

Utility of *DIBELS*[®] Next for Predicting GRADE Outcomes

Roland H. Good III, Kelly A. Powell-Smith, Dynamic Measurement Group

Amy Murdoch, College of Mount St. Joseph

Rachael Latimer, Dynamic Measurement Group

Introduction

DIBELS Next is a set of brief assessments designed to measure critical skills in early reading. A powerful utility of the measures is for early identification and progress monitoring of students who may not meet reading standards. This poster examines the extent to which the *DIBELS Next* measures administered at the beginning and middle of the year predict outcomes on the Group Reading Assessment and Diagnostic Evaluation (GRADE) assessment administered at the end of the year. We present correlations between *DIBELS Next* measures and the Group Reading Assessment and Diagnostic Evaluation (GRADE). We also present correlations between individual *DIBELS Next* measures and the *DIBELS* Composite Score. Finally, we present inter-rater, test-retest, and alternate-form reliability information. Implications for practice and research are discussed.

Research Questions

1. What is the alternate form reliability of *DIBELS Next* measures?
2. What is the test-retest reliability of *DIBELS Next* measures?
3. What is the inter-rater reliability for *DIBELS Next* measures?
4. What are the correlations between the *DIBELS Next* measures and the Group Reading Assessment and Diagnostic Evaluation (GRADE), a criterion measure of reading proficiency that includes comprehension?
5. What are the correlations between *DIBELS Next* measures and the *DIBELS* Composite score?

Method

The technical adequacy data reported here were from a study designed for the purpose of developing benchmark goals for the *DIBELS Next* assessments.

Participants

- Students were recruited from 13 schools in five school districts representing five US regions.
- School-level demographics from the NCES website for the 2008–2009 school year were aggregated across participating schools (NCES, 2008, <http://nces.ed.gov/>). These data indicate a predominantly white student body (94% white, 4% Hispanic) with a free/reduced lunch rate of 16%.
- School districts had a median of 10 years experience using *DIBELS*.
- K–6th grade students participated in *DIBELS Next* assessments ($n = 3,816$ total; 433 to 569 per grade). The percentage of this sample earning scores at or above benchmark ranged from 65%–79% across grades and times of year.
- Four distinct subsamples of these students participated in data collection to examine correlations with the Group Reading Assessment and Diagnostic Evaluation (GRADE), test-retest reliability, alternate-form reliability, or inter-rater reliability.
- GRADE subsample ($n = 1257$ total; 103 to 219 per grade): A stratified sampling approach was used to select this subsample based upon beginning-of-year *DIBELS* benchmark assessment. Attempts were made to accommodate each sites' requests for the manner in which students were to be selected (e.g., by in-tact classrooms). The GRADE sub-sample was 50% female on average across grades. Additional demographic data on this subsample is shown in *Figures 1–3*.
- Alternate form subsample: A stratified random sample from a single school district was selected based on student performance from the beginning-of-year *DIBELS* benchmark assessment ($n = 166$ total; 20 to 30 per grade).

-
- Test-retest subsample: A stratified random sample from a single school district was selected based upon student *DIBELS* performance from the beginning-of-year benchmark assessment ($n = 152$ total; 21 to 28 per grade). Data are not reported for kindergarten and sixth-grade measures, or for first-grade Retell, due to insufficient sample sizes leaving a final sample of 120 students.
 - Inter-rater reliability subsample: Students across all grades were randomly selected in five schools for shadow-scoring ($n = 264$ total; 20 to 28 per grade).

Measures

Measures in this study included all *DIBELS Next* measures and the Group Reading Assessment and Diagnostic Evaluation (GRADE).

DIBELS Next (except for Daze, all are individually-administered one-minute assessments) include:

- Letter Naming Fluency
- First Sound Fluency
- Phoneme Segmentation Fluency
- Nonsense Word Fluency
- Oral Reading Fluency (includes Retell)
- Daze (*DIBELS*-maze) (group-administered; 3 minutes)
- *DIBELS* Composite Score

Group Reading Assessment and Diagnostic Evaluation (GRADE):

- Un-timed and group-administered
- Appropriate for students in preschool through grade 12
- Five components and 16 subtests that combine to form the following composites: Phonemic Awareness, Early Literacy Skills, Comprehension, Vocabulary, and Total Test. The GRADE Total Test score is comprised of scores across subtests of the GRADE that vary by grade level.
- Reliability ranges from .77 to .98.
- Correlation coefficients range from .69 to .86 with other group- and individually-administered achievement tests.

Procedures

All Data were collected during the 2009–2010 school year.

