NWF-CLS	.72	183	.64–.78
		DIBELS Next Composite	.82	154	.76–.86
First	1	DIBELS Next ORF-WRC	.63	148	.52–.72
		DIBELS Next Composite	.64	132	.53–.73
		Iowa Total Reading	.57	80	.40–.70
		Iowa Word Analysis	.57	80	.40–.70
	2	DIBELS Next ORF-WRC	.71	191	.63–.77
		DIBELS Next Composite	.70	170	.62–.77
		Iowa Total Reading	.57	115	.43–.68
		Iowa Word Analysis	.52	117	.37–.64

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.21 Predictive Validity Coefficients for DIBELS 8th Edition Phoneme Segmentation Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next PSF	.37	97	.18–.53
		DIBELS Next NWF-CLS	.24	97	.05–.42
		DIBELS Next Composite	.44	72	.23–.61
	2	DIBELS Next PSF	.75	29	.52–.88
		DIBELS Next NWF-CLS	.10	29	-.27–.45
		DIBELS Next Composite	.57	29	.26–.78
First	1	DIBELS Next ORF-WRC	.12	145	.00–.24
		DIBELS Next Composite	.11	129	-.02–.23
		Iowa Total Reading	.12	79	-.05–.28
		Iowa Word Analysis	.02	79	-.14–.19
	2	DIBELS Next ORF-WRC	.21	189	.07–.34
		DIBELS Next Composite	.23	168	.05–.31
		Iowa Total Reading	.29	113	.12–.45
		Iowa Word Analysis	.23	115	.08–.36

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.22 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next NWF-CLS	.45	113	.29–.59
		DIBELS Next Composite	.43	89	.24–.59
		Iowa Total Reading	.54	78	.36–.68
		Iowa Word Analysis	.29	80	.08–.48
	2	DIBELS Next NWF-CLS	.79	157	.72–.84
		DIBELS Next Composite	.72	130	.63–.80
		Iowa Total Reading	.54	111	.39–.66
		Iowa Word Analysis	.38	95	.19–.54
First	1	DIBELS Next NWF-CLS	.79	169	.73–.84
		DIBELS Next ORF	.76	167	.68–.81
		DIBELS Next Composite	.76	154	.69–.82
		Iowa Total Reading	.55	153	.43–.65
		Iowa Word Analysis	.51	153	.38–.62
	2	DIBELS Next NWF-CLS	.82	213	.77–.86
		DIBELS Next ORF	.77	214	.71–.82
		DIBELS Next Composite	.76	194	.70–.82
		Iowa Total Reading	.60	197	.50–.68
		Iowa Word Analysis	.49	199	.38–.59

Table 3.22 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Correct Letter Sounds

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.77	123	.69–.83
		DIBELS Next Composite	.72	97	.61–.80
		Iowa Total Reading	.66	49	.47–.79
	2	DIBELS Next ORF	.75	194	.68–.81
		DIBELS Next Composite	.67	117	.55–.76
		Iowa Total Reading	.56	76	.38–.70
		Iowa Word Analysis	.47	49	.22–.66
Third	3	DIBELS Next ORF	.66	222	.58–.73
		DIBELS Next Composite	.66	107	.54–.76
		Iowa Total Reading	.39	89	.19–.55
		Iowa Word Analysis	.27	72	.05–.48

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.23 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly with End of Year Measures

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next NWF-CLS	.65	97	.52–.75
		DIBELS Next Composite	.71	73	.58–.81
		Iowa Total Reading	.47	75	.27–.63
		Iowa Word Analysis	.18	77	-.05–.39
	2	DIBELS Next NWF-CLS	.74	153	.66–.80
		DIBELS Next Composite	.71	126	.61–.79
		Iowa Total Reading	.55	106	.40–.67
		Iowa Word Analysis	.26	90	.06–.44
First	1	DIBELS Next NWF-CLS	.71	163	.63–.78
		DIBELS Next ORF	.69	163	.59–.76
		DIBELS Next Composite	.70	163	.61–.77
		Iowa Total Reading	.51	153	.39–.62
		Iowa Word Analysis	.50	153	.37–61
	2	DIBELS Next NWF-CLS	.76	213	.70–.81
		DIBELS Next ORF	.75	214	.68–.80
		DIBELS Next Composite	.75	185	.68–.80
		Iowa Total Reading	.58	197	.47–.66
		Iowa Word Analysis	.51	199	.40–.61