- *DIBELS Next* assessments were administered at regular benchmark intervals.
- GRADE testing was conducted across two to three sessions in the spring. Testing time ranged from 60 to 90 minutes.
- Test-retest data were collected by testing in the two weeks following the middle-of-year benchmark assessment for all measures except Daze. Daze was not administered due to time constraints. Inter-rater reliability data for all *DIBELS Next* measures was gathered during beginning-of-year benchmark administration using a shadow-scoring process. All *DIBELS Next* measures were included in this portion of the study. In third through sixth grade, students were divided into two groups; one group had shadow-scoring for DORF and the other for Daze.
- Prior to data analysis, data for students with missing or duplicate IDs were removed. We also removed data for scores that were invalid due to known data collection errors, invalid score ranges, or significant univariate or bivariate outliers.
- To obtain the *DIBELS* Composite Score, for each grade and time of year, the *DIBELS Next* measures that correlate highly with later outcomes were combined. Each measure is weighted so that all contribute approximately equally to the *DIBELS* Composite Score.

Results

Figure 1

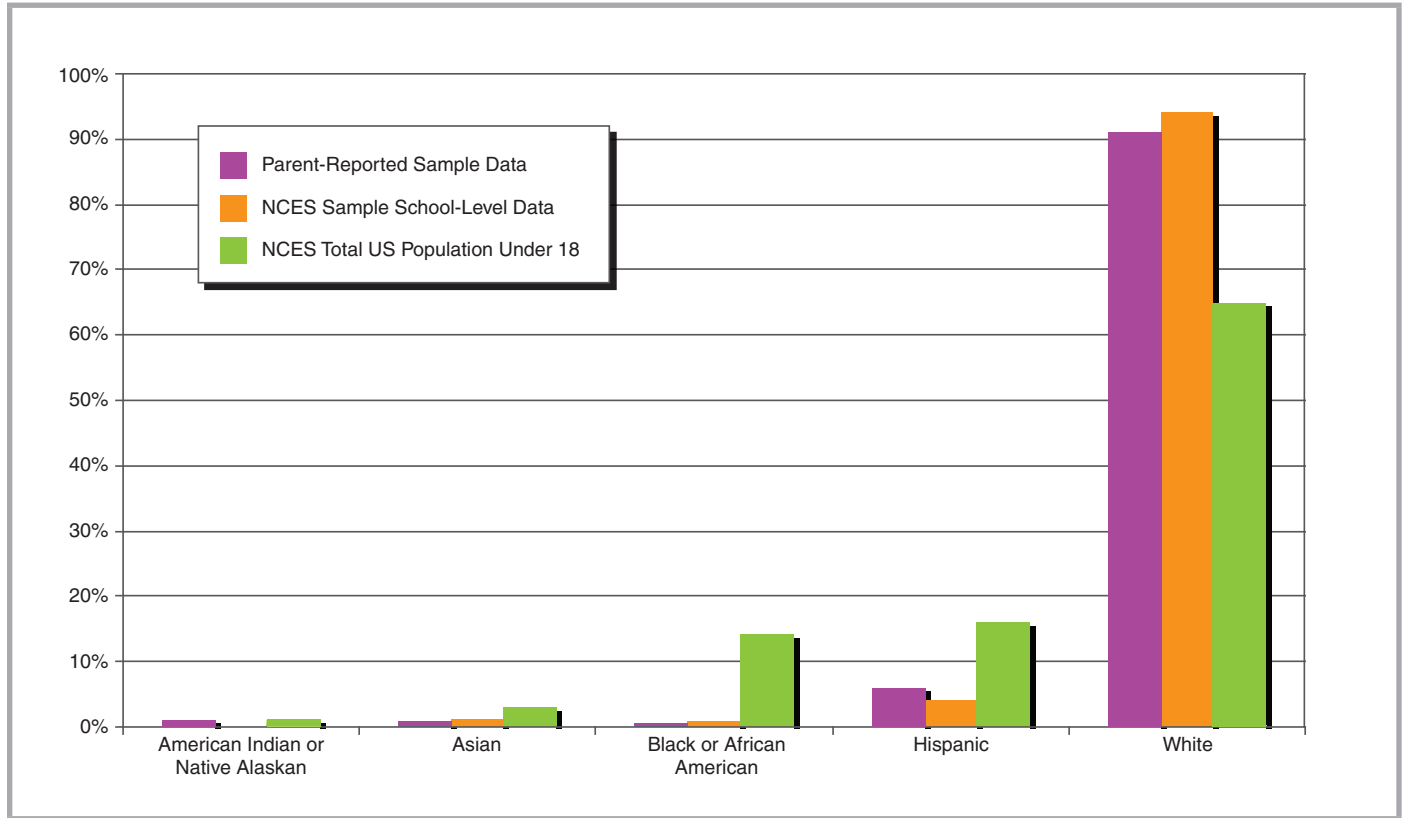


Figure 2

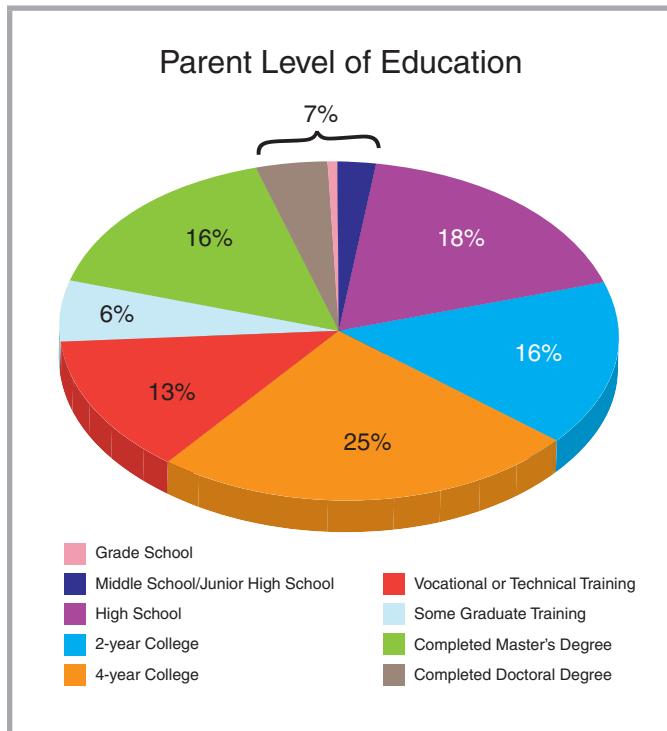
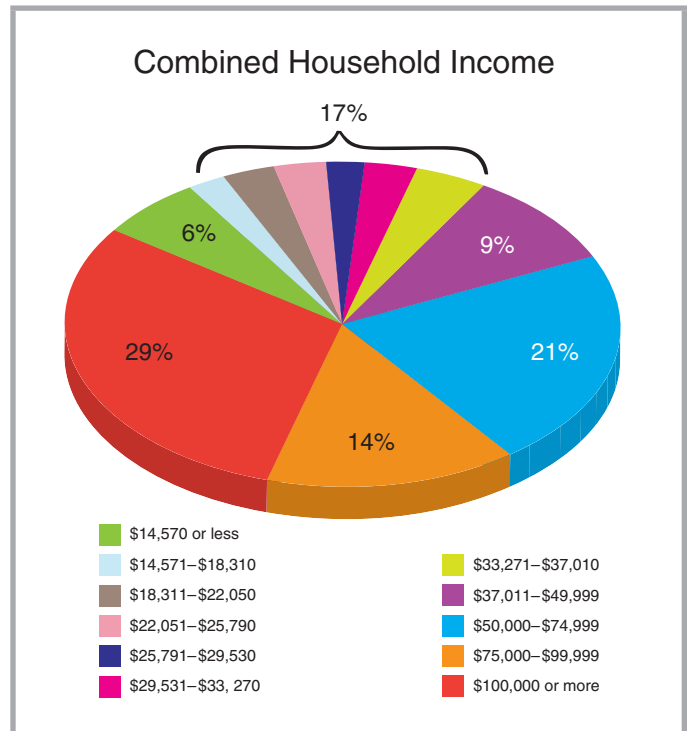


Figure 3



Reliability

Salvia, Ysseldyke, & Bolt's (2007) standards for reliability were used to evaluate the reliability data for *DIBELS Next*. According to these standards, a minimum of .60 is required for administrative purposes and scores that are reported for groups of individuals, a minimum of .80 is required for screening decisions, and a minimum of .90 is required for important educational decisions concerning an individual student.

Reliability estimates are reported for individual test administrations of each measure as well as for the aggregate (mean or median) of three alternate forms. For DORF, the reliability of three-form triads is reported. For other measures, the reliability of three-form aggregates is estimated using the Spearman-Brown Prophecy Formula. Reliability estimates for the *DIBELS Composite Score* represent the reliability of an aggregate of multiple different measures administered at one time.

Alternate form Reliability. Results are presented in *Tables 1–4*. For individual scores, most coefficients are above .80, indicating sufficient reliability for screening decisions. Several coefficients are above .90, indicating sufficient reliability for important individual educational decisions. For the *DIBELS Composite Score*, reliability is consistently high across first through fifth grade.