Table 3.23 Predictive Validity Coefficients for DIBELS 8th Edition Nonsense Word Fluency-Words Read Correctly with End of Year Measures

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.76	123	.67–.83
		DIBELS Next Composite	.71	97	.60–.85
		Iowa Total Reading	.64	49	.44–.78
	2	DIBELS Next ORF	.76	194	.69–.81
		DIBELS Next Composite	.70	117	.58–.78
		Iowa Total Reading	.60	76	.43–.73
		Iowa Word Analysis	.56	49	.33–.73
Third	2	DIBELS Next ORF	.66	222	.58–.73
		DIBELS Next Composite	.69	107	.65–.82
		Iowa Total Reading	.41	89	.22–.57
		Iowa Word Analysis	.31	72	.09–.51

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.24 Predictive Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
Kindergarten	1	DIBELS Next LNF	.72	112	.61–.80
		DIBELS Next Composite	.65	86	.50–.76
	2	DIBELS Next NWF-CLS	.71	177	.63–.78
		DIBELS Next Composite	.67	121	.56–.76
First	1	DIBELS Next NWF-CLS	.84	148	.79–.88
		DIBELS Next Composite	.78	132	.71–.84
		Iowa Total Reading	.65	80	.50–.76
		Iowa Word Analysis	.59	80	.43–.72
	2	DIBELS Next ORF-WRC	.87	239	.84–.90
		DIBELS Next Composite	.85	170	.80–.89
		Iowa Total Reading	.74	115	.64–.81
		Iowa Word Analysis	.63	117	.50–.73
Second	1	DIBELS Next ORF-WRC	.88	156	.84–.91
		DIBELS Next Composite	.87	130	.82–.91
		Iowa Total Reading	.78	51	.64–.87
		Iowa Word Analysis	.66	36	.42–.81
	2	DIBELS Next ORF-WRC	.90	189	.87–.92
		DIBELS Next Composite	.87	150	.82–.90
		Iowa Total Reading	.72	77	.59–.81
		Iowa Word Analysis	.69	49	.50–.81

Table 3.24 Predictive Validity Coefficients for DIBELS 8th Edition Word Reading Fluency

Grade	Period	Criterion	r	N	CI
Third	1	DIBELS Next ORF-WRC	.82	192	.77–.87
		DIBELS Next Composite	.84	131	.78–.88
		Iowa Total Reading	.61	71	.44–.74
		Iowa Word Analysis	.47	53	.23–.66
	2	DIBELS Next ORF-WRC	.82	228	.77–.86
		DIBELS Next Composite	.82	154	.77–.87
		Iowa Total Reading	.59	89	.43–.71
		Iowa Word Analysis	.46	72	.25–.62

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.25 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next ORF	.86	156	.82–.90
		DIBELS Next Composite	.82	141	.76–.87
		Iowa Total Reading	.73	59	.58–.83
		Iowa Word Analysis	.60	59	.40–.74
	2	DIBELS Next ORF	.91	223	.88–.93
		DIBELS Next Composite	.88	203	.84–.91
		Iowa Total Reading	.79	115	.71–.85
		Iowa Word Analysis	.69	117	.58–.77