Table 1: Two-Week Alternate-Form Reliability for Kindergarten and First Grade *DIBELS* Measures

<i>DIBELS Measures by Grade</i>	N	<i>First Form</i>		<i>Second Form</i>		<i>Reliability</i>	
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Single-Form</i>	<i>Estimated Three-Form</i>
Kindergarten							
First Sound Fluency	29	32.34	10.67	32.79	6.65	.52**	.76
Letter Naming Fluency	29	39.76	15.90	45.48	15.64	.86	.95
Phoneme Segmentation Fluency	29	25.45	14.46	29.97	11.43	.44	.70
NWF Correct Letter Sounds	27	17.37	10.78	21.89	14.82	.71	.88
NWF Whole Words Read	27	0.74	1.81	2.04	3.78	.92	.97
First Grade							
NWF Correct Letter Sounds	28	53.25	23.91	54.18	25.96	.85	.94
NWF Whole Words Read	28	9.50	12.00	10.29	12.52	.90	.96

Note. Based on middle of year data. The estimated three-form reliability is based on the Spearman-Brown Prophecy Formula. Unless marked, correlations significant, $p < .001$; ** $p < .01$.

Table 2: Two-Week Alternate-Form Reliability for Three-Passage Groups (Triads) of *DIBELS Next* Oral Reading Fluency Passages

DORF Scores by Grade	N	First Triad		Second Triad		Triad Reliability
		Mean	SD	Mean	SD	
DORF Words Correct						
First Grade	28	37.39	40.06	46.00	41.57	.98
Second Grade	24	75.08	42.06	82.46	38.01	.97
Third Grade	30	91.87	39.93	95.80	35.21	.96
Fourth Grade	30	104.47	39.48	110.43	37.86	.96
Fifth Grade	25	113.56	27.96	120.48	27.98	.95
DORF Accuracy						
First Grade	28	77%	15	84%	11	.88
Second Grade	24	91%	9	93%	8	.83
Third Grade	30	96%	4	95%	5	.80
Fourth Grade	30	96%	5	97%	4	.85
Fifth Grade	25	97%	2	98%	2	.76
DORF Retell						
Second Grade	20	26.6	13.32	29.73	17.22	.68
Third Grade	27	32.11	20.00	27.80	16.33	.81
Fourth Grade	30	34.17	18.16	38.50	18.74	.80
Fifth Grade	25	37.24	15.86	36.04	18.55	.65

Note. Based on middle of year data. ‘Triad’ refers to a group of three DORF passages, and the mean scores reported in this table represent the mean of the student-level median scores based upon a standardized benchmark administration of the triad. Data is unavailable for first-grade DORF Retell and all sixth-grade measures due to insufficient sample sizes. DORF passages are administered in triads, thus the alternate-form reliability is reported for triads. All correlations significant, $p < .001$.

Table 3: Two-Week Alternate-Form Reliability for Daze

Grade	N	First Form		Second Form		Reliability	
		Mean	SD	Mean	SD	Single-Form	Estimated Three-Form
Third Grade	24	11.13	7.83	13.75	7.93	.86	.95
Fourth Grade	29	16.34	5.92	20.93	7.28	.67	.86
Fifth Grade	20	13.15	5.96	23.35	8.41	.49*	.74

Note. Based on from middle of year data. The estimated three-form reliability is based on the Spearman-Brown Prophecy Formula. Unless marked, correlations significant, $p < .001$; * $p < .05$.

Table 4: Two-Week Alternate-Form Reliability for *DIBELS* Composite Score

Grade	N	First Composite		Second Composite		Reliability
		Mean	SD	Mean	SD	
Kindergarten	27	119.04	36.47	132.63	36.21	.66
First Grade	28	156.07	92.35	177.18	95.37	.95
Second Grade	24	183.08	108.63	209.08	99.35	.92
Third Grade	20	271.40	137.57	273.95	121.86	.97
Fourth Grade	25	317.80	118.60	359.56	123.74	.95
Fifth Grade	20	327.60	87.21	376.50	95.65	.91

Note. Based on middle of year data. The first composite was calculated from middle-of-year benchmark assessment data. The second composite was calculated from alternate forms that were administered two weeks after middle-of-year benchmark assessment. All correlations significant, $p < .001$.

Test-retest reliability. Test-retest reliability is presented in Tables 5–7. Test-retest reliability coefficients appear to be conservative estimates in light of the alternate-form reliability coefficients presented. For NWF, reliability coefficients are sufficient for screening decisions. In general, for DORF Words Correct and the *DIBELS* Composite Score, reliability coefficients are sufficient for making important individual educational decisions.

Table 5: Test-Retest Reliability for First Grade Nonsense Word Fluency

NWF Scores	N	First Administration		Second Administration		Reliability	
		Mean	SD	Mean	SD	Single-Form	Estimated Three-Form
NWF Correct Letter Sounds	27	58.63	22.27	69.00	22.83	.76	.90
NWF Whole Words Read	27	12.63	10.58	17.11	11.54	.70	.88

Note. Based on middle of year data. Data not available for kindergarten due to insufficient sample size. The estimated three-form test-retest reliability is based on the Spearman-Brown Prophecy Formula. All correlations are significant, $p < .001$.

Table 6: Test-Retest Reliability for *DIBELS* Oral Reading Fluency

<i>DORF Scores by Grade</i>	<i>N</i>	<i>First Form</i>		<i>Second Form</i>		<i>Reliability</i>
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
<i>DORF Words Correct</i>						
First Grade	28	35.86	26.22	44.29	28.66	.95
Second Grade	21	102.38	27.74	113.76	28.37	.91
Third Grade	27	104.93	35.03	123.37	38.51	.93
Fourth Grade	21	121.14	38.49	140.14	37.09	.97
Fifth Grade	23	124.43	42.71	134.13	43.56	.97
<i>DORF Accuracy</i>						
First Grade	28	77%	22	83%	4	.84
Second Grade	21	97%	3	99%	1	.57**
Third Grade	27	97%	2	99%	2	.68
Fourth Grade	21	97%	3	99%	2	.91
Fifth Grade	23	96%	5	97%	8	.94
<i>DORF Retell</i>						
Second Grade	21	48.33	15.21	49.86	17.81	.27†
Third Grade	27	57.07	20.22	58.89	19.78	.69
Fourth Grade	21	57.57	22.11	52.90	15.18	.36†
Fifth Grade	22	52.32	19.15	60.27	15.75	.58**

Note. Based on middle of year data. Data not available for first-grade DORF Retell and sixth-grade measures due to insufficient sample size. DORF passages are administered in triads, thus the test-retest reliability is reported as three-form. Unless marked, all correlations significant, $p < .001$; ** $p < .01$; † Not significant.

Table 7: Test-Retest Reliability for *DIBELS* Composite Score

<i>Grade</i>	<i>N</i>	<i>First Form</i>		<i>Second Form</i>		<i>Reliability</i>
		<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	
First Grade	27	163.63	78.57	194.44	82.56	.94
Second Grade	21	298.86	60.79	321.67	64.48	.81

Note. Based on middle of year data. Test-retest reliability for *DIBELS* Composite Score for third through sixth grade is unavailable, because information about Daze was not available. All correlations significant, $p < .001$.

Inter-rater reliability. Inter-rater reliability coefficients are presented in Tables 8–11. Mean scores across all grades are different by approximately 1 point or less. Correlations for most measures are above .90. Inter-rater reliability is high for all measures indicating that scoring directions were applied in a consistent manner across assessors in this study.

Table 8: Inter-Rater Reliability for Kindergarten, First and Second Grade *DIBELS* Measures

<i>DIBELS</i> Measures by Grade	N	First Rater		Second Rater		Inter-Rater Reliability
		Mean	SD	Mean	SD	
Kindergarten						
First Sound Fluency	25	12.36	11.98	11.56	12.17	.94
Letter Naming Fluency	25	20.52	14.31	20.12	14.50	.99
First Grade						
Letter Naming Fluency	25	48.52	19.79	48.68	19.90	.99
Phoneme Segmentation Fluency	25	38.76	17.16	37.20	16.29	.95
NWF Correct Letter Sounds	25	41.32	32.18	40.80	32.41	.99
NWF Whole Words Read	25	8.00	12.19	7.60	12.14	.99
Second Grade						
NWF Correct Letter Sounds	25	64.08	32.63	64.00	33.39	.90
NWF Whole Words Read	25	16.72	14.69	16.56	14.36	.99

Note. Based on beginning of year data. The estimated three-form reliability based on the Spearman-Brown Prophecy Formula for all measures was above .98. All correlations were significant, $p < .001$.

Table 9: Inter-Rater Reliability for *DIBELS* Oral Reading Fluency (DORF)

<i>DORF</i> Scores by Grade	N	First Rater		Second Rater		Triad Reliability
		Mean	SD	Mean	SD	
DORF Words Correct						
Second Grade	25	58.72	28.67	58.32	29.37	.99
Third Grade	25	95.24	37.97	94.68	37.79	.99
Fourth Grade	24	98.71	32.44	98.38	31.92	.99
Fifth Grade	28	110.04	40.24	110.25	40.51	.99
Sixth Grade	20	140.80	32.30	141.25	32.34	.99
DORF Accuracy						
Second Grade	25	90%	10	90%	10	.99
Third Grade	25	95%	6	95%	6	.85
Fourth Grade	24	96%	4	96%	4	.93
Fifth Grade	28	96%	4	95%	4	.95
Sixth Grade	20	98%	2	98%	2	.91
DORF Retell						
Second Grade	20	26.60	12.65	26.75	13.35	.98
Third Grade	24	36.96	14.95	37.29	15.80	.92
Fourth Grade	24	39.17	18.13	39.75	19.25	.98
Fifth Grade	28	35.79	16.96	35.07	18.26	.96
Sixth Grade	20	41.10	19.60	42.50	19.31	.99

Note. Based on Study C beginning of year data. DORF passages are administered in triads, thus the inter-rater reliability is reported as three-form. All correlations were significant, $p < .001$.