Table 3.25 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.84	148	.78–.88
		DIBELS Next Composite	.79	121	.71–.85
		Iowa Total Reading	.63	137	.51–.72
		Iowa Word Analysis	.55	129	.41–.66
	2	DIBELS Next ORF	.93	180	.91–.95
		DIBELS Next Composite	.89	140	.85–.92
		Iowa Total Reading	.74	163	.66–.80
		Iowa Word Analysis	.64	137	.53–.73
Third	1	DIBELS Next ORF	.94	84	.91–.96
		DIBELS Next Composite	.93	81	.89–.95
		Iowa Total Reading	.74	61	.61–.84
	2	DIBELS Next ORF	.90	203	.87–.92
		DIBELS Next Composite	.86	126	.80–.90
		Iowa Total Reading	.69	132	.59–.77
		Iowa Word Analysis	.50	96	.33–.64
Fourth	1	DIBELS Next ORF	.91	28	.82–.96
		Iowa Total Reading	.53	55	.30–.70
	2	Iowa Total Reading	.67	55	.49–.79
Fifth	1	Iowa Total Reading	.63	128	.52–.73
	2	DIBELS Next ORF	.91	85	.86–.94
		Iowa Total Reading	.69	99	.57–.78

Table 3.25 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency

Grade	Period	Criterion	r	N	CI
Sixth	1	Iowa Total Reading	.50	49	.25–.68
	2	Iowa Total Reading	.65	86	.50–.75
Seventh	1	Iowa Total Reading	.52	59	.31–.69
	2	Iowa Total Reading	.57	101	.42–.69
Eighth	1	Iowa Total Reading	.69	46	.50–.82
	2	Iowa Total Reading	.48	85	.30–.63

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

Table 3.26 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
First	1	DIBELS Next ORF	.82	156	.77–.87
		DIBELS Next Composite	.87	141	.82–.90
		Iowa Total Reading	.82	59	.71–.89
		Iowa Word Analysis	.72	59	.56–.82
	2	DIBELS Next ORF	.75	223	.69–.81
		DIBELS Next Composite	.83	203	.78–.87
		Iowa Total Reading	.78	115	.70–.84
		Iowa Word Analysis	.72	117	.62–.80

Table 3.26 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
Second	1	DIBELS Next ORF	.62	148	.51–.71
		DIBELS Next Composite	.70	121	.60–.78
		Iowa Total Reading	.61	137	.49–.70
		Iowa Word Analysis	.46	129	.31–.59
	2	DIBELS Next ORF	.61	180	.51–.69
		DIBELS Next Composite	.69	140	.59–.77
		Iowa Total Reading	.51	163	.38–.61
		Iowa Word Analysis	.54	137	.41–.65
Third	1	DIBELS Next ORF	.73	84	.61–.82
		DIBELS Next Composite	.76	81	.65–.84
		Iowa Total Reading	.67	61	.51–.79
	2	DIBELS Next ORF	.56	203	.46–.65
		DIBELS Next Composite	.70	126	.70–.78
		Iowa Total Reading	.53	132	.53–.64
		Iowa Word Analysis	.44	96	.44–.59
Fourth	1	Iowa Total Reading	.44	55	.20–.63
	2	Iowa Total Reading	.37	55	.11–.58
Fifth	1	Iowa Total Reading	.46	128	.31–.59
	2	DIBELS Next ORF	.60	85	.44–.72
		Iowa Total Reading	.41	99	.23–.56

Table 3.26 Predictive Validity Coefficients for DIBELS 8th Edition Oral Reading Fluency-Accuracy

Grade	Period	Criterion	r	N	CI
Sixth	1	Iowa Total Reading	.46	49	.20–.65
	2	Iowa Total Reading	.53	86	.36–.67
Seventh	1	Iowa Total Reading	.29	59	.04–.51
	2	Iowa Total Reading	.36	101	.17–.52
Eighth	1	Iowa Total Reading	.44	46	.17–.65
	2	Iowa Total Reading	.23	85	.02–.42

Note. Period 1 = Beginning of year. 2 = Middle of year. 3 = End of year.

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. ROC analyses 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 describe whether DIBELS 8th Edition scores correctly predicted performance on a criterion measure of reading: DIBELS Next composite score percentile ranks in kindergarten or Iowa Assessment Total Reading percentile ranks in all other grades. ROC analyses yield an area under the curve (AUC) estimate, which describes a test's classification accuracy. An AUC of .5 indicates the test predicts no better than chance. An AUC of 1.0 indicates that a test has perfect predictive power (Habibzadeh, Habibzadeh, & Yadollahie, 2016).