Table 10: Inter-Rater Reliability for Daze

Grade	N	First Rater		Second Rater		Single-Form Reliability
		Mean	SD	Mean	SD	
Third Grade	25	10.60	6.64	10.56	6.60	.99
Fourth Grade	25	15.92	6.20	15.96	6.33	.98
Fifth Grade	26	20.81	9.87	21.23	9.95	.99
Sixth Grade	20	22.55	8.61	22.40	8.75	.99

Note. Based on beginning of year data. The estimated three-form reliability of Daze based on the Spearman-Brown Prophecy Formula was above .99. All correlations were significant, $p < .001$.

Table 11: Inter-Rater Reliability for DIBELS Composite Score

Grade	N	First Rater		Second Rater		Reliability
		Mean	SD	Mean	SD	
Kindergarten	25	32.88	21.47	31.68	22.25	.97
First Grade	25	128.60	55.93	126.68	55.37	.99
Second Grade	25	169.32	80.13	150.32	99.69	.98

Note. Based on middle of year data. Reliability for third- through sixth-grade is unavailable, because students in this portion of the study received only DORF or Daze, and not both. All correlations significant, $p < .001$.

Validity

Predictive and concurrent validity. Predictive and concurrent validity data are reported for *DIBELS Next* measures with respect to the GRADE Total Test and for the *DIBELS* Composite Score. Descriptors from Hopkins (2002) are used to categorize the strength of the relations.

Correlation coefficients indicating the strength of the relation between the *DIBELS Next* measures and GRADE Total Test are reported in *Table 12* (predictive and concurrent validity). Overall, the validity of all *DIBELS* measures is well supported by the correlations with the GRADE Total Test. The *DIBELS* Composite Score in kindergarten and first grade is moderately to strongly correlated with the GRADE Total Test. For second through sixth grade, predictive validity coefficients for the *DIBELS* Composite Score indicate moderate-strong to strong relations with the GRADE Total Test. When examining individual measures, predictive and concurrent validity coefficients are moderate to strong for second- through sixth-grade measures with the GRADE Total Test.

Table 12: Criterion-Related Validity for *DIBELS Next* Measures with GRADE Total Test

<i>DIBELS Measure by Time of Year</i>	<i>Grade Level</i>						
	<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
<i>Predictive Validity Coefficients</i>							
Beginning of year							
First Sound Fluency	.52	--	--	--	--	--	--
Letter Naming Fluency	.39	.54	--	--	--	--	--
Phoneme Segmentation Fluency	--	.33	--	--	--	--	--
NWF Correct Letter Sounds	--	.43	.51	--	--	--	--
NWF Whole Words Read	--	.39	.51	--	--	--	--
DORF Words Correct	--	--	.69	.66	.77	.69	.64
DORF Accuracy	--	--	.75	.68	.62	.53	.55
Retell	-	--	.53	.48	.56	.61	.55
Daze Adjusted Score	--	--	--	.65	.67	.56	.60
DIBELS Composite Score	.50	.55	.75	.73	.80	.76	.71
Middle of year							
First Sound Fluency	.40	--	--	--	--	--	--
Letter Naming Fluency	.35	--	--	--	--	--	--
Phoneme Segmentation Fluency	.34	--	--	--	--	--	--
NWF Correct Letter Sounds	.47	.51	--	--	--	--	--
NWF Whole Words Read	.19*	.52	--	--	--	--	--
DORF Words Correct	--	.64	.76	.67	.77	.65	.59
DORF Accuracy	--	.80	.78	.71	.62	.49	.47
Retell	--	.55	.52	.56	.63	.63	.59
Daze Adjusted Score	--	--	--	.61	.61	.59	.56
DIBELS Composite Score	.48	.71	.80	.78	.80	.76	.68
<i>Concurrent Validity Coefficients</i>							
End of year							
Letter Naming Fluency	.35	--	--	--	--	--	--
Phoneme Segmentation Fluency	.24	--	--	--	--	--	--
NWF Correct Letter Sounds	.40	.56	--	--	--	--	--
NWF Whole Words Read	.35	.56	--	--	--	--	--
DORF Words Correct	--	.75	.73	.66	.74	.65	.61
DORF Accuracy	--	.73	.67	.59	.54	.49	.55
Retell	--	.40	.48	.53	.62	.65	.56
Daze Adjusted Score	--	--	--	.67	.68	.66	.64
DIBELS Composite Score	.40	.77	.75	.75	.80	.77	.73

Note. GRADE Total Test = Group Reading Assessment and Diagnostic Evaluation Total Test raw composite scores. Total sample size = 1306. GRADE administered at end of year. Unless marked, all correlations significant, $p < .001$; * $p < .05$.