In addition to reporting the AUC for each benchmark subtest, sensitivity and specificity are reported. Sensitivity provides information about 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 is set for a subtest. DIBELS 8 balances sensitivity and specificity because of 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 benchmarks where no cut scores satisfied these criteria, the cut was set at the score that minimized the difference between sensitivity and specificity.

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 in the end of year served as the criterion measure, and in all other grades the criterion measure was the Iowa Assessment Total Reading Score for the end of year. The only exception to this is for NWF in the beginning of third grade and Maze, where the Iowa was administered in the beginning of 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, at some risk or not meeting end of year learning goals, and in need of some support.

In kindergarten, AUCs are uniformly high with the majority falling at .8 or above and sensitivity and specificity are also routinely high (see Table 3.27). AUC, sensitivity, and specificity statistics are not reported for the NWF-WRC 40th percentile cut-score in beginning and middle of year benchmarks in kindergarten because the cut was identical to the 20th percentile cut. LNF and PSF are the strongest indicators for the beginning of kindergarten, while NWF-CLS and WRF increase in strength over the kindergarten year.

In first grade, for NWF scores, WRF, and ORF scores, AUCs are uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Table 3.28). They are somewhat lower for LNF and PSF, which is likely 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. While NWF is a robust predictor, the strongest predictors in first grade are WRF and ORF.

In second grade, AUCs are again uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Table 3.29). While NWF remains a robust predictor, the strongest predictors in second grade are WRF, ORF, and Maze. However, ORF Accuracy shows a declining trend in its predictive power.

In third grade, AUCs are again uniformly high with the majority falling at .8 or above, and sensitivity and specificity are also routinely high (see Table 3.30). The strongest predictors in third grade are ORF and Maze, although NWF and WRF remain good predictors. ORF Accuracy again shows a declining trend in its predictive power.

In fourth, fifth, and sixth grade, AUCs, sensitivity, and specificity for ORF and Maze remain quite high (see Table 3.31, 3.32, and 3.33 respectively). ORF Accuracy shows the same declining trend in predictive power, and is particularly pronounced in fifth grade.

In seventh and eighth grade, AUCs, sensitivity, and specificity for ORF remain quite high (see Table 3.34 and 3.35 respectively). ORF Accuracy shows the same declining trend in predictive power as in other upper elementary grades. Maze ROC analyses are not yet available in seventh and eighth grade, but results will be released during 2019 in an addendum to this manual.

Table 3.27 ROC Curve Results for DIBELS 8 Kindergarten Subtests Predicting DIBELS Next Composite Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
LNF	20th	1	123	.89	.82–.96	.79	.81
		2	154	.92	.88–.96	.89	.81
		3	98	.94	.89–.99	.86	.83
	40th	1	123	.90	.84–.96	.88	.81
		2	154	.93	.89–.96	.86	.83
		3	98	.96	.92–.99	.94	.86
PSF	20th	1	72	.86	.75–.98	.80	.83
		2	29	.98	.92–.99+	.99+	.96
		3	95	.92	.85–.98	.90	.90
	40th	1	72	.79	.67–.91	.70	.76
		2	29	.88	.69–.99+	.83	.91
		3	95	.84	.74–.93	.83	.79
NWF-CLS	20th	1	89	.73	.63–.84	.83	.70
		2	130	.88	.81–.95	.77	.82
		3	109	.90	.84–.96	.82	.84
	40th	1	89	.73	.63–.83	.78	.69
		2	130	.85	.77–.92	.80	.83
		3	109	.93	.88–.98	.84	.87

Table 3.27 ROC Curve Results for DIBELS 8 Kindergarten Subtests Predicting DIBELS Next Composite Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-WRC	20th	1	73	.73	.59–.88	.67	.66
		2	136	.82	.77–.88	.94	.68
		3	108	.86	.79–.93	.77	.82
	40th	1	NA	NA	NA	NA	NA
		2	NA	NA	NA	NA	NA
		3	109	.82	.74–.90	.72	.72
WRF	20th	1	86	.81	.75–.87	.99+	.62
		2	121	.86	.84–.92	.96	.76
		3	92	.95	.90–.99	.88	.83
	40th	1	86	.86	.80–.91	.99+	.71
		2	121	.86	.80–.92	.83	.81
		3	92	.94	.89–.99	.85	.85

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.