Correlation coefficients indicating the strength of the relation between the *DIBELS Next* measures and the *DIBELS* Composite score at a later time (predictive validity) are reported in *Table 13*. Overall, the predictive validity of all *DIBELS* measures is well supported by correlations with the *DIBELS* Composite Score at a later time. With the exception of PSF, the *DIBELS Next* measures in kindergarten and first grade are moderately to strongly correlated with the later *DIBELS* Composite Scores. For second through sixth grade, predictive validity coefficients of all measures with later *DIBELS* Composite Scores are moderate-strong to strong.

Table 13: Predictive Criterion-Related Validity for all *DIBELS Next* Measures with the *DIBELS* Composite Score

<i>DIBELS Next Measure</i>	<i>DIBELS Composite Score by Grade and Time of Year</i>													
	<i>Middle of Year</i>							<i>End of Year</i>						
	<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
	<i>Predictive Validity Coefficients - Beginning of Year</i>													
FSF	.57	--	--	--	--	--	--	.43	--	--	--	--	--	--
LNF	.60	.65	--	--	--	--	--	.49	.65	--	--	--	--	--
PSF	--	.25	--	--	--	--	--	--	.26	--	--	--	--	--
NWF-CLS	--	.82	.69	--	--	--	--	--	.71	.65	--	--	--	--
NWF-WWR	--	.79	.65	--	--	--	--	--	.66	.62	--	--	--	--
DORF Words Correct	--	--	.85	.88	.90	.89	.87	--	--	.81	.86	.86	.85	.86
DORF Accuracy	--	--	.75	.71	.72	.69	.66	--	--	.71	.70	.71	.66	.65
Retell	--	--	.63	.64	.62	.58	.61	--	--	.62	.64	.62	.58	.62
Daze	--	--	--	.79	.76	.74	.78	--	--	--	.74	.76	.69	.77
	<i>Predictive Validity Coefficients - Middle of Year</i>													
FSF	--	--	--	--	--	--	--	.47	--	--	--	--	--	--
LNF	--	--	--	--	--	--	--	.60	--	--	--	--	--	--
PSF	--	--	--	--	--	--	--	.47	--	--	--	--	--	--
NWF-CLS	--	--	--	--	--	--	--	.65	.78	--	--	--	--	--
NWF-WWR	--	--	--	--	--	--	--	.52	.78	--	--	--	--	--
DORF Words Correct	--	--	--	--	--	--	--	--	.83	.87	.86	.87	.87	.87
DORF Accuracy	--	--	--	--	--	--	--	--	.81	.75	.69	.68	.64	.62
Retell	--	--	--	--	--	--	--	--	.67	.70	.65	.68	.67	.72
Daze	--	--	--	--	--	--	--	--	--	--	.72	.75	.77	.77

Note. Approximate pair-wise sample sizes: kindergarten ≈ 465; first grade ≈ 440; second grade ≈ 540; third grade ≈ 480; fourth grade ≈ 570; fifth grade ≈ 520; sixth grade ≈ 510. All correlations significant, $p < .001$.

Discriminant validity. Discriminant validity was examined for the *DIBELS* Composite Score relative to two levels of performance on the GRADE Total Test, below the 40th percentile on the GRADE's national norms and at or above the 40th percentile. *DIBELS* Composite Score descriptive statistics were calculated for each group and compared. Results are reported in *Table 14*. Differences in means were examined using a between-groups t-test for each grade; all yielded significant results. The t-statistics are reported to illustrate the magnitude of the differences in means. The effect size of the *DIBELS* Composite Score based on Cohen's *d* is large across all grades. Overall, the *DIBELS* Composite Score adequately discriminates between these two distinct levels of reading skill at kindergarten through sixth grade levels.