Table 3.28 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
LNF	20th	1	80	.83	.73–.93	.76	.82
		2	115	.80	.72–.88	.71	.82
		3	117	.77	.68–.86	.69	.68
	40th	1	80	.82	.73–.91	.72	.76
		2	115	.78	.70–.87	.71	.72
		3	117	.78	.69–.86	.69	.73
PSF	20th	1	79	.53	.38–.68	.54	.55
		2	113	.66	.54–.77	.62	.66
		3	117	.58	.46–.69	.62	.60
	40th	1	79	.54	.41–.67	.52	.51
		2	113	.61	.50–.71	.55	.58
		3	117	.54	.43–.65	.57	.56
NWF-CLS	20th	1	153	.83	.76–.90	.77	.75
		2	196	.86	.80–.92	.75	.81
		3	198	.88	.83–.93	.86	.80
	40th	1	153	.76	.72–.87	.73	.71
		2	196	.82	.76–.88	.75	.77
		3	198	.87	.82–.92	.83	.82

Table 3.28 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-WRC	20th	1	153	.75	.67–.83	.72	.68
		2	196	.82	.76–.88	.75	.74
		3	198	.86	.80–.90	.75	.80
	40th	1	153	.78	.70–.85	.74	.71
		2	196	.81	.75–.87	.73	.71
		3	198	.84	.79–.90	.71	.80
WRF	20th	1	80	.88	.79–.96	.83	.80
		2	115	.89	.82–.95	.86	.84
		3	117	.90	.84–.97	.88	.88
	40th	1	80	.93	.88–.99	.88	.84
		2	115	.96	.92–.99	.92	.88
		3	117	.95	.90–.99+	.92	.90
ORF	20th	1	59	.87	.77–.98	.85	.85
		2	115	.91	.85–.97	.88	.85
		3	116	.89	.83–.96	.85	.89
	40th	1	59	.96	.90–.99+	.90	.97
		2	115	.97	.93–.99+	.92	.92
		3	116	.94	.90–.99	.88	.90

Table 3.28 ROC Curve Results for DIBELS 8 First Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF-ACC	20th	1	59	.87	.77–.98	.85	.87
		2	115	.90	.84–.96	.86	.86
		3	116	.88	.82–.95	.83	.84
	40th	1	59	.95	.89–.99+	.87	.93
		2	115	.95	.92–.99	.89	.92
		3	116	.94	.89–.99	.94	.89

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.

Table 3.29 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1	49	.87	.77–.97	.76	.82
		2	76	.81	.71–.91	.76	.77
		3	84	.79	.69–.89	.70	.72
	40th	1	49	.84	.70–.97	.78	.83
		2	76	.82	.69–.95	.86	.80
		3	84	.83	.73–.92	.79	.73

Table 3.29 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-WRC	20th	1	49	.86	.75–.97	.76	.86
		2	76	.83	.74–.92	.73	.77
		3	84	.77	.66–.87	.70	.75
	40th	1	49	.85	.72–.97	.70	.83
		2	76	.81	.69–.93	.80	.75
		3	84	.79	.67–.90	.69	.77
WRF	20th	1	51	.93	.85–.99+	.91	.93
		2	77	.90	.82–.98	.91	.84
		3	87	.88	.79–.96	.83	.91
	40th	1	51	.90	.82–.99	.82	.83
		2	77	.84	.73–.95	.82	.81
		3	87	.86	.77–.94	.79	.83
ORF	20th	1	116	.89	.81–.98	.87	.82
		2	163	.91	.87–.95	.89	.82
		3	87	.93	.87–.99	.92	.85
	40th	1	116	.86	.79–.93	.77	.81
		2	163	.88	.83–.93	.81	.81
		3	87	.91	.83–.98	.90	.85

Table 3.29 ROC Curve Results for DIBELS 8 Second Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF-ACC	20th	1	116	.87	.77–.97	.83	.81
		2	163	.86	.79–.92	.73	.84
		3	87	.74	.58–.92	.67	.75
	40th	1	116	.82	.74–.90	.71	.81
		2	163	.83	.77–.90	.74	.79
		3	87	.73	.62–.84	.72	.67
Maze	20th	1*	119	.91	.86–.96	.88	.86
		2*	87	.95	.91–.99+	.90	.91
		3*	60	.86	.77–.95	.99+	.49
	40th	1*	119	.96	.94–.99	.80	.97
		2*	87	.92	.85–.99	.84	.93
		3*	60	.94	.88–.99	.99+	.45

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 3.30 ROC Curve Results for DIBELS 8 Third Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
NWF-CLS	20th	1*	95	.84	.76–.93	.70	.82
		2	89	.71	.57–.85	.70	.74
		3	90	.77	.65–.89	.65	.63
	40th	1*	95	.84	.75–.93	.72	.82
		2	89	.71	.60–.82	.65	.65
		3	90	.75	.64–.85	.73	.63
NWF-WRC	20th	1*	95	.84	.76-.93	.70	.82
		2	89	.71	.57–.85	.70	.65
		3	90	.76	.64–.88	.65	.61
	40th	1*	95	.84	.75-.93	.72	.82
		2	89	.71	.61–.82	.64	.68
		3	90	.73	.68–.83	.65	.61
WRF	20th	1	71	.83	.70–.96	.69	.71
		2	89	.82	.72–.92	.70	.75
		3	90	.80	.70–.90	.70	.67
	40th	1	71	.79	.68–.90	.71	.80
		2	89	.80	.71–.89	.71	.70
		3	90	.79	.70–.88	.71	.71

Table 3.30 ROC Curve Results for DIBELS 8 Third Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	43	.91	.80–.99+	.99+	.83
		2	132	.86	.78–.94	.78	.86
		3	90	.82	.72–.92	.75	.77
	40th	1	43	.83	.70–.95	.72	.80
		2	132	.84	.77–.90	.74	.82
		3	90	.77	.67–.87	.73	.71
ORF-ACC	20th	1	43	.92	.80–.99+	.99+	.83
		2	132	.84	.73–.94	.83	.77
		3	90	.77	.69–.89	.65	.73
	40th	1	43	.86	.74–.99+	.78	.84
		2	132	.81	.75–.89	.77	.73
		3	90	.70	.62–.82	.77	.50
Maze	20th	1*	123	.91	.86–.97	.88	.80
		2*	100	.92	.86–.97	.89	.73
		3*	65	.90	.82–.98	.90	.74
	40th	1*	123	.81	.73–.89	.71	.80
		2*	100	.83	.74–.91	.77	.79
		3*	65	.81	.68–.93	.96	.29

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 3.31 ROC Curve Results for DIBELS 8 Fourth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	55	.80	.67–.93	.83	.69
		2	55	.84	.68–.99+	.83	.78
		3	91	.81	.72–.89	.71	.70
	40th	1	55	.75	.61–.88	.77	.73
		2	55	.82	.71–.93	.68	.73
		3	91	.80	.70–.91	.73	.71
ORF-ACC	20th	1	55	.70	.43–.97	.67	.74
		2	55	.81	.66–.98	.67	.76
		3	91	.70	.60–.81	.73	.59
	40th	1	55	.73	.57–.86	.59	.79
		2	55	.41	.46–.76	.68	.49
		3	91	.71	.62–.82	.66	.71
Maze	20th	1*	133	.93	.88–.98	.86	.88
		2*	97	.92	.86–.99	.91	.78
		3*	64	.85	.75–.95	.96	.35
	40th	1*	133	.81	.73–.89	.85	.49
		2*	97	.83	.74–.91	.83	.70
		3*	64	.81	.68–.93	.99+	.18