Table 14: Discriminant Validity of the *DIBELS* Composite Score Based on the 40th Percentile Rank on GRADE Total Test Raw Score

Grade by Time of Year	<i>DIBELS Composite Score Descriptive Statistics by GRADE Total Test Percentile Rank</i>						<i>Difference Statistics</i>	
	<i>Below 40th Percentile</i>			<i>Above 40th Percentile</i>			<i>t-stat</i>	<i>Cohen's d</i>
	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>		
Kindergarten								
Beginning	54	22.31	19.65	112	45.42	23.76	6.41	1.03
Middle	55	111.90	54.96	113	156.10	43.16	5.45	0.94
End	53	132.10	40.78	113	156.50	39.09	3.67	0.62
First Grade								
Beginning	54	105.00	29.68	139	145.90	39.54	7.33	1.11
Middle	55	96.51	48.69	140	220.50	88.12	11.17	1.58
End	54	115.10	65.34	139	228.00	59.81	11.26	1.85
Second Grade								
Beginning	61	111.20	61.32	153	219.80	60.88	11.74	1.79
Middle	61	136.70	83.90	158	282.10	60.87	13.26	2.15
End	60	194.00	82.49	157	309.90	67.27	10.19	1.62
Third Grade								
Beginning	49	168.80	96.65	135	327.60	85.88	10.43	1.80
Middle	51	221.50	94.03	136	390.30	83.82	11.56	1.96
End	51	279.80	99.64	136	442.00	79.58	11.00	1.91
Fourth Grade								
Beginning	64	200.00	110.10	119	360.60	82.62	10.68	1.73
Middle	65	250.20	102.00	120	400.90	73.24	11.06	1.79
End	66	316.30	106.30	120	467.60	76.42	10.70	1.73
Fifth Grade								
Beginning	93	311.70	95.38	101	454.30	77.61	11.41	1.66
Middle	92	346.70	82.12	102	477.50	73.50	11.67	1.69
End	92	377.80	92.00	101	527.80	80.28	12.06	1.75
Sixth Grade								
Beginning	19	292.30	98.61	84	442.30	77.95	6.70	1.85
Middle	17	330.90	112.70	85	483.80	93.18	5.61	1.60
End	19	334.90	101.40	86	502.40	84.85	7.11	1.92

Note. All *t*-tests were performed under both equal and unequal variance assumptions, both of which yielded highly significant results; the reported *t*-statistic is the average between the two tests under different assumptions. A pooled standard deviation was calculated for Cohen's *d*.

Discussion

Conclusions

Reliability coefficients are consistently high across all three forms of reliability. Alternate-form reliability for individual DORF passages is particularly strong, indicating high consistency between passages. With repeated assessment across multiple forms, reliability increases substantially, as noted where the estimated three-form reliability is reported. Reliability estimates increase substantially to be sufficient for important individual decisions for most measures and grade levels when three-form aggregates are examined.

In addition, the aggregate of multiple different measures using the *DIBELS* Composite Score provides highly reliable information for educational decisions. The *DIBELS* Composite Score provides the best estimate of the student's overall reading proficiency, and reliability for this score is above .90 for first through sixth grades, indicating sufficient reliability for important individual educational decisions. In general, the results presented here suggests that *DIBELS Next* possesses little test error and that users can have confidence in test results.

With respect to validity, moderate to strong concurrent and predictive validity coefficients with the GRADE Total Test were found for individual *DIBELS Next* measures across grades. Further, the *DIBELS* Composite Score demonstrates good discriminant validity with respect to GRADE Total Test results.

Implications for Practice and Research

The reliability data suggest confidence in the stability of scores earned on the *DIBELS Next* measures. The validity data suggest that the results from *DIBELS Next* measures and the *DIBELS* Composite Score provide meaningful information regarding students' development of critical reading skills. Taken together, these results suggest educators can be confident when making decisions using *DIBELS Next* for universal screening and ongoing progress monitoring. Further study to replicate these findings with additional samples of students and other outcome measures would be helpful.

Note: Data presented in this poster are from the Benchmark Goals Study. Additional studies that contributed technical data regarding *DIBELS Next* are not presented here. For complete information on the technical adequacy of *DIBELS Next* see the *DIBELS Next Technical Manual* available at www.dibels.org.

References

- Hopkins, W. G. (2002). A scale of magnitudes for the effect statistics. In *A review of statistics*. Retrieved Aug 11, 2010 from <http://www.sportsci.org/resource/stats/effectmag.html>
- Salvia, J., Ysseldyke, J., & Bolt, S. (2007). *Assessment in special and inclusive education* (Tenth Edition). Boston: Houghton Mifflin Company.

Contact information:

Roland H. Good, III / rhgood@dibels.org
Ruth A. Kaminski / rkamin@dibels.org
Kelly A. Powell-Smith / kpowellsmith@dibels.org
General information / info@dibels.org

Or visit our website at:
<http://www.dibels.org/>

DIBELS is a registered trademark of Dynamic Measurement Group, Inc.