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 3.32 ROC Curve Results for DIBELS 8 Fifth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	128	.83	.76–.90	.72	.74
		2	99	.82	.74–.90	.72	.75
		3	59	.86	.70–.99+	.78	.82
	40th	1	128	.81	.73–.88	.72	.71
		2	99	.80	.70–.89	.65	.68
		3	59	.89	.80–.97	.71	.84
ORF-ACC	20th	1	128	.70	.61–.81	.58	.63
		2	99	.69	.60–.81	.69	.60
		3	59	.39	.41–.86	.44	.78
	40th	1	128	.65	.56–.75	.74	.57
		2	99	.71	.61–.82	.62	.74
		3	59	.50	.35–.66	.57	.42
Maze	20th	1*	137	.89	.83–.95	.74	.84
		2*	95	.81	.71–.92	.81	.76
		3*	68	.78	.67–.90	.96	.59
	40th	1*	137	.92	.88–.96	.81	.82
		2*	95	.87	.79–.94	.90	.78
		3*	68	.80	.69–.92	.95	.40

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 3.33 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	49	.77	.62–.91	.79	.72
		2	86	.82	.73–.91	.78	.80
		3	82	.82	.73–.92	.74	.82
	40th	1	49	.82	.62–.91	.71	.79
		2	86	.82	.73–.91	.70	.86
		3	82	.84	.73–.92	.74	.86
ORF-ACC	20th	1	49	.77	.63–.91	.71	.68
		2	86	.77	.67–.88	.68	.82
		3	82	.71	.59–.82	.58	.82
	40th	1	49	.67	.52–.83	.57	.64
		2	86	.79	.70–.88	.55	.96
		3	82	.80	.72–.88	.69	.81
Maze	20th	1*	59	.86	.69–.99+	.83	.79
		2*	57	.90	.77–.99+	.60	.92
		3*	TBD	TBD	TBD	TBD	TBD
	40th	1*	59	.80	.66–.93	.73	.73
		2*	57	.81	.64–.98	.70	.81
		3*	TBD	TBD	TBD	TBD	TBD

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 3.34 ROC Curve Results for DIBELS 8 Seventh Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	59	.80	.68–.92	.85	.72
		2	109	.75	.66–.85	.69	.68
		3	99	.78	.70–.88	.71	.71
	40th	1	59	.76	.64–.89	.71	.76
		2	109	.83	.74–.91	.79	.75
		3	99	.79	.70–.89	.72	.72
ORF-ACC	20th	1	59	.79	.69–.92	.67	.84
		2	109	.71	.60–.80	.67	.60
		3	99	.69	.58–.80	.65	.59
	40th	1	59	.73	.65–.89	.53	.86
		2	109	.82	.75–.92	.65	.82
		3	99	.76	.68–.87	.64	.80

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 3.35 ROC Curve Results for DIBELS 8 Eighth Grade Subtests Predicting Iowa Total Reading Scores

Measure	Criterion	Period	N	AUC	AUC CI	Sensitivity	Specificity
ORF	20th	1	46	.82	.69–.94	.71	.73
		2	85	.73	.68–.85	.75	.66
		3	77	.74	.63–.86	.64	.70
	40th	1	46	.89	.79–.98	.78	.86
		2	85	.81	.69–.94	.80	.80
		3	77	.89	.80–.99	.87	.81
ORF-ACC	20th	1	46	.79	.65–.92	.63	.73
		2	85	.70	.59–.82	.68	.68
		3	77	.68	.57–.80	.64	.64
	40th	1	46	.82	.70–.97	.81	.71
		2	85	.76	.65–.90	.79	.65
		3	77	.71	.58–.84	.57	.69

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.

Summary

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

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

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